Algebra & Number Theory

Volume 7
2013
No. 9

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2013 is US $200/year for the electronic version, and $350/year (+$40, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscribers address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2059

Multiplicities associated to graded families of ideals

Steven Dale Cutkosky

We prove that limits of multiplicities associated to graded families of ideals exist under very general conditions. Most of our results hold for analytically unramified equicharacteristic local rings with perfect residue fields. We give a number of applications, including a "volume = multiplicity" formula, generalizing the formula of Lazarsfeld and Mustaţă, and a proof that the epsilon multiplicity of Ulrich and Validashti exists as a limit for ideals in rather general rings, including analytic local domains. We prove a generalization of this to generalized symbolic powers of ideals proposed by Herzog, Puthenpurakal and Verma. We also prove an asymptotic "additivity formula" for limits of multiplicities and a formula on limiting growth of valuations, which answers a question posed by the author, Kia Dalili and Olga Kashcheyeva. Our proofs are inspired by a philosophy of Okounkov for computing limits of multiplicities as the volume of a slice of an appropriate cone generated by a semigroup determined by an appropriate filtration on a family of algebraic objects.

1. Introduction

In a series of papers, Okounkov interprets the asymptotic multiplicity of graded families of algebraic objects in terms of the volume of a slice of a corresponding cone (the Okounkov body). Okounkov's method employs an asymptotic version of a result of Khovanskii [1992] for finitely generated semigroups. One of his realizations of this philosophy [Okounkov 1996; 2003] gives a construction that computes the volume of a family of graded linear systems. This method was systematically developed by Lazarsfeld and Mustaţă [2009], who give many interesting consequences, including a new proof of Fujita approximation (see [Fujita 1994] for the original proof) and the fact that the volume of a big divisor on an irreducible projective variety over an algebraically closed field is a limit, which was earlier proven in [Lazarsfeld 2004] using Fujita approximation. More recently, Fulger [2011] has extended this result to compute local volumes of divisors on a log resolution of a normal variety over

The author is partially supported by the NSF.
MSC2010: primary 13H15; secondary 14B05.
Keywords: multiplicity, graded family of ideals, Okounkov body.

an algebraically closed field. Kaveh and Khovanskii [2012] have recently greatly generalized the theory of Newton–Okounkov bodies and applied this to general graded families of linear systems.

The method used in these papers is to choose a nonsingular closed point β on the d-dimensional variety X and then using a flag, a sequence of subvarieties

$$\{\beta\} = X_0 \subset X_1 \subset \cdots \subset X_{d-1} \subset X$$

that are nonsingular at β, to determine a rank-d valuation of the function field $k(X)$ that dominates the regular local ring $\mathbb{O}_{X,\beta}$. This valuation gives a very simple filtration of $\mathbb{O}_{X,\beta}$ represented by monomials in a regular system of parameters of $\mathbb{O}_{X,p}$, which are local equations of the flag. Since the residue field is algebraically closed, this allows us to associate a set of points in $\mathbb{Z}^d$ to a linear system on X (by means of a k-subspace of $k(X)$ giving the linear system) so that the number of these points is equal to the dimension of the linear system. In this way, a semigroup in $\mathbb{Z}^{d+1}$ is associated to a graded family of linear systems.

One of their applications is to prove a formula of equality of volume and multiplicity for a graded family $\{I_i\}_{i\in\mathbb{N}}$ of m_R-primary ideals in a local ring (R, m_R) such that R is a local domain that is essentially of finite type over an algebraically closed field k with $R/m_R = k$ [Lazarsfeld and Mustaţă 2009, Theorem 3.8]. These assumptions on R are all necessary for their proof. The proof involves interpreting the problem in terms of graded families of linear systems on a projective variety X on which R is the local ring of a closed point α. Then a valuation as above is constructed that is centered at a nonsingular point $\beta \in X$, and the cone methods are used to prove the limit. The formula "volume = multiplicity" for graded families of ideals was first proven by Ein, Lazarsfeld and Smith [Ein et al. 2003] for valuation ideals associated to an Abhyankar valuation in a regular local ring that is essentially of finite type over a field. Mustaţă [2002] proved the formula for regular local rings containing a field. In all of these cases, the volume $\mathrm{vol}(I_*)$ of the family, which is defined as a lim sup, is shown to be a limit.

Let $\{I_i\}$ be a graded family of ideals in a d-dimensional (Noetherian) local ring (R, m_R); that is, the family is indexed by the natural numbers with $I_0 = R$ and $I_i I_j \subset I_{i+j}$ for all i and j. Suppose that the ideals are m_R-primary (for $i > 0$). Let $\ell_R(N)$ denote the length of an R-module N. We find very general conditions on R under which the "volume"

$$\mathrm{vol}(I_*) = \limsup \frac{\ell_R(R/I_n)}{n^d/d!}$$

is actually a limit. For instance, we show that this limit exists if R is analytically unramified and equicharacteristic with perfect residue field (Theorem 5.8) or if R is regular (Theorem 4.6).

We thank the referee for pointing out that our basic result Theorem 4.2 is valid without our original assumption of excellence.

Our proof involves reducing to the case of a complete domain and then finding a suitable valuation that dominates R to construct an Okounkov body. The valuation that we use is of rank 1 and rational rank d. There are two issues that require special care in the proof. The first issue is to reduce to the case of an analytically irreducible domain. Analytic irreducibility is necessary to handle the boundedness restriction on the corresponding cone (condition (2)). The proof of boundedness is accomplished by using the linear Zariski subspace theorem of Hübl [2001] (which is valid if R is assumed excellent) or, as was pointed out by the referee, by an application of the version of Rees [1989] of Izumi's theorem, for which excellence is not required. The second issue is to handle the case of a nonclosed residue field. Our method for converting the problem into a problem of cones requires that the residue field of the valuation ring be equal to the residue field of R. Care needs to be taken when the base field is not algebraically closed. The perfect condition in Theorem 5.8 on the residue field is to prevent the introduction of nilpotents upon base change.

The limit $\lim_{n\to\infty} \ell_R(R/I^n)/(n^d/d!)$ is just the Hilbert–Samuel multiplicity $e(I)$, which is a positive integer, in the case when $I_n = I^n$ with I an m_R-primary ideal. In general, when working with the kind of generality allowed by a graded family of m_R-primary ideals, the limit will be irrational. For instance, given $\lambda \in \mathbb{R}_+$, the graded family of m_R-primary ideals I_n generated by the monomials $x^i y^j$ such that $(1/2\lambda)i + j \geq n$ in the power-series ring $R = k[\![x, y]\!]$ in two variables will give us the limit $\lim_{n\to\infty} \ell_R(R/I_n)/n^2 = \lambda$.

We also obtain irrational limits for more classical families of ideals. Suppose that R is an excellent d-dimensional local domain with perfect residue field and ν is a discrete valuation dominating R (the value group is $\mathbb{Z}$). Then the valuation ideals $I_n = \{f \in R \mid \nu(f) \geq n\}$ form a graded family of m_R-primary ideals, so Theorem 5.8 tells us that the limit $\lim_{n\to\infty} \ell_R(R/I_n)/n^d$ exists. This limit will however in general not be rational. [Cutkosky and Srinivas 1993, Example 6] gives such an example in a three-dimensional normal local ring.

We give a number of applications of this formula and these techniques to the computation of limits in commutative algebra.

We prove the formula "$\mathrm{vol}(I_*) = \mathrm{multiplicity}(I_*)$" for local rings R and graded families of m_R-primary ideals such that either R is regular or R is analytically unramified and equicharacteristic with perfect residue field in Theorem 6.5. In our proof, we use a critical result on volumes of cones, which is derived in [Lazarsfeld and Mustaţă 2009]. We generalize this result to obtain an asymptotic additivity formula for multiplicities of an arbitrary graded family of ideals (not required to be m_R-primary) in Theorem 6.10.

Another application is to show that the epsilon multiplicity of Ulrich and Validashti [2011], defined as a lim sup, is actually a limit in some new situations. We prove that this limit exists for graded families of ideals in a local ring R such that one of the following holds: R is regular, R is analytically irreducible and excellent with algebraically closed residue field or R is normal, excellent and equicharacteristic with perfect residue field. As an immediate consequence, we obtain the existence of the limit for graded families of ideals in an analytic local domain, which is of interest in singularity theory. In [Cutkosky et al. 2005], an example is given showing that this limit is in general not rational. Previously, the limit was shown to exist in some cases in [Cutkosky et al. 2010b], and the existence of the limit was proven (for more general modules) in some cases in [Kleiman 2010] and over a domain R that is essentially of finite type over a perfect field in [Cutkosky 2011]. The proof in the latter paper used Fujita approximation on a projective variety on which the ring R was the local ring of a closed point.

We prove in Corollary 6.4 a formula on asymptotic multiplicity of generalized symbolic powers proposed by Herzog, Puthenpurakal and Verma [Herzog et al. 2008, beginning of Introduction].

We also prove that a question raised in [Cutkosky et al. 2010a] about the growth of the semigroup of a valuation semigroup has a positive answer for very general valuations and domains. We prove in Theorem 7.1 that if R is a d-dimensional regular local ring or an analytically unramified local domain with algebraically closed residue field and ω is a zero-dimensional rank-1 valuation dominating R with value group contained in $\mathbb{R}$ and if $\varphi(n)$ is the number of elements in the semigroup of values attained on R that are less than n, then

$$\lim_{n\to\infty} \frac{\varphi(n)}{n^d}$$

exists. This formula was established if R is a regular local ring of dimension 2 with algebraically closed residue field in [Cutkosky et al. 2010a] and if R is an arbitrary regular local ring of dimension 2 in [Cutkosky and Vinh 2011] using a detailed analysis of a generating sequence associated to the valuation. Our proof of this result in general dimension follows, as an application of the existence of limits for graded families of m_R-primary ideals, from the fact that $\varphi(n) = \ell_R(R/I_n)$, where $I_n = \{f \in m_R \mid \nu(f) \geq n\}$ [Cutkosky et al. 2010a; Cutkosky and Teissier 2010]. It is shown in [Cutkosky et al. 2010a] that the limits $\lim_{n\to\infty} \varphi(n)/n^2$ that can be attained on a regular local ring of dimension 2 are the real numbers β with $0 \leq \beta < \frac{1}{2}$.

2. Notation

Let m_R denote the maximal ideal of a local ring R. $Q(R)$ will denote the quotient field of a domain R and $\ell_R(N)$ the length of an R-module N. $\mathbb{Z}_+$ denotes the

positive integers and $\mathbb{N}$ the nonnegative integers. Suppose that $x \in \mathbb{R}$. Then $\lceil x \rceil$ is the smallest integer n such that $x \le n$ and $\lfloor x \rfloor$ the largest integer n such that $n \le x$.

We recall some notation on multiplicity from [Zariski and Samuel 1960, Chapter VIII, §10; Serre 1965, p. V-2; Bruns and Herzog 1993, § 4.6]. Suppose that (R, m_R) is a (Noetherian) local ring, N is a finitely generated R-module with $r = \dim N$ and a is an ideal of definition of R. Then

$$e_a(N) = \lim_{k\to\infty} \frac{\ell_R(N/a^k N)}{k^r/r!}.$$

We write $e(a) = e_a(R)$.

If $s \ge r = \dim N$, then we define

$$e_s(a, N) = \begin{cases} e_a(N) & \text{if } \dim N = s, \\ 0 & \text{if } \dim N < s. \end{cases}$$

A local ring is analytically unramified if its completion is reduced. In particular, a reduced excellent local ring is analytically unramified.

3. Semigroups and cones

Suppose that $\Gamma \subset \mathbb{N}^{d+1}$ is a semigroup. Set

$$\Sigma = \Sigma(\Gamma) = \text{closed convex cone}(\Gamma) \subset \mathbb{R}^{d+1},$$
$$\Delta = \Delta(\Gamma) = \Sigma \cap (\mathbb{R}^d \times \{1\}).$$

For $m \in \mathbb{N}$, put

$$\Gamma_m = \Gamma \cap (\mathbb{N}^d \times \{m\}),$$

which can be viewed as a subset of $\mathbb{N}^d$. Consider the following three conditions on Γ:

(1) $\Gamma_0 = \{0\}$.

(2) There exist finitely many vectors $(v_i, 1)$ spanning a semigroup $B \subset \mathbb{N}^{d+1}$ such that $\Gamma \subset B$.

Let $G(\Gamma)$ is the subgroup of $\mathbb{Z}^{d+1}$ generated by Γ.

(3) $G(\Gamma) = \mathbb{Z}^{d+1}$.

We will use the convention that $\{e_i\}$ is the standard basis of $\mathbb{Z}^{d+1}$.

Assuming the boundedness condition (2), condition (1) simply states that 0 is in the semigroup Γ.

Theorem 3.1 [Okounkov 2003, §3; Lazarsfeld and Mustaţă 2009, Proposition 2.1]. *Suppose that Γ satisfies (1)–(3). Then*

$$\lim_{m\to\infty} \frac{\#\Gamma_m}{m^d} = \operatorname{vol}(\Delta(\Gamma)).$$

Recently, it has been shown that limits exist under much weaker conditions by Kaveh and Khovanskii [2012].

Theorem 3.2 [Lazarsfeld and Mustaţă 2009, Proposition 3.1]. *Suppose that* Γ *satisfies (1)–(3). Fix* $\varepsilon > 0$. *Then there is an integer* $p_0 = p_0(\varepsilon)$ *such that if* $p \geq p_0$, *then the limit*

$$\lim_{k\to\infty} \frac{\#(k\Gamma_p)}{k^d p^d} \geq \mathrm{vol}(\Delta(\Gamma)) - \varepsilon$$

exists, where

$$k\Gamma_p = \{x_1 + \cdots + x_k \mid x_1, \ldots, x_k \in \Gamma_p\}.$$

4. An asymptotic theorem on lengths

Definition 4.1. A graded family of ideals $\{I_i\}$ in a ring R is a family of ideals indexed by the natural numbers such that $I_0 = R$ and $I_i I_j \subset I_{i+j}$ for all i and j. If R is a local ring and I_i is m_R-primary for $i > 0$, then we will say that $\{I_i\}$ is a graded family of m_R-primary ideals.

In this section, we prove the following theorem:

Theorem 4.2. *Suppose that* R *is an analytically irreducible local domain of dimension* $d > 0$ *and* $\{I_n\}$ *is a graded family of ideals in* R *such that*

$$\text{there exists } c \in \mathbb{Z}_+ \text{ such that } m_R^c \subset I_1. \tag{4}$$

Suppose that there exists a regular local ring S *such that* S *is essentially of finite type and birational over* R (R *and* S *have the same quotient field*) *and the residue field map* $R/m_R \to S/m_S$ *is an isomorphism. Then*

$$\lim_{i\to\infty} \frac{\ell_R(R/I_i)}{i^d}$$

exists.

We remark that the assumption $m_R^c \subset I_1$ implies that either I_n is m_R-primary for all positive n or there exists $n_0 > 1$ such that $I_{n_0} = R$. In this case, $m_R^{cn_0} \subset I_n$ for all $n \geq n_0$, so $\ell_R(R/I_i)$ is actually bounded.

Let assumptions be as in Theorem 4.2. Let $y_1, \ldots, y_d$ be a regular system of parameters in S. Let $\lambda_1, \ldots, \lambda_d$ be rationally independent real numbers such that

$$\lambda_i \geq 1 \quad \text{for all } i. \tag{5}$$

We define a valuation ν on $Q(R)$ that dominates S by prescribing

$$\nu(y_1^{a_1} \cdots y_d^{a_d}) = a_1\lambda_1 + \cdots + a_d\lambda_d$$

for $a_1, \ldots, a_d \in \mathbb{Z}_+$ and $\nu(\gamma) = 0$ if $\gamma \in S$ has nonzero residue in S/m_S.

Let C be a coefficient set of S. Since S is a regular local ring, for $r \in \mathbb{Z}_+$ and $f \in S$, there is a unique expression

$$f = \sum s_{i_1,\dots,i_d} y_1^{i_1} \cdots y_d^{i_d} + g_r$$

with $g_r \in m_S^r$, $s_{i_1,\dots,i_d} \in S$ and $i_1 + \cdots + i_d < r$ for all $i_1, \dots, i_d$ appearing in the sum. Take r so large that $r > i_1\lambda_1 + \cdots + i_d\lambda_d$ for some term with $s_{i_1,\dots,i_d} \neq 0$. Then define

$$\nu(f) = \min\{i_1\lambda_1 + \cdots + i_d\lambda_d \mid s_{i_1,\dots,i_d} \neq 0\}. \tag{6}$$

This definition is well-defined, and we calculate that $\nu(f+g) \geq \min\{\nu(f), \nu(g)\}$ and $\nu(fg) = \nu(f) + \nu(g)$ (by the uniqueness of the expansion (6)) for all $0 \neq f, g \in S$. Thus, ν is a valuation. Let V_ν be the valuation ring of ν (in $Q(R)$). The value group of V_ν is the (nondiscrete) ordered subgroup $\mathbb{Z}\lambda_1 + \cdots + \mathbb{Z}\lambda_d$ of $\mathbb{R}$. Since there is a unique monomial giving the minimum in (6), we have that the residue field of V_ν is $S/m_S = R/m_R$.

For $\lambda \in \mathbb{R}$, define ideals K_λ and K_{λ^+} in V_ν by

$$K_\lambda = \{f \in Q(R) \mid \nu(f) \geq \lambda\},$$
$$K_{\lambda^+} = \{f \in Q(R) \mid \nu(f) > \lambda\}.$$

We follow the usual convention that $\nu(0) = \infty$ is larger than any element of $\mathbb{R}$.

Lemma 4.3. *There exists $\alpha \in \mathbb{Z}_+$ such that $K_{\alpha n} \cap R \subset m_R^n$ for all $n \in \mathbb{N}$.*

Proof. Let $\rho = \lceil \max\{\lambda_1, \dots, \lambda_d\} \rceil \in \mathbb{Z}_+$. Suppose that $\lambda \in \mathbb{R}_+$. K_λ is generated by the monomials $y_1^{i_1} \cdots y_d^{i_d}$ such that $i_1\lambda_1 + \cdots + i_d\lambda_d \geq \lambda$, which implies that

$$\frac{\lambda}{\rho} \leq i_1 + \cdots + i_d$$

so that

$$K_\lambda \cap S \subset m_S^{\lceil \lambda/\rho \rceil}. \tag{7}$$

We now establish the following equation: there exists $a \in \mathbb{Z}_+$ such that

$$m_S^{a\ell} \cap R \subset m_R^\ell \tag{8}$$

for all $\ell \in \mathbb{N}$.

In the case when R is excellent, this is immediate from the linear Zariski subspace theorem [Hübl 2001, Theorem 1].

We now give a proof of (8) that was provided by the referee, which is valid without assuming that R is excellent. Let ω be the m_S-adic valuation. Let ν_i be the Rees valuations of m_R. The ν_i extend uniquely to the Rees valuations of $m_{\hat{R}}$. By the version of Rees [1989] of Izumi's theorem, the topologies defined on R by ω

and the ν_i are linearly equivalent. Let $\bar{\nu}_{m_R}$ be the reduced order of m_R. By the Rees valuation theorem (recalled in [Rees 1989]),

$$\bar{\nu}_{m_R}(x) = \min_i \left\{ \frac{\nu_i(x)}{\nu_i(m_R)} \right\}$$

for all $x \in R$, so the topology defined by ω on R is linearly equivalent to the topology defined by $\bar{\nu}_{m_R}$. The $\bar{\nu}_{m_R}$ topology is linearly equivalent to the m_R-topology by [Rees 1956] since R is analytically unramified. Thus, (8) is established.

Let $\alpha = \rho a$, where ρ is the constant of (7) and a is the constant of (8).

$$K_{\alpha n} \cap S = K_{\rho a n} \cap S \subset m_S^{an}$$

by (7), and thus,

$$K_{\alpha n} \cap R \subset m_S^{an} \cap R \subset m_R^n$$

by (8). □

Remark 4.4. The conclusions of Lemma 4.3 fail if R is not analytically irreducible as can be seen from the example

$$R = (k[x, y]/y^2 - x^2(x+1))_{(x,y)} \to S = k[s]_{(s)},$$

where $s = y/x - 1$.

For $0 \neq f \in R$, define

$$\varphi(f) = (n_1, \dots, n_d) \in \mathbb{N}^d$$

if $\nu(f) = n_1\lambda_1 + \cdots + n_d\lambda_d$.

Lemma 4.5. *Suppose that $I \subset R$ is an ideal and $\lambda \in \mathbb{R}_+$. Then there are isomorphisms of R/m_R-modules*

$$K_\lambda \cap I / K_{\lambda^+} \cap I \cong \begin{cases} k & \text{if there exists } f \in I \text{ such that } \nu(f) = \lambda, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Suppose that $f, g \in K_\lambda \cap I$ are such that $\nu(f) = \nu(g) = \lambda$. Then $\nu(f/g) = 0$. Let $\bar{\alpha}$ be the class of f/g in $V_\nu/m_\nu \cong R/m_R$. Let $\alpha \in R$ be a lift of $\bar{\alpha}$ to R. Then $\nu(f - \alpha g) > \lambda$, and the class of f in $K_\lambda \cap I / K_{\lambda^+} \cap I$ is equal to $\bar{\alpha}$ times the class of g in $K_\lambda \cap I / K_{\lambda^+} \cap I$. □

Suppose that $I \subset R$ is an ideal and $K_\beta \cap R \subset I$ for some $\beta \in \mathbb{R}_+$. Then

$$\begin{aligned} \ell_R(R/I) &= \ell_R(R/K_\beta \cap R) - \ell_R(I/K_\beta \cap R) \\ &= \dim_k \Big(\bigoplus_{\lambda<\beta} K_\lambda \cap R / K_{\lambda^+} \cap R \Big) - \dim_k \Big(\bigoplus_{\lambda<\beta} K_\lambda \cap I / K_{\lambda^+} \cap R \Big) \\ &= \#\{(n_1, \dots, n_d) \in \varphi(R) \mid n_1\lambda_1 + \cdots + n_d\lambda_d < \beta\} \\ &\qquad - \#\{(n_1, \dots, n_d) \in \varphi(I) \mid n_1\lambda_1 + \cdots + n_d\lambda_d < \beta\}. \end{aligned} \tag{9}$$

Let $\beta = \alpha c \in \mathbb{Z}_+$, where c is the constant of (4) and α is the constant of Lemma 4.3 so that, for all $i \in \mathbb{Z}_+$,

$$K_{\beta i} \cap R = K_{\alpha c i} \cap R \subset m_R^{ic} \subset I_i. \tag{10}$$

We have from (9) that

$$\begin{aligned}\ell_R(R/I_i) = {} & \#\{(n_1, \dots, n_d) \in \varphi(R) \mid n_1\lambda_1 + \cdots + n_d\lambda_d < \beta i\} \\ & - \#\{(n_1, \dots, n_d) \in \varphi(I_i) \mid n_1\lambda_1 + \cdots + n_d\lambda_d < \beta i\}. \end{aligned} \tag{11}$$

Now $(n_1, \dots, n_d) \in \varphi(R)$ and $n_1 + \cdots + n_d \geq \beta i$ imply $n_1\lambda_1 + \cdots + n_d\lambda_d \geq \beta i$ by (5) so that $(n_1, \dots, n_d) \in \varphi(I_i)$ by (10). Thus,

$$\begin{aligned}\ell_R(R/I_i) = {} & \#\{(n_1, \dots, n_d) \in \varphi(R) \mid n_1 + \cdots + n_d \leq \beta i\} \\ & - \#\{(n_1, \dots, n_d) \in \varphi(I_i) \mid n_1 + \cdots + n_d \leq \beta i\}. \end{aligned} \tag{12}$$

Let $\Gamma \subset \mathbb{N}^{d+1}$ be the semigroup

$$\Gamma = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \varphi(I_i) \text{ and } n_1 + \cdots + n_d \leq \beta i\}.$$

$I_0 = R$ (and $\nu(1) = 0$) implies (1) holds. The semigroup

$$B = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \mathbb{N}^d \text{ and } n_1 + \cdots + n_d \leq \beta i\}$$

is generated by $B \cap (\mathbb{N}^d \times \{1\})$ and contains Γ, so (2) holds.

Write $y_i = f_i/g_i$ with $f_i, g_i \in R$ for $1 \leq i \leq d$. Let $0 \neq h \in I_1$. Then $hf_i, hg_i \in I_1$. There exists $c' \in \mathbb{Z}_+$ such that $c' \geq c$ and $hf_i, hg_i \notin m_R^{c'}$ for $1 \leq i \leq d$. We may replace c with c' in (4). Then $\varphi(hf_i), \varphi(hg_i) \in \Gamma_1 = \Gamma \cap (\mathbb{N}^d \times \{1\})$ for $1 \leq i \leq d$ since hf_i and hg_i all have values $n_1\lambda_1 + \cdots + n_d\lambda_d < \beta i$ so that $n_1 + \cdots + n_d < \beta i$. We have that $\varphi(hf_i) - \varphi(hg_i) = e_i$ for $1 \leq i \leq d$. Thus,

$$(e_i, 0) = (\varphi(hf_i), 1) - (\varphi(hg_i), 1) \in G(\Gamma)$$

for $1 \leq i \leq d$. Since $(\varphi(hf_i), 1) \in G(\Gamma)$, we have that $(0, 1) \in G(\Gamma)$, so $G(\Gamma) = \mathbb{Z}^{d+1}$ and (3) holds. By Theorem 3.1,

$$\lim_{i \to \infty} \frac{\#\Gamma_i}{i^d} = \mathrm{vol}(\Delta(\Gamma)). \tag{13}$$

Let $\Gamma' \subset \mathbb{N}^{d+1}$ be the semigroup

$$\Gamma' = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \varphi(R) \text{ and } n_1 + \cdots + n_d \leq \beta i\}.$$

Our calculation for Γ shows that (1)–(3) hold for Γ'. By Theorem 3.1,

$$\lim_{i \to \infty} \frac{\#\Gamma'_i}{i^d} = \mathrm{vol}(\Delta(\Gamma')). \tag{14}$$

We obtain the conclusions of Theorem 4.2 from Equations (12), (13) and (14).

The following is an immediate consequence of Theorem 4.2, taking $S = R$:

Theorem 4.6. *Suppose that R is a regular local ring of dimension $d > 0$ and $\{I_n\}$ is a graded family of m_R-primary ideals in R. Then the limit*

$$\lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d}$$

exists.

5. A theorem on asymptotic lengths in more general rings

Lemma 5.1. *Suppose that R is a d-dimensional reduced local ring and $\{I_n\}$ is a graded family of m_R-primary ideals in R. Let $p_1, \dots, p_s$ be the minimal primes of R, set $R_i = R/p_i$, and let S be the ring $S = \bigoplus_{i=1}^s R_i$. Then there exists $\alpha \in \mathbb{Z}_+$ such that for all $n \in \mathbb{Z}_+$,*

$$\left| \sum_{i=1}^s \ell_{R_i}(R_i/I_nR_i) - \ell_R(R/I_n) \right| \le \alpha n^{d-1}.$$

Proof. There exists $c \in \mathbb{Z}_+$ such that $m_R^c \subset I_1$. Since S is a finitely generated R-submodule of the total ring of fractions $T = \bigoplus_{i=1}^s Q(R_i)$ of R, there exists a nonzero divisor $x \in R$ such that $xS \subset R$.

The natural inclusion $R \to S$ induces exact sequences of R-modules

$$0 \to R \cap I_nS/I_n \to R/I_n \to S/I_nS \to N_n \to 0. \tag{15}$$

We also have exact sequences of R-modules

$$0 \to A_n \to R/I_n \xrightarrow{x} R/I_n \to M_n \to 0. \tag{16}$$

We have that $x(R \cap I_nS) \subset I_n$ and $A_n = I_n : x/I_n$ so that

$$\ell_R(R \cap I_nS/I_n) \le \ell_R(A_n). \tag{17}$$

Now $M_n \cong (R/x)/I_n(R/x)$, so

$$\ell_R(M_n) \le \ell_R((R/x)/m_R^{nc}(R/x)) \le \beta(nc)^{d-1}$$

for some β, computed from the Hilbert–Samuel polynomial of R/x and the finitely many values of the Hilbert–Samuel function of R/x that do not agree with this polynomial. Thus,

$$\ell_R(A_n) = \ell(M_n) \le \beta c^{d-1} n^{d-1} \tag{18}$$

by (16).

Since $xS \subset R$, we have that

$$N_n \cong (S/R + I_nS) = S/(R + I_nS + xS).$$

Thus,

$$\ell_R(N_n) \le \ell_R((S/xS)/m_R^{nc}(S/xS)) \le \gamma(nc)^{d-1} \tag{19}$$

for some γ, computed from the Hilbert–Samuel polynomial of the semilocal ring S/x with respect to the ideal of definition $m_R(S/xS)$. Thus,

$$|\ell_R(R/I_n) - \ell_R(S/I_nS)| \le \max\{\beta, \gamma\}c^{d-1}n^{d-1}.$$

The lemma now follows since

$$\ell_R(S/I_nS) = \sum \ell_{R_i}(R_i/I_nR_i). \qquad \square$$

Theorem 5.2. *Suppose that R is an analytically unramified local ring with algebraically closed residue field. Let $d > 0$ be the dimension of R. Suppose that $\{I_n\}$ is a graded family of m_R-primary ideals in R. Then*

$$\lim_{i\to\infty} \frac{\ell_R(R/I_i)}{i^d}$$

exists.

Proof. Let $\hat{R}$ be the m_R-adic completion of R, which is reduced and excellent. Since the I_n are m_R-primary, we have that $R/I_n \cong \hat{R}/I_n\hat{R}$ and $\ell_R(R/I_n) = \ell_{\hat{R}}(\hat{R}/I_n\hat{R})$ for all n. Let $\{q_1, \dots, q_s\}$ be the minimal primes of $\hat{R}$. By Lemma 5.1, we reduce to proving the theorem for the families of ideals $\{I_n\hat{R}/q_i\}$ in $\hat{R}/q_i$ for $1 \le i \le s$. We may thus assume that R is a complete domain. Let $\pi : X \to \operatorname{Spec}(R)$ be the normalization of the blow-up of m_R. X is of finite type over R since R is excellent. Since $\pi^{-1}(m_R)$ has codimension 1 in X and X is normal, there exists a closed point $x \in X$ such that the local ring $\mathcal{O}_{X,x}$ is a regular local ring. Let S be this local ring. $S/m_S = R/m_R$ since S/m_S is finite over R/m_R, which is an algebraically closed field. $\square$

Lemma 5.3. *Suppose that R is a Noetherian local domain that contains a field k. Suppose that k' is a finite separable field extension of k such that $k \subset R/m_R \subset k'$. Let $S = R \otimes_k k'$. Then S is a reduced Noetherian semilocal ring. Let $p_1, \dots, p_r$ be the maximal ideals of S. Then $m_RS = p_1 \cap \cdots \cap p_r$.*

Proof. Let K be the quotient field of R. Then $K \otimes_k k'$ is reduced [Zariski and Samuel 1958, Theorem 39, p. 195]. Since k' is flat over k, we have an inclusion $R \otimes_k k' \subset K \otimes_k k'$, so $S = R \otimes_k k'$ is reduced. $S/m_RS \cong (R/m_R) \otimes_k k'$ is also reduced by [Zariski and Samuel 1958, Theorem 39]. Thus, $m_RS = p_1 \cap \cdots \cap p_r$. $\square$

Remark 5.4. In the case that R is a regular local ring, we have that $S = R \otimes_k k'$ is a regular ring.

Proof. Since R is a regular local ring, m_R is generated by $d = \dim R$ elements. For $1 \le i \le r$, we thus have that $p_iS_{p_i} = m_RS_{p_i}$ is generated by $d = \dim R = \dim S_{p_i}$ elements. Thus, S_{p_i} is a regular local ring. $\square$

Remark 5.5. If k' is Galois over k, then $S/p_i \cong k'$ for $1 \le i \le r$.

Proof. Let $\tilde{k} = R/m_R$. By our assumption, $\tilde{k}$ is a finite separable extension of k. Thus, $\tilde{k} = k[\alpha]$ for some $\alpha \in k'$. Let $f(x) \in k[x]$ be the minimal polynomial of α. Since k' is a normal extension of k containing α, $f(x)$ splits into linear factors in $k'[x]$. Thus,

$$\bigoplus_{i=1}^{r} R/p_i \cong S/m_R S \cong \tilde{k} \otimes_k k' \cong k'[x]/(f(x)) \cong (k')^r. \qquad \square$$

Remark 5.6. If R is complete in the m_R-adic topology, then $R \otimes_k k'$ is complete in the $m_R R \otimes_k k'$-adic topology [Zariski and Samuel 1960, Theorem 16, p. 277]. If $p_1, \dots, p_r$ are the maximal ideals of $R \otimes_k k'$, then $R \otimes_k k' \cong \bigoplus_{i=1}^{r} (R \otimes_k k')_{p_i}$ [Matsumura 1986, Theorem 8.15]. Thus, each $(R \otimes_k k')_{p_i}$ is a complete local ring.

Lemma 5.7. *Let assumptions and notation be as in Lemma 5.3, and suppose that I is an m_R-primary ideal in R. Then*

$$[k' : k]\ell_R(R/I) = \sum_{i=1}^{r} [S/p_i : R/m_R]\ell_{S_{p_i}}((S/IS)_{p_i}).$$

Proof. We have

$$\dim_k R/I = [R/m_R : k]\ell_R(R/I),$$
$$\dim_k S/IS = \dim_k (R/I) \otimes_k k' = [k' : k] \dim_k (R/I).$$

S/IS is an Artin local ring so that $S/IS \cong \bigoplus_{i=1}^{r} (S/IS)_{p_i}$. Thus,

$$\dim_k (S/IS) = \sum_{i=1}^{r} [S/p_i : k]\ell_{S_{p_i}}((S/IS)_{p_i}). \qquad \square$$

We will need the following definition. A commutative ring A containing a field k is said to be geometrically irreducible over k if $A \otimes_k k'$ has a unique minimal prime for all finite extensions k' of k.

Theorem 5.8. *Suppose that R is an analytically unramified equicharacteristic local ring with perfect residue field. Let $d > 0$ be the dimension of R. Suppose that $\{I_n\}$ is a graded family of m_R-primary ideals in R. Then*

$$\lim_{i \to \infty} \frac{\ell_R(R/I_i)}{i^d}$$

exists.

Proof. There exists $c \in \mathbb{Z}_+$ such that $m_R^c \subset I_1$. Let $\hat{R}$ be the m_R-adic completion of R. Since the I_n are m_R-primary, we have that $R/I_n \cong \hat{R}/I_n\hat{R}$ and $\ell_R(R/I_n) = \ell_{\hat{R}}(\hat{R}/I_n\hat{R})$ for all n. $\hat{R}$ is reduced since R is analytically unramified.

Let $\{q_1, \dots, q_s\}$ be the minimal primes of $\hat{R}$. By Lemma 5.1, we reduce to proving the theorem for the families of ideals $\{I_n\hat{R}/q_i\}$ in $\hat{R}/q_i$ for $1 \le i \le s$. In the case of a minimal prime q_i of R such that $\dim R/q_i < d$, the limits

$$\lim_{n\to\infty} \frac{\ell_R(R_i/I_nR_i)}{n^d}$$

are all zero since $\ell_R(R_i/I_nR_i) \le \ell_R(R_i/m_R^{nc}R_i)$ for all n.

We may thus assume that R is a complete domain. $\hat{R}$ contains a coefficient field $k \cong R/m_R$ by the Cohen structure theorem as R is complete and equicharacteristic. Let k' be the separable closure of k in $Q(R)$, and let $\bar{R}$ be the integral closure of R in $Q(R)$. We have that $k' \subset \bar{R}$. $\bar{R}$ is a finitely generated R-module since R is excellent. Let $n \subset \bar{R}$ be a maximal ideal lying over m_R. Then the residue field extension $R/m_R \to \bar{R}/n$ is finite. Since $k' \subset \bar{R}/n$, we have that k' is a finite extension of k. By [Grothendieck 1965, Corollary 4.5.11], there exists a finite extension L of k (which can be taken to be Galois over k) such that if $q_1, \dots, q_r$ are the minimal primes of $R \otimes_k L$, then each ring $R \otimes_k L/q_i$ is geometrically irreducible over L.

$R \otimes_k L$ is a reduced semilocal ring by Lemma 5.3, and by Remark 5.5, the residue field of all maximal ideals of $R \otimes_k L$ is L, which is a perfect field. By Remark 5.6 and Lemmas 5.1 and 5.7, we reduce to the case where R is a complete local domain with perfect coefficient field k such that R is geometrically irreducible over k. Let $\pi : X \to \operatorname{Spec}(R)$ be the normalization of the blow-up of m_R. Since R is excellent, π is projective and birational. Since $m_R\mathcal{O}_X$ is locally principal, $\pi^{-1}(m_R)$ has codimension 1 in X. Since X is normal, it is regular in codimension 1, so there exists a closed point $q \in X$ such that $\pi(q) = m_R$ and $S = \mathcal{O}_{X,q}$ is a regular local ring. Let $k' = S/m_S$. Then k' is finite over k and is thus a separable extension of the perfect field k.

Let k'' be a finite Galois extension of k containing k'. Let $R' = R \otimes_k k''$. R' is a local domain with residue field k''. R' is complete by Remark 5.6. $S \otimes_k k''$ is regular and semilocal by Remark 5.4. Let $p \in S \otimes_k k''$ be a maximal ideal. Let $S' = (S \otimes_k k'')_p$. There exist $f_0, \dots, f_t \in Q(R)$ such that S is a localization of $R[f_1/f_0, \dots, f_t/f_0]$ at a maximal ideal that necessarily contracts to m_R. Thus, S' is essentially of finite type and birational over R' since we can regard $f_0, \dots, f_t \in R'$. Since S' is a regular local ring and $k'' = S'/m_{S'} = R'/m_{R'}$ by Remark 5.5, we have that Theorem 5.8 follows from Lemma 5.7 and Theorem 4.2. □

6. Some applications to asymptotic multiplicities

Theorem 6.1. *Suppose that R is a local ring of dimension $d > 0$ such that one of the following holds*:

(1) *R is regular or*

(2) *R is analytically irreducible with algebraically closed residue field or*

(3) *R is normal, excellent and equicharacteristic with perfect residue field.*

Suppose that $\{I_i\}$ and $\{J_i\}$ are graded families of nonzero ideals in R. Further suppose that $I_i \subset J_i$ for all i and there exists $c \in \mathbb{Z}_+$ such that

$$m_R^{ci} \cap I_i = m_R^{ci} \cap J_i \tag{20}$$

for all i. Then the limit

$$\lim_{i\to\infty} \frac{\ell_R(J_i/I_i)}{i^d}$$

exists.

Remark 6.2. An analytic local domain R satisfies the hypotheses of Theorem 6.1(2). The fact that R is analytically irreducible ($\hat{R}$ is a domain) follows from [Grothendieck 1965, Corollary 18.9.2].

Proof of Theorem 6.1. We will apply the method of Theorem 4.2. When R is regular, we take $S = R$, and in case (2), we construct S by the argument of the proof of Theorem 5.2. We will consider case (3) at the end of the proof.

Let ν be the valuation of $Q(R)$ constructed from S in the proof of Theorem 4.2 with associated valuation ideals K_λ in the valuation ring V_ν of ν.

Apply Lemma 4.3 if R is not regular to find $\alpha \in \mathbb{Z}_+$ such that

$$K_{\alpha n} \cap R \subset m_R^n$$

for all $n \in \mathbb{Z}_+$. When R is regular so that $R = S$, the existence of such an α follows directly from (7). We will use the function $\varphi : R \setminus \{0\} \to \mathbb{N}^{d+1}$ of the proof of Theorem 4.2. We have that

$$K_{\alpha cn} \cap I_n = K_{\alpha cn} \cap J_n$$

for all n. Thus,

$$\ell_R(J_n/I_n) = \ell_R(J_n/K_{\alpha cn} \cap J_n) - \ell_R(I_n/K_{\alpha cn} \cap I_n) \tag{21}$$

for all n. Let $\beta = \alpha c$ and

$$\Gamma(J_*) = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \varphi(J_i) \text{ and } n_1 + \cdots + n_d \le \beta i\},$$
$$\Gamma(I_*) = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \varphi(I_i) \text{ and } n_1 + \cdots + n_d \le \beta i\}.$$

We have that

$$\ell_R(J_n/I_n) = \#\Gamma(J_*)_n - \#\Gamma(I_*)_n \tag{22}$$

as explained in the proof of Theorem 4.2. As in the proof of Theorem 4.2, we have that $\Gamma(J_*)$ and $\Gamma(I_*)$ satisfy the conditions (1)–(3). Thus,

$$\lim_{n\to\infty}\frac{\#\Gamma(J_*)_n}{n^d}=\mathrm{vol}(\Delta(\Gamma(J_*)))\quad\text{and}\quad\lim_{n\to\infty}\frac{\#\Gamma(I_*)_n}{n^d}=\mathrm{vol}(\Delta(\Gamma(I_*)))$$

by Theorem 3.1. The theorem (in cases (1) or (2)) now follows from (22).

Now suppose that R satisfies the assumptions of case (3). Then the m_R-adic completion $\hat{R}$ satisfies the assumptions of case (3).

Suppose that R satisfies the assumptions of case (3) and R is m_R-adically complete. Let k be a coefficient field of R. The algebraic closure of k in $Q(R)$ is contained in R, so it is contained in $R/m_R=k$. Thus, k is algebraically closed in $Q(R)$. Suppose that k' is a finite Galois extension of k. $Q(R)\otimes_k k'$ is a field by [Zariski and Samuel 1958, Corollary 2, p. 198], and thus, $R'=R\otimes_k k'$ is a domain. R' is a local ring with residue field k' since $R'/m_R R'\cong R/m_R\otimes_k k'\cong k'$. R' is normal by [Grothendieck 1965, Corollary 6.14.2]. Thus, R' satisfies the assumptions of case (3).

Thus, in the reductions in the proof of Theorem 5.8 to Theorem 4.2, the only extensions that we need to consider are local homomorphisms $R\to R'$ that are either m_R-adic completion or a base extension by a Galois field extension. These extensions are all flat, and $m_R R'=m_{R'}$. Thus,

$$m_S^{nc}\cap I_nS=m_R^{nc}S\cap I_nS=(m_R^{nc}\cap I_n)S=(m_R^{nc}\cap J_n)S=m_R^{nc}S\cap J_nS=m_S^{nc}\cap J_nS$$

for all n. Thus, the condition (20) is preserved, so we reduce to the case (2) of this theorem and conclude that Theorem 6.1 is true in case (3). □

If R is a local ring and I is an ideal in R, then the saturation of I is

$$I^{\mathrm{sat}}=I:m_R^\infty=\bigcup_{k=1}^{\infty}I:m_R^k.$$

Corollary 6.3. *Suppose that R is a local ring of dimension $d>0$ such that one of the following holds*:

(1) *R is regular or*

(2) *R is analytically irreducible with algebraically closed residue field or*

(3) *R is normal, excellent and equicharacteristic with perfect residue field.*

Suppose that I is an ideal in R. Then the limit

$$\lim_{i\to\infty}\frac{\ell_R((I^i)^{\mathrm{sat}}/I^i)}{i^d}$$

exists.

Since $(I^n)^{\mathrm{sat}}/I^n \cong H^0_{m_R}(R/I^n)$, the epsilon multiplicity of Ulrich and Validashti [2011]

$$\varepsilon(I) = \limsup \frac{\ell_R(H^0_{m_R}(R/I^n))}{n^d/d!}$$

exists as a limit under the assumptions of Corollary 6.3.

Corollary 6.3 is proven for more general families of modules when R is a local domain that is essentially of finite type over a perfect field k such that R/m_R is algebraic over k in [Cutkosky 2011]. The limit in Corollary 6.3 can be irrational as shown in [Cutkosky et al. 2005].

Proof of Corollary 6.3. By [Swanson 1997, Theorem 3.4], there exists $c \in \mathbb{Z}_+$ such that each power I^n of I has an irredundant primary decomposition

$$I^n = q_1(n) \cap \cdots \cap q_s(n),$$

where $q_1(n)$ is m_R-primary and $m_R^{nc} \subset q_1(n)$ for all n. As $(I^n)^{\mathrm{sat}} = q_2(n) \cap \cdots \cap q_s(n)$, we have that

$$I^n \cap m_R^{nc} = m_R^{nc} \cap q_2(n) \cap \cdots \cap q_s(n) = m_R^{nc} \cap (I^n)^{\mathrm{sat}}$$

for all $n \in \mathbb{Z}_+$. Thus, the corollary follows from Theorem 6.1, taking $I_i = I^i$ and $J_i = (I^i)^{\mathrm{sat}}$. □

A stronger version of the previous corollary is true. The following corollary proves a formula proposed by Herzog et al. [2008, Introduction].

Suppose that R is a ring and I and J are ideals in R. Then the n-th symbolic power of I with respect to J is

$$I_n(J) = I^n : J^\infty = \bigcup_{i=1}^{\infty} I^n : J^i.$$

Corollary 6.4. *Suppose that R is a local domain of dimension d such that one of the following holds*:

(1) *R is regular or*

(2) *R is normal and excellent of equicharacteristic* 0 *or*

(3) *R is essentially of finite type over a field of characteristic* 0.

Suppose that I and J are ideals in R. Let s be the constant limit dimension of $I_n(J)/I^n$ for $n \gg 0$. Then

$$\lim_{n\to\infty} \frac{e_{m_R}(I_n(J)/I^n)}{n^{d-s}}$$

exists.

Proof. There exists a positive integer n_0 such that the set of associated primes of R/I^n stabilizes for $n \geq n_0$ by [Brodmann 1979]. Let $\{p_1, \dots, p_t\}$ be this set of associated primes. We thus have irredundant primary decompositions for $n \geq n_0$

$$I^n = q_1(n) \cap \cdots \cap q_t(n), \tag{23}$$

where $q_i(n)$ are p_i-primary.

We further have that

$$I^n : J^\infty = \bigcap_{J \not\subset p_i} q_i(n). \tag{24}$$

Thus, $\dim I_n(J)/I^n$ is constant for $n \geq n_0$. Let s be this limit dimension. The set

$$A = \left\{ p \in \bigcup_{n \geq n_0} \operatorname{Ass}(I_n(J)/I^n) \;\middle|\; n \geq n_0 \text{ and } \dim R/p = s \right\}$$

is a finite set. Moreover, every such prime is in $\operatorname{Ass}(I_n(J)/I^n)$ for all $n \geq n_0$. For $n \geq n_0$, we have by the additivity formula [Serre 1965, p. V-2; Bruns and Herzog 1993, Corollary 4.6.8, p. 189] that

$$e_{m_R}(I_n(J)/I^n) = \sum_p \ell_{R_p}((I_n(J)/I^n)_p) e(m_{R/p}),$$

where the sum is over the finite set of primes $p \in \operatorname{Spec}(R)$ such that $\dim R/p = s$. This sum is thus over the finite set A.

Suppose that $p \in A$ and $n \geq n_0$. Then

$$I_p^n = \bigcap q_i(n)_p,$$

where the intersection is over the $q_i(n)$ such that $p_i \subset p$, and

$$I_n(J) = \bigcap q_i(n)_p,$$

where the intersection is over the $q_i(n)$ such that $J \not\subset p_i$ and $p_i \subset p$. Thus, there exists an index i_0 such that $p_{i_0} = p$ and

$$I_p^n = q_{i_0}(n)_p \cap I_n(J)_p.$$

By (23),

$$(I_p^n)^{\mathrm{sat}} = I_n(J)_p$$

for $n \geq n_0$. Since R_p satisfies one of the cases (1) or (3) of Theorem 6.1 or the conditions of [Cutkosky 2011, Corollary 1.5] and $\dim R_p = d - s$ (as R is universally catenary), the limit

$$\lim_{n \to \infty} \frac{\ell_R((I_n(J)/I_n)_p)}{n^{d-s}}$$

exists. $\square$

Theorem 6.5. *Suppose that R is a d-dimensional local ring such that either*

(1) *R is regular or*

(2) *R is analytically unramified and equicharacteristic with perfect residue field.*

Suppose that $\{I_i\}$ is a graded family of m_R-primary ideals in R. Then

$$\lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d/d!} = \lim_{p\to\infty} \frac{e(I_p)}{p^d}.$$

Here $e(I_p)$ is the multiplicity

$$e(I_p) = e_{I_p}(R) = \lim_{k\to\infty} \frac{\ell_R(R/I_p^k)}{k^d/d!}.$$

Theorem 6.5 is proven for valuation ideals associated to an Abhyankar valuation in a regular local ring that is essentially of finite type over a field in [Ein et al. 2003], for general families of m_R-primary ideals when R is a regular local ring containing a field in [Mustaţă 2002] and when R is a local domain that is essentially of finite type over an algebraically closed field k with $R/m_R = k$ in [Lazarsfeld and Mustaţă 2009, Theorem 3.8].

Proof of Theorem 6.5. There exists $c \in \mathbb{Z}_+$ such that $m_R^c \subset I_1$.

We first prove the theorem when R satisfies the assumptions of Theorem 4.2. Let ν be the valuation of $Q(R)$ constructed from S in the proof of Theorem 4.2 with associated valuation ideals K_λ in the valuation ring V_ν of ν.

Apply Lemma 4.3 if R is not regular to find $\alpha \in \mathbb{Z}_+$ such that

$$K_{\alpha n} \cap R \subset m_R^n$$

for all $n \in \mathbb{N}$. When R is regular so that $R = S$, the existence of such an α follows directly from (7). We will use the function $\varphi : R \setminus \{0\} \to \mathbb{N}^{d+1}$ of the proof of Theorem 4.2.

We have that

$$K_{\alpha cn} \cap R \subset m_R^{cn} \subset I_n$$

for all n.

Let

$$\Gamma(I_*) = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \varphi(I_i) \text{ and } n_1 + \cdots + n_d \le \alpha c i\},$$
$$\Gamma(R) = \{(n_1, \dots, n_d, i) \mid (n_1, \dots, n_d) \in \varphi(R) \text{ and } n_1 + \cdots + n_d \le \alpha c i\}.$$

As in the proof of Theorem 4.2, $\Gamma(I_*)$ and $\Gamma(R)$ satisfy the conditions (1)–(3). For fixed $p \in \mathbb{Z}_+$, let

$$\Gamma(I_*)(p) = \{(n_1, \dots, n_d, kp) \mid (n_1, \dots, n_d) \in \varphi(I_p^k) \text{ and } n_1 + \cdots + n_d \le \alpha c k p\}.$$

We have inclusions of semigroups

$$k\Gamma(I_*)_p \subset \Gamma(I_*)(p)_{kp} \subset \Gamma(I_*)_{kp}$$

for all p and k.

By Theorem 3.2, given $\varepsilon > 0$, there exists p_0 such that $p \geq p_0$ implies

$$\mathrm{vol}(\Delta(\Gamma(I_*))) - \varepsilon \leq \lim_{k\to\infty} \frac{\#k\Gamma(I_*)_p}{k^d p^d}.$$

Thus,

$$\mathrm{vol}(\Delta(\Gamma(I_*))) - \varepsilon \leq \lim_{k\to\infty} \frac{\#\Gamma(I_*)(p)_{kp}}{k^d p^d} \leq \mathrm{vol}(\Delta(\Gamma(I_*))).$$

Again by Theorem 3.2, we can choose p_0 sufficiently large so that we also have

$$\mathrm{vol}(\Delta(\Gamma(R))) - \varepsilon \leq \lim_{k\to\infty} \frac{\#\Gamma(R)_{kp}}{k^d p^d} \leq \mathrm{vol}(\Delta(\Gamma)).$$

Now

$$\ell_R(R/I_p^k) = \#\Gamma(R)_{pk} - \#\Gamma(I_*)(p)_{kp},$$
$$\ell_R(R/I_n) = \#\Gamma(R)_n - \#\Gamma(I_*)_n.$$

By Theorem 3.1,

$$\lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d} = \mathrm{vol}(\Delta(\Gamma(R))) - \mathrm{vol}(\Delta(\Gamma(I_*))).$$

Thus,

$$\lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d} - \varepsilon \leq \lim_{k\to\infty} \frac{\ell_R(R/I_p^k)}{k^d p^d} = \frac{e(I_p)}{d!\, p^d} \leq \lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d} + \varepsilon.$$

Taking the limit as $p \to \infty$, we obtain the conclusions of the theorem.

Now assume that R is general, satisfying the assumptions of the theorem. We reduce to the above case by a series of reductions, first taking the completion of R, then modding out by minimal primes and taking a base extension by a finite Galois extension.

The proof thus reduces to showing that

$$\lim_{p\to\infty} \frac{e_d(I_p, R)}{p^d} = \lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d/d!}$$

in each of the following cases:

(a) $$\lim_{p\to\infty} \frac{e_d(I_p\hat{R}, \hat{R})}{p^d} = \lim_{n\to\infty} \frac{\ell_{\hat{R}}(\hat{R}/I_n\hat{R})}{n^d/d!}.$$

(b) Suppose that the minimal primes of (the reduced ring) R are $\{q_1, \dots, q_s\}$. Let $R_i = R/q_i$, and suppose that

$$\lim_{p\to\infty} \frac{e_d(I_p R_i, R_i)}{p^d} = \lim_{n\to\infty} \frac{\ell_{R_i}(R_i/I_n R_i)}{n^d/d!}$$

for all i.

(c) Suppose that $k \subset R$ is a field and k' is a finite Galois extension of k containing R/m_R. Let $\{p_1, \dots, p_r\}$ be the maximal ideals of the semilocal ring $S = R \otimes_k k'$. Suppose that

$$\lim_{p\to\infty} \frac{e_d(I_p S_{p_i}, S_{p_i})}{p^d} = \lim_{n\to\infty} \frac{\ell_{S_{p_i}}(S_{p_i}/I_n S_{p_i})}{n^d/d!}$$

for all i.

Recall that

$$\frac{e_d(I_p, R)}{d!} = \lim_{k\to\infty} \frac{\ell_R(R/I_p^k)}{k^d}.$$

Case (a) follows since

$$\ell_R(R/I_p^k) = \ell_{\hat{R}}(\hat{R}/I_p^k \hat{R})$$

for all p and k.

In case (b), we have that

$$\frac{e_d(I_p, R)}{p^d} = \sum_{i=1}^{s} \frac{e_d(I_p R_i, R_i)}{p^d}$$

by the additivity formula [Serre 1965, §V-3; Bruns and Herzog 1993, Corollary 4.6.8, p. 189] or directly from Lemma 5.1. Case (b) thus follows from the fact that

$$\lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d} = \sum_{i=1}^{s} \lim_{k\to\infty} \frac{\ell_R(R_i/I_n R_i)}{n^d}$$

by Lemma 5.1.

In case (c), we have that k' is Galois over k so that $S/p_i \cong k'$ for all i by Remark 5.5. Thus, Lemma 5.7 becomes

$$\ell_R(R/I_p^k) = \sum_{i=1}^{r} \ell_{S_{p_i}}(S_{p_i}/I_p^k S_{p_i})$$

for all p and k, from which this case follows. □

Suppose that R is a Noetherian ring and $\{I_i\}$ is a graded family of ideals in R. Let

$$s = s(I_*) = \limsup \dim R/I_i.$$

Let $i_0 \in \mathbb{Z}_+$ be the smallest integer such that

$$\dim R/I_i \le s \quad \text{for } i \ge i_0. \tag{25}$$

For $i \ge i_0$ and p a prime ideal in R such that $\dim R/p = s$, we have that $(I_i)_p = R_p$ or $(I_i)_p$ is p_p-primary.

In general, s is not a limit as is shown by the following simple example:

Example 6.6. Suppose that R is a Noetherian ring and $p \subset q \subset R$ are prime ideals. Let

$$I_i = \begin{cases} p & \text{if } i \text{ is odd,} \\ q & \text{if } i \text{ is even.} \end{cases}$$

We have that

$$I_i I_j = \begin{cases} p^2 \text{ or } q^2 & \text{if } i+j \text{ is even,} \\ pq & \text{if } i+j \text{ is odd.} \end{cases}$$

Thus, $I_i I_j \subset I_{i+j}$ for all i and j and

$$\dim R/I_i = \begin{cases} \dim R/p & \text{if } i \text{ is odd,} \\ \dim R/q & \text{if } i \text{ is even.} \end{cases}$$

Let

$$T = T(I_*) = \big\{ p \in \operatorname{Spec}(R) \\ \mid \dim R/p = s \text{ and there exist arbitrarily large } j \text{ such that } (I_j)_p \ne R_p \big\}.$$

Lemma 6.7. *$T(I_*)$ is a finite set.*

Proof. Let U be the set of prime ideals p of R that are an associated prime of some I_i with $i_0 \le i \le 2i_0 - 1$ and $\operatorname{ht} p = s$. Suppose that $q \in T$. There exists $j \ge i_0$ such that $(I_j)_q \ne R_q$. We can write $j = ai_0 + (i_0 + k)$ with $0 \le k \le i_0 - 1$ and $a \ge 0$. Thus, $I_{i_0}^a I_{i_0+k} \subset I_j$. Thus, $q \in U$ since $(I_{i_0}^a I_{i_0+k})_q \ne R_q$. □

Lemma 6.8. *There exist $c = c(I_*) \in \mathbb{Z}_+$ such that if $j \ge i_0$ and $p \in T(I_*)$, then*

$$p^{jc} R_p \subset I_j R_p.$$

Proof. There exists $a \in \mathbb{Z}_+$ such that for all $p \in T$, $p_p^a \subset (I_i)_p$ for $i_0 \le i \le 2i_0 - 1$.

Write $j = ti_0 + (i_0 + k)$ with $t \ge 0$ and $0 \le k \le i_0 - 1$. Then

$$p_p^{(t+1)a} \subset I_{i_0}^t I_{i_0+k} R_p \subset I_j R_p.$$

Let $c = \lceil a/i_0 \rceil + a$. Then

$$jc \ge a + j\frac{a}{i_0} = a + (ti_0 + i_0 + k)\frac{a}{i_0} \ge (t+1)a.$$

Thus, $p_p^{jc} \subset p_p^{(t+1)a} \subset (I_j)_p$. □

Let

$$A(I_*) = \{q \in T(I_*) \mid I_n R_q \text{ is } q_q\text{-primary for } n \geq i_0\}.$$

Lemma 6.9. *Suppose that $q \in T(I_*) \setminus A(I_*)$. Then there exists $b \in \mathbb{Z}_+$ such that $q_q^b \subset (I_n)_q$ for all $n \geq i_0$.*

Proof. There exists $n_0 \in \mathbb{Z}_+$ such that $n_0 \geq i_0$ and $(I_{n_0})_q = R_q$. Let $b \in \mathbb{Z}_+$ be such that $q_q^b \subset (I_n)_q$ for $0 \leq n < n_0$. Suppose that $n_0 \leq n$. Write $n = \beta n_0 + \alpha$ with $\beta \geq 0$ and $0 \leq \alpha < n_0$. Then

$$q_q^b \subset (I_{n_0}^\beta I_\alpha)_q \subset (I_n)_q. \qquad \square$$

We obtain the following asymptotic additivity formula:

Theorem 6.10. *Suppose that R is a d-dimensional local ring such that either*

(1) *R is regular or*

(2) *R is analytically unramified of equicharacteristic* 0.

Suppose that $\{I_i\}$ is a graded family of ideals in R. Let $s = s(I_*) = \limsup \dim R/I_i$ and $A = A(I_*)$. Suppose that $s < d$. Then

$$\lim_{n\to\infty} \frac{e_s(m_R, R/I_n)}{n^{d-s}/(d-s)!} = \sum_{q\in A} \left(\lim_{k\to\infty} \frac{e((I_k)_q)}{k^{d-s}} \right) e(m_{R/q}).$$

Proof. Let i_0 be the (smallest) constant satisfying (25). By the additivity formula [Serre 1965, p. V-2; Bruns and Herzog 1993, Corollary 4.6.8, p. 189], for $i \geq i_0$,

$$e_s(m_R, R/I_i) = \sum_p \ell_{R_p}(R_p/(I_i)_p) e_{m_R}(R/p),$$

where the sum is over all prime ideals p of R with $\dim R/p = s$. By Lemma 6.7, for $i \geq i_0$, the sum is actually over the finite set $T(I_*)$ of prime ideals of R.

For $p \in T(I_*)$, R_p is a local ring of dimension at most $d - s$. Further, R_p is analytically unramified [Rees 1961; Huneke and Swanson 2006, Proposition 9.1.4]. By Lemma 6.8 and by Theorem 4.6 or 5.8, replacing $(I_i)_p$ with p_p^{ic} if $i < i_0$, we have that

$$\lim_{i\to\infty} \frac{\ell_{R_p}(R_p/(I_i)_p)}{i^{d-s}}$$

exists. Further, this limit is zero if $p \in T(I_*) \setminus A(I_*)$ by Lemma 6.9 and since $s < d$. Finally, we have

$$\lim_{i\to\infty} \frac{\ell_{R_q}(R_q/(I_i)_q)}{i^{d-s}/(d-s)!} = \lim_{k\to\infty} \frac{e_{(I_k)_q}(R_q)}{k^{d-s}}$$

for $q \in A(I_*)$ by Theorem 6.5. $\square$

7. An application to growth of valuation semigroups

As a consequence of our main result, we obtain the following theorem, which gives a positive answer to a question raised in [Cutkosky et al. 2010a]. This formula was established if R is a regular local ring of dimension 2 with algebraically closed residue field in [Cutkosky et al. 2010a] and if R is an arbitrary regular local ring of dimension 2 in [Cutkosky and Vinh 2011] using a detailed analysis of a generating sequence associated to the valuation. A valuation ω dominating a local domain R is zero-dimensional if the residue field of ω is algebraic over R/m_R.

Theorem 7.1. *Suppose that R is a regular local ring or an analytically unramified local domain. Further suppose that R has an algebraically closed residue field. Let $d > 0$ be the dimension of R. Let ω be a zero-dimensional rank-1 valuation of the quotient field of R that dominates R. Let $S^R(\omega)$ be the semigroup of values of elements of R, which can be regarded as an ordered subsemigroup of $\mathbb{R}_+$. For $n \in \mathbb{Z}_+$, define*

$$\varphi(n) = |S^R(\omega) \cap (0, n)|.$$

Then

$$\lim_{n\to\infty} \frac{\varphi(n)}{n^d}$$

exists.

Proof. Let $I_n = \{f \in R \mid \omega(f) \geq n\}$ and $\lambda = \omega(m_R) = \min\{\omega(f) \mid f \in m_R\}$. Let $c \in \mathbb{Z}_+$ be such that $c\lambda > 1$. Then $m_R^c \subset I_1$. By Theorem 4.6 or 5.2, we have that

$$\lim_{n\to\infty} \frac{\ell_R(R/I_n)}{n^d}$$

exists.

Since R has an algebraically closed residue field, we have by [Cutkosky et al. 2010a; Cutkosky and Teissier 2010] that

$$\#\varphi(n) = \ell_R(R/I_n) - 1.$$

Thus, the theorem follows. □

In [Cutkosky et al. 2010a], it is shown that the real numbers β with $0 \leq \beta < \frac{1}{2}$ are the limits attained on a regular local ring R of dimension 2.

Acknowledgment

We thank the referee for their careful reading of this paper and for suggesting that we present the theorems with the less restrictive assumption of analytically unramified instead of reduced and excellent.

References

[Brodmann 1979] M. Brodmann, "Asymptotic stability of $\mathrm{Ass}(M/I^n M)$", *Proc. Amer. Math. Soc.* **74**:1 (1979), 16–18. MR 80c:13012 Zbl 0372.13010

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR 95h:13020 Zbl 0788.13005

[Cutkosky 2011] S. D. Cutkosky, "Asymptotic growth of saturated powers and epsilon multiplicity", *Math. Res. Lett.* **18**:1 (2011), 93–106. MR 2012b:13060 Zbl 1238.13012

[Cutkosky and Srinivas 1993] S. D. Cutkosky and V. Srinivas, "On a problem of Zariski on dimensions of linear systems", *Ann. of Math.* (2) **137**:3 (1993), 531–559. MR 94g:14001 Zbl 0822.14006

[Cutkosky and Teissier 2010] S. D. Cutkosky and B. Teissier, "Semigroups of valuations on local rings, II", *Amer. J. Math.* **132**:5 (2010), 1223–1247. MR 2011k:13005 Zbl 1213.13010

[Cutkosky and Vinh 2011] S. D. Cutkosky and P. A. Vinh, "Valuation semigroups of two dimensional local rings", preprint, 2011. arXiv 1105.1448

[Cutkosky et al. 2005] S. D. Cutkosky, H. T. Hà, H. Srinivasan, and E. Theodorescu, "Asymptotic behavior of the length of local cohomology", *Canad. J. Math.* **57**:6 (2005), 1178–1192. MR 2006f:13014 Zbl 1095.13015

[Cutkosky et al. 2010a] S. D. Cutkosky, K. Dalili, and O. Kashcheyeva, "Growth of rank 1 valuation semigroups", *Comm. Algebra* **38**:8 (2010), 2768–2789. MR 2011j:13007 Zbl 1203.13003

[Cutkosky et al. 2010b] S. D. Cutkosky, J. Herzog, and H. Srinivasan, "Asymptotic growth of algebras associated to powers of ideals", *Math. Proc. Cambridge Philos. Soc.* **148**:1 (2010), 55–72. MR 2011a:13009 Zbl 1200.13010

[Ein et al. 2003] L. Ein, R. Lazarsfeld, and K. E. Smith, "Uniform approximation of Abhyankar valuation ideals in smooth function fields", *Amer. J. Math.* **125**:2 (2003), 409–440. MR 2003m:13004 Zbl 1033.14030

[Fujita 1994] T. Fujita, "Approximating Zariski decomposition of big line bundles", *Kodai Math. J.* **17**:1 (1994), 1–3. MR 95c:14053 Zbl 0814.14006

[Fulger 2011] M. Fulger, "Local volumes on normal algebraic varieties", preprint, 2011. arXiv 1105.2981

[Grothendieck 1965] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, II", *Inst. Hautes Études Sci. Publ. Math.* **24** (1965), 5–231. MR 33 #7330 Zbl 0135.39701

[Herzog et al. 2008] J. Herzog, T. J. Puthenpurakal, and J. K. Verma, "Hilbert polynomials and powers of ideals", *Math. Proc. Cambridge Philos. Soc.* **145**:3 (2008), 623–642. MR 2009i:13028 Zbl 1157.13013

[Hübl 2001] R. Hübl, "Completions of local morphisms and valuations", *Math. Z.* **236**:1 (2001), 201–214. MR 2002d:13003 Zbl 1058.13014

[Huneke and Swanson 2006] C. Huneke and I. Swanson, *Integral closure of ideals, rings, and modules*, London Mathematical Society Lecture Note Series **336**, Cambridge University Press, 2006. MR 2008m:13013 Zbl 1117.13001

[Kaveh and Khovanskii 2012] K. Kaveh and A. G. Khovanskii, "Newton–Okounkov bodies, semigroups of integral points, graded algebras and intersection theory", *Ann. of Math. (2)* **176**:2 (2012), 925–978. MR 2950767 Zbl 06093944

[Khovanskii 1992] A. G. Khovanskii, "Newton polyhedron, Hilbert polynomial and sums of finite sets", *Funct. Anal. Appl.* **26**:4 (1992), 276–281. MR 94e:14068 Zbl 0809.13012

[Kleiman 2010] S. Kleiman, "The ϵ-multiplicity as a limit", communication to the author, 2010.

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, I: Classical setting, line bundles and linear series*, Ergeb. Math. Grenzgeb. (3) **48**, Springer, Berlin, 2004. MR 2005k:14001a Zbl 1066.14021

[Lazarsfeld and Mustaţă 2009] R. Lazarsfeld and M. Mustaţă, "Convex bodies associated to linear series", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:5 (2009), 783–835. MR 2011e:14012 Zbl 1182.14004

[Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics **8**, Cambridge University Press, 1986. MR 88h:13001 Zbl 0603.13001

[Mustaţă 2002] M. Mustaţă, "On multiplicities of graded sequences of ideals", *J. Algebra* **256**:1 (2002), 229–249. MR 2003k:13030 Zbl 1076.13500

[Okounkov 1996] A. Okounkov, "Brunn–Minkowski inequality for multiplicities", *Invent. Math.* **125**:3 (1996), 405–411. MR 99a:58074 Zbl 0893.52004

[Okounkov 2003] A. Okounkov, "Why would multiplicities be log-concave?", pp. 329–347 in *The orbit method in geometry and physics* (Marseille, 2000), edited by C. Duval et al., Progr. Math. **213**, Birkhäuser, Boston, MA, 2003. MR 2004j:20022 Zbl 1063.22024

[Rees 1956] D. Rees, "Valuations associated with a local ring, II", *J. London Math. Soc.* **31** (1956), 228–235. MR 18,8c Zbl 0074.26401

[Rees 1961] D. Rees, "A note on analytically unramified local rings", *J. London Math. Soc.* **36** (1961), 24–28. MR 23 #A3761 Zbl 0115.26202

[Rees 1989] D. Rees, "Izumi's theorem", pp. 407–416 in *Commutative algebra* (Berkeley, CA, 1987), edited by M. Hochster et al., Math. Sci. Res. Inst. Publ. **15**, Springer, New York, 1989. MR 90g:13010 Zbl 0741.13011

[Serre 1965] J.-P. Serre, *Algèbre locale: multiplicités*, 2nd ed., Lecture Notes in Mathematics **11**, Springer, Berlin, 1965. MR 34 #1352 Zbl 0142.28603

[Swanson 1997] I. Swanson, "Powers of ideals: primary decompositions, Artin–Rees lemma and regularity", *Math. Ann.* **307**:2 (1997), 299–313. MR 97j:13005 Zbl 0869.13001

[Ulrich and Validashti 2011] B. Ulrich and J. Validashti, "Numerical criteria for integral dependence", *Math. Proc. Cambridge Philos. Soc.* **151**:1 (2011), 95–102. MR 2012k:13011 Zbl 1220.13006

[Zariski and Samuel 1958] O. Zariski and P. Samuel, *Commutative algebra*, vol. 1, Van Nostrand Company, Princeton, NJ, 1958. MR 19,833e Zbl 081.26501

[Zariski and Samuel 1960] O. Zariski and P. Samuel, *Commutative algebra*, vol. 2, Van Nostrand Company, Princeton, NJ, 1960. MR 22 #11006 Zbl 0121.27801

Communicated by Craig Huneke
Received 2012-07-20 Revised 2012-10-11 Accepted 2012-11-17

cutkoskys@missouri.edu *Department of Mathematics, University of Missouri, Columbia, MO, 65211, United States*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2085

Normal coverings of linear groups

John R. Britnell and Attila Maróti

For a noncyclic finite group G, let $\gamma(G)$ denote the smallest number of conjugacy classes of proper subgroups of G needed to cover G. In this paper, we show that if G is in the range $\mathrm{SL}_n(q) \le G \le \mathrm{GL}_n(q)$ for $n > 2$, then $n/\pi^2 < \gamma(G) \le (n+1)/2$. This result complements recent work of Bubboloni, Praeger and Spiga on symmetric and alternating groups. We give various alternative bounds and derive explicit formulas for $\gamma(G)$ in some cases.

1. Introduction

Normal coverings. Let G be a noncyclic finite group. We write $\gamma(G)$ for the smallest number of conjugacy classes of proper subgroups of G needed to cover it. In other words, $\gamma(G)$ is the least k for which there exist subgroups $H_1, \dots, H_k < G$ such that

$$G = \bigcup_{i=1}^{k} \bigcup_{g \in G} {H_i}^g.$$

We say that the set of conjugacy classes $\{{H_i}^G \mid i = 1, \dots, k\}$ is a *normal covering* for G.

Bubboloni and Praeger [2011] have recently investigated $\gamma(G)$ in the case that G is a finite symmetric or alternating group. They show, for example, that if n is an odd composite number then

$$\frac{\phi(n)}{2} + 1 \le \gamma(S_n) \le \frac{n-1}{2},$$

where ϕ is Euler's totient function. Similar results are established for all values of n and for both S_n and A_n. Part of the motivation for their work comes from an application in number theory.

Britnell was supported by a Research Fellowship at the Heilbronn Institute for Mathematical Research. The research of Maróti was supported by a Marie Curie International Reintegration Grant within the seventh European Community Framework Programme, by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by OTKA K84233.
MSC2010: primary 20D60; secondary 20G40.
Keywords: covering, normal covering, linear group, finite group.

It is a well-known theorem of Jordan that no finite group is covered by the conjugates of any proper subgroup. To paraphrase, $\gamma(G) \neq 1$ for any finite group G. It is known that there exists a finite solvable group G with $\gamma(G) = k$ for every $k > 1$ [Crestani and Lucchini 2012]. It has been shown in [Bubboloni and Lucido 2002] that if G is one of the groups $\mathrm{GL}_n(q)$, $\mathrm{SL}_n(q)$, $\mathrm{PGL}_n(q)$ or $\mathrm{PSL}_n(q)$, then $\gamma(G) = 2$ if and only if $n \in \{2, 3, 4\}$. (Notice that γ is undefined for $n = 1$ since the groups are cyclic in this case.) Other groups of Lie type possessing a normal covering of size 2 have been studied in [Bubboloni et al. 2006; 2011].

In this paper, we give bounds on $\gamma(G)$, where $\mathrm{SL}_n(q) \leq G \leq \mathrm{GL}_n(q)$, for all values of n. In some cases, we are able to give an exact value. Our bounds extend without change to $G/Z(G)$.

We introduce some notation. We write $\lfloor x \rfloor$ for the integer part of a real number x. As already noted above, ϕ denotes Euler's function. We shall also use Lehmer's *partial totient function*, which we define here.

Definition. Let k and t be such that $0 \leq t < k < n$. We define the partial totient $\phi(k, t, n)$ to be the number of integers x, coprime with n, such that

$$\frac{nt}{k} < x < \frac{n(t+1)}{k}.$$

We give two separate upper bounds on $\gamma(G)$.

Theorem 1.1. *Let $n \in \mathbb{N}$, and let $\nu = \nu(n)$ be the number of prime factors of n. Let $p_1, \ldots, p_\nu$ be the distinct prime factors of n with $p_1 < p_2 < \cdots < p_\nu$. Let G be a group such that $\mathrm{SL}_n(q) \leq G \leq \mathrm{GL}_n(q)$.*

(1) *If $\nu \geq 2$, then*

$$\gamma(G) \leq \left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\frac{n}{2} + 2.$$

(2) *If $n > 6$, then*

$$\gamma(G) \leq \left\lfloor \frac{n}{3} \right\rfloor + \phi(6, 2, n) + \nu.$$

A great deal of information is given in [Lehmer 1955, §6] about the function $\phi(6, t, n)$, from which the following statement can be derived:

$$\frac{\phi(n)}{6} - \phi(6, 2, n) = \begin{cases} 0 & \text{if } n \text{ is divisible either by 9 or by a prime} \\ & \qquad \text{of the form } 3k+1 \text{ for } k \in \mathbb{N}, \\ \frac{1}{12}\lambda(n)2^\nu & \text{otherwise, if } n \text{ is divisible by 3,} \\ \frac{1}{6}\lambda(n)2^\nu & \text{otherwise, if } n \text{ is not divisible by 3,} \end{cases}$$

in which $\lambda(n) = (-1)^\ell$, where ℓ is the number of prime divisors of n counted with multiplicity.

Independent sets of conjugacy classes. Let $\kappa(G)$ be the size of the largest set of conjugacy classes of G such that any pair of elements from distinct classes generates G. We call such a set an *independent set of classes.* Guralnick and Malle [2012] have shown that $\kappa(G) \geq 2$ for any finite simple group G. It is clear that whenever $\gamma(G)$ is defined, we have the inequality

$$\kappa(G) \leq \gamma(G)$$

since if $\mathscr{C}$ is a normal covering of G, and if $\mathscr{I}$ is an independent set of classes, then each element of $\mathscr{C}$ covers at most one element of $\mathscr{I}$.

We establish two lower bounds for $\kappa(G)$. By the observation of the previous paragraph, these also operate as lower bounds for $\gamma(G)$.

Theorem 1.2. *Let $n \in \mathbb{N}$, and let $\nu = \nu(n)$ be the number of prime factors of n. Let $p_1, \ldots, p_\nu$ be the distinct prime factors of n with $p_1 < p_2 < \cdots < p_\nu$. Let G be a group such that $\mathrm{SL}_n(q) \leq G \leq \mathrm{GL}_n(q)$.*

(1) *If $\nu \geq 2$, then*

$$\frac{\phi(n)}{2} + \nu(n) \leq \kappa(G).$$

(2) *If $\nu \geq 3$, and if n is not equal to $6p$ or $10p$ for any prime p, then*

$$\left\lfloor \frac{n+6}{12} \right\rfloor + \phi(12, 1, 3n) + \nu \leq \kappa(G).$$

Furthermore, if $\mathrm{hcf}(n, 6) = 1$, *then*

$$\left\lfloor \frac{n+6}{12} \right\rfloor + \phi(12, 1, 3n) + \phi(12, 0, n) + \nu \leq \kappa(G).$$

The values $t = 0, 1$ are not amongst those for which the function $\phi(12, t, n)$ is evaluated explicitly in [Lehmer 1955]. However, Theorem 10 of the same work gives the general estimate

$$|\phi(n) - k\phi(k, t, n)| \leq (k-1)2^\nu,$$

where ν is the number of prime divisors of n. This yields the lower bound

$$\phi(12, t, n) \geq \frac{\phi(n)}{12} - \frac{11}{12}2^\nu.$$

There are certain cases in which an upper bound for $\gamma(G)$ coincides with a lower bound for $\kappa(G)$. In these cases, we must have $\gamma(G) = \kappa(G)$, and we obtain a precise formula.

Theorem 1.3. *Let G be a group such that $\mathrm{SL}_n(q) \leq G \leq \mathrm{GL}_n(q)$.*

(1) *If $n = p^a$, where p is a prime and $a \in \mathbb{N}$, and if $n > 2$, then*

$$\gamma(G) = \kappa(G) = \left(1 - \frac{1}{p}\right)\frac{n}{2} + 1.$$

(2) *If* $n = p^a q^b$, *where* p *and* q *are distinct primes and* $a, b \in \mathbb{N}$, *then*

$$\gamma(G) = \kappa(G) = \Big(1 - \frac{1}{p}\Big)\Big(1 - \frac{1}{q}\Big)\frac{n}{2} + 2.$$

(3) *If* $n = 6p$, *where* p *is a prime, then* $\gamma(G) = \kappa(G) = p + 2$.

(4) *If* $n = 10p$, *where* p *is a prime, then* $\gamma(G) = \kappa(G) = 2p + 2$.

Certain cases of Theorem 1.3 will require independent treatment as they arise as exceptional cases in the proof of Theorem 1.2.

Linear bounds. Theorems 1.1(1), 1.2(2) and 1.3, taken together, imply that

$$\frac{n}{12} < \kappa(G) \le \gamma(G) \le \frac{n+1}{2} \tag{1}$$

for all $n > 2$. The upper bound is exact when n is an odd prime. (When $n = 2$, it is known that $\gamma(G) = 2$; see [Bubboloni and Lucido 2002] or the remark after Proposition 4.1 below. It is also easy to show that $\kappa(G) = 2$ in this case.) It follows immediately that

$$\limsup \frac{\gamma(G)}{n} = \frac{1}{2}. \tag{2}$$

The lower bound for γ can be improved as the following theorem indicates:

Theorem 1.4. *If* G *is a group such that* $\mathrm{SL}_n(q) \le G \le \mathrm{GL}_n(q)$, *then* $n/\pi^2 < \gamma(G)$.

From the first part of Theorem 1.1 and from Theorem 1.4, it is easy to show that

$$\frac{1}{\pi^2} \le \liminf \frac{\gamma(G)}{n} \le \frac{1}{6}. \tag{3}$$

It follows from the theorems that we have stated that $\gamma(G)$ and $\kappa(G)$ are bounded above and below by monotonic functions that grow linearly with n. It appears that the situation for symmetric groups is similar. It was announced in [Bubboloni et al. 2012, §1.1] and demonstrated in [Bubboloni et al. 2013] that $\gamma(S_n)$ and $\gamma(A_n)$ are bounded above and below by linear functions of n. In fact, the numbers $\gamma(S_n)$ and $\gamma(\mathrm{GL}_n(q))$ seem to be closely related; in all cases where both are known exactly, they differ by at most 1. It is not hard to show, and it is worth remarking in this connection, that the upper bounds stated for $\gamma(G)$ in Theorem 1.1 are also upper bounds for $\gamma(S_n)$ improving marginally on those of [Bubboloni and Praeger 2011, Theorem A]. It should also be noted that all of our bounds are independent of the field size q.

We establish the upper bounds of Theorem 1.1 in Section 2 by exhibiting explicit normal coverings of the necessary sizes. This builds on work described in [Britnell et al. 2008], in which coverings of $\mathrm{GL}_n(q)$ by proper subgroups are constructed. The two lower bounds of Theorem 1.2 are proved in Section 3. Both are proved by exhibiting an independent set of classes. This requires an account of overgroups of

certain special elements in $\mathrm{GL}_n(q)$. For such an account, we rely on [Guralnick et al. 1999], which provides a classification of subgroups whose orders are divisible by primitive prime divisors of $q^d - 1$ for all $d > n/2$. The remaining cases of Theorem 1.3 are brought together in Section 4. Finally, Theorem 1.4 is established in Section 5. Its proof relies on work from the doctoral thesis of Joseph DiMuro [2007], which extends the classification of [Guralnick et al. 1999] to cover all $d \geq n/3$.

The classes of subgroups in our normal covering remain distinct, proper and nontrivial in the quotient of G by $Z(G)$. This is true also of the classes of maximal overgroups that cover the conjugacy classes in our independent sets. It follows that the bounds that we have stated for $\gamma(G)$ and for $\kappa(G)$ hold equally for $\gamma(G/Z(G))$ and for $\kappa(G/Z(G))$.

2. Normal coverings of G

We shall write V for the space ${\mathbb{F}_q}^n$. We assume that $\mathrm{SL}(V) \leq G \leq \mathrm{GL}(V)$ throughout the paper.

We begin by introducing the classes of subgroups that we shall need for our coverings. Proposition 2.1 below contains standard information about certain subgroups of $\mathrm{GL}_n(q)$, and we shall not prove it here.

Proposition 2.1. (1) *Let d be a divisor of n. There exist embeddings of $\mathrm{GL}_{n/d}(q^d)$ into $\mathrm{GL}_n(q)$. All such embeddings are conjugate by elements of $\mathrm{SL}_n(q)$, and each has index d in its normalizer in $\mathrm{GL}_n(q)$. If d is prime, then the normalizer is a maximal subgroup of $\mathrm{GL}_n(q)$.*

(2) *Suppose that $1 \leq k < n$, and let U be a k-dimensional subspace of V. Then the set stabilizer G_U of U in G is a maximal subgroup of G. If W is another k-dimensional subspace, then G_U and G_W are conjugate in G.*

It will be convenient to have concise notation for these subgroups.

Definition. (1) We refer to the maximal subgroups of Proposition 2.1(1) as *extension field subgroups* of degree d, and we write $\mathrm{efs}(d)$ for the conjugacy class consisting of the intersections of all such subgroups with the group G.

(2) We refer to the subgroups of Proposition 2.1(2) as *subspace stabilizers* of dimension k, and we write $\mathrm{ss}(k)$ for the conjugacy class consisting of all such subgroups.

The following technical lemma will be useful:

Lemma 2.2. (1) *Suppose that $X \in \mathrm{GL}(V)$ and that X stabilizes a k-dimensional subspace of V. Then X stabilizes a subspace whose dimension is $n - k$.*

(2) *Let $X \in \mathrm{GL}(V)$, and let p be a prime dividing n. If X lies in no extension field subgroup of degree p, then it stabilizes a subspace of V whose dimension is coprime with p.*

Proof. (1) Suppose X stabilizes a space U of dimension k. Then the transpose X^t acts on the dual space V^* and stabilizes the annihilator of U, which has dimension $n-k$.

(2) If X stabilizes no subspace whose dimension is coprime with p, then every irreducible divisor of its characteristic polynomial has degree divisible by p and must therefore split into p factors over $\mathbb{F}_{q^p}$. Suppose that the elementary divisors of X are $f_1^{a_1}, \dots, f_t^{a_t}$. For each i, let g_i be an irreducible factor of f_i over $\mathbb{F}_{q^p}$, and let $Y \in \mathrm{GL}_{n/p}(q^p)$ have elementary divisors $g_1^{a_1}, \dots, g_t^{a_t}$. Then it is not hard to see that any embedding of $\mathrm{GL}_{n/p}(q^p)$ into $\mathrm{GL}_n(q)$ must map Y to a conjugate of X. □

We are now in a position to exhibit some normal coverings of G.

Lemma 2.3. (1) *Let p be a prime dividing n. Then there is a normal covering $\mathscr{C}_p$ for G given by*

$$\mathscr{C}_p = \{\mathrm{efs}(p)\} \cup \{\mathrm{ss}(k) \mid 1 \le k \le n/2,\ \ p \nmid k\}.$$

The size of $\mathscr{C}_p$ is

$$|\mathscr{C}_p| = \left\lfloor \left(1 - \frac{1}{p}\right)\frac{n}{2} \right\rfloor + 1 + \epsilon,$$

where

$$\epsilon = \begin{cases} 1 & \text{if } p = 2 \text{ and } n/2 \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

This is minimized when p is the smallest prime divisor of n.

(2) *Let p_1 and p_2 be distinct prime divisors of n. Then there is a normal covering $\mathscr{C}_{p_1,p_2}$ for G given by*

$$\mathscr{C}_{p_1,p_2} = \{\mathrm{efs}(p_1), \mathrm{efs}(p_2)\} \cup \{\mathrm{ss}(k) \mid 1 \le k < n/2,\ \ p_1, p_2 \nmid k\}.$$

The size of $\mathscr{C}_{p_1,p_2}$ is

$$|\mathscr{C}_{p_1,p_2}| = \left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\frac{n}{2} + 2.$$

This is minimized when p_1 and p_2 are the two smallest prime divisors of n.

Proof. The sizes of the sets $\mathscr{C}_p$ and $\mathscr{C}_{p_1,p_2}$ are easily seen to be as stated. That $\mathscr{C}_p$ is a normal covering follows immediately from Lemma 2.2. So it remains only to prove that C_{p_1,p_2} is a normal covering.

Let $X \in G$, let f_X be the characteristic polynomial of X, and let $g_1, \dots, g_s$ be the irreducible factors of f_X over $\mathbb{F}_q$ with degrees $d_1, \dots, d_s$, respectively. Then clearly there exist X-invariant subspaces $U_1, \dots, U_s$ such that $\dim U_i = d_i$ for all i and such that $U_i \cap U_j = \{0\}$ whenever $i \ne j$. If any d_i is divisible by neither of the primes p_1 and p_2, then X is contained in a subspace stabilizer from one of

the classes in $\mathscr{C}_{p_1,p_2}$. So we assume that each d_i is divisible by at least one of p_1 or p_2. Suppose that d_a is divisible by p_1 but not by p_2 and that d_b is divisible by p_2 but not by p_1. Then $U_a \oplus U_b$ is an X-invariant subspace, and its dimension is coprime with p_1 and p_2; so again, X is in a subspace stabilizer from $\mathscr{C}_{p_1,p_2}$. But if no such d_a and d_b can be found, then either all of the d_i are divisible by p_1 or they are all divisible by p_2. In this case, X lies in an extension field subgroup either of degree p_1 or of degree p_2. □

We note that the argument of the last paragraph of this proof does not extend to the case of three primes, p_1, p_2 and p_3. It is possible to find matrices whose invariant subspaces all have dimensions divisible by one of those primes but which lie in no extension field subgroup. In the case that the primes are 2, 3 and 5, for instance, there are 30-dimensional matrices whose irreducible invariant spaces have dimensions 2, 3 and 25. (Another example is used in the proof of Proposition 4.4 below.) This is the explanation for the appearance of the two smallest prime divisors of n in the first upper bound of Theorem 1.1, which may at first seem a little curious.

The second upper bound of Theorem 1.1 is proved in a somewhat similar fashion.

Lemma 2.4. *Let $p_1, \dots, p_\nu$ be the distinct primes dividing n. Then there is a normal covering $\mathfrak{D}$ of G given by*

$$\begin{aligned}\mathfrak{D} = {} & \{\mathrm{ss}(k) \mid 1 \le k \le n/3\} \\ & \cup \{\mathrm{ss}(k) \mid n/3 < k \le n/2,\ \mathrm{hcf}(k,n) = 1\} \\ & \cup \{\mathrm{efs}(p_i) \mid 1 \le i \le \nu\}.\end{aligned}$$

For $n > 6$, the size of $\mathfrak{D}$ is

$$\left\lfloor \frac{n}{3} \right\rfloor + \phi(6, 2, n) + \nu.$$

Proof. Let $X \in G$. Suppose that X is reducible and that its smallest nontrivial invariant subspace has dimension k. If $k > n/3$, then it is not hard to see (for instance, by considering the irreducible factors of the characteristic polynomial) that X stabilizes at most one other proper nontrivial subspace of dimension $n - k$. It follows that if p is a prime dividing both n and k, then X is contained in an element of $\mathrm{efs}(p)$. It is now a straightforward matter to show that $\mathfrak{D}$ is a normal covering, and we omit further details. The size of $\mathfrak{D}$ follows immediately from its definition. □

3. Lower bounds for $\kappa(G)$

Recall that $\mathrm{GL}_n(q)$ contains elements of order $q^n - 1$, known as Singer elements. Such elements stabilize no nontrivial proper subspace of V. The determinant of a Singer element generates the multiplicative group of $\mathbb{F}_q$.

In order to handle all groups G in the range $\mathrm{SL}_n(q) \le G \le \mathrm{GL}_n(q)$ together, we define a parameter $\alpha \in \mathbb{N}$ by

$$\alpha = \begin{cases} 0 & \text{if } G = \mathrm{SL}_n(q), \\ -|\mathrm{GL}_n(q) : G| & \text{otherwise.} \end{cases}$$

Let ζ be a generator of the multiplicative group of $\mathbb{F}_q$. Then we have

$$\frac{G}{\mathrm{SL}_d(q)} \cong \langle \zeta^\alpha \rangle.$$

Definition. (1) For $d = 1, \dots, n$, let Γ_d be a Singer element with determinant ζ in $\mathrm{GL}_d(q)$.

(2) For $k < n/2$, define

$$\Sigma_k = \mathrm{diag}({\Gamma_k}^{\alpha-1}, \Gamma_{n-k}).$$

(3) For $j < (n-2)/4$, define

$$T_j = \mathrm{diag}({\Gamma_j}^{\alpha-2}, \Gamma_{j+1}, \Gamma_{n-2j-1}).$$

The reasons for defining α as above will be clear from the following remark:

Remark. (1) Since $\det \Sigma_k = \det T_j = \zeta^\alpha$, we have $\Sigma_k, T_j \in G$.

(2) It is clear from the definition of α that $(1-q) < \alpha \le 0$ and hence that $|\alpha - 2| < q + 1$. It follows easily that the actions of the matrices ${\Gamma_k}^{\alpha-1}$ and ${\Gamma_j}^{\alpha-2}$ are irreducible for all k and j. Therefore, the module $\mathbb{F}_q\langle \Sigma_k \rangle$ decomposes into precisely two irreducible summands, and $\mathbb{F}_q\langle T_j \rangle$ decomposes into precisely three irreducible summands.

Lemma 3.1. *Suppose that $n > 4$. Let $k < n/2$, and if $q = 2$, then suppose that $n - k \ne 6$. Let $j < (n-2)/4$, and if $q = 2$, then suppose that $n - 2j - 1 \ne 6$.*

(1) *If M is a maximal subgroup of G containing Γ_n, then M is an extension field subgroup of prime degree.*

(2) *If M is a maximal subgroup of G containing Σ_k, then M is either an extension field subgroup whose degree is a prime divisor of* $\gcd(k, n)$ *or else the stabilizer of a subspace of dimension k or $n - k$.*

(3) *Let n have at least three distinct prime divisors. If M is a maximal subgroup of G containing T_j, then M is the stabilizer of a subspace whose dimension is one of j, $j+1$, $2j+1$, $n-2j-1$, $n-j-1$ or $n-j$.*

Proof. Part (1) of the lemma is a result of Kantor [1980].

For $(n, q) \neq (11, 2)$, part (2) of the lemma follows from [Britnell et al. 2008, Theorem 4.1(2)]. However, a few comments are to be made about this assertion. The matrix that we have called Σ_k is referred to as GLk in [Britnell et al. 2008]. The result in [Britnell et al. 2008] is stated only for the groups $\mathrm{GL}_n(q)$ and $\mathrm{SL}_n(q)$, but the proof given there applies equally to intermediate subgroups. Finally, the proof in [Britnell et al. 2008] relies on the existence of primitive prime divisors of $q^{n-k} - 1$ (where $n - k > 2$), which is given by the theorem of Zsigmondy [1892] for all pairs $(q, n - k)$ except $(2, 4)$ and $(2, 6)$; the second of these exceptions accounts for the excluded case in the statement of the present lemma. The argument uses the classification in [Guralnick et al. 1999] of subgroups of $\mathrm{GL}_n(q)$ whose order is divisible by a prime divisor of $q^e - 1$, where $e > n/2$.

To finish the proof of part (2) of the present lemma, we must consider the exceptional case of the group $\mathrm{GL}_{11}(2)$. In this case, we require a reference directly to the lists of [Guralnick et al. 1999]. We find that there are several irreducible subgroups whose order is divisible by a primitive prime divisor 11 of $2^{10} - 1$; we must show that none of these contains Σ_1. All of these subgroups are almost simple and have a socle that is isomorphic either to one of the Mathieu groups M_{23} or M_{24} or to the unitary group $\mathrm{PSU}_5(2)$ or to a linear group $\mathrm{SL}_2(11)$ or $\mathrm{SL}_2(23)$. (These subgroups may be found in Table 5 (lines 12 and 14) and Table 8 (lines 2, 7 and 9) of [Guralnick et al. 1999].) Information about these groups may be found in [Conway et al. 1985]. None of these groups themselves, nor any of their outer automorphism groups, have order divisible by 31. Therefore, an almost simple group of one of these types can contain no element of order $2^{10} - 1 = 3 \cdot 11 \cdot 31$, which is the order of the element Σ_1.

For the proof of part (3) of the lemma, we refer once again to the classification of [Guralnick et al. 1999], this time for matrix groups whose order is divisible by a primitive prime divisor of $q^{n-2j-1} - 1$. It is not hard to see that T_j has no overgroups of classical type. The condition that n has three distinct prime divisors rules out the small dimensional sporadic examples contained in Tables 1–7. Other examples are ruled out because their order is less than $q^{n-2j-1} - 1$, which is the order of the summand Γ_{n-2j-1} of T_j. □

We define a set of classes that will help us to establish the first of our lower bounds for $\kappa(G)$.

Definition. Define a set Φ of classes of G by

$$\Phi = \big\{[\Sigma_p] \;\big|\; p \mid n,\ p \text{ prime},\ p < n/2\big\} \cup \big\{[\Sigma_k] \;\big|\; k < n/2,\ \mathrm{hcf}(n, k) = 1\big\},$$

where $[g]$ denotes the conjugacy class of g.

Lemma 3.2. *Let $n > 2$, and let $\nu(n)$ be the number of prime factors of n. Then*

$$|\Phi| = \phi(n)/2 + \nu(n) - \epsilon,$$

where

$$\epsilon = \begin{cases} 1 & \text{if } n = 2p \text{ for some odd prime } p, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. This is immediate from the definition of Φ. □

Lemma 3.2, together with the following two lemmas, will imply the first part of Theorem 1.2:

Lemma 3.3. *Φ is an independent set of classes.*

Proof. Suppose that $q \neq 2$ or that $[\Sigma_{n-6}] \notin \Phi$. Then Lemma 3.1 provides full information about the maximal subgroups of G that contain elements of Φ, and it is easy to check that the result holds in this case.

Next suppose that $q = 2$ and $[\Sigma_{n-6}] \in \Phi$. (This implies that $n \in \{7, 8, 9, 11\}$.) Lemma 3.1 gives full information about the maximal subgroups of G covering elements of the classes in Φ other than $[\Sigma_{n-6}]$. No class of subgroups contains elements of more than one such class, and it is easy to check that none covers the element Σ_{n-6} itself. □

Lemma 3.4. *Let $n = 2p$, where $p > 2$ is a prime. Then $\kappa(G) \geq |\Phi| + 1$.*

Proof. The proof of Lemma 3.3 shows that in any normal covering of G, the distinct classes in Φ are covered by distinct classes of subgroups. We add an extra conjugacy class to Φ, namely the class represented by $\Sigma_p = \mathrm{diag}({\Gamma_p}^{\alpha-1}, \Gamma_p)$, where Γ_p is a Singer element in $\mathrm{GL}_p(q)$. This element stabilizes no subspace of dimension k for any k coprime with n nor does it stabilize a subspace of dimension 2 or $n-2$. Therefore, by part (2) of Lemma 3.1, if $\Phi \cup \{[\Sigma_p]\}$ is not an independent set of classes, then Σ_p must lie in a subgroup in efs(2).

Note that since 2 and p are coprime, ${\Sigma_p}^2$ has two irreducible summands of dimension p. It is not hard to show that these submatrices are not conjugate, and neither of them is reducible over $\mathbb{F}_{q^2}$; it follows that ${\Sigma_p}^2$ is not contained in any embedding of $\mathrm{GL}_p(q^2)$ into G. Hence, Σ_p itself is not contained in an embedding of $\mathrm{GL}_p(q^2) \cdot 2$. □

Lemmas 3.2, 3.3 and 3.4 complete the proof of part (1) of Theorem 1.2.

We define a second independent set of classes that yields the second lower bound of Theorem 1.2. We shall require the following lemma:

Lemma 3.5. *Let p be a prime divisor of n. Suppose that n has at least three distinct prime divisors and that n is not equal to $6q$ or $10q$ for any prime q. Then there exists an integer w_p such that $(n-2)/4 \leq w_p < n/2$ and such that w_p is divisible*

by p and by no other prime divisor of n. If $p \neq 3$, then w_p may be chosen so that it is not divisible by 3.

Proof. Bertrand's postulate states that for every $k > 3$ there is a prime r such that $k < r < 2k - 2$. The conditions on n imply that $n \geq 12p$. So there is a prime $r > 3$ such that

$$\frac{n}{4p} < r < \frac{n}{2p}.$$

If r is not itself a prime divisor of n or if it is equal to p, then we may take $w_p = pr$. On the other hand, if r is a prime divisor of n other than p, then clearly $n = 3pr$, and since we have assumed that $n \geq 12p$, we have $r \geq 5$. Now we see that there exists m equal either to $r + 1$ or to $r + 2$ such that m is not divisible by 3, and we may take $w_p = pm$. □

Definition. Let n be a number with at least three distinct prime divisors and not equal to $6p$ or $10p$ for any prime p. We define a set Ψ of classes of G by

$$\begin{aligned}\Psi = \{[T_j] \mid j < (n-2)/4,\ \ j \equiv 1 \bmod 3\}\\ \cup \{[\Sigma_k] \mid n/4 < k < n/2,\ \mathrm{hcf}(3n, k) = 1\}\\ \cup \{[\Sigma_{6b}] \mid b < n/12,\ \mathrm{hcf}(n, 6b) = 1\}\\ \cup \big\{[\Sigma_{w_p}] \bigm| p \mid n,\ \ p \text{ prime}\big\},\end{aligned}$$

where w_p is as constructed in Lemma 3.5 and where $[g]$ denotes the conjugacy class of g.

To describe the size of the set Ψ, we use Lehmer's partial totient function $\phi(k, t, n)$, which was defined before the statement of Theorem 1.1 above.

Lemma 3.6. *Let n have ν distinct prime divisors, where $\nu \geq 3$, and suppose that n is not equal to $6p$ or $10p$ for any prime p.*

(1) *If* 2 *or* 3 *divides n, then*

$$|\Psi| = \left\lfloor \frac{n+6}{12} \right\rfloor + \phi(12, 1, 3n) + \nu.$$

(2) *If* $\mathrm{hcf}(n, 6) = 1$, *then*

$$|\Psi| = \left\lfloor \frac{n+6}{12} \right\rfloor + \phi(12, 1, 3n) + \phi(12, 0, n) + \nu.$$

Proof. We write $\lceil x \rceil$ for the least integer not less than x. The size X of the set $\{[T_j] \mid j < (n-2)/4,\ \ j \equiv 1 \bmod 3\}$ is $\lceil N/3 \rceil$, where $N = \lfloor (n-2)/4 \rfloor$. By examining residues modulo 12, it is not hard to show that $X = \lfloor (n+6)/12 \rfloor$, the first term in our sum.

It is immediate from the definition of the function $\phi(k, t, n)$ that the size of the set $\{[\Sigma_k] \mid n/4 < k < n/2,\ \mathrm{hcf}(3n, k) = 1\}$ is $\phi(12, 1, 3n)$. We observe that the set

$\{[\Sigma_{6b}] \mid b < n/12,\ \mathrm{hcf}(n, 6b) = 1\}$ is empty if $\mathrm{hcf}(n, 6) \neq 1$; otherwise, it has size $\phi(12, 0, n)$. And clearly the set $\{[\Sigma_{w_p}] \mid p|n,\ p \text{ prime}\}$ has size ν as required. □

To establish the second lower bound in Theorem 1.2, it will suffice to show that any normal covering for G has size at least $|\Psi|$. This is done in the following lemma:

Lemma 3.7. *Let n have at least three distinct prime divisors and not be equal to $6p$ or $10p$ for any prime p. Then Ψ is an independent set of classes.*

Proof. Lemma 3.1 describes the maximal subgroups of G that contain elements of the classes in Ψ. The elements T_j lie only in members of $\mathrm{ss}(\ell)$ or $\mathrm{ss}(n-\ell)$, where $\ell \in \{j, j+1, 2j+1\}$. Notice that if $\ell > n/4$, then $\ell = 2j+1$, and hence, $\ell \equiv 3 \bmod 6$. The elements Σ_k, where k is coprime with n, lie only in members of $\mathrm{ss}(k)$ or $\mathrm{ss}(n-k)$. And the elements Σ_{w_p} lie in subspace stabilizers and also in elements of $\mathrm{efs}(p)$. It is easy to check that the values permitted for j, k, b and w_p ensure that no two elements of distinct classes in Ψ stabilize subspaces of the same dimension. Therefore, no two classes in Ψ can be covered by a single class of subgroups. □

4. Several equalities

In this section, we establish the various claims of Theorem 1.3. We do this simply by comparing upper and lower bounds from earlier parts of the paper.

Proposition 4.1. *If $n = p^a$, where p is a prime and $a \in \mathbb{N}$, and if $n > 2$, then*

$$\gamma(G) = \kappa(G) = \left(1 - \frac{1}{p}\right)\frac{n}{2} + 1.$$

Proof. Lemmas 2.3 and 3.3 together tell us that

$$|\Phi| \leq \kappa(G) \leq \gamma(G) \leq |\mathscr{C}_p|.$$

But it is easy to check, using Lemma 3.2, that $|\Phi| = |\mathscr{C}_p|$ and that this number is as claimed in the proposition. □

Remark. If $n = 2$, then the covering $\mathscr{C}_2$ has size 2. Since no finite group is covered by a single class of proper subgroups, it follow that $\gamma(G) = 2$ in this case.

Proposition 4.2. *If $n = p^a q^b$, where p and q are distinct primes and $a, b \in \mathbb{N}$, then*

$$\gamma(G) = \kappa(G) = \left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right)\frac{n}{2} + 2.$$

Proof. As in the proof above, Lemma 2.3 with Lemmas 3.3 and 3.4 yields that

$$|\Phi| + \epsilon \leq \kappa(G) \leq \gamma(G) \leq |\mathscr{C}_{p,q}|,$$

where $\epsilon = 1$ if $n = 2p$ (or $n = 2q$) and $\epsilon = 0$ otherwise. But we see that $|\Phi| + \epsilon = |\mathscr{C}_{p,q}|$ and that this number is as claimed in the proposition. □

Proposition 4.3. *If $n = 6p$, where p is a prime, then*

$$\gamma(G) = \kappa(G) = p + 2.$$

Proof. In this case, we have

$$|\Phi| \leq \kappa(G) \leq \gamma(G) \leq |\mathscr{C}_{2,3}|,$$

and it is easy to calculate that $|\Phi| = |\mathscr{C}_{2,3}| = p + 2$. □

Proposition 4.4. *If $n = 10p$, where p is a prime, then*

$$\gamma(G) = \kappa(G) = 2p + 2.$$

Proof. If p is 2 or 5, then the result follows from Proposition 4.2; if $p = 3$, then it follows from Proposition 4.3. So we may assume that $p > 5$. Then we have

$$|\Phi| \leq \kappa(G) \leq \gamma(G) \leq |\mathscr{C}_{2,5}|,$$

but in this case, we see that $|\Phi| = 2p + 1$ whereas $|\mathscr{C}_{2,5}| = 2p + 2$. To prove that the upper bound is sharp for $\kappa(G)$, it will be sufficient to exhibit an element Y of G that cannot be covered by any class of subgroups containing an element of any conjugacy class in Φ. We define

$$Y = \mathrm{diag}({\Gamma_p}^{\alpha-2}, \Gamma_5, \Gamma_{n-p-5}).$$

Notice that $n - p - 5$ is even and coprime with 5 and with p. It follows that Y does not stabilize a subspace of dimension coprime with n. But certainly Y lies in no extension field subgroup, and so it satisfies the required condition. □

5. Proof of Theorem 1.4

For a positive integer n, let $f(n)$ be the number of partitions of n with exactly three parts. By an elementary counting argument, the following formula can be found for $f(n)$:

Lemma 5.1. $$f(n) = \begin{cases} \frac{1}{12}(n-1)(n-2) + \frac{1}{2}\lfloor (n-1)/2 \rfloor & \text{if } 3 \nmid n, \\ \frac{1}{12}(n-1)(n-2) + \frac{1}{2}\lfloor (n-1)/2 \rfloor + \frac{1}{3} & \text{if } 3 \mid n. \end{cases}$$

It follows from Lemma 5.1 that

$$\left| f(n) - \frac{n^2}{12} \right| \leq \tfrac{1}{3}.$$

We define $\epsilon_n = f(n) - n^2/12$.

Let $P(n)$ be the set of partitions of n into three parts having no common divisor greater than 1. Let $g(n) = |P(n)|$. Then we have $f(n) = \sum_{d|n} g(d)$. By the Möbius

inversion formula, we obtain

$$\begin{aligned} g(n) &= \sum_{d|n} \mu(d) f(n/d) = \sum_{d|n} \mu(d) \tfrac{1}{12}(n/d)^2 + \sum_{d|n} \mu(d)\epsilon_{n/d} \\ &> \frac{n^2}{12} \sum_{d|n} \frac{\mu(d)}{d^2} + \sum_{d|n} \mu(d)\epsilon_{n/d} \\ &> \frac{n^2}{12} \prod_{p \text{ prime}} \Big(1 - \frac{1}{p^2}\Big) + \sum_{d|n} \mu(d)\epsilon_{n/d}. \end{aligned}$$

Since

$$\prod_{p \text{ prime}} \Big(1 - \frac{1}{p^2}\Big) = \frac{6}{\pi^2},$$

we have

$$g(n) > \frac{n^2}{2\pi^2} + \sum_{d|n} \mu(d)\epsilon_{n/d}.$$

Now since the number of divisors of n is less than $2\sqrt{n}$, we obtain the following lemma:

Lemma 5.2. *We have*

$$\frac{n^2}{2\pi^2} - \frac{2}{3}\sqrt{n} < g(n).$$

The next lemma is the principal step in our proof. It gives information about the maximal overgroups in G of an element of the form $\mathrm{diag}({\Gamma_a}^{\alpha-2}, \Gamma_b, \Gamma_c)$, where the degrees a, b and c are coprime. The proof relies on knowledge of the subgroups of $\mathrm{GL}_n(q)$ whose order is divisible by a primitive prime divisor of $q^d - 1$, where $d > n/3$. An account of such subgroups has been given in the doctoral dissertation of Joseph DiMuro [2007]; this work extends the classification of [Guralnick et al. 1999], which deals with the case $d > n/2$.[1]

Lemma 5.3. *Let $\nu(n) \geq 3$, and let $n \geq 98$. For $\lambda = (a, b, c) \in P(n)$ with $a \leq b \leq c$ and with a, b and c coprime, let*

$$g = g_\lambda = \mathrm{diag}({\Gamma_a}^{\alpha-2}, \Gamma_b, \Gamma_c).$$

Then every maximal overgroup M of g in G is a subspace stabilizer except possibly in the following cases:

(i) $2 \mid n$, $c = n/2$ *and* $M \cong G \cap (\mathrm{GL}_{n/2}(q) \wr C_2)$.

[1] DiMuro's dissertation aims to classify elements of $\mathrm{GL}_n(q)$ of prime power order that act faithfully and irreducibly on a subspace of dimension $n/3$ or greater. However, we have been informed by its author that there is at present a gap in the argument concerning those elements whose orders are prime powers but not prime. For our purposes, only the results concerning elements of prime order are required.

(ii) $4 \mid n$, $(a, b, c) = (2, (n-2)/2, (n-2)/2)$ *and either* $M \cong G \cap (\mathrm{GL}_{n/2}(q) \wr C_2)$ *or* $M \cong G \cap (\mathrm{GL}_{n/2}(q) \circ \mathrm{GL}_2(q))$.

(Here $\circ$ is used to denote a central product.)

Proof. We observe that V may be decomposed as $V_a \oplus V_b \oplus V_c$, where V_a, V_b and V_c are g-invariant subspaces of dimensions a, b and c, respectively. The action of g on each of these summands is irreducible. It follows that g lies in the stabilizers of proper subspaces of at least four different dimensions, and so g is covered by the class $\mathrm{ss}(k)$ for at least four values of k.

Note that $c > n/3$ and that $q^c - 1$ divides the order of g. Hence, a maximal overgroup M of g must belong to one of the classes of groups mentioned in [DiMuro 2007, §1.2]. We observe firstly that owing to our assumption that $\nu \geq 3$ and $n \geq 98$, the subgroup M cannot be any of those in [DiMuro 2007, Tables 1.1–1.9]; this immediately rules out several of the Examples listed there. We shall go through the remaining Examples.

Example 1. *Classical examples.* The determinant of g is a generator of the quotient $G/\mathrm{SL}_n(q)$, and so M cannot contain $\mathrm{SL}_n(q)$.

Any element of a symplectic or orthogonal group is similar to its own inverse; an element g of a unitary group is similar to its conjugate inverse $g^{-\tau}$, where τ is induced by an involutory field automorphism. (See [Wall 1963, §2.6 or (3.7.2)] for groups in characteristic 2.)

If M normalizes a symplectic or orthogonal group H, then g^{q-1} lies in H itself, and so g^{q-1} is similar to its own inverse. Then it is clear that ${\Gamma_c}^{q-1}$ is similar to its own inverse (it does not matter here whether $b = c$). But this cannot be the case since $c > 2$.

Similarly, if M normalizes a unitary group U, then g^{q+1} lies in U, and it follows that g^{q+1} is similar to its conjugate inverse. But then it follows that ${\Gamma_c}^{q+1}$ is similar to its conjugate inverse, and it is easy to show that this is not the case.

Example 3. *Imprimitive examples.* Here M preserves a decomposition $V = U_1 \oplus \cdots \oplus U_t$ for $t \geq 2$. Let $\dim U_i = m$ so that $n = mt$. Recall that the $\langle g \rangle$-module V is the direct sum of three irreducible submodules V_a, V_b and V_c of dimensions a, b and c, respectively. So $\langle g \rangle$ has at most three orbits on the set of spaces U_i.

Let r be the smallest integer such that V_c is contained in the direct sum of r of the spaces U_i. We observe that $n/3 < c \leq rm$, and so $m > n/3r$. Without loss of generality, we may assume that $V_c \leq W = U_1 \oplus \cdots \oplus U_r$. It is clear that W is g-invariant. Let $\bar{g}$ be the restriction of g to W. Then $\langle \bar{g} \rangle$ acts transitively on $\{U_1, \ldots, U_r\}$. Since $\bar{g}^r$ acts in the same way on each U_i for $i \leq r$, an upper bound for the order of $\bar{g}$ is $(q^m - 1)r$. But since $m \leq n/r$ and since $n \geq 98$ by assumption,

we see that $(q^m - 1)r < q^{n/3} - 1$ if $r \geq 4$. Therefore, we must have $r \leq 3$.

It follows that V_c is a simple $\mathbb{F}_q\langle \bar{g}^r \rangle$-module. Now since $\bar{g}^r$ commutes with the projections of W onto its summands U_i, we see that at least one of the spaces U_i contains an $\bar{g}^r$-invariant subspace of dimension c. So $m > n/3$, and hence, $r \neq 3$.

Suppose that $r = 2$. Since $\bar{g}^r$ has two fixed spaces of dimension m, we see that $b = c = m$ and that $V_b \oplus V_c \leq W$. If $W < V$, then $W = V_b \oplus V_c$. Now we see that m divides each of a and $b + c = 2c$. Since a, b and c are coprime, it follows that $m = 2$. But this implies that $n < 6$, which contradicts the assumption that $n \geq 98$. So we may suppose that $W = V$. Then it is not hard to show that V_a has two irreducible summands as a $\langle \bar{g}^2 \rangle$-module. But this can occur only when $a = 2$, and this accounts for the first of the exceptional cases of the lemma.

Finally, if $r = 1$, then $m \geq c > n/3$, and so $t = 2$. It is easy to see, in this case, that $c = m = n/2$, and this accounts for the second exceptional case of the lemma.

Example 4. *Extension field examples.* If g stabilizes an $\mathbb{F}_{q^r}$-structure on V, then g^r lies in the image of an embedding of $\mathrm{GL}_{n/r}(q^r)$ into $\mathrm{GL}_n(q)$. Now if this is the case, then it is not hard, by considering the degrees of the eigenvalues of g over the fields $\mathbb{F}_q$ and $\mathbb{F}_{q^r}$, to show that r must divide each of a, b and c. But this implies that $r = 1$ since a, b and c are coprime.

Example 5. *Tensor product decomposition examples.* Here M stabilizes a nontrivial tensor product decomposition $V = V_1 \otimes V_2$. There is an embedding of the central product $\mathrm{GL}(V_1) \circ \mathrm{GL}(V_2)$ into $\mathrm{GL}_n(q)$, and M is the intersection of this group with G. For $x_1 \in \mathrm{GL}(V_1)$ and $x_2 \in \mathrm{GL}(V_2)$, we write (x_1, x_2) for the corresponding element of $\mathrm{GL}(V_1) \circ \mathrm{GL}(V_2)$.

We shall suppose that V_1 and V_2 have dimensions n_1 and n_2, respectively, with $n_1 \leq n_2$. Then since $c > n/3$, it is not hard to see that we have $n_1 = 2$.

Suppose $g \in M$, and let $g_1 \in \mathrm{GL}(V_1)$ and $g_2 \in \mathrm{GL}(V_2)$ be such that $g = (g_1, g_2)$. Let $h = g^{q^2-1}$. Since the order of g is coprime with q, we see that the element $g_1^{q^2-1}$ is the identity on V_1, and so $h = (1, h_2)$ for some $h_2 \in \mathrm{GL}(V)$.

The largest dimension of an irreducible $\langle h \rangle$-subspace of V is c, and there are at most two such subspaces. We obtain the $\langle h \rangle$-subspace decomposition of V up to isomorphism by taking two copies of each summand of the $\langle h_2 \rangle$-subspace decomposition of V_2. It follows that there must be at least two summands of dimension c and hence that $b = c$ and that $a < b$. It follows also that the a-dimensional summand of g splits into two summands as an $\mathbb{F}_q\langle h \rangle$-module. But it is not hard to see that this can occur only if $a = 2$, and so we have $a = 2$ and $b = c = (n-2)/2$. This is the second exceptional case of the lemma.

Example 6. *Subfield examples.* These cannot occur since g is built up using Singer cycles, which do not preserve any proper subfield structure.

Example 7. *Symplectic type examples.* This class of groups exists only in prime-power dimension and cannot occur in the cases we are considering since we have assumed that $\nu \geq 3$.

Example 8(a). *Permutation module examples.* In this case, S is an alternating group A_m for some $m \geq 5$. Then it is known that the order of an element in M is at most $(q-1) \cdot e^{\vartheta \sqrt{m \log m}}$, where $\vartheta = 1.05314$, by a result of Massias [1984]. Here $n = m-1$ or $m-2$. But a routine calculation shows that the inequality $e^{\vartheta \sqrt{(n+2)\log(n+2)}} < (q^{n/3}-1)/(q-1)$ holds for all $q \geq 2$ and for all $n \geq 98$. (This inequality fails when $q = 2$ and $n = 97$.)

Example 11. *Cross-characteristic groups of Lie type.* The examples not yet ruled out are contained in [DiMuro 2007, Table 1.10]. But the order of an element of M is less than n^3, which is less than $q^{n/3} - 1$ for $n \geq 98$. □

Proof of Theorem 1.4. Define a set Ω of classes of G by

$$\Omega = \left\{[\Gamma_n^{\alpha+q-1}]\right\} \cup \left\{[g_\lambda] \;\middle|\; \lambda \in P(n)\right\}.$$

Let $\mathscr{C}$ be a set of conjugacy classes of subgroups of G that covers Ω of the smallest size such that this is possible. Then clearly $|\mathscr{C}| \leq \gamma(G)$. By the theorem of Kantor [1980] mentioned in the proof of Lemma 3.1 above and by Lemma 5.3, we see that $\mathscr{C}$ must contain a single class of extension field subgroups. If $n \geq 98$ and $\nu \geq 3$, then each remaining element of $\mathscr{C}$ is either a class of subspace stabilizers or else one of the classes of subgroups mentioned in the exceptional cases of Lemma 3.1. Each subspace stabilizer contains at most $n/2$ of the elements g_λ, and each of the exceptional classes contains at most $n/4$. Now, using Lemma 5.2, we see that

$$\gamma(G) \geq |\mathscr{C}| \geq 1 + \frac{2g(n)}{n} > \frac{n}{\pi^2}$$

as required for the theorem.

To remove the conditions that $n \geq 98$ and that $\nu \geq 3$, it is enough to observe that the lower bound for $\kappa(G)$ given by Theorem 1.2 is larger than n/π^2 in any case where either of these conditions fails. □

References

[Britnell et al. 2008] J. R. Britnell, A. Evseev, R. M. Guralnick, P. E. Holmes, and A. Maróti, "Sets of elements that pairwise generate a linear group", *J. Combin. Theory Ser. A* **115**:3 (2008), 442–465. MR 2009b:20089 Zbl 1140.05033

[Bubboloni and Lucido 2002] D. Bubboloni and M. S. Lucido, "Coverings of linear groups", *Comm. Algebra* **30**:5 (2002), 2143–2159. MR 2003h:20088 Zbl 1007.20030

[Bubboloni and Praeger 2011] D. Bubboloni and C. E. Praeger, "Normal coverings of finite symmetric and alternating groups", *J. Combin. Theory Ser. A* **118**:7 (2011), 2000–2024. MR 2012e:20006 Zbl 1248.20005

[Bubboloni et al. 2006] D. Bubboloni, M. S. Lucido, and T. Weigel, "Generic 2-coverings of finite groups of Lie type", *Rend. Sem. Mat. Univ. Padova* **115** (2006), 209–252. MR 2007f:20087 Zbl 1156.20039

[Bubboloni et al. 2011] D. Bubboloni, M. S. Lucido, and T. Weigel, "2-Coverings of classical groups", preprint, 2011. arXiv 1102.0660v1

[Bubboloni et al. 2012] D. Bubboloni, F. Luca, and P. Spiga, "Compositions of n satisfying some coprimality conditions", *J. Number Theory* **132**:12 (2012), 2922–2946. MR 2965200 Zbl 1251.05011

[Bubboloni et al. 2013] D. Bubboloni, C. E. Praeger, and P. Spiga, "Normal coverings and pairwise generation of finite alternating and symmetric groups", *J. Algebra* **390** (2013), 199–215. MR 3072118

[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, Eynsham, 1985. MR 88g:20025 Zbl 0568.20001

[Crestani and Lucchini 2012] E. Crestani and A. Lucchini, "Normal coverings of solvable groups", *Arch. Math.* (*Basel*) **98**:1 (2012), 13–18. MR 2012m:20043 Zbl 1255.20027

[DiMuro 2007] J. DiMuro, *On prime power elements of* $\mathrm{GL}_d(q)$ *acting irreducibly on large subspaces*, Ph.D. thesis, University of Southern California, 2007, Available at http://digitallibrary.usc.edu/assetserver/controller/item/etd-DiMuro-20071119.pdf.

[Guralnick and Malle 2012] R. Guralnick and G. Malle, "Simple groups admit Beauville structures", *J. Lond. Math. Soc.* (2) **85**:3 (2012), 694–721. MR 2927804 Zbl 1255.20009

[Guralnick et al. 1999] R. Guralnick, T. Penttila, C. E. Praeger, and J. Saxl, "Linear groups with orders having certain large prime divisors", *Proc. London Math. Soc.* (3) **78**:1 (1999), 167–214. MR 99m:20113 Zbl 1041.20035

[Kantor 1980] W. M. Kantor, "Linear groups containing a Singer cycle", *J. Algebra* **62**:1 (1980), 232–234. MR 81g:20089 Zbl 0429.20004

[Lehmer 1955] D. H. Lehmer, "The distribution of totatives", *Canad. J. Math.* **7** (1955), 347–357. MR 16,998i Zbl 0064.27902

[Massias 1984] J.-P. Massias, "Majoration explicite de l'ordre maximum d'un élément du groupe symétrique", *Ann. Fac. Sci. Toulouse Math.* (5) **6**:3-4 (1984), 269–281. MR 87a:11093 Zbl 0574.10043

[Wall 1963] G. E. Wall, "On the conjugacy classes in the unitary, symplectic and orthogonal groups", *J. Austral. Math. Soc.* **3** (1963), 1–62. MR 27 #212 Zbl 0122.28102

[Zsigmondy 1892] K. Zsigmondy, "Zur Theorie der Potenzreste", *Monatsh. Math. Phys.* **3**:1 (1892), 265–284. MR 1546236 Zbl 24.0176.02

Communicated by David Benson
Received 2012-07-28 Revised 2012-11-01 Accepted 2013-01-14

j.britnell@imperial.ac.uk *Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*

maroti.attila@renyi.mta.hu *Hungarian Academy of Sciences, Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13–15, H-1053, Budapest, Hungary*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2103

Modularity of the concave composition generating function

George E. Andrews, Robert C. Rhoades and Sander P. Zwegers

A composition of an integer constrained to have decreasing then increasing parts is called concave. We prove that the generating function for the number of concave compositions, denoted $v(q)$, is a mixed mock modular form in a more general sense than is typically used.

We relate $v(q)$ to generating functions studied in connection with "Moonshine of the Mathieu group" and the smallest parts of a partition. We demonstrate this connection in four different ways. We use the elliptic and modular properties of Appell sums as well as q-series manipulations and holomorphic projection.

As an application of the modularity results, we give an asymptotic expansion for the number of concave compositions of n. For comparison, we give an asymptotic expansion for the number of concave compositions of n with strictly decreasing and increasing parts, the generating function of which is related to a false theta function rather than a mock theta function.

1. Introduction

A composition of an integer n is a sum of positive integers adding to n, in which order matters. The study of compositions has a long history dating back to MacMahon [1893]. The book of Heubach and Mansour [2010] contains more on the history of compositions. It is natural to impose restrictions on the ascents or descents of consecutive parts of a composition. For instance, compositions with no ascents correspond to integer partitions.

A concave composition of n is a sum of integers of the form

$$\sum_{i=1}^{L} a_i + c + \sum_{i=1}^{R} b_i = n$$

where $a_1 \geq \cdots \geq a_L > c < b_1 \leq \cdots \leq b_R$, where $c \geq 0$ is the central part of the composition. Let $V(n)$ be the number of concave compositions of n. For

MSC2010: primary 05A17; secondary 11P82, 11F03.
Keywords: concave composition, partition, unimodal sequences, mock theta function, mixed mock modular form.

example, $V(3) = 13$ since $\{0,3\}$, $\{3,0\}$, $\{0,1,2\}$, $\{2,1,0\}$, $\{0,1,1,1\}$, $\{1,1,1,0\}$, $\{1,2\}$, $\{2,1\}$, $\{1,0,2\}$, $\{2,0,1\}$, $\{1,0,1,1\}$, $\{1,1,0,1\}$, and $\{3\}$ are all concave sequences. In [Andrews 2011] it was shown that the generating function for concave compositions with further restrictions is related to statistics for spiral self-avoiding random walks as well as other partition problems.

Standard techniques show that the generating function for the sequence $\{V(n)\}_{n=0}^{\infty}$ is given by

$$v(q) := \sum_{n=0}^{\infty} V(n)q^n = \sum_{n=0}^{\infty} \frac{q^n}{(q^{n+1};q)_\infty^2}$$

where $(x)_n = (x;q)_n := \prod_{j=0}^{n-1}(1-xq^j)$ and $(x)_\infty = (x;q)_\infty := \prod_{j=0}^{\infty}(1-xq^j)$; see [Andrews 2013]. We establish the modularity properties of $v(q)$ as a mixed mock modular form.

Theorem 1.1. *Let $q = e^{2\pi i\tau}$ and $\tau \in \mathbb{H}$. Define $f(\tau) = q(q)_\infty^3 v(q)$ and*

$$\widehat{f}(\tau) := f(\tau) - \frac{i}{2}\eta(\tau)^3 \int_{-\overline{\tau}}^{i\infty} \frac{\eta(z)^3}{(-i(z+\tau))^{\frac{1}{2}}}\,dz + \frac{\sqrt{3}}{2\pi i}\,\eta(\tau) \int_{-\overline{\tau}}^{i\infty} \frac{\eta(z)}{(-i(z+\tau))^{\frac{3}{2}}}\,dz$$

where the Dedekind η-function is given by $\eta(\tau) = q^{1/24}(q)_\infty$. The function $\widehat{f}$ transforms as a modular form of weight 2 for $\mathrm{SL}_2(\mathbb{Z})$.

Remark. Theorem 1.1 was used by Bryson, Ono, Pitman and the second author [Bryson et al. 2012] to show that the modular form $f(\tau)/(q)_\infty = q(q)_\infty^2 v(q)$ produces a quantum modular form. The q-hypergeometric series defining $q(q)_\infty^2 v(q)$ was shown to be dual to Kontsevich's strange function

$$F(q) := \sum_{n=0}^{\infty} (q;q)_n,$$

which is defined only when q is a root of unity and was studied by Zagier [2001].

Following Zagier [2009], a *mock theta function of weight* $k \in \{\frac{1}{2}, \frac{3}{2}\}$ is a q-series $H(q) = \sum_{n=0}^{\infty} a_n q^n$ such that there exists a rational number λ and a unary theta function of weight $2-k$, $g(\tau) = \sum_{n\in\mathbb{Q}^+} b_n q^n$, where $q = e^{2\pi i\tau} = e^{2\pi i(x+iy)}$, such that $h(\tau) = q^\lambda H(q) + g^*(\tau)$ is a nonholomorphic modular form of weight k, where

$$g^*(\tau) = (i/2)^{k-1} \sum_{n\in\mathbb{Q}^+} n^{k-1}\overline{b}_n \Gamma(1-k, 4\pi n y) q^{-n} \tag{1-1}$$

and $\Gamma(k,t) = \int_t^\infty u^{k-1}e^{-u}\,du$ is the incomplete gamma function. Such a nonholomorphic modular form is called a harmonic weak Maass form (see Section 2 for a definition and Ono's surveys [2008; 2009] for history). The theta function g is called the *shadow* of the mock theta function H.

In [Andrews 2013] the following identity is established, which is crucial in establishing Theorem 1.1. We have

$$v(q) = q^{-1}(v_1(q) + v_2(q) + v_3(q)) \tag{1-2}$$

where

$$v_1(q) := \frac{1}{(q)_\infty^3}\left(\sum_{n\neq 0} \frac{(-1)^{n+1} q^{3n(n+1)/2}}{(1-q^n)^2} - 3\sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)^2} + \tfrac{1}{12}\right), \tag{1-3}$$

$$v_2(q) := \frac{1}{(q)_\infty^3}\left(\sum_{n\neq 0} \frac{(-1)^{n+1} n q^{n(n+1)/2}}{1-q^n} - \tfrac{1}{4} - 2\sum_{n=1}^{\infty} \frac{q^n}{(1+q^n)^2}\right), \tag{1-4}$$

$$v_3(q) := \frac{1}{(q)_\infty^3}\left(\tfrac{1}{6} + 2\sum_{n=1}^{\infty} q^n\left(\frac{1}{(1+q^n)^2} + \frac{1}{(1-q^n)^2}\right)\right). \tag{1-5}$$

Theorem 1.1 may be recast in the following terms.

Theorem 1.2. *With $q = e^{2\pi i z}$ and $z \in \mathbb{H}$ we have*:

(1) *$q^{-1/24}(q)_\infty^2 v_1(q)$ is a mock theta function of weight $\frac{3}{2}$ with shadow proportional to $\eta(z)$. Consequentially, $v_1(q)$ is a mixed mock modular form.*

(2) *$q^{-1/8} v_2(q)$ is a mock theta function of weight $\frac{1}{2}$ with shadow proportional to $\eta(z)^3$.*

(3) *$q^{-1/8} v_3(q)$ is a modular form of weight $\frac{1}{2}$.*

Remark. Theorem 4.1 gives the level and shadow for each of the corresponding harmonic weak Maass form.

The mock theta function $v_2(q)$ has appeared in a number of interesting places. For example, it arises in the work of Eguchi, Ooguri, and Tachikawa [Eguchi et al. 2011] and Cheng [2010] concerned with the character table of the Mathieu group M_{24} and "Moonshine of the Mathieu group". It also appears in [Malmendier and Ono 2012], which deals with Donaldson invariants of $\mathbb{C}P^2$. In that work $v_2(q)$ arises in a different form, which is equivalent to the following identity. Moreover, the claim for $v_2(q)$ in Theorem 1.2 follows from the next theorem and the results in [Zwegers 2002].

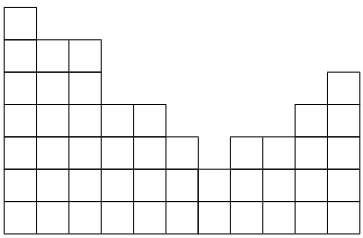
Theorem 1.3. *For $|q| < 1$ we have*

$$\begin{aligned}\widetilde{F}(q) &:= \frac{1}{(q)_\infty(-q)_\infty^2}\sum_{n\in\mathbb{Z}} \frac{q^{n(n+1)/2}}{1+q^n} \\ &= \frac{1}{(q)_\infty^3}\left(\tfrac{1}{2} + 4\sum_{n\geq 1} \frac{q^n}{(1+q^n)^2} - 2\sum_{n\neq 0} \frac{(-1)^{n+1} n q^{n(n+1)/2}}{1-q^n}\right).\end{aligned}$$

We give three different ways of obtaining this identity. Each proof relies on different symmetries of Appell sums. The first uses an elliptic transformation property and Taylor expansions of the Jacobi theta function. The second proof uses the modular properties of the Appell sums. The third is via q-series manipulations. Finally, we sketch a connection with the holomorphic projection construction for mock modular forms.

Remark. Theorem 11.1 gives an analogous result for the Appell sum in the definition of $v_1(q)$. Moreover, it relates $v_1(q)$ to the smallest parts generating function studied in [Andrews 2008].

There is a convenient graphical representation of a composition, where each part is represented by a vertical column of boxes; for example,



represents a concave composition of 47. Considering each composition of n, possibly from a restricted subset, it is natural to ask about the average limiting behavior of the graphical representation as $n \to \infty$. For example, there is a great deal of literature about the limiting shape of integer partitions; see, for instance, [Fristedt 1993; Pittel 1997; Vershik 1995].

Properties of the typical representation are often studied via probabilistic models. However, when the generating functions are modular forms, very strong theorems can be proved for the statistics of interest. For instance, the modularity of the generating function for the number of partitions and the circle method yield the following asymptotic expansion for the number of partitions of n, denoted $p(n)$:

$$p(n) = \frac{2\sqrt{3}}{(24n-1)} \exp\left(\frac{\pi}{6}\sqrt{24n-1}\right)\left(1 - \frac{6}{\pi\sqrt{24n-1}} + O\left(\frac{1}{n^T}\right)\right)$$

for any $T \geq 0$.

As an application of the modular properties of $v(q)$ we give an asymptotic expansion for $V(n)$. Since the generating function is a mixed mock modular form, the circle method may be applied to obtain asymptotics for $V(n)$. Care must be taken to deal with contributions from the holomorphic Fourier series as well as the nonholomorphic period integrals. We use methods of Bringmann and Mahlburg [2011; 2013] for dealing with these difficulties.

Theorem 1.4. *For any $T \geq 1$ as $n \to \infty$, we have the asymptotic expansion*

$$V(n) = \frac{2\sqrt{6}}{(12n+\frac{21}{2})^{3/4}} \exp\Bigl(\frac{\pi}{3}\sqrt{12n+\tfrac{21}{2}}\Bigr) \times \Bigl(\sum_{t=1}^{T} \frac{(2t-1)!!\,3^t\alpha_t}{2^{2t}\pi^t} \frac{1}{(12n+\frac{21}{2})^{t/2}} + O\Bigl(\frac{1}{n^{(T+1)/2}}\Bigr)\Bigr)$$

where α_t *are defined by*

$$\sum_{k=1}^{\infty} \alpha_k x^{2k} := \exp\bigl(\sqrt{1-2x^2}-1+x^2\bigr)x\,\frac{\sinh\bigl(\frac{2\pi x}{3}\bigr)}{\cosh(\pi x)}.$$

In particular, $\alpha_1 = 2\pi/3$.

Remark. It is possible to obtain an asymptotic with a polynomial error term; see [Bringmann and Mahlburg 2011; 2013; Rhoades 2012].

Remark. A concave composition corresponds to a triple (λ, μ, c) where λ and μ are partitions, and $c \geq 0$ is an integer strictly smaller than the smallest parts in λ and μ, such that $n = |\lambda| + |\mu| + c$. Fristedt's results [1993] imply that partitions of size n have a part of size 1 with probability roughly equal to $1-(\pi/\sqrt{6n})$. We expect at least one of the partitions μ or λ to have size not much smaller than $n/2$. That partition will almost surely contain a part of size 1. Thus we expect $c = 0$ for most triples. Therefore, we expect that $V(n)$ will agree to leading order with the asymptotic for the number of pairs of partitions (μ, λ) with $|\mu| + |\lambda| = n$. Standard circle method calculations show this number is

$$\frac{\sqrt{6}}{(12n-1)^{5/4}} \exp\Bigl(\frac{\pi}{3}\sqrt{12n-1}\Bigr)\Bigl(1+O\Bigl(\frac{1}{\sqrt{n}}\Bigr)\Bigr).$$

As expected this agrees with the leading order asymptotic of $V(n)$.

For comparison we introduce the notion of strongly concave compositions. A strongly concave composition of n is a sum of integers of the form

$$\sum_{i=1}^{L} a_i + c + \sum_{i=1}^{R} b_i = n$$

where $a_1 > \cdots > a_L > c < b_1 < \cdots < b_R$ and where $c \geq 0$. Let $V_d(n)$ be the number of strongly concave compositions of n. We have

$$\begin{aligned} v_d(q) := \sum_{n=0}^{\infty} V_d(n)\,q^n &= \sum_{n=0}^{\infty} q^n(-q^{n+1})_\infty(-q^{n+1})_\infty \\ &= -\sum_{n=0}^{\infty}(-1)^n q^{n(n+1)/2} + 2(-q)_\infty^2 \sum_{n\geq 0}\Bigl(\frac{-12}{n}\Bigr) q^{(n^2-1)/24}, \end{aligned} \tag{1-6}$$

where the second equality follows from standard techniques and the third equality is established in [Andrews 2013] and $(\frac{\cdot}{\cdot})$ is the Kronecker symbol. The function $\sum_{n\geq 0}(-12/n)q^{(n^2-1)/24}$ is called a partial theta function. The asymptotic behavior of the partial theta function is discussed in Section 10 and used to derive the following theorem for $V_d(n)$.

Theorem 1.5. *For any $T \geq 0$ as $n \to \infty$, we have*

$$V_d(n) = \frac{\sqrt{3}}{(24n+2)^{3/4}} \exp\Bigl(\frac{\pi}{6}\sqrt{24n+2}\Bigr) \times \Bigl(\sum_{m=0}^{T} \frac{(-1)^m}{2^{2m} m!\,(24n+2)^{m/2}}\,\gamma(m) + O\Bigl(\frac{1}{n^{(T+1)/2}}\Bigr)\Bigr),$$

where

$$\gamma(m) = \sum_{a+b=m} \binom{m}{a} L(-2a)\Bigl(\frac{\pi}{3}\Bigr)^{a-b} p(b,a)$$

with $p(b,a) = \prod_{j=0}^{b}(4(a-1)^2 - j^2)$ and

$$L(-r) = -\frac{6^r}{r+1}\bigl(B_{r+1}(\tfrac{1}{6}) - B_{r+1}(\tfrac{5}{6})\bigr),$$

where $B_r(x)$ is the r-th Bernoulli polynomial.

Remark. Let $\mathrm{sm}(\lambda)$ be the smallest part in the partition λ. Strongly concave compositions are characterized by a pair of partitions into distinct parts (λ, μ) with $\mathrm{sm}(\lambda) \neq \mathrm{sm}(\mu)$. Let $\widetilde{u}(n)$ be the number of pairs of partitions into distinct parts with sizes summing to n. Therefore, we expect the asymptotic of

$$V_d(n) \sim (1 - \mathrm{Prob}\{\mathrm{sm}(\lambda) = \mathrm{sm}(\mu) : |\lambda| + |\mu| = n\})\,\widetilde{u}(n).$$

We have

$$V_d(n) \sim \tfrac{2}{3}\frac{\sqrt{3}}{(24n)^{3/4}} \exp\Bigl(\frac{\pi}{6}\sqrt{24n}\Bigr) \quad \text{and} \quad \widetilde{u}(n) \sim \frac{\sqrt{3}}{(24n)^{3/4}} \exp\Bigl(\frac{\pi}{6}\sqrt{24n}\Bigr).$$

It follows from Fristedt [1993, Theorem 9.1] that the smallest part of a partition into distinct parts has size j with probability roughly equal to $1/2^j$. Therefore, a pair of partitions into distinct parts will have the same smallest part with probability roughly equal to $\frac{1}{3}$, which agrees with the prediction.

It would be of interest to address some of the following questions as $n \to \infty$:

(1) What is the distribution of the center part of a (strongly) concave composition?

(2) How many parts does a typical (strongly) concave composition of n have?

(3) What is the distribution of the number of parts to the left of the center part minus the number of parts to the right of the center part?

(4) What is the distribution of the perimeter of the (strongly) concave composition of n?

Some of these questions can be answered by modeling a concave composition as the convolution of two random partitions (discussed above), while others can be treated via modular techniques of Bringmann, Mahlburg and the second author [Bringmann et al. 2012].

It is convenient to introduce a two-variable generating function. Let $R-L$ be the *rank* of the concave composition. This quantity measures the position of the central part. Let $V(m,n)$ and $V_d(m,n)$ be the numbers of concave compositions and strongly concave compositions, respectively, of n with rank equal to m. Standard techniques give

$$v(x,q) := \sum_{\substack{n\geq 0\\ m\in\mathbb{Z}}} V(m,n)x^m q^n = \sum_{n=0}^{\infty} \frac{q^n}{(xq^{n+1};q)_\infty (x^{-1}q^{n+1};q)_\infty}, \tag{1-7}$$

$$v_d(x,q) := \sum_{\substack{n\geq 0\\ m\in\mathbb{Z}}} V_d(m,n)x^m q^n = \sum_{n=0}^{\infty} q^n(-xq^{n+1};q)_\infty(-x^{-1}q^{n+1};q)_\infty. \tag{1-8}$$

The following identities are deduced in a similar manner to Theorem 1 of [Andrews 2013].

Theorem 1.6. *In the notation above,*

$$v_d(x,q) = -x\sum_{n=0}^{\infty}(-1)^n x^{2n} q^{\frac{n(n+1)}{2}} + (x)_\infty (x^{-1}q)_\infty \sum_{n=0}^{\infty}\left(\frac{12}{n}\right) x^{\frac{n-1}{2}} q^{\frac{n^2-1}{24}},$$

$$qv(x,q) = -\frac{1}{(x)_\infty (x^{-1})_\infty}\sum_{n=0}^{\infty}\frac{q^{n^2}}{(xq)_n (x^{-1}q)_n} + \frac{(1-x)}{(q)_\infty (x)_\infty (x^{-1})_\infty}\sum_{n\in\mathbb{Z}}\frac{(-1)^n q^{(n^2+n)/2}x^{-n}}{1-xq^n}.$$

One should compare these generating functions to the rank of a partition studied in [Bringmann et al. 2012]. As is the case in [Bringmann et al. 2012] the moments of this rank statistic can be calculated precisely. We expect this statistic to be asymptotically asymptotically distributed; see [Diaconis et al. 2013] for a similar result in the case of the partition rank.

In Section 2 we recall some basic facts about holomorphic and nonholomorphic modular forms. In Section 3 we prove Theorem 1.1. In Section 4 we prove Theorem 1.2. In Section 5 we give the first proof of Theorem 1.3, which proceeds

via elliptic properties of Appell sums. In Section 6 we use q-series manipulations to prove Theorem 1.3. In Section 7 we prove Theorem 1.3 via "modular methods". Section 8 contains a calculation of the holomorphic projection operation which relates to Theorem 1.3. In Sections 9 and 10 we prove Theorems 1.4 and 1.5. Both use the circle method, modular properties of the generating functions of interest, and asymptotic analysis. Finally, in Section 11 we discuss the analog of Theorem 1.3 for the function $v_1(q)$.

Throughout the remainder of the text we let $q := e^{2\pi i\tau}$, with $\tau = x + iy$ for $x, y \in \mathbb{R}$ and $y > 0$. We let $z \in \mathbb{R}^+$, $h, k \in \mathbb{N}_0$, $0 \le h \le k$ with $(h, k) = 1$. Moreover, we denote by $[a]_b$ the inverse of a modulo b.

2. Holomorphic modular forms and harmonic weak Maass forms

In this section we define and give some basic properties of harmonic weak Maass forms. Before turning to nonholomorphic modular forms we describe the classic holomorphic modular forms of half integral weight.

We follow Shimura [1973], see also [Ono 2004], by setting

$$\left(\frac{a}{b}\right) = \eta\left(\frac{a}{|b|}\right)$$

where $\eta = -1$ if $a, b, < 0$ and $\eta = 1$ if $a > 0$ or $b > 0$. For an odd integer m, we put $\epsilon_m = 1$ if $m \equiv 3 \pmod 4$ and $\epsilon_m = i$ if $m \equiv 3 \pmod 4$. For all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$, let

$$j(\gamma, \tau) := \left(\frac{c}{d}\right) \epsilon_d^{-1} (c\tau + d)^{1/2}. \tag{2-1}$$

The Dedekind η-function is defined by $\eta(\tau) := q^{1/24} \prod_{n=1}^{\infty} (1 - q^n)$ and define $\chi(h, [-h]_k, k)$ to be the multiplier so that

$$\eta\left(\frac{1}{k}(h + iz)\right) = \sqrt{\frac{i}{z}}\, \chi(h, [-h]_k, k)\, \eta\left(\frac{1}{k}\left([-h]_k + \frac{i}{z}\right)\right). \tag{2-2}$$

By [Ono 2004, Theorem 1.60], for instance,

$$\eta(8\tau)^3 = \sum_{n\in\mathbb{Z}} (-1)^n (2n+1) e^{2\pi i (2n+1)^2 \tau} = \sum_{n=1}^{\infty} \left(\frac{-4}{n}\right) n q^{n^2}. \tag{2-3}$$

The next result follows from Theorem 1.44 of [Ono 2004].

Lemma 2.1. *$\eta(8\tau)^3$ is a weight-$\frac{3}{2}$ modular form on $\Gamma_0(64)$ with trivial Nebentypus.*

Before discussing harmonic weak Maass forms we introduce the quasimodular form

$$E_2(\tau) := 1 - 24 \sum_{n=1}^{\infty} \sigma_1(n) q^n, \tag{2-4}$$

where $\sigma_1(n) := \sum_{d|n} d$. In particular it satisfies $E_2(\tau+1) = E_2(\tau)$ and

$$E_2(\tau) = \tau^{-2} E_2\Big(-\frac{1}{\tau}\Big) - \frac{6}{\pi i \tau}. \tag{2-5}$$

Therefore, the completed form of E_2 defined by

$$\widehat{E}_2(\tau) := E_2(\tau) - \frac{3}{\pi y} \tag{2-6}$$

transforms as a modular form of weight 2 for $\mathrm{SL}_2(\mathbb{Z})$.

We turn to harmonic weak Maass forms. Define the weight-k hyperbolic Laplacian by

$$\Delta_k := -y^2\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big) + iky\Big(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\Big). \tag{2-7}$$

Definition 2.2. Suppose that $k \in \frac{1}{2}\mathbb{Z}$, N is a positive integer, and that ψ is a Dirichlet character with modulus $4N$. A *harmonic weak Maass* form of weight k on $\Gamma_0(4N)$ with Nebentypus character ψ is a smooth function $f : \mathbb{H} \to \mathbb{C}$ satisfying the following:

(1) For all $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4N)$ and all $\tau \in \mathbb{H}$, we have

$$f(A\tau) = \psi(d) j(A,\tau)^{2k} f(\tau).$$

(2) We have $\Delta_k f = 0$.

(3) The function f has at most linear exponential growth at all the cusps of $\mathbb{H}/\Gamma_0(N)$.

In [Zwegers 2002] a general class of harmonic weak Maass forms was constructed by "completing" certain Appell sums. The Appell sum is defined for $u, v \in \mathbb{C} \setminus (\mathbb{Z} + \mathbb{Z}\tau)$ by

$$\mu(u,v;\tau) := \frac{e^{\pi i u}}{\vartheta(v;\tau)} \sum_{n\in\mathbb{Z}} \frac{(-1)^n e^{2\pi i n v} q^{n(n+1)/2}}{1 - e^{2\pi i u} q^n}, \tag{2-8}$$

where

$$\vartheta(v;\tau) := \sum_{\nu \in \frac{1}{2}+\mathbb{Z}} e^{\pi i \nu^2 \tau + 2\pi i \nu (v + \frac{1}{2})}$$

is the Jacobi theta function. The Jacobi theta function satisfies the triple product identity

$$\vartheta(v;\tau) = -i q^{1/8} \zeta^{-1/2} \prod_{n=1}^{\infty} (1-q^n)(1-\zeta q^{n-1})(1-\zeta^{-1} q^n) \tag{2-9}$$

with $\zeta = e^{2\pi i v}$, and the transformation

$$\vartheta\Big(-i\,vz;\frac{1}{k}(h+iz)\Big)=\chi^3(h,[-h]_k,k)\sqrt{\frac{i}{z}}\,e^{\pi kzv^2}\vartheta\Big(v,\frac{1}{k}\Big([-h]_k+\frac{i}{z}\Big)\Big). \tag{2-10}$$

The nonholomorphic correction term of the Appell sum requires the definition

$$R(u;\tau):=\sum_{\nu\in\mathbb{Z}+\frac{1}{2}}(-1)^{\nu-\frac{1}{2}}\Big\{\operatorname{sgn}(\nu)-E\Big(\Big(\nu+\frac{\operatorname{Im}(u)}{\operatorname{Im}(\tau)}\Big)\sqrt{2\operatorname{Im}(\tau)}\Big)\Big\}e^{-2\pi i\nu u}e^{-\pi i\nu^2\tau},$$

with $E(x)$ defined by

$$E(x):=2\int_0^x e^{-\pi u^2}\,du=\operatorname{sgn}(x)(1-\beta(x^2)), \tag{2-11}$$

where for positive real x we let $\beta(x):=\int_x^\infty u^{-1/2}e^{-\pi u}\,du$. We have the following useful properties of R.

Proposition 2.3 [Zwegers 2002, Propositions 1.9 and 1.10]. *If $u\in\mathbb{C}$ and $\operatorname{Im}(\tau)>0$, then*

(1) $R(u+1;\tau)=R(-u;\tau)=-R(u;\tau)$,

(2) $R(u;\tau+1)=e^{-\pi i/4}R(u;\tau)$,

(3) $R(u;\tau)=-\sqrt{\frac{i}{\tau}}e^{i\pi u^2/\tau}\Big(R\Big(\frac{u}{\tau};-\frac{1}{\tau}\Big)-H\Big(\frac{u}{\tau};-\frac{1}{\tau}\Big)\Big)$,

where the Mordell integral is defined by

$$H(u;z):=\int_{-\infty}^{\infty}\frac{e^{\pi izx^2-2\pi xu}}{\cosh(\pi x)}\,dx.$$

Moreover, we need the following "dissection" property of R.

Proposition 2.4 [Bringmann and Folsom 2013, Proposition 2.3]. *For $n\in\mathbb{N}$, we have*

$$R\Big(u;\frac{z}{n}\Big)=\sum_{\ell=0}^{n-1}q^{-\frac{1}{2n}(\ell-\frac{n-1}{2})^2}e^{-2\pi i(\ell-\frac{n-1}{2})(u+\frac{1}{2})}R\Big(nu+\Big(\ell-\frac{n-1}{2}\Big)z+\frac{n-1}{2};nz\Big).$$

The completion of μ is defined by

$$\widehat{\mu}(u,v;\tau):=\mu(u,v;\tau)+\frac{i}{2}R(u-v;\tau). \tag{2-12}$$

This function satisfies the following elliptic and modular transformation laws.

Theorem 2.5 [Zwegers 2002, Theorem 1.11]. *Assume all of the notation and hypotheses from above. If $k,\ell,m,n\in\mathbb{Z}$, then we have*

$$\widehat{\mu}(u+k\tau+\ell,v+m\tau+n;\tau)=(-1)^{k+\ell+m+n}e^{\pi i\tau(k-m)^2+2\pi i(k-m)(u-v)}\widehat{\mu}(u,v;\tau)$$

and

$$\widehat{\mu}\Big(-iuz,-ivz;\frac{h+iz}{k}\Big)=\chi^{-3}(h,[-h]_k,k)\sqrt{\frac{i}{z}}e^{-\pi kz(u-v)^2}\widehat{\mu}\Big(u,v;\frac{1}{k}\Big([-h]_k+\frac{i}{z}\Big)\Big).$$

Finally, we have a result that is useful in determining the shadow of a mock theta function.

Theorem 2.6 [Zwegers 2002, Theorem 1.16]. *For $a\in(-\frac{1}{2},\frac{1}{2})$ and $b\in\mathbb{R}$ we have*

$$\int_{-\overline{\tau}}^{i\infty}\frac{g_{a+\frac{1}{2},b+\frac{1}{2}}(z)}{\sqrt{-i(\tau+z)}}\,dz=-e^{-\pi ia^2\tau+2\pi ia(b+\frac{1}{2})}R(a\tau-b;\tau)$$

where $g_{a,b}(z):=\sum_{\nu\in a+\mathbb{Z}}\nu e^{\pi i\nu^2z+2\pi i\nu b}$.

3. Proof of Theorem 1.1

In this section we use the Jacobi properties of the Appell sums

$$\begin{aligned}A(u,v;\tau)&:=e^{\pi iu}\sum_{n\in\mathbb{Z}}\frac{(-1)^ne^{2\pi inv}q^{(n^2+n)/2}}{1-e^{2\pi iu}q^n},\\ A_3(u;\tau)&:=e^{\pi iu}\sum_{n\in\mathbb{Z}}\frac{(-1)^nq^{(3n^2+n)/2}}{1-e^{2\pi iu}q^n},\end{aligned}\tag{3-1}$$

to deduce Theorem 1.1. The function $A_3(u;\tau)$ was studied by Bringmann [2008].

Direct computation gives

$$f_2(\tau):=\sum_{n\neq0}\frac{(-1)^{n+1}q^{\frac{3n(n+1)}{2}}}{(1-q^n)^2}=\Big(\frac{1}{2\pi i}\frac{\partial}{\partial u}(e^{-\pi iu}A_3(u,\tau)-\frac{1}{1-e^{2\pi iu}})\Big)\Big|_{u=0},$$

$$f_1(\tau):=\sum_{n\neq0}\frac{(-1)^{n+1}nq^{\frac{n(n+1)}{2}}}{1-q^n}=-\frac{1}{2\pi i}\frac{\partial}{\partial v}\Big|_{u=v=0}A(u,v;\tau).$$

Thus the modular properties of the Appell sums will dictate the modular properties of our functions $v_1(q)$ and $v_2(q)$. In analogy to (2-12) the completed forms of these sums are

$$\widehat{A}(u,v;\tau):=A(u,v;\tau)+\frac{i}{2}\vartheta(v;\tau)R(u-v;\tau),$$

$$\widehat{A}_3(u;\tau):=A_3(u;\tau)+\tfrac{1}{2}\eta(\tau)q^{-\frac{1}{6}}(e^{2\pi iu}R(3u-\tau;3\tau)-e^{-2\pi iu}R(3u+\tau;3\tau)).$$

Using the transformation properties of μ found in Chapter 1 of [Zwegers 2002], we have the following modular properties of $\widehat{A}$. For $\widehat{A}_3$ we will use Theorem 3.1 of [Bringmann 2008].

Proposition 3.1. *For all* $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ *we have*

$$\widehat{A}\Big(\frac{u}{c\tau+d}, \frac{v}{c\tau+d}; \frac{a\tau+b}{c\tau+d}\Big) = (c\tau+d)e^{\pi i \frac{c}{c\tau+d}(2uv-v^2)}\widehat{A}(u,v;\tau).$$

We then have the following theorem which together with the transformation properties of $\widehat{E}_2(\tau)$ and (1-2) yields Theorem 1.1.

Theorem 3.2. *With* $\tau = x+iy$ *for* $x, y \in \mathbb{R}$ *and* $y > 0$ *we let*

$$\widehat{f}_1(\tau) = f_1(\tau) - \frac{1}{4\pi y} - \frac{i}{2}\eta(\tau)^3 \int_{-\overline{\tau}}^{i\infty} \frac{\eta(z)^3}{(-i(z+\tau))^{\frac{1}{2}}}\, dz,$$

$$\widehat{f}_2(\tau) = f_2(\tau) - \tfrac{1}{24} + \frac{3}{8\pi y} + \frac{\sqrt{3}}{2\pi i}\eta(\tau) \int_{-\overline{\tau}}^{i\infty} \frac{\eta(z)}{(-i(z+\tau))^{\frac{3}{2}}}\, dz.$$

Both $\widehat{f}_1$ *and* $\widehat{f}_2$ *transform as a weight* 2 *modular form for all of* $\mathrm{SL}_2(\mathbb{Z})$ *with trivial Nebentypus.*

Proof. We establish the claim for f_1 first. We have

$$\begin{aligned}\frac{1}{2\pi i}\frac{\partial}{\partial v}\Big|_{u=v=0}\widehat{A}(u,v;\tau) &= -f_1(\tau) + \frac{1}{4\pi}\vartheta'(0)R(0;\tau)\\ &= -f_1(\tau) + \frac{i}{2}\eta(\tau)^3\int_{-\overline{\tau}}^{i\infty}\frac{\eta(z)^3}{(-i(z+\tau))^{\frac{1}{2}}}\,dz\\ &= -\widehat{f}_1(\tau) - \frac{1}{4\pi y},\end{aligned}$$

where we have used Theorem 2.6 in the second equality and the fact that $\vartheta(0) = 0$.

Additionally, from Proposition 3.1 we have (by taking $(1/2\pi i)\partial/\partial v\big|_{u=v=0}$ on both sides)

$$\begin{aligned}&\frac{1}{(c\tau+d)}\Big(-\widehat{f}_1\Big(\frac{a\tau+b}{c\tau+d}\Big) - \frac{1}{4\pi}\cdot\frac{1}{\mathrm{Im}\big(\frac{a\tau+b}{c\tau+d}\big)}\Big)\\ &\qquad = (c\tau+d)\Big(-\widehat{f}_1(\tau) - \frac{1}{4\pi y}\Big) + c\lim_{u,v\to 0} u e^{\pi i\frac{c}{c\tau+d}(2uv-u)^2}\widehat{A}(u,v;\tau)\end{aligned}$$

Using

$$\lim_{u\to 0} u\widehat{A}(u,v;\tau) = -\frac{1}{2\pi i} \quad \text{or} \quad \lim_{v\to 0}\widehat{A}(u,v;\tau) = -\frac{1}{2\pi i}\frac{\vartheta'(0)}{\vartheta(u)},$$

we find that

$$\lim_{u,v\to 0} u e^{\pi i\frac{c}{c\tau+d}(2uv-u^2)}\widehat{A}(u,v;\tau) = -\frac{1}{2\pi i}.$$

Additionally, since

$$\operatorname{Im}\left(\frac{a\tau+b}{c\tau+d}\right)=\frac{y}{|c\tau+d|^2},$$

we have

$$\begin{aligned}(c\tau+d)\widehat{f}_1(\tau)-\frac{1}{(c\tau+d)}\widehat{f}_1\left(\frac{a\tau+b}{c\tau+d}\right)&=\frac{1}{c\tau+d}\frac{1}{4\pi\operatorname{Im}\left(\frac{a\tau+b}{c\tau+d}\right)}-\frac{c\tau+d}{4\pi y}-\frac{c}{2\pi i}\\&=\frac{c\bar{\tau}+d}{4\pi y}-\frac{c\tau+d}{4\pi y}-\frac{c}{2\pi i}=0,\end{aligned}$$

which gives the result for $\widehat{f}_1$.

We define

$$\widetilde{A}_3(u;\tau)=e^{-\pi iu}A_3(u,\tau)-\frac{1}{1-e^{2\pi iu}}.$$

By Theorem 3.1 of [Bringmann 2008] we have

$$\begin{aligned}\widetilde{A}_3(u,\tau)=\frac{1}{1-e^{2\pi iu}}-\frac{ie^{\frac{3\pi u^2}{\tau}-\pi iu-\frac{\pi u}{\tau}}}{\tau\left(1-e^{-\frac{2\pi u}{\tau}}\right)}+\frac{i}{\tau}e^{\frac{3\pi u^2}{\tau}-\pi iu-\frac{\pi u}{\tau}}\widetilde{A}_3\left(\frac{iu}{\tau},-\frac{1}{\tau}\right)\\-(-i\tau)^{-\frac{1}{2}}\eta\left(-\frac{1}{\tau}\right)\int_{\mathbb{R}}e^{3\pi i\tau x^2}\left(\sum_{\pm}\frac{e^{\pm\frac{\pi i}{6}\mp\pi i\tau x}}{1-e^{2\pi iu}e^{\pm\frac{\pi i}{3}\mp 2\pi i\tau x}}\right)dx\end{aligned}\tag{3-2}$$

Remark. This may also be derived by using

$$\begin{aligned}e^{3\pi iu/2}\sum_{n\in\mathbb{Z}}\frac{(-1)^n q^{3n(n+1)/2}e^{2\pi inv}}{1-e^{2\pi iu}q^n}\\=\sum_{k=0}^{2}e^{2\pi iku}\vartheta(v+k\tau+1;3\tau)\mu(3u,v+k\tau+1;3\tau)\end{aligned}$$

and the results of Chapter 1 of [Zwegers 2002].

Lemma 4.2 of [Bringmann 2008] gives

$$\begin{aligned}\frac{1}{2\pi i}\frac{\partial}{\partial u}\Big|_{u=0}\int_{\mathbb{R}}e^{3\pi i\tau x^2}\left(\sum_{\pm}\frac{e^{\pm\frac{\pi i}{6}\mp\pi i\tau x}}{1-e^{2\pi iu}e^{\pm\frac{\pi i}{3}\mp 2\pi i\tau x}}\right)dx\\=\frac{\sqrt{3}(-i\tau)^2}{2\pi}\int_0^\infty\frac{\eta(iw)}{(-i(iw-i\tau))^{\frac{3}{2}}}\,dw.\end{aligned}$$

Using the transformation $\eta(-1/\tau)=\sqrt{-i\tau}\eta(\tau)$ and applying $(1/2\pi i)\partial/\partial u\big|_{u=0}$ to (3-2), we have

$$\begin{aligned}
&f_2(\tau)\\
&=\frac{1}{2\pi i}\frac{\partial}{\partial u}\left(\frac{1}{1-e^{2\pi iu}}+\frac{e^{\frac{3\pi iu^2}{\tau}-\pi iu-\frac{\pi iu}{\tau}}}{\tau\left(1-e^{-\frac{2\pi iu}{\tau}}\right)}+\frac{1}{\tau}e^{\frac{3\pi iu^2}{\tau}-\pi iu-\frac{\pi iu}{\tau}}\widetilde{A}_3\left(\frac{u}{\tau},-\frac{1}{\tau}\right)\right)\Bigg|_{u=0}\\
&\qquad\qquad+\tau^2\eta(\tau)\frac{\sqrt{3}}{2\pi}\int_0^\infty\frac{\eta(iw)}{(-i(iw-i\tau))^{\frac{3}{2}}}\,dw\\
&=\frac{1}{2\pi i}\frac{\partial}{\partial u}\left(\frac{1}{1-e^{2\pi iu}}+\frac{e^{\frac{3\pi iu^2}{\tau}-\pi iu-\frac{\pi u}{\tau}}}{\tau(1-e^{-\frac{2\pi iu}{\tau}})}\right)\Bigg|_{u=0}+\tau^{-2}f_2\left(-\frac{1}{\tau}\right)\\
&\qquad\qquad+\tau^2\eta(\tau)\frac{\sqrt{3}}{2\pi}\int_0^\infty\frac{\eta(iw)}{(-i(iw-i\tau))^{\frac{3}{2}}}\,dw\\
&=\left(\tfrac{1}{24}-\frac{1}{24\tau^2}+\frac{3}{4\pi i\tau}\right)+\tau^{-2}f_2\left(-\frac{1}{\tau}\right)\\
&\qquad\qquad+\tau^2\eta(\tau)\frac{\sqrt{3}}{2\pi}\int_0^\infty\frac{\eta(iw)}{(-i(iw-i\tau))^{\frac{3}{2}}}\,dw,
\end{aligned}\tag{3-3}$$

where we have used

$$\sum_{n\neq 0}\frac{(-1)^n q^{\frac{3n^2+n}{2}}}{1-q^n}=0.$$

Next define

$$\mathcal{S}(\tau):=\frac{\sqrt{3}}{2\pi i}\eta(\tau)\int_{-\bar{\tau}}^{i\infty}\frac{\eta(w)}{(-i(w+\tau))^{\frac{3}{2}}}\,dw.$$

Then the modular transformation of $\eta(\tau)$ implies that

$$\mathcal{S}(\tau)=\tau^2\mathcal{S}\left(-\frac{1}{\tau}\right)-\tau^2\frac{\sqrt{3}}{2\pi}\int_0^\infty\frac{\eta(iw)}{(-i(iw-i\tau))^{\frac{3}{2}}}\,dw.$$

Combining this with (3-3) gives $\widehat{f}_2(\tau)=\tau^{-2}\widehat{f}_2(-1/\tau)$. Using $\widehat{f}_2(\tau+1)=\widehat{f}_2(\tau)$, we obtain the result. □

Proof of Theorem 1.1. By (1-2) and

$$\sum_{n=1}^\infty\frac{q^n}{(1-q^n)^2}=\sum_{n=1}^\infty\frac{nq^n}{1-q^n}=\sum_{n\geq 1}\sigma_1(n)q^n=\frac{1-E_2(\tau)}{24},\tag{3-4}$$

we have $q(q)_\infty^3 v(q)=f_1(\tau)+f_2(\tau)+\frac{1}{24}(E_2(\tau)-1)$. The result now follows from Theorem 3.2 and (2-5). □

4. Proof of Theorem 1.2

In this section we give the following more precise version of Theorem 1.2.

Theorem 4.1. *With $q = e^{2\pi i\tau}$ we have*

$$q^{-1}(q^{24};q^{24})_\infty^2 v_1(q^{24}) + \frac{i}{4\sqrt{2}\pi}\int_{-\bar\tau}^{i\infty}\frac{\eta(24z)}{(-i(\tau+z))^{\frac{3}{2}}}\,dz$$

is a harmonic weak Maass form of weight $\frac{3}{2}$ on $\Gamma_0(576)$ with Nebentypus $\left(\frac{12}{\cdot}\right)$. Also,

$$q^{-1}v_2(q^8) - i\sqrt{2}\int_{-\bar\tau}^{i\infty}\frac{\eta(8z)^3}{\sqrt{-i(z+\tau)}}\,dz$$

is a harmonic weak Maass form of weight $\frac{1}{2}$ on $\Gamma_0(64)$ with trivial Nebentypus. Finally, let

$$E_2^{\mathrm{odd}}(\tau) := 1 + 24\sum_{n>0}\Bigg(\sum_{\substack{d\mid n\\ d\,odd}} d\Bigg)q^n = 2E_2(2\tau) - E_2(\tau).$$

Then

$$q^{-1}v_3(q^8) = \frac{E_2^{\mathrm{odd}}(8\tau)}{6\eta(8\tau)^3}$$

is a weight-$\frac{1}{2}$ weakly holomorphic modular form on $\Gamma_0(64)$ with trivial Nebentypus.

The modular properties of $v_1(q)$ and $v_3(q)$ are straightforward or follow from known results. Similar to (3-4) we have

$$\sum_{n=1}^{\infty}\frac{q^n}{(1+q^n)^2} = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}nq^n}{1-q^n} = -\sum_{n\geq 1}\Bigg(\sum_{d\mid n}(-1)^d d\Bigg)q^n. \tag{4-1}$$

Hence

$$\sum_{n=1}^{\infty} q^n\Bigg(\frac{1}{(1+q^n)^2} + \frac{1}{(1-q^n)^2}\Bigg) = \frac{E_2^{\mathrm{odd}}(\tau)-1}{12}. \tag{4-2}$$

The claim for $v_2(q)$ follows from Lemma 2.1 and the fact that $E^{\mathrm{odd}}(\tau)$ is a holomorphic modular form of weight 2 on $\Gamma_0(4)$. The proof of the claims about $v_1(q)$ follows from (3-4) and the following theorem of Bringmann [2008].

Theorem 4.2 [Bringmann 2008, Theorem 1.1]. *Let $q = e^{2\pi i\tau}$ and*

$$\mathcal{R}(\tau) := \frac{q^{-1}}{(q^{24};q^{24})_\infty}\sum_{n\neq 0}\frac{(-1)^{n+1}q^{36n(n+1)}}{(1-q^{24n})^2}.$$

Then

$$\mathcal{M}(\tau) := \mathcal{R}(\tau) - \frac{i}{4\sqrt{2}\pi}\int_{-\bar{\tau}}^{i\infty} \frac{\eta(24z)}{(-i(\tau+z))^{\frac{3}{2}}}\,dz - \frac{1}{24\eta(24\tau)} + \frac{E_2(24\tau)}{8\eta(24\tau)}$$

is a harmonic weak Maass form of weight $\frac{3}{2}$ *on* $\Gamma_0(576)$ *with Nebentypus* $(\frac{12}{\cdot})$.

The modularity of $v_2(q)$ follows easily from the identity of Theorem 1.3 and the results of Chapter 1 of [Zwegers 2002], namely, the following theorem.

Theorem 4.3. *With* $q = e^{2\pi i\tau}$ *we have* $\widetilde{F}(q) = -2iq^{1/8}\mu(\frac{1}{2}, \frac{1}{2}; \tau)$ *is a mock theta function with shadow proportional to* $\eta^3(z)$. *More precisely,*

$$q^{-1}\widetilde{F}(q^8) + 2i\sqrt{2}\int_{-\bar{\tau}}^{i\infty} \frac{\eta(8z)^3}{\sqrt{-i(z+\tau)}}\,dz$$

is a harmonic weak Maass form of weight $\frac{1}{2}$ *on* $\Gamma_0(64)$ *with trivial Nebentypus.*

Proof. Since this function is written in terms of μ the result follows from Chapter 1 of [Zwegers 2002]. In particular, to compute the shadow we use Theorem 2.6 and Lemma 2.1. For additional details, see [Eguchi and Hikami 2009; Rhoades 2012]. □

We have used Theorem 1.3 to establish the modularity of $v_2(q)$. In the next three sections we will give different proofs of this theorem. Each proof uses different techniques and highlights different aspects of the Appell sums appearing in Theorem 1.3.

5. Elliptic proof of Theorem 1.3

In this section we prove the identity of Theorem 1.3 via a transformation property of $\mu(u, v; \tau)$ with respect to the elliptic variables u and v.

Proposition 5.1 [Zwegers 2002, Proposition 1.4(7)]. *For* $u, v, u+z, v+z \notin \mathbb{Z}\tau + \mathbb{Z}$, *we have*

$$\mu(u+z, v+z; \tau) - \mu(u, v; \tau) = \frac{1}{2\pi i}\frac{\vartheta'(0)\vartheta(u+v+z)\vartheta(z)}{\vartheta(u)\vartheta(v)\vartheta(u+z)\vartheta(v+z)},$$

where we write $\vartheta(u) = \vartheta(u; \tau)$ *when* τ *is understood.*

Let

$$F_{2k}(\tau) := \sum_{n=1}^{\infty}\sum_{d|n}(-1)^d d^{2k-1}q^n \quad \text{and} \quad \Phi_{2k}(\tau) := \sum_{n=1}^{\infty}\sum_{d|n} d^{2k-1}q^n. \tag{5-1}$$

We will need the following lemma of [Rhoades 2013].

Lemma 5.2. *With $Z = 2\pi i u$ we have*

$$\vartheta(u;\tau) = -2\sin(\pi u)\eta^3(\tau)\exp\Bigg(-2\sum_{\substack{\ell \text{ even}\\ \ell>0}} \frac{Z^\ell}{\ell!}\Phi_\ell(\tau)\Bigg),$$

$$\vartheta(u+\tfrac{1}{2};\tau) = -2\cos(\pi u)\frac{\eta(2\tau)^2}{\eta(\tau)}\exp\Bigg(-2\sum_{\substack{\ell>0\\ \ell \text{ even}}} \frac{Z^\ell}{\ell!}F_\ell(\tau)\Bigg).$$

With this lemma we turn to the proof of Theorem 1.3 using Proposition 5.1.

"Elliptic" proof of Theorem 1.3. Throughout this proof let $x = e^{2\pi i u}$. From Proposition 5.1 and using $\vartheta'(0) = -2\pi\eta(\tau)^3$, we have

$$\vartheta(u)\vartheta(-u)\mu(u+\tfrac{1}{2},-u+\tfrac{1}{2};\tau) - \vartheta(u)\vartheta(-u)\mu(u,-u;\tau) = \frac{i\eta^3(\tau)\vartheta(\frac{1}{2})^2}{\vartheta(u+\frac{1}{2})^2}. \tag{5-2}$$

Taylor expanding each of the three terms in this identity around $u = 0$ by applying Lemma 5.2, we have

$$\vartheta(u)\vartheta(-u)\mu(u+\tfrac{1}{2},-u+\tfrac{1}{2};\tau) = -4i\pi^2\Bigg(\frac{\eta(\tau)^6}{\vartheta(\frac{1}{2})}\sum_{n\in\mathbb{Z}}\frac{q^{\frac{n^2+n}{2}}}{1+q^n}\Bigg)u^2 + O(u^4), \tag{5-3}$$

$$\frac{i\eta^3(\tau)\vartheta(\frac{1}{2})}{\vartheta(u+\frac{1}{2})^2} = i\eta(\tau)^3 + 4i\pi^2(\tfrac{1}{4} - 2F_2(\tau))u^2 + O(u^4). \tag{5-4}$$

The remaining term is more interesting. From the definition of $\mu(u,-u;\tau)$, we have

$$\vartheta(u)\vartheta(-u)\mu(u,-u;\tau) = \frac{i\vartheta(u)}{2\sin(\pi u)} + \vartheta(u;\tau)e^{\pi i u}\sum_{n\neq 0}\frac{(-1)^n x^{-n} q^{(n^2+n)/2}}{1-xq^n}.$$

Since

$$\sum_{n\neq 0}\frac{(-1)^n q^{(n^2+n)/2}}{1-q^n} = 0,$$

we have

$$\sum_{n\neq 0}\frac{(-1)^n x^{-n} q^{(n^2+n)/2}}{1-xq^n} = 2\pi i u\Bigg(\sum_{n\neq 0}\frac{(-1)^{n+1} n q^{(n^2+n)/2}}{1-q^n} + \sum_{n\neq 0}\frac{(-1)^n q^{(n^2+n)/2}}{(1-q^n)^2}\Bigg) + O(u^2).$$

Again applying Lemma 5.2 we have

$$
\begin{aligned}
&-\vartheta(u)\vartheta(-u)\mu(u,-u;\tau)\\
&= i\eta^3(\tau) - 4i\pi^2\eta^3(\tau)\\
&\quad\times\Bigg(\sum_{n\neq 0}\frac{(-1)^n n q^{(n^2+n)/2}}{1-q^n} - \sum_{n\neq 0}\frac{(-1)^n q^{(n^2+n)/2}}{(1-q^n)^2} - \Phi_1(\tau)\Bigg)u^2 + O(u^3). \qquad (5\text{-}5)
\end{aligned}
$$

Using (5-2), (5-3), (5-4), and (5-5) to compare the u^2 coefficient of the Taylor expansion around $u=0$ of both sides of (5-2), we have

$$
\begin{aligned}
\frac{1}{\vartheta(\frac{1}{2})}\sum_{n\in\mathbb{Z}}\frac{q^{\frac{n^2+n}{2}}}{1+q^n} - \frac{1}{\eta^3(\tau)}\Bigg(\sum_{n\neq 0}\frac{(-1)^{n+1} n q^{\frac{n^2+n}{2}}}{1-q^n} + \sum_{n\neq 0}\frac{(-1)^n q^{\frac{n^2+n}{2}}}{(1-q^n)^2} + \Phi_2(\tau)\Bigg)&\\
= \frac{1}{\eta^3(z)}(-\tfrac{1}{4} + 2F_2(\tau)).&
\end{aligned}
$$

The identity follows from $\vartheta(\frac{1}{2}) = -2q^{1/8}(q)_\infty(-q)_\infty^2$, $\eta^3(\tau) = q^{1/8}(q)_\infty^3$,

$$
\sum_{n\neq 0}\frac{(-1)^n q^{(n^2+n)/2}}{(1-q^n)^2} = -\Phi_2(\tau), \quad \text{and} \quad \sum_{n=1}^{\infty}\frac{q^n}{(1+q^n)^2} = -F_2(\tau). \qquad \square
$$

6. q-Series proof of Theorem 1.3

In this section we will give a q-series proof of the identity in Theorem 1.3. We will divide the proof into several lemmas.

For negative values of n we define $(x)_n = (x)_\infty/(xq^n)_\infty$. For simplicity we write $(a,b,c,d;q)_n = (a;q)_n(b;q)_n(c;q)_n(d;q)_n$. We will use the equations

$$
\vartheta_4 := \sum_{n\in\mathbb{Z}}(-1)^n q^{n^2} = \frac{(q)_\infty}{(-q)_\infty}, \qquad (6\text{-}1)
$$

which is [Andrews 1998, (2.2.12)], and

$$
\begin{aligned}
&\sum_{n\in\mathbb{Z}}\frac{(e)_n(f)_n(\frac{aq}{ef})^n}{(\frac{aq}{c})_n(\frac{aq}{d})_n}\\
&\quad= \frac{(\frac{q}{c},\frac{q}{d},\frac{aq}{e},\frac{aq}{f};q)_\infty}{(aq,a^{-1}q,\frac{aq}{cd},\frac{aq}{ef};q)_\infty}\sum_{n\in\mathbb{Z}}\frac{(1-aq^{2n})(c,d,e,f;q)_n(\frac{qa^3}{cdef})^n q^{n^2}}{(1-a)(\frac{aq}{c},\frac{aq}{d},\frac{aq}{e},\frac{aq}{f};q)_n}, \qquad (6\text{-}2)
\end{aligned}
$$

which can be found in [Bailey 1950] or page 135 of [Gasper and Rahman 2004].

Lemma 6.1.

$$
2\sum_{n\in\mathbb{Z}}\frac{q^{\binom{n+1}{2}}}{1+q^n} = \frac{1}{\vartheta_4^2}\Bigg(1 + \sum_{n=1}^{\infty}\Bigg(\frac{8q^{2n^2+n}}{(1+q^n)^2} - \frac{16nq^{2n^2}(1-q^n)}{1+q^n}\Bigg)\Bigg).
$$

Proof. In (6-2) set $e = c = -1$, $a = 1$, and let $d, f \to \infty$ to obtain

$$
\begin{aligned}
&2\sum_{n\in\mathbb{Z}} \frac{q^{n(n+1)/2}}{1+q^n} \\
&\quad = \frac{1}{\vartheta_4^2} \lim_{a\to 1} \sum_{n\in\mathbb{Z}} \frac{(1-aq^2n)}{1-a} \frac{(-1)_n^2}{(-aq)_n^2} a^3 n q^{2n^2} \\
&\quad = \frac{1}{\vartheta_4^2}\Bigg(1 - \sum_{n=1}^{\infty} \frac{d}{da}\Big|_{a=1} \Bigg(\frac{(1-aq^{2n})(-1)_n^2 a^{3n} q^{2n^2}}{(-aq)_n^2} \\
&\qquad\qquad + \frac{(1-aq^{-2n})(-1)_{-n}^2 a^{-3n} q^{2n^2}}{(-aq)_{-n}^2}\Bigg)\Bigg) \\
&\quad = \frac{1}{\vartheta_4^2}\Bigg(1 - \frac{d}{da}\Big|_{a=1} \sum_{n=1}^{\infty} (1-aq^{2n}) \left(\frac{(-1)_n}{(-aq)_n}\right)^2 a^{3n} q^{2n^2} \\
&\quad - \frac{d}{da}\Big|_{a=1} \sum_{n=1}^{\infty} (1-aq^{-2n}) \left(\frac{(a+1)(a+q)\cdots(a+q^{n-1})}{(-q)_n^2}\right)^2 a^{-3n} q^{2n^2+2n}\Bigg)
\end{aligned}
$$

where we use (6-1) and L'Hopital's rule in the second displayed equation. Continuing, we have

$$
\begin{aligned}
&2\sum_{n\in\mathbb{Z}} \frac{q^{n(n+1)/2}}{1+q^n} \\
&= \frac{1}{\vartheta_4^2}\Bigg(1 + \sum_{n=1}^{\infty} \frac{4q^{2n^2+2n}}{(1+q^n)^2} - \sum_{n=1}^{\infty} (1-q^{2n})(-1)_n^2 (-2(-q)_n^{-3}) \sum_{j=1}^{n} \frac{q^j}{1+q^j} (-q)_n q^{2n^2} \\
&\quad - \sum_{n=1}^{\infty} (1-q^{2n}) \frac{12nq^{2n^2}}{(1+q^n)^2} + \sum_{n=1}^{\infty} \frac{4q^{2n^2}}{(1+q^n)^2} \\
&\quad - \sum_{n=1}^{\infty} (1-q^{-2n}) q^{2n^2+2n} \frac{2(-1)_n^2}{(-q)_n^2} \sum_{j=0}^{n-1} \frac{1}{1+q^j} + \sum_{n=1}^{\infty} (1-q^{-2n}) q^{2n^2+2n} \frac{12n}{(1+q^n)^2}\Bigg) \\
&= \frac{1}{\vartheta_4^2}\Bigg(1 + 4\sum_{n=1}^{\infty} \frac{q^{2n^2+2n}}{(1+q^n)^2} + 8\sum_{n=1}^{\infty} (1-q^{2n}) \frac{q^{2n^2}}{(1+q^n)^2} \sum_{j=1}^{n} \frac{q^j}{1+q^j} \\
&\quad - 12\sum_{n=1}^{\infty} \frac{nq^{2n^2}(1-q^n)}{(1+q^n)} + 4\sum_{n=1}^{\infty} \frac{q^{2n^2}}{(1+q^n)^2} + 8\sum_{n=1}^{\infty} (1-q^{2n}) \frac{q^{2n^2}}{(1+q^n)^2} \sum_{j=0}^{n-1} \frac{1}{1+q^j} \\
&\qquad\qquad - 12\sum_{n=1}^{\infty} \frac{(1-q^n)nq^{2n^2}}{(1+q^n)}\Bigg)
\end{aligned}
$$

Noting that

$$\sum_{j=1}^{n} \frac{q^j}{1+q^j} + \sum_{j=0}^{n-1} \frac{1}{1+q^j} = n + \frac{q^n}{1+q^n} - \tfrac{1}{2},$$

we see that the right hand side equals

$$\frac{1}{\vartheta_4^2}\Bigg(1 + 4\sum_{n=1}^{\infty} \frac{(1+q^{2n})q^{2n^2}}{(1+q^n)^2} - 16\sum_{n=1}^{\infty} \frac{n(1-q^n)q^{2n^2}}{(1+q^n)} + 8\sum_{n=1}^{\infty} \frac{(1-q^n)q^{2n^2}}{(1+q^n)}\left(\frac{q^n}{1+q^n} - \tfrac{1}{2}\right)\Bigg)$$
$$= \frac{1}{\vartheta_4^2}\left(1 + 4\sum_{n=1}^{\infty} \frac{q^{2n^2+n}}{(1+q^n)^2} - 16\sum_{n=1}^{\infty} \frac{n(1-q^n)q^{2n^2}}{1+q^n}\right),$$

which gives the result. □

Lemma 6.2.

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1} n q^{n(n+1)/2}}{1-q^n} = \sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)^2} - 2\sum_{n=1}^{\infty} \frac{n(1+q^n)q^{2n^2}}{1-q^n} + \sum_{n=1}^{\infty} \frac{q^{2n^2+n}}{(1-q^n)^2}$$

Proof. Recall from the proof of (1.2) in [Andrews 2013] that

$$\lim_{a,b\to 1} \frac{1}{(1-a)(1-b)}\left(S_1(a,b;q) - \frac{1}{(ab;q)_\infty}\right)$$
$$= \frac{1}{(q)_\infty}\left(-\sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)^2} + 2\sum_{n=1}^{\infty} \frac{(-1)^{n-1} n q^{n(n+1)/2}}{1-q^n}\right) \tag{6-3}$$

where

$$S_1(a,b;q) = \frac{(a^{-1}q;q)_\infty}{(bq,q;q)_\infty} \sum_{n\in\mathbb{Z}} \frac{(1-a)(-b)^n q^{n(n+1)/2}}{1-aq^n}.$$

We now apply (6-2) with $f, d \to \infty$ and setting a equal to ab. Then letting $e \to a$ we obtain

$$S_1(a,b;q) = \frac{(\frac{q}{a},\frac{q}{b};q)_\infty}{(abq,\frac{q}{ab},q;q)_\infty} \sum_{n\in\mathbb{Z}} \frac{(1-abq^{2n})}{(1-ab)} \frac{(a,b;q)_n}{(aq,bq;q)_n} (ab)^{2n} q^{2n^2}.$$

Now

$$\lim_{a,b\to 1}\frac{1}{(1-a)(1-b)}\left(\frac{(\frac{q}{a},\frac{q}{b};q)_\infty}{(abq,\frac{q}{ab},q;q)_\infty}-\frac{1}{(abq)_\infty}\right)$$
$$=\lim_{a,b\to 1}\frac{1}{(1-a)(1-b)}\left(\frac{1}{(abq)_\infty}\sum_{n=0}^{\infty}\frac{(a)_n(b)_n(\frac{q}{ab})^n}{(q)_n^2}-\frac{1}{(abq)_\infty}\right)$$
$$=\lim_{a,b\to 1}\frac{1}{(abq)_\infty}\sum_{n=1}^{\infty}\frac{(aq)_{n-1}(bq)_{n-1}(\frac{q}{ab})^n}{(q)_n^2}=\frac{1}{(q)_\infty}\sum_{n=1}^{\infty}\frac{q^n}{(1-q^n)^2} \tag{6-4}$$

where we use Corollary 2.3 of [Andrews 1998] in the second equality.

Next we have

$$\lim_{a,b\to 1}\frac{(\frac{q}{a},\frac{q}{b};q)_\infty}{(1-a)(1-b)(abq,\frac{q}{ab},q;q)_\infty}\sum_{n\neq 0}\frac{(1-abq^{2n})}{(1-ab)}\frac{(a,b;q)_n}{(bq,aq;q)_n}q^{2n^2}(ab)^{2n}$$
$$=\frac{1}{(q)_\infty}\lim_{a,b\to 1}\sum_{n\neq 0}\frac{(1-abq^{2n})q^{2n^2}(ab)^{2n}}{(1-ab)(1-aq^n)(1-bq^n)}$$
$$=\frac{1}{(q)_\infty}\lim_{b\to 1}\sum_{n=1}^{\infty}\left(\frac{(1-bq^{2n})b^{2n}q^{2n^2}}{(1-b)(1-q^n)(1-bq^n)}+\frac{(1-bq^{-2n})b^{-2n}q^{2n^2}}{(1-b)(1-q^{-n})(1-bq^{-n})}\right)$$
$$=-\frac{1}{(q)_\infty}\sum_{n=1}^{\infty}\frac{d}{db}\Big|_{b=1}\left(\frac{(1-bq^{2n})b^{2n}q^{2n^2}}{(1-b)(1-q^n)(1-bq^n)}+\frac{(q^{2n}-b)b^{-2n}q^{2n^2}}{(1-b)(1-q^n)(b-q^n)}\right)$$
$$=-\frac{4}{(q)_\infty}\sum_{n=1}^{\infty}\frac{n(1+q^n)q^{2n^2}}{1-q^n}+\frac{2}{(q)_\infty}\sum_{n=1}^{\infty}\frac{q^{2n^2+n}}{1-q^n} \tag{6-5}$$

Comparing (6-3), (6-4), and (6-5) and multiplying by $(q)_\infty/2$ we obtain the result. □

Lemma 6.3.

$$\sum_{n=1}^{\infty}\frac{q^n}{(1-q^n)^2}-\sum_{n=1}^{\infty}\frac{q^n}{(1+q^n)^2}=4\sum_{n=1}^{\infty}\frac{q^{2n}}{(1-q^{2n})^2}$$
$$=\sum_{n=1}^{\infty}\frac{4n(1+q^{2n})q^{2n^2}}{1-q^{2n}}+4\sum_{n=1}^{\infty}\frac{q^{2n^2+2n}}{(1-q^{2n})^2}$$

Proof. The first equality is immediate and follows by combining the initial two sums term by term. Finally,

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{q^{2n}}{(1-q^{2n})^2} &= \sum_{n,m=1}^{\infty} mq^{2nm} = \sum_{m=1}^{\infty}\Big(\sum_{n=1}^{m} + \sum_{m+1}^{\infty}\Big) mq^{2nm} \\
&= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} mq^{2nmn} + \sum_{m=1}^{\infty} \frac{mq^{2m^2+2m}}{1-q^2m} \\
&= \sum_{n=1}^{\infty}\sum_{m=0}^{\infty} (m+n)q^{2n(m+n)} + \sum_{m=1}^{\infty} \frac{mq^{2m^2+2m}}{1-q^{2m}} \\
&= \sum_{n=1}^{\infty} \frac{q^{2n^2+2n}}{(1-q^{2n})^2} + \sum_{n=1}^{\infty} \frac{nq^{2n^2}}{1-q^{2n}} + \sum_{m=1}^{\infty} \frac{mq^{2m^2+2m}}{1-q^{2m}} \\
&= \sum_{n=1}^{\infty} \frac{nq^{2n^2}(1+q^{2n})}{1-q^{2n}} + \sum_{n=1}^{\infty} \frac{q^{2n^2+2n}}{(1-q^{2n})^2}. \qquad \square
\end{aligned}$$

We are now ready to prove the main result.

q-Series proof of Theorem 1.3. Comparing the statement of Theorem 1.3 with the assertion of Lemma 1, we see that the theorem is equivalent to the following:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{q^{2n^2+n}}{(1+q^n)^2} - 2\sum_{n=1}^{\infty} \frac{nq^{2n^2}(1-q^n)}{1+q^n} \\
= \sum_{n=1}^{\infty} \frac{q^n}{(1+q^n)^2} + \sum_{n=1}^{\infty} \frac{(-1)^n nq^{n(n+1)/2}}{1-q^n}. \qquad (6\text{-}6)
\end{aligned}$$

Now in (6-6) we replace the final sum on the right-hand side by the negative of the right-hand side of the identity given by Lemma 6.2. Hence Theorem 1.3 is equivalent to the assertion that

$$\begin{aligned}
&\sum_{n=1}^{\infty} \frac{q^{2n^2+n}}{(1+q^n)^2} - 2\sum_{n=1}^{\infty} \frac{nq^{2n^2}(1-q^n)}{1+q^n} \\
&\quad = \sum_{n=1}^{\infty} \frac{q^n}{(1+q^n)^2} - \sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)^2} + 2\sum_{n=1}^{\infty} \frac{nq^{2n^2}(1+q^n)}{1-q^n} + \sum_{n=1}^{\infty} \frac{q^{2n^2+n}}{(1-q^n)^2}.
\end{aligned}$$

We now combine the first sum on the left-hand side with the final sum on the right-hand side and apply Lemma 6.3 to the first two sums on the right-hand side.

Hence Theorem 1.3 is equivalent to the assertion that

$$-4\sum_{n=1}^{\infty}\frac{q^{2n^2+2n}}{(1-q^{2n})^2}-4\sum_{n=1}^{\infty}\frac{nq^{2n^2}(1-q^n)}{1+q^n}=\sum_{n=1}^{\infty}\frac{q^n}{(1+q^n)^2}-\sum_{n=1}^{\infty}\frac{q^n}{(1-q^n)^2},$$

which is merely a restatement of Lemma 6.3. □

7. Modular proof of Theorem 1.3

In this section we establish Theorem 1.3 by computing the modular transformations of the right hand side and demonstrate that they match those of the left hand side. Finally checking that enough of the Fourier coefficients are equal yields the result.

"Modular" proof of Theorem 1.3. We use Theorem 2.6 together with (2-12) to find

$$\begin{aligned}\widehat{\mu}(u,u;\tau)=\mu(u,u;\tau)+\frac{i}{2}R(0;\tau)&=\mu(u,u;\tau)-\frac{i}{2}\int_{-\overline{\tau}}^{i\infty}\frac{g_{\frac{1}{2},\frac{1}{2}}(z)}{\sqrt{-i(z+\tau)}}\,dz\\&=\mu(u,u;\tau)+\tfrac{1}{2}\int_{-\overline{\tau}}^{i\infty}\frac{\eta(z)^3}{\sqrt{-i(z+\tau)}}\,dz.\end{aligned}$$

Combining this with the definition of $\widehat{f}_1$ in Theorem 3.2 and the definition of $\widehat{E}_2$ in (2-6), we see that if we define

$$M(u;\tau):=f_1(\tau)+i\eta(\tau)^3\mu(u,u;\tau)-\tfrac{1}{12}E_2(\tau),$$

then this also equals $\widehat{f}_1(\tau)+i\eta(\tau)^3\widehat{\mu}(u,u;\tau)-\frac{1}{12}\widehat{E}_2(\tau)$. Now using the transformation properties of $\widehat{f}_1$ from Theorem 3.2, $\widehat{\mu}$ from Theorem 2.5, and $\widehat{E}_2$, we get that M transforms as

$$M\Big(\frac{u}{c\tau+d};\frac{a\tau+b}{c\tau+d}\Big)=(c\tau+d)^2M(u;\tau)$$

for all $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbb{Z})$. Further, by the first part of Theorem 2.5 we have the elliptic transformation property $M(u+k\tau+l;\tau)=M(u;\tau)$, for all $k,l\in\mathbb{Z}$. Using these we easily see that

$$m(\tau):=\begin{pmatrix}M(\frac{1}{2};\tau)\\M(\tau/2;\tau)\\M((\tau+1)/2;\tau)\end{pmatrix}$$

transforms as a vector-valued modular form of weight 2 on $\mathrm{SL}_2(\mathbb{Z})$:

$$m(\tau+1)=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}m(\tau)\quad\text{and}\quad m\Big(-\frac{1}{\tau}\Big)=\tau^2\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}m(\tau).$$

From this we see that $M(\frac{1}{2};\tau)$ transforms as a modular form of weight 2 on $\Gamma_0(2)$.

On the other hand,

$$g(\tau) := \begin{pmatrix} E_2(\tau) - 2E_2(2\tau) \\ E_2(\tau) - \frac{1}{2}E_2(\tau/2) \\ E_2(\tau) - \frac{1}{2}E_2((\tau+1)/2) \end{pmatrix} = \begin{pmatrix} \widehat{E}_2(\tau) - 2\widehat{E}_2(2\tau) \\ \widehat{E}_2(\tau) - \frac{1}{2}\widehat{E}_2(\tau/2) \\ \widehat{E}_2(\tau) - \frac{1}{2}\widehat{E}_2((\tau+1)/2) \end{pmatrix}$$

satisfies the same modular transformation properties as m. Further we get directly from the definitions

$$g(\tau) = \begin{pmatrix} -1 + O(q) \\ \frac{1}{2} + O(q^{\frac{1}{2}}) \\ \frac{1}{2} + O(q^{\frac{1}{2}}) \end{pmatrix} \quad \text{and} \quad m(\tau) = \begin{pmatrix} \frac{1}{6} + O(q) \\ -\frac{1}{12} + O(q^{\frac{1}{2}}) \\ -\frac{1}{12} + O(q^{\frac{1}{2}}) \end{pmatrix},$$

and so

$$m(\tau) + \tfrac{1}{6}g(\tau) = \begin{pmatrix} O(q) \\ O(q^{\frac{1}{2}}) \\ O(q^{\frac{1}{2}}) \end{pmatrix}.$$

If we take the product of the three components of $m + \frac{1}{6}g$, then we get a holomorphic cusp form on $\mathrm{SL}_2(\mathbb{Z})$ of weight 6, and hence this equals zero. This then implies that all three components are zero and so we get $m + \frac{1}{6}g = 0$.

The first component equals

$$\frac{1}{6}\left(-1 - 24\sum_{n>0}\sigma_1^{\mathrm{odd}}(n)q^n\right) - \frac{1}{12}E_2(\tau) + \sum_{n\neq 0}\frac{(-1)^{n+1}nq^{n(n+1)/2}}{1-q^n} + 2\frac{(q)_\infty^2}{(-q)_\infty^2}\sum_{n\in\mathbb{Z}}\frac{q^{n(n+1)/2}}{1+q^n}$$

where we have used

$$\vartheta(\tfrac{1}{2};\tau) = -2q^{\frac{1}{8}}(q)_\infty(-q)_\infty^2 \quad \text{and} \quad \sigma_1^{\mathrm{odd}}(n) = \sum_{\substack{d\mid n \\ d \text{ odd}}} d.$$

Using (4-1), (4-2), and (3-4) to rewrite the Eisenstein series terms gives the result. □

Remark. The calculations above show that $M(u;\tau)$ is a multiple of the Weierstrass $\wp$-function. Thus the conclusion follows from known facts about $\wp$ at half periods. We have included the details for completeness.

8. Holomorphic projection

In this section we demonstrate that the function

$$B(\tau) = \frac{1}{\eta(\tau)^3} \sum_{n \neq 0} \frac{(-1)^n n q^{n(n+1)/2}}{1-q^n}$$

arises from the holomorphic projection of $\eta^3(\tau)(\eta^3(\tau))^*$. The holomorphic projection operator is the unique linear map $\pi_{\mathrm{hol}} = \pi_{\mathrm{hol},k}$ mapping the space of nonholomorphic modular forms of weight k to the space of cusp forms of weight k and level N satisfying $(h, \Phi) = (h, \pi(\Phi))$ for all cusp forms h of weight k and level N and $(\cdot, \cdot)$ is the Petersson inner product.

If $\Phi(\tau) := \sum_{n \in \mathbb{Z}} a_n(y) q^n$ is a modular form of weight k, not necessarily holomorphic, on $\mathrm{SL}_2(\mathbb{Z})$ such that for $\tau \to i\infty$ there exists an $\epsilon > 0$ with

$$(\Phi \mid_k \gamma)(\tau) = O(y^{-\epsilon}), \tag{8-1}$$

then $\pi_{\mathrm{hol}}(\Phi)(\tau) := \sum_{n=1}^{\infty} c_n(\Phi) q^n$ with

$$c_n(\Phi) := \frac{(4\pi n)^{k-1}}{\Gamma(k-1)} \int_0^{\infty} a_n(y) e^{-4\pi n y} y^{k-2}\, dy. \tag{8-2}$$

See for instance [Sturm 1980] or Proposition 5.1 of [Gross and Zagier 1986].

Proposition 6.2 of [Gross and Zagier 1986] (see also Proposition 11 of [Coates 1986]) suggests that $(1/\eta^3)\pi_{\mathrm{hol}}(\eta^3(\eta^3)^*)$ is, after the addition of an Eisenstein series, the holomorphic part of a harmonic weak Maass form (see [Ono 2008] for discussion of holomorphic parts of harmonic weak Maass forms). Precisely, we have the following result.

Proposition 8.1. *With the notation from above,*

$$\pi_{\mathrm{hol}}(\eta^3(\eta^3)^*)(\tau) = \sum_{n \neq 0} \frac{(-1)^n n q^{n(n+1)/2}}{1-q^n}.$$

Proof. For simplicity let $g(\tau) = \eta^3(\tau)$. From (2-3) we have

$$g(\tau) = \sum_{n=1}^{\infty} \left(\frac{-4}{n}\right) n q^{n^2/8}.$$

Thus

$$(g(\tau))^* = \pi^{-\frac{1}{2}} \sum_{n=1}^{\infty} \left(\frac{-4}{n}\right) \Gamma\left(\tfrac{1}{2}; \frac{\pi n^2 y}{2}\right) q^{-n^2/8},$$

so that

$$g(\tau)g^*(\tau) = \pi^{-\frac{1}{2}} \sum_{n,m>0} m\left(\frac{-4}{nm}\right)\Gamma\left(\tfrac{1}{2}; \frac{\pi n^2 y}{2}\right) q^{(m^2-n^2)/8}.$$

With $\ell = m^2 - n^2$ we need to compute

$$\begin{aligned}
&\int_0^\infty \Gamma\left(\tfrac{1}{2}; \frac{\pi n^2 y}{2}\right) e^{-\frac{\pi \ell y}{2}}\, dy \\
&\quad = \int_0^\infty \int_{\frac{\pi n^2 y}{2}}^\infty e^{-t} t^{\frac{1}{2}} e^{-\frac{\pi \ell y}{2}} \frac{dt}{t}\, dy = \int_0^\infty \int_1^\infty e^{-(wn^2+\ell)\frac{\pi}{2} y} (yw)^{\frac{1}{2}} \frac{dw}{w}\, dy \\
&\quad = n\sqrt{\frac{\pi}{2}} \int_1^\infty w^{\frac{1}{2}} \frac{dw}{w} \int_0^\infty e^{-(wn^2+\ell)\frac{\pi}{2} y} y^{\frac{3}{2}} \frac{dy}{y} \\
&\quad = \frac{n}{\sqrt{\pi}} \int_1^\infty \frac{1}{w^{\frac{1}{2}}(wn^2+\ell)^{\frac{3}{2}}}\, dw = \frac{n}{\sqrt{\pi}} \left(\frac{2\sqrt{w}}{\ell\sqrt{wn^2+\ell}} \Bigg|_1^\infty \right) \\
&\quad = \frac{2n}{\sqrt{\pi}\ell}\left(\frac{1}{n} - \frac{1}{\sqrt{n^2+\ell}}\right) = \frac{2}{\sqrt{\pi} m(m+n)},
\end{aligned}$$

where we use $\int_0^\infty e^{-\alpha y} y^{1/2} dy/y = \alpha^{-3/2}\sqrt{\pi}/2$ in the third equality.

Inserting this into (8-2) gives

$$\pi_{\mathrm{hol}}(gg^*)(\tau) = \sum_{m>n>0} \left(\frac{-4}{nm}\right)(m-n) q^{(m^2-n^2)/8}.$$

Now $(-4/n)(-4/m) = 0$ unless both n and m are odd, in which case we have $n = 2a+1$ and $m = 2b+1$. Thus $(-4/nm) = (-1)^{a+b}$ and we have

$$\begin{aligned}
\pi_{\mathrm{hol}}(gg^*)(\tau) &= 2 \sum_{a>b\geq 0} (-1)^{a+b}(a-b) q^{\frac{a^2+a-b^2-b}{2}} \\
&= 2 \sum_{b\geq 0, h>0} (-1)^h h q^{\frac{h^2+h}{2}+bh} = 2\sum_{h>0} \frac{(-1)^h h q^{h(h+1)/2}}{1-q^h},
\end{aligned}$$

which yields the result. □

9. Proof of Theorem 1.4

In this section we compute an asymptotic expansion for $V(n)$. We follow a circle method argument used by Bringmann and Mahlburg [2011] (see also [Bringmann and Mahlburg 2013]) to calculate an asymptotic for coefficients of mixed modular forms. It is convenient to work with the two variable rank generating function for concave compositions given in (1-7) and (1-8). We begin with the proof of Theorem 1.6.

Proof of Theorem 1.6. The identity for $v_d(x,q)$ follows from an application of (3.6) of [Andrews 1981] and then (13.3) and (6.1) of [Fine 1988]. The second identity is a corollary of Theorem 4 of [Choi 2011]. □

9A. ***Circle method and the proof of Theorem 1.4.*** By Cauchy's theorem we have

$$V(n-1) = \frac{1}{2\pi i}\int_C \frac{qv(q)}{q^{n+1}}\,dq,$$

where C is an arbitrary path inside the unit circle that loops around 0 in the counterclockwise direction. We choose the circle with radius $r = e^{-\pi/N^2}$ with $N := \lfloor n^{1/2}\rfloor$, and use the parametrization $q = e^{-2\pi/N^2+2\pi it}$ with $0 \le t \le 1$. As usual in the circle method, we define

$$\vartheta'_{h,k} := \frac{1}{k(k_1+k)} \quad \text{and} \quad \vartheta''_{h,k} := \frac{1}{k(k_2+k)},$$

where $h_1/k_1 < h/k < h_2/k_2$ are adjacent Farey fractions in the Farey sequence of order $N := \lfloor n^{1/2}\rfloor$. So $1/(k+k_j) \le 1/(N+1)$ for $j = 1, 2$. Next, decompose the path of integration into paths along the Farey arcs $-\vartheta'_{h,k} \le \Phi \le \vartheta''_{h,k}$, where Φ is defined by $z = (k/N^2) - ki\Phi$ with $-\vartheta'_{h,k} \le \Phi \le \vartheta''_{h,k}$ and $0 \le h \le k \le N$ with $(h,k) = 1$. Hence,

$$V(n-1) = \sum_{\substack{1\le k<N\\(h,k)=1}} e^{-\frac{2\pi ihn}{k}} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} v\big(e^{\frac{2\pi i}{k}(h+iz)}\big) e^{\frac{2\pi i}{k}(h+iz)} e^{\frac{2\pi nz}{k}}\,d\Phi. \tag{9-1}$$

For computing the asymptotic nature of $v(e^{(2\pi i/k)(h+iz)})$ as $z \to 0$, the useful form of Theorem 1.6 is

$$qv(x,q) = q^{\frac{1}{12}} \frac{\eta(\tau)}{2\sin(\pi u)\vartheta(u;\tau)} R(u,\tau) - iq^{\frac{1}{8}}\mu(u,-u;\tau), \tag{9-2}$$

where we have set $q = e^{2\pi i\tau}$, $x = e^{2\pi iu}$, and

$$R(u,\tau) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(xq)_n(x^{-1}q)_n}$$

is the Rank generating function; see [Ono 2008]. The idea is to compute the asymptotics as $z \to 0^+$ of the two variable version of the generating function and then set $u = 0$. We are interested in exponential growth and will freely ignore terms once they are determined to have smaller growth.

The proof of Proposition 3.5 of [Bringmann et al. 2012] gives the following asymptotic evaluation for $R(u,\tau)$. Let $\widetilde{h} \in \{-1, 0, 1\}$ defined by $\widetilde{h} \equiv h \pmod 3$.

Moreover, if $0 \le \ell < k$, then we write

$$\alpha^{\pm}(\ell,k) := \frac{1}{k}\Big(\pm\tfrac{1}{3} - \Big(\ell - \frac{k-1}{2}\Big)\Big), \quad \widetilde{k} = \begin{cases} 3k & \text{if } 3 \nmid k, \\ k/3 & \text{if } 3 \mid k, \end{cases}$$

$$\begin{aligned} &\widetilde{H}^{\pm}(u,h,\ell,k;z) \\ &\quad := \pm e^{\frac{\widetilde{h}^2\pi}{3kz}} \xi_\ell\Big(\frac{3h}{(3,k)}, \frac{k}{(3,k)}\Big) H\Big(\frac{(3,k)iu}{z} + \alpha^{\pm}\Big(\ell, \frac{k}{(3,k)}\Big) \mp \frac{\widetilde{h}i}{kz}; \frac{i}{\widetilde{k}z}\Big), \end{aligned} \tag{9-3}$$

where

$$\xi_\ell(h,k) := (-1)^{\ell+1} e^{-\frac{\pi i h}{4k}(2\ell+1)^2 \mp \frac{\pi i}{k}(\frac{\widetilde{h}-h}{3})(2\ell+1) + \frac{2\pi i \widetilde{h}}{9k}}. \tag{9-4}$$

Proposition 9.1. *For $\tau = (1/k)(h+iz)$ with $0 \le h < k$ with $(h,k) = 1$ set $q = e^{(2\pi i/k)(h+iz)}$ and $q_1 = e^{(2\pi i/k)([-h]_k + (i/z))}$. With this notation, we have*

$$\begin{aligned} R(u,\tau) &= -i^{\frac{3}{2}} \chi^{-1}\big(h,[-h]_k,k\big)(qq_1^{-1})^{\frac{1}{24}} z^{-\frac{1}{2}} e^{\frac{3k\pi u^2}{z}} \frac{\sin(\pi u)}{\sinh(\frac{\pi u}{z})} \\ &\quad + \sum_{\ell=1}^{\infty} a_\ell(z) \frac{(2\pi i u)^\ell}{\ell!} - \frac{i\sin(\pi u)}{\sqrt{\widetilde{k}z}} e^{-\frac{\pi z}{12k} - \frac{\pi i h}{4k} + \frac{3\pi k u^2}{z}} \sum_{\pm} \sum_{\ell=0}^{\frac{k}{(3,k)}-1} \widetilde{H}^{\pm}(u,h,\ell,k;z) \end{aligned}$$

with $a_\ell(z) \ll_\ell |z|^{1-\ell} e^{-(23\pi/(12k))\operatorname{Re}(1/z)}$.

Additionally, we have

$$\frac{q^{\frac{1}{12}}\eta(\tau)}{\vartheta(u;\tau)} = q^{\frac{1}{12}} \chi^{-2}(h,[-h]_k,k) e^{k\pi u^2/z} \frac{\eta(\tau')}{\vartheta(\frac{iu}{z};\tau')} = \frac{q^{\frac{1}{12}} q_1^{\frac{1}{24}}(1+O(q_1)) e^{k\pi u^2/z}}{\chi^2(h,[-h]_k,k)\vartheta(\frac{iu}{z};\tau')}.$$

Thus

$$\begin{aligned} &\frac{q^{\frac{1}{12}}\eta(\tau)}{2\sin(\pi u)\vartheta(u;\tau)} R(u,\tau) \\ &\quad = -i^{\frac{3}{2}} \chi^{-3}(h,[-h]_k,k) q^{\frac{1}{8}} \frac{z^{-\frac{1}{2}} e^{\frac{4\pi k u^2}{z}}}{2\sinh(\frac{\pi u}{z})\vartheta(\frac{iu}{z};\tau')} + O\big(e^{-\frac{\alpha}{k}\operatorname{Re}(\frac{1}{z})}\big) \\ &\qquad - i\chi^{-2}(h,[-h]_k,k) \frac{q^{\frac{1}{12}} e^{-\frac{\pi z}{12k} - \frac{\pi i h}{4k}} q_1^{\frac{1}{24}}(1+O(q_1))}{2\sqrt{\widetilde{k}z}\vartheta(\frac{iu}{z};\tau')} \\ &\qquad\qquad \times \sum_{\pm} \sum_{\ell=0}^{\frac{k}{(3,k)}-1} \widetilde{H}^{\pm}(u,h,\ell,k;z) \end{aligned} \tag{9-5}$$

for some $\alpha > 0$ independent of k as $z \to 0$ and $u \to 0$.

Turning to the other term we have the following proposition.

Proposition 9.2. *Let* $\tau = \frac{1}{k}(h + iz)$ *and* $\tau' = (1/k)([-h]_k + (i/z))$ *then*

$$\mu(u,-u;\tau) = i^{\frac{1}{2}}\chi^{-3}(h,[-h]_k,k)z^{-\frac{1}{2}}e^{\frac{4\pi k u^2}{z}}\mu\Big(\frac{iu}{z},-\frac{iu}{z};\tau'\Big)$$
$$+\frac{(-1)^{h+1}i}{2}e^{\frac{4\pi k u^2}{z}}\sum_{\ell=0}^{k-1}e^{-\frac{\pi i}{k}(\ell+\frac{1}{2})^2}H\Big(\frac{2iu}{z}-\frac{1}{k}\Big(\ell-\frac{k-1}{2}\Big);\frac{i}{kz}\Big).$$

Proof. Theorem 2.5 implies that

$$\mu(u,-u;\tau) = \chi^{-3}\big(h,[-h]_k,k\big)\sqrt{\frac{i}{z}}e^{\frac{4\pi k u^2}{z}}\mu\Big(\frac{iu}{z},-\frac{iu}{z};\tau'\Big)$$
$$+\frac{i}{2}\chi^{-3}\big(h,[-h]_k,k\big)\sqrt{\frac{i}{z}}e^{\frac{4\pi k u^2}{z}}R\Big(\frac{2iu}{z};\tau'\Big)-\frac{i}{2}R(2u;\tau). \quad (9\text{-}6)$$

Propositions 2.3 and 2.4 yield

$$R(2u;\tau) = -\sum_{\ell=0}^{k-1}e^{-\frac{\pi i}{k}(h+iz)(\ell-\frac{k-1}{2})^2-4\pi iu(\ell-\frac{k-1}{2})}e^{-\pi i(\ell-\frac{k-1}{2})-\frac{\pi ikh}{4}}$$
$$\times(-1)^{(\ell-\frac{k-1}{2})(h+1)}\frac{e^{\frac{\pi}{kz}(2uk+(\ell-\frac{k-1}{2})iz)^2}}{\sqrt{kz}}$$
$$\times\bigg(R\Big(\frac{2iu}{z}-\frac{1}{k}\Big(\ell-\frac{k-1}{2}\Big);\frac{i}{kz}\Big)-H\Big(\frac{2iu}{z}-\frac{1}{k}\Big(\ell-\frac{k-1}{2}\Big);\frac{i}{kz}\Big)\bigg).$$

The nonholomorphic R-functions above will exactly cancel with the other term of (9-6); this can be shown as in [Bringmann and Mahlburg 2013; Bringmann et al. 2012]. Simplifying the factors and multiplying the H-functions gives the result. □

We will evaluate the asymptotic nature of the two terms on the right hand side of Proposition 9.2. First consider the terms with the H-function. We have

$$H\Big(\frac{2iu}{z}-\frac{1}{k}\Big(\ell-\frac{k-1}{2}\Big);\frac{i}{kz}\Big) = O\bigg(\int_{-\infty}^{\infty}e^{-\frac{\pi x^2}{k}\operatorname{Re}(\frac{1}{z})}\,dx\bigg) = O\Big(\sqrt{k}\operatorname{Re}\Big(\frac{1}{z}\Big)^{-\frac{1}{2}}\Big)$$

as $u \to 0$ and $z \to 0$. Next turning to the other term, we have

$$\mu\Big(\frac{iu}{z},-\frac{iu}{z};\tau'\Big) = \frac{1}{2\sinh(\frac{\pi u}{z})\vartheta(-\frac{iu}{z};\tau')}+\frac{e^{-\frac{\pi u}{z}}}{\vartheta(-\frac{iu}{z};\tau')}\sum_{n\neq 0}\frac{(-1)^n q_1^{\frac{n(n+1)}{2}}e^{\frac{2\pi nu}{z}}}{1-e^{-\frac{2\pi u}{z}}q_1^n}.$$

The summation is $O(uq_1)$, since it is 0 when $u = 0$ by symmetry. Consequently, using Lemma 5.2 we have $\vartheta(-iu/z;\tau')^{-1} = O((z/u)q_1^{-1/8})$, and we see that the term with the summation is $O(q_1^{7/8}) = O(e^{-7\pi/(4kz)})$ after setting $u = 0$. So it is

of exponential decay and will not contribute to the asymptotic expansion of our generating function. Hence

$$\begin{aligned}
&-iq^{\frac{1}{8}}\mu(u,-u;\tau)\\
&\qquad = -i^{\frac{3}{2}}\chi^{-3}(h,[-h]_k,k)q^{\frac{1}{8}}z^{-\frac{1}{2}}e^{4\pi k u^2/z}\mu\Big(\frac{iu}{z},-\frac{iu}{z};\tau'\Big)+O_k(1)\\
&\qquad = \frac{i^{\frac{3}{2}}q^{\frac{1}{8}}\chi^{-3}(h,[-h]_k,k)z^{-\frac{1}{2}}\,e^{4\pi k u^2/z}}{2\sinh(\frac{\pi u}{z})\vartheta(\frac{iu}{z};\tau')}+O\big(z^{-\frac{1}{2}}q_1^{-\frac{7}{8}}\big)+O_k(1) \qquad (9\text{-}7)
\end{aligned}$$

as $u\to 0$ when z is chosen as in the circle method calculations.

Using (9-2), (9-5), and (9-7) we see that as $u\to 0$ and $z\to 0$ we have the following asymptotic evaluation for $v(q)$ with $q=e^{(2\pi i/k)(h+iz)}$.

$$\begin{aligned}
qv(q) &\sim -i\chi^{-2}(h,[-h]_k,k)\\
&\qquad\times\frac{q^{\frac{1}{12}}e^{-\frac{\pi z}{12k}-\frac{\pi i h}{4k}}q_1^{\frac{1}{24}}(1+O(q_1))}{2\sqrt{\tilde{k}z}\vartheta(\frac{iu}{z};\tau')}\sum_{\pm}\sum_{\ell=0}^{\frac{k}{(3,k)}-1}\tilde{H}^{\pm}(u,h,\ell,k;z).
\end{aligned}$$

It is evident that the largest growth comes from the case when $k=1$ and thus the largest contribution to (9-1) will come from the case $k=1$. In this case we set

$$\begin{aligned}
\sum_{\pm}\tilde{H}^{\pm}(u,0,0,1;z) &= -H\Big(\frac{iu}{z}+\tfrac{1}{3};\frac{i}{3z}\Big)+H\Big(\frac{iu}{z}-\tfrac{1}{3};\frac{i}{3z}\Big)\\
&= -\int_{-\infty}^{\infty}\frac{e^{-\pi x^2/(3z)}}{\cosh(\pi x)}e^{-2\pi i x u/z}\big(e^{-2\pi x/3}-e^{2\pi x/3}\big)\,dx\\
&= -2i\int_{-\infty}^{\infty}e^{-\pi x^2/(3z)}\sin\Big(\frac{2\pi x u}{z}\Big)\frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)}\,dx\\
&= -2i\cdot\frac{2\pi u}{z}\int_{-\infty}^{\infty}xe^{-\pi x^2/(3z)}\frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)}\,dx+O(u^3)
\end{aligned}$$

Therefore, with $q=e^{-2\pi z}$ we have

$$\begin{aligned}
qv(q) &\sim \lim_{u\to 0}-e^{-\frac{\pi z}{4}}e^{-\frac{\pi}{12z}}\frac{1}{2\sqrt{3z}\vartheta(\frac{iu}{z};\tau')}\\
&\qquad\times\Big(-\frac{4\pi i u}{z}\int_{-\infty}^{\infty}xe^{-\frac{\pi x^2}{3z}}\frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)}\,dx\Big)\\
&\sim \frac{e^{\frac{\pi}{6}(\frac{1}{z}-\frac{3}{2}z)}}{\sqrt{3z}}\int_{-\infty}^{\infty}xe^{-\frac{\pi x^2}{3z}}\frac{\sinh(\frac{2\pi x}{3})}{\cos(\pi x)}\,dx.
\end{aligned}$$

Therefore, we have

$$V(n-1) \sim -\int_{-\vartheta'_{0,1}}^{\vartheta''_{0,1}} \frac{e^{\frac{\pi}{6}(\frac{1}{z}+(12n-\frac{3}{2})z)}}{\sqrt{3z}} \int_{-\infty}^{\infty} x e^{-\frac{\pi x^2}{3z}} \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)}\, dx\, d\Phi$$
$$\sim -\frac{1}{\sqrt{3}} \int_{-\infty}^{\infty} x \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)} \int_{-\frac{1}{N}}^{\frac{1}{N}} z^{-\frac{1}{2}} e^{\frac{\pi}{6}((1-2x^2)\frac{1}{z}+(12n-\frac{3}{2})z)}\, d\Phi\, dx$$

where $z = (1/N^2) - i\Phi$ and we symmetrize the integral by using

$$\int_{-\vartheta'_{0,1}}^{\vartheta''_{0,1}} = \int_{-\frac{1}{N}}^{\frac{1}{N}} - \int_{-\frac{1}{N}}^{-\vartheta'_{0,1}} - \int_{\vartheta''_{0,1}}^{\frac{1}{N}}$$

and $\sim$ represents the asymptotic expansion with respect to n. The final two sums contribute a polynomially bounded error with respect to N. We handle the resulting integral with respect to Φ exactly as in Proposition 3.2 of [Bringmann and Mahlburg 2011] (see Lemma 4.2 of [Rhoades 2012] for an analogous calculation). This results in

$$\int_{-\frac{1}{N}}^{\frac{1}{N}} z^{-\frac{1}{2}} e^{\frac{\pi}{6}((1-2x^2)\frac{1}{z}+(12n-\frac{3}{2})z)}\, d\Phi$$
$$= -2\pi \Big(\frac{1-2x^2}{12n-\frac{3}{2}}\Big)^{\frac{1}{4}} I_{\frac{1}{2}}\Big(\frac{\pi}{3}\sqrt{(12n-\tfrac{3}{2})(1-2x^2)}\Big)$$
$$= -2\sqrt{6}(12n-\tfrac{3}{2})^{-\frac{1}{2}} \sinh\Big(\frac{\pi}{3}\sqrt{(12n-\tfrac{3}{2})(1-2x^2)}\Big)$$

Therefore,

$$V(n-1) \sim \frac{2\sqrt{2}}{(12n-\frac{3}{2})^{\frac{1}{2}}} \int_{-\infty}^{\infty} x \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)} \sinh\Big(\frac{\pi}{3}\sqrt{(12n-\tfrac{3}{2})(1-2x^2)}\Big)\, dx$$
$$\sim \frac{2\sqrt{2}}{(12n-\frac{3}{2})^{\frac{1}{2}}} \int_{-\frac{1}{\sqrt{2}}}^{\frac{1}{\sqrt{2}}} x \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)} \sinh\Big(\frac{\pi}{3}\sqrt{(12n-\tfrac{3}{2})(1-2x^2)}\Big)\, dx$$
$$\sim \frac{2\sqrt{2}}{(12n-\frac{3}{2})^{\frac{1}{2}}} \int_{0}^{\frac{1}{\sqrt{2}}} x \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)} \exp\Big(\frac{\pi}{3}\sqrt{(12n-\tfrac{3}{2})(1-2x^2)}\Big)\, dx \tag{9-8}$$

where the second equality follows because

$$\sinh\Big(\frac{\pi}{3}\sqrt{(12n-\tfrac{3}{2})(1-2x^2)}\Big) = O(1) \quad \text{for } |x| > \frac{1}{\sqrt{2}}.$$

We have the following lemma which gives the asymptotic expansion of the remaining integral.

Lemma 9.3. *Let α_k for $k \geq 1$ be defined by*

$$\sum_{k=1}^{\infty} \alpha_k x^{2k} = \exp\bigl(\sqrt{1-2x^2} - 1 + x^2\bigr)\, x\, \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)}.$$

As $y \to \infty$ we have

$$\int_0^{\frac{1}{\sqrt{2}}} f e^{y\sqrt{1-2x^2}} x \frac{\sinh(\frac{2\pi x}{3})}{\cosh(\pi x)} \sim e^y \frac{\sqrt{\pi}}{2\sqrt{y}} \sum_{t=1}^{\infty} \frac{(2t-1)!!}{2^{2t}} \alpha_t \frac{1}{y^t}.$$

From (9-8) and Lemma 9.3 we have

$$V(n-1) \sim \frac{\sqrt{6}}{(12n-\frac{3}{2})^{\frac{3}{4}}} \exp\Bigl(\frac{\pi}{3}\sqrt{12n-\tfrac{3}{2}}\Bigr) \sum_{t=1}^{\infty} \frac{(2t-1)!!\, 3^t}{2^{2t}\pi^t} \alpha_t (12n-\tfrac{3}{2})^{-t/2},$$

which gives Theorem 1.4.

Note we have

$$\sum_{k=1}^{\infty} \alpha_k x^{2k} = \frac{2\pi}{3}x^2 - \frac{23}{81}\pi^3 x^4 - \frac{4860\pi - 1681\pi^5}{14580} x^6 + \frac{1837080\pi + 257543\pi^7 - 782460\pi^3}{5511240} x^8 + \cdots$$

10. Proof of Theorem 1.5

In this section we compute an asymptotic for $V_d(n)$. We follow the standard circle method set-up as in Section 9. As above, we have

$$V_d(n) = \sum_{\substack{1 \leq k < N \\ (h,k)=1}} e^{-2\pi i h n/k} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} v_d\bigl(e^{(2\pi i/k)(h+iz)}\bigr) e^{2\pi n z/k}\, d\Phi. \tag{10-1}$$

Lemma 10.1. *Let $c_{h,k}(n) := (-12/n) e^{(\pi i h/12k)n^2}$. Then if $h \neq 0$ then $c_{h,k}(n)$ is $24k$ periodic and has mean value 0.*

Applying this lemma with the proposition in Section 3 of [Lawrence and Zagier 1999], we have the following.

Lemma 10.2. *Let $q = e^{(2\pi i/k)(h+iz)}$. Then*

$$\sum_{n=0}^{\infty} \Bigl(\frac{-12}{n}\Bigr) q^{n^2/24} \sim \sum_{r=0}^{\infty} L(-2r, h, k) \Bigl(\frac{-\pi}{12k}\Bigr)^r \frac{z^r}{r!}$$

where

$$L(-r,h,k) := -\frac{(24k)^r}{r+1}\sum_{n=1}^{24k} c_{h,k}(n) B_{r+1}\Big(\frac{n}{24k}\Big)$$

where $B_r(x)$ denotes the r-th Bernoulli polynomial and c_k is defined in Lemma 10.1. If $h=0$ we have

$$L(-r) := L(-r,0,1) = -\frac{6^r}{r+1}\big(B_{r+1}(\tfrac{1}{6}) - B_{r+1}(\tfrac{5}{6})\big).$$

By (2-2) we have

$$(-q)_\infty^2 = \begin{cases} \frac{1}{2} e^{\frac{\pi}{12k}(\frac{1}{z}+2z)} i^{\frac{1-k}{2}} \omega_{h,k} e^{-\frac{\pi i h}{4k}+\frac{\pi i k h}{4}} (h/k)\big(1+O\big(e^{-\frac{\pi}{kz}}\big)\big) & \text{if } 2 \nmid k, \\ O\big(e^{-\frac{\pi}{6kz}}\big) & \text{if } 2 \mid k, \end{cases}$$

so that we have the following results.

Proposition 10.3. *Let $q = e^{(2\pi i/k)(h+iz)}$ with $0 \le h < k$ and $(h,k)=1$. If $2 \mid k$, then $v_d(q) + \sum_{n\ge 0}(-1)^n q^{(n+1)(n+2)/2} = O(e^{-\pi/(6kz)})$. When $2 \nmid k$, we have*

$$v_d(q) + \sum_{n\ge 0}(-1)^n q^{\frac{(n+1)(n+2)}{2}} \sim \tfrac{1}{2} e^{\frac{\pi}{12k}(\frac{1}{z}+2z)} i^{\frac{1-k}{2}} \omega_{h,k} e^{-\frac{\pi i h}{4k}+\frac{\pi i k h}{4}} \Big(\frac{h}{k}\Big) \sum_{r=0}^{\infty} L(-2r,k)\Big(\frac{-\pi}{12k}\Big)^r \frac{z^r}{r!}$$

Using these asymptotics with (10-1) and the integral evaluation

$$\int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} e^{\frac{\pi}{12k}(mz+\frac{1}{z})} z^{\frac{1}{2}-j}\, d\Phi = \frac{2\pi}{k} m^{\frac{j}{2}-\frac{3}{4}} I_{\frac{3}{2}-j}\Big(\frac{\pi\sqrt{m}}{6k}\Big) + O(N^{-\frac{1}{4}}) \tag{10-2}$$

(see [Lehner 1964] for details), we have

$$V_d(n) \sim (-1)^{n+1}\delta(n) + \pi \sum_{\substack{2\nmid k<N \\ (h,k)=1}} \frac{\mathcal{A}_k(n)}{k} \sum_{r=0}^{\infty} L(-2r,k)\Big(\frac{-\pi}{12k}\Big)^r \frac{1}{r!}(24n+2)^{-\frac{r+1}{2}} I_{r-1}\Big(\frac{\pi}{6k}\sqrt{24n+2}\Big)$$

where

$$\mathcal{A}_k(n) = \sum_{(h,k)=1} i^{\frac{1-k}{2}} \Big(\frac{h}{k}\Big) \omega_{h,k} e^{-\frac{\pi i h}{4}(\frac{1}{k}+k)-\frac{2\pi i h n}{k}},$$

$$\delta(m) = \begin{cases} 1 & \text{if } m = (n+1)(n+2)/2 \text{ for some } n, \\ 0 & \text{else.} \end{cases}$$

The main term comes from $k = 1$ and so we have

$$\begin{aligned}
V_d(n) &\sim (-1)^{n+1}\delta(n) \\
&\qquad + \pi \sum_{r=0}^{\infty} L(-2r)\Big(\frac{-\pi}{12}\Big)^r \frac{1}{r!}(24n+2)^{-\frac{r+1}{2}} I_{r-1}\Big(\frac{\pi}{6}\sqrt{24n+2}\Big) \\
&\sim (-1)^{n+1}\delta(n) + \frac{\sqrt{3}e^{\frac{\pi}{6}\sqrt{24n+2}}}{(24n+2)^{\frac{3}{4}}} \sum_{r=0}^{\infty} L(-2r)\Big(\frac{\pi}{12}\Big)^r \frac{(-1)^r}{r!} \\
&\qquad\qquad \times \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!}\Big(\frac{3}{4\pi}\Big)^\ell p(\ell, r)(24n+2)^{-\frac{r+\ell}{2}} \\
&\sim (-1)^{n+1}\delta(n) + \frac{\sqrt{3}e^{\frac{\pi}{6}\sqrt{24n+2}}}{(24n+2)^{\frac{3}{4}}} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^{2m} m!\,(24n+2)^{\frac{m}{2}}}\gamma(m).
\end{aligned}$$

11. The q-series $v_1(q)$

We have used a variety of different methods to establish the modular properties of the q-series defined by $v_2(q)$ or

$$\sum_{n\neq 0} \frac{(-1)^{n+1} n q^{(n^2+n)/2}}{1-q^n}.$$

On the other hand, we quoted the results of Bringmann [2008] for the modularity of $v_1(q)$. But this q-series is susceptible to similar methods of the ones discussed here. Analogous to Theorem 1.3 and Proposition 8.1 we have the following result for the Appell sum appearing in the definition in $v_1(q)$ and the Appell-like sum arising from the holomorphic projection operation.

Theorem 11.1. *In the notation above*,

$$\begin{aligned}
\pi_{\mathrm{hol}}(\eta\eta^*) &= \sum_{n\neq 0} (-1)^n \Big(\frac{-3}{n-1}\Big) \frac{n q^{n(n+1)/6}}{1-q^n} \\
&= \sum_{n\neq 0} \frac{(-1)^n q^{3n(n+1)/2}}{(1-q^n)^2} + \sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)^2} = \sum_{n>0} \frac{(q;q)_{n-1} q^n}{(1-q^n)}.
\end{aligned}$$

Remark. It follows that

$$\sum_{n=1}^{\infty} \mathrm{spt}(n) q^n = \frac{1}{(q)_\infty}\pi_{\mathrm{hol}}(\eta^*\eta) = \frac{1}{(q)_\infty} \sum_{n\neq 0} (-1)^n \Big(\frac{-3}{n-1}\Big) \frac{n q^{n(n+1)/6}}{1-q^n}$$

where $\mathrm{spt}(n)$ is the smallest parts function and is equal to the sum of the number of smallest parts in the partitions of n. See [Andrews 2008] for more on the spt-function.

Sketch of proof of Theorem 11.1. The first equality is given in Zagier's Bourbaki lecture [2009], while the second equality can be proved via "modular" methods discussed above. It would be interesting to establish a q-series proof of the second equality.

Finally, the last equality is derived in [Andrews 2008]. It may also be derived from the recent results of the first author, Garvan, and Liang [Andrews et al. 2012]. Namely, letting $z = 1$ in [ibid., Theorem 2.4] gives

$$\sum_{n\geq 1} \frac{(q)_{n-1}q^n}{(1-q^n)} = \sum_{n\neq 0} \frac{(-1)^{n-1}q^{n(n+1)/2}}{(1-q^n)^2} - \sum_{n\neq 0} \frac{(-1)^{n-1}q^{n(3n+1)/2}}{(1-q^n)^2}.$$

The result follows from the well known identity

$$\sum_{n\neq 0} \frac{(-1)^{n-1}q^{n(n+1)/2}}{(1-q^n)^2} = \sum_{n=1}^{\infty} \frac{q^n}{(1-q^n)^2}. \qquad \square$$

References

[Andrews 1981] G. E. Andrews, "Ramanujan's "lost" notebook, I: Partial θ-functions", *Adv. in Math.* **41**:2 (1981), 137–172. MR 83m:10034a Zbl 0477.33001

[Andrews 1998] G. E. Andrews, *The theory of partitions*, Cambridge Mathematical Library, Cambridge University Press, 1998. MR 99c:11126 Zbl 0996.11002

[Andrews 2008] G. E. Andrews, "The number of smallest parts in the partitions of n", *J. Reine Angew. Math.* **624** (2008), 133–142. MR 2009m:11172 Zbl 1153.11053

[Andrews 2011] G. E. Andrews, "Concave compositions", *Electron. J. Combin.* **18**:2 (2011), Paper 6, 13. MR 2012e:05039 Zbl 1229.05029

[Andrews 2013] G. E. Andrews, "Concave and convex compositions", *Ramanujan J.* **31**:1-2 (2013), 67–82. MR 3048655 Zbl 06181374

[Andrews et al. 2012] G. E. Andrews, F. G. Garvan, and J. Liang, "Combinatorial interpretations of congruences for the spt-function", *Ramanujan J.* **29**:1-3 (2012), 321–338. MR 2994105 Zbl 06120459

[Bailey 1950] W. N. Bailey, "On the basic bilateral hypergeometric series ${}_2\Psi_2$", *Quart. J. Math., Oxford Ser.* (2) **1** (1950), 194–198. MR 12,178e Zbl 0038.05001

[Bringmann 2008] K. Bringmann, "On the explicit construction of higher deformations of partition statistics", *Duke Math. J.* **144**:2 (2008), 195–233. MR 2009e:11203 Zbl 1154.11034

[Bringmann and Folsom 2013] K. Bringmann and A. Folsom, "On the asymptotic behavior of Kac-Wakimoto characters", *Proc. Amer. Math. Soc.* **141**:5 (2013), 1567–1576. MR 3020844 Zbl 06146504

[Bringmann and Mahlburg 2011] K. Bringmann and K. Mahlburg, "An extension of the Hardy–Ramanujan circle method and applications to partitions without sequences", *Amer. J. Math.* **133**:4 (2011), 1151–1178. MR 2012g:11184 Zbl 1251.11072

[Bringmann and Mahlburg 2013] K. Bringmann and K. Mahlburg, "Asymptotic formulas for coefficients of Kac-Wakimoto characters", *Math. Proc. Cambridge Philos. Soc.* **155**:1 (2013), 51–72. MR 3065259 Zbl 06191889

[Bringmann et al. 2012] K. Bringmann, K. Mahlburg, and R. C. Rhoades, "Taylor coefficients of mock-Jacobi forms and moments of partition statistics", preprint, 2012, Available at math.stanford.edu/~rhoades/FILES/RCZ.pdf.

[Bryson et al. 2012] J. Bryson, K. Ono, S. Pitman, and R. C. Rhoades, "Unimodal sequences and quantum and mock modular forms", *Proc. Natl. Acad. Sci. USA* **109**:40 (2012), 16063–16067. MR 2994899

[Cheng 2010] M. C. N. Cheng, "$K3$ surfaces, $N = 4$ dyons and the Mathieu group M_{24}", *Commun. Number Theory Phys.* **4**:4 (2010), 623–657. MR 2012e:11076 Zbl 1231.14031

[Choi 2011] Y.-S. Choi, "The basic bilateral hypergeometric series and the mock theta functions", *Ramanujan J.* **24**:3 (2011), 345–386. MR 2012e:33042 Zbl 1225.33019

[Coates 1986] J. Coates, "The work of Gross and Zagier on Heegner points and the derivatives of L-series", pp. 57–72 in *Séminaire Bourbaki,* 1984/85, Astérisque **133-134**, Société Mathématique de France, Paris, 1986. MR 87j:11058 Zbl 0608.14020

[Diaconis et al. 2013] P. Diaconis, S. Janson, and R. C. Rhoades, "Note on a partition limit theorem for rank and crank", *Bull. Lond. Math. Soc.* **45**:3 (2013), 551–553. MR 3065024 Zbl 06176878

[Eguchi and Hikami 2009] T. Eguchi and K. Hikami, "Superconformal algebras and mock theta functions, II: Rademacher expansion for $K3$ surface", *Commun. Number Theory Phys.* **3**:3 (2009), 531–554. MR 2011b:58048 Zbl 1189.81191

[Eguchi et al. 2011] T. Eguchi, H. Ooguri, and Y. Tachikawa, "Notes on the $K3$ surface and the Mathieu group M_{24}", *Exp. Math.* **20**:1 (2011), 91–96. MR 2012e:58039

[Fine 1988] N. J. Fine, *Basic hypergeometric series and applications*, Mathematical Surveys and Monographs **27**, American Mathematical Society, Providence, RI, 1988. MR 91j:33011 Zbl 0647.05004

[Fristedt 1993] B. Fristedt, "The structure of random partitions of large integers", *Trans. Amer. Math. Soc.* **337**:2 (1993), 703–735. MR 93h:11090 Zbl 0795.05009

[Gasper and Rahman 2004] G. Gasper and M. Rahman, *Basic hypergeometric series*, 2nd ed., Encyclopedia of Mathematics and its Applications **96**, Cambridge University Press, 2004. MR 2006d:33028 Zbl 1129.33005

[Gross and Zagier 1986] B. H. Gross and D. B. Zagier, "Heegner points and derivatives of L-series", *Invent. Math.* **84**:2 (1986), 225–320. MR 87j:11057 Zbl 0608.14019

[Heubach and Mansour 2010] S. Heubach and T. Mansour, *Combinatorics of compositions and words*, Discrete Mathematics and its Applications, CRC Press, Boca Raton, FL, 2010. MR 2010i:68122 Zbl 1184.68373

[Lawrence and Zagier 1999] R. Lawrence and D. Zagier, "Modular forms and quantum invariants of 3-manifolds", *Asian J. Math.* **3**:1 (1999), 93–107. MR 2000j:11057 Zbl 1024.11028

[Lehner 1964] J. Lehner, "On automorphic forms of negative dimension", *Illinois J. Math.* **8** (1964), 395–407. MR 29 #2230 Zbl 0129.05801

[MacMahon 1893] P. A. MacMahon, "Memoir on the theory of compositions of numbers", *London R. S. Phil. A* **184** (1893), 835–901. JFM 25.0258.01

[Malmendier and Ono 2012] A. Malmendier and K. Ono, "Moonshine and Donaldson invariants of $\mathbb{C}P^2$", preprint, 2012. arXiv 1207.5139

[Ono 2004] K. Ono, *The web of modularity: Arithmetic of the coefficients of modular forms and q-series*, CBMS Regional Conference Series in Mathematics **102**, Amer. Math. Soc., Providence, 2004. MR 2005c:11053 Zbl 1119.11026

[Ono 2008] K. Ono, "Mock theta functions, ranks, and Maass forms", pp. 119–141 in *Surveys in number theory*, edited by K. Alladi, Dev. Math. **17**, Springer, New York, 2008. MR 2009j:11170 Zbl 1183.11064

[Ono 2009] K. Ono, "Unearthing the visions of a master: harmonic Maass forms and number theory", pp. 347–454 in *Current developments in mathematics, 2008*, edited by D. Jerison et al., International Press, 2009. MR 2010m:11060 Zbl 1229.11074

[Pittel 1997] B. Pittel, "On a likely shape of the random Ferrers diagram", *Adv. in Appl. Math.* **18**:4 (1997), 432–488. MR 98m:60013 Zbl 0894.11039

[Rhoades 2012] R. C. Rhoades, "Strongly unimodal sequences and mixed mock modular forms", preprint, 2012, Available at math.stanford.edu/~rhoades/FILES/unimodal.pdf.

[Rhoades 2013] R. C. Rhoades, "Families of quasimodular forms and Jacobi forms: the crank statistic for partitions", *Proc. Amer. Math. Soc.* **141**:1 (2013), 29–39. MR 2988708

[Shimura 1973] G. Shimura, "On modular forms of half integral weight", *Ann. of Math.* (2) **97** (1973), 440–481. MR 48 #10989 Zbl 0266.10022

[Sturm 1980] J. Sturm, "Projections of C^∞ automorphic forms", *Bull. Amer. Math. Soc.* (*N.S.*) **2**:3 (1980), 435–439. MR 81g:10039 Zbl 0433.10013

[Vershik 1995] A. M. Vershik, "Asymptotic combinatorics and algebraic analysis", pp. 1384–1394 in *Proceedings of the International Congress of Mathematicians* (Zürich, 1994), vol. 2, edited by S. D. Chatterji, Birkhäuser, Basel, 1995. MR 98e:60018 Zbl 0843.05003

[Zagier 2001] D. Zagier, "Vassiliev invariants and a strange identity related to the Dedekind eta-function", *Topology* **40**:5 (2001), 945–960. MR 2002g:11055 Zbl 0989.57009

[Zagier 2009] D. Zagier, "Ramanujan's mock theta functions and their applications (after Zwegers and Ono–Bringmann)", pp. 143–164 (exposé 986) in *Séminaire Bourbaki* 2007/2008, Astérisque **326**, Société Mathématique de France, Paris, 2009. MR 2011h:11049 Zbl 1198.11046

[Zwegers 2002] S. Zwegers, *Mock theta functions*, Ph.D. thesis, Universiteit Utrecht, 2002, Available at igitur-archive.library.uu.nl/dissertations/2003-0127-094324/full.pdf.

Communicated by Georgia Benkart
Received 2012-07-30 Revised 2012-09-10 Accepted 2012-10-22

andrews@math.psu.edu *Department of Mathematics, The Pennsylvania State University, 109 McAllister Building, University Park PA 16802-7000, United States*

rhoades@math.stanford.edu *Department of Mathematics, Stanford University, Bldg 380, Stanford CA 94305, United States*

szwegers@uni-koeln.de *Mathematical Institute, University of Cologne, Weyertal 86-90, D-50931 Cologne, Germany*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2141

Moduli of elliptic curves via twisted stable maps

Andrew Niles

Abramovich, Corti and Vistoli have studied modular compactifications of stacks of curves equipped with abelian level structures arising as substacks of the stack of twisted stable maps into the classifying stack of a finite group, provided the order of the group is invertible on the base scheme. Recently Abramovich, Olsson and Vistoli extended the notion of twisted stable maps to allow arbitrary base schemes, where the target is a tame stack, not necessarily Deligne–Mumford. We use this to extend the results of Abramovich, Corti and Vistoli to the case of elliptic curves with level structures over arbitrary base schemes; we prove that we recover the compactified Katz–Mazur regular models, with a natural moduli interpretation in terms of level structures on Picard schemes of twisted curves. Additionally, we study the interactions of the different such moduli stacks contained in a stack of twisted stable maps in characteristics dividing the level.

1. Introduction

Abramovich and Vistoli [2002] introduced the stack $\mathcal{H}_{g,n}(\mathcal{X})$ of *n-pointed genus-g twisted stable maps* into $\mathcal{X}$, where $\mathcal{X}$ is a proper tame Deligne–Mumford stack over a base scheme S with a projective coarse moduli space X/S. They proved that

MSC2010: primary 11G18; secondary 14K10, 14H10, 14D23, 14H52.
Keywords: generalized elliptic curve, twisted curve, Drinfeld structure, moduli stack.

$\mathcal{K}_{g,n}(\mathcal{X})$ is an algebraic stack, proper over the Kontsevich stack $\overline{\mathcal{M}}_{g,n}(X)$ of stable maps into X, giving an appropriate analogue of the usual Kontsevich stack of stable maps when the target is allowed to be a tame stack instead of merely a scheme or algebraic space. In particular, taking $\mathcal{X} = \mathcal{B}G$ for a finite group G naturally yields a modular compactification of the stack of n-pointed genus-g curves with certain level structure, where the level structure on a curve corresponds to giving a G-torsor over the curve; this is studied extensively in [Abramovich et al. 2003]. This of course differs from the approach in [Abramovich and Romagny 2012], where the coverings of curves are assumed themselves to be stable curves; here the coverings are generally not stable curves. In [Petersen 2012] it is shown that in good characteristics we recover the usual compactification of $\mathcal{Y}(N)$ in $\mathcal{K}_{1,1}(\mathcal{B}(\mathbb{Z}/(N))^2)$.

However, the algebraic stack $\mathcal{B}G$ is not tame in characteristics dividing $|G|$, so a priori the results of [Abramovich et al. 2003] only hold over $\mathbb{Z}[1/|G|]$. The notion of a tame stack is generalized in [Abramovich et al. 2008b] to algebraic stacks that are not necessarily Deligne–Mumford, and in [Abramovich et al. 2011] the stack $\mathcal{K}_{g,n}(\mathcal{X})$ of twisted stable maps into certain proper tame algebraic stacks is introduced and shown to be a proper algebraic stack. This naturally leads us to attempt replacing the finite group G with a group scheme $\mathcal{G}$, agreeing with G over $\mathbb{Z}[1/|G|]$ but such that $\mathcal{B}\mathcal{G}$ is a tame algebraic stack over $\operatorname{Spec}(\mathbb{Z})$; see, for example, [Abramovich 2012]. In particular, the group scheme μ_N is (noncanonically) isomorphic to $\mathbb{Z}/(N)$ over $\mathbb{Z}[1/N, \zeta_N]$, and the classifying stack $\mathcal{B}\mu_N$ (unlike $\mathcal{B}(\mathbb{Z}/(N))$) is a tame stack over $\operatorname{Spec}(\mathbb{Z})$.

The purpose of this paper is to record how the results of [Abramovich et al. 2003] extend to moduli of elliptic curves with level structure over arbitrary base schemes, using the group scheme μ_N in place of $\mathbb{Z}/(N)$. For example, the moduli stack $\mathcal{Y}_1(N)$ arises as an open substack in the rigidification $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$ (defined below) of $\mathcal{K}_{1,1}(\mathcal{B}\mu_N)$. Explicitly, given an elliptic curve E/S and a $[\Gamma_1(N)]$-structure P on E, via the canonical group scheme isomorphism $E \cong \operatorname{Pic}^0_{E/S}$ over S, we may view P as a "point of exact order N" on $\operatorname{Pic}^0_{E/S}$, determining a group scheme homomorphism $\phi : \mathbb{Z}/(N) \to \operatorname{Pic}^0_{E/S}$. The stack $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$ will be seen to classify certain pairs $(\mathscr{C}, \phi)$ where $\mathscr{C}/S$ is a 1-marked genus-1 twisted stable curve and $\phi : \mathbb{Z}/(N) \to \operatorname{Pic}^0_{\mathscr{C}/S}$ is a group scheme homomorphism, so this construction defines the immersion $\mathcal{Y}_1(N) \to \overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$. We will define the notion of a $[\Gamma_1(N)]$-structure on a 1-marked genus-1 twisted stable curve, and write $\mathcal{X}_1^{\mathrm{tw}}(N)$ for the stack classifying such structures. Our first main result is:

Theorem 1.1. *Let S be a scheme and let $\mathcal{X}_1^{\mathrm{tw}}(N)$ be the stack over S classifying $[\Gamma_1(N)]$-structures on 1-marked genus-1 twisted stable curves with nonstacky marking. Then $\mathcal{X}_1^{\mathrm{tw}}(N)$ is a closed substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$, which contains $\mathcal{Y}_1(N)$ as an open dense substack.*

This will be proved as Theorem 4.6. The main point is to verify the valuative criterion for properness. We accomplish this by using properness of $\mathcal{X}_1(N)$ and the e_N-pairing on generalized elliptic curves. To complete a family of twisted curves with level structure when the generic fiber is smooth, we first complete it to a family of generalized elliptic curves with level structure. Then we use a quotient construction involving the e_N-pairing to produce a completed family of twisted curves with level structure, modifying our strategy in characteristics dividing N by exploiting the relationship between cyclotomic torsors and line bundles (Lemma 3.19). The same techniques immediately give a corresponding result for $\mathcal{Y}(N)$, which is an open substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$; its closure $\mathcal{X}^{\mathrm{tw}}(N)$ classifies $[\Gamma(N)]$-structures on 1-marked genus-1 twisted stable curves, as we show in Theorem 5.3.

Unsurprisingly, it turns out that these closures are isomorphic as algebraic stacks to the stacks $\mathcal{X}_1(N)$ and $\mathcal{X}(N)$ classifying $[\Gamma_1(N)]$-structures and $[\Gamma(N)]$-structures on generalized elliptic curves, as studied in [Conrad 2007]:

Theorem 1.2. *Over the base $S = \mathrm{Spec}(\mathbb{Z})$, there is a canonical isomorphism of algebraic stacks $\mathcal{X}_1^{\mathrm{tw}}(N) \cong \mathcal{X}_1(N)$ extending the identity map on $\mathcal{Y}_1(N)$, and a canonical isomorphism of algebraic stacks $\mathcal{X}^{\mathrm{tw}}(N) \cong \mathcal{X}(N)$ extending the identity map on $\mathcal{Y}(N)$.*

We will prove this as Theorem 6.1. We first verify that $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are finite over $\overline{\mathcal{M}}_{1,1}$. Some commutative algebra then tells us they are Cohen–Macaulay over $\mathrm{Spec}(\mathbb{Z})$, at which point we may proceed as in [Conrad 2007, §4.1] to identify these stacks with the normalizations of $\overline{\mathcal{M}}_{1,1}$ in $\mathcal{X}_1(N)|_{\mathbb{Z}[1/N]}$ and $\mathcal{X}(N)|_{\mathbb{Z}[1/N]}$. These isomorphisms have a natural moduli interpretation, as discussed in Corollaries 6.9 and 6.10.

The techniques of this paper also yield new moduli interpretations of various moduli stacks of elliptic curves that are not (apparently) contained in a stack of twisted stable maps. These results may be known to experts, but are not recorded in the literature; for completeness we include a careful proof of the modular interpretation of the closure of $\mathcal{Y}_1(N)$ in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$. We also study how the different moduli stacks of elliptic curves contained in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$ and $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$ interact in characteristics dividing N; this easily generalizes an example in [Abramovich et al. 2011] but otherwise does not appear in the literature.

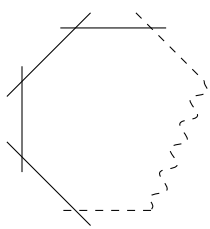
In the Appendix we recall an example of [Chai and Norman 1990] which implies that the techniques of this paper do not generalize nicely to curves of higher genus: using the Katz–Mazur notion of a "full set of sections" we can define a stack over $\mathrm{Spec}(\mathbb{Z})$ classifying genus-g curves with full level-N structures on their Jacobians, but this stack is not even flat over $\mathrm{Spec}(\mathbb{Z})$. The corresponding stack over $\mathbb{Z}[1/N]$ is well-behaved, and arises as an open substack of $\overline{\mathcal{K}}_{g,0}(\mathcal{B}\mu_N^{2g})$, but no moduli interpretation for its closure in characteristics dividing N appears to be known.

2. Review of moduli of generalized elliptic curves

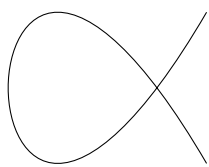
Drinfeld level structures on generalized elliptic curves. For the convenience of the reader, we recall the definitions and results from the theory of generalized elliptic curves that we require in this paper.

Definition 2.1. A *Deligne–Rapoport* (DR) *semistable curve of genus* 1 over a scheme S is a proper, flat, separated, finitely presented morphism of schemes $f : C \to S$, all of whose geometric fibers are nonempty, connected semistable curves with trivial dualizing sheaves.

By [Deligne and Rapoport 1973, §II.1], an equivalent definition for $f : C \to S$ to be a DR semistable curve of genus 1 is that f is a proper flat morphism of finite presentation and relative dimension 1, such that every geometric fiber is either a smooth connected genus-1 curve or a Néron polygon. Recall [loc. cit.] that over a base scheme S, the *standard Néron N-gon* C_N/S (for any $N \geq 1$) is obtained from $\widetilde{C}_N := \mathbb{P}^1_S \times \mathbb{Z}/(N)$ by "gluing" the section 0 in the i-th copy of $\mathbb{P}^1_S$ to the section ∞ in the $(i+1)$-th copy of $\mathbb{P}^1_S$:



Standard Néron N-gon



Standard Néron 1-gon

The natural multiplication action of $\mathbb{G}_m$ on $\mathbb{P}^1_S$, together with the action of $\mathbb{Z}/(N)$ on itself via the group law, determines an action of the group scheme $\mathbb{G}_m \times \mathbb{Z}/(N)$ on $\mathbb{P}^1_S \times \mathbb{Z}/(N)$, descending uniquely to an action of $\mathbb{G}_m \times \mathbb{Z}/(N) = C_N^{\mathrm{sm}}$ on C_N [Deligne and Rapoport 1973, II.1.9].

Definition 2.2. A *generalized elliptic curve* over a scheme S is a DR semistable curve E/S of genus 1, equipped with a morphism $E^{\mathrm{sm}} \times E \to E$ and a section $0_E \in E^{\mathrm{sm}}(S)$ such that the restriction $E^{\mathrm{sm}} \times E^{\mathrm{sm}} \to E^{\mathrm{sm}}$ makes E^{sm} a commutative group scheme over S with identity 0_E, and such that on any singular geometric fiber $E_{\bar{s}}$, the translation action by a rational point on $E^{\mathrm{sm}}_{\bar{s}}$ acts by a rotation on the graph $\Gamma(E_{\bar{s}})$ [Deligne and Rapoport 1973, I.3.5] of the irreducible components of $E_{\bar{s}}$.

By [Deligne and Rapoport 1973, II.1.15], over an algebraically closed field a generalized elliptic curve is either a smooth elliptic curve or a Néron N-gon (for some $N \geq 1$) with the action described above. In fact this result says more: for any generalized elliptic curve E/S, there is a locally finite family $(S_N)_{N \geq 1}$ of closed

subschemes of S such that $\bigcup S_N$ is the nonsmooth locus in S of $E \to S$, and $E \times_S S_N$ is fppf locally on S_N isomorphic to the standard Néron N-gon over S_N.

Recall that for an S-scheme X, a *relative effective Cartier divisor* in X over S is an effective Cartier divisor in X which is flat over S.

Definition 2.3. Let E/S be a generalized elliptic curve. A $[\Gamma_1(N)]$-*structure* on E is a section $P \in E^{\mathrm{sm}}(S)$ such that:

- $N \cdot P = 0_E$.
- The relative effective Cartier divisor
$$D := \sum_{a \in \mathbb{Z}/(N)} [a \cdot P]$$
in E^{sm} is a subgroup scheme.
- For any geometric point $\bar{p} \to S$, $D_{\bar{p}}$ meets every irreducible component of $E_{\bar{p}}$.

We write $\mathcal{X}_1(N)$ for the stack over $\mathrm{Spec}(\mathbb{Z})$ associating to a scheme S the groupoid of pairs (E, P), where E/S is a generalized elliptic curve and P is a $[\Gamma_1(N)]$-structure on E. We write $\mathcal{Y}_1(N)$ for the substack classifying such pairs, where E/S is a smooth elliptic curve.

Definition 2.4. Let E/S be a generalized elliptic curve. A $[\Gamma(N)]$-*structure* on E is an ordered pair (P, Q) of sections in $E^{\mathrm{sm}}[N](S)$ such that:

- The relative effective Cartier divisor
$$D := \sum_{a,b \in \mathbb{Z}/(N)} [a \cdot P + b \cdot Q]$$
in E^{sm} is an N-torsion subgroup scheme, hence $D = E^{\mathrm{sm}}[N]$.
- For any geometric point $\bar{p} \to S$, $D_{\bar{p}}$ meets every irreducible component of $E_{\bar{p}}$.

We write $\mathcal{X}(N)$ for the stack over $\mathrm{Spec}(\mathbb{Z})$ associating to a scheme S the groupoid of tuples $(E, (P, Q))$, where E/S is a generalized elliptic curve and (P, Q) is a $[\Gamma(N)]$-structure on E. We write $\mathcal{Y}(N)$ for the substack classifying such tuples, where E/S is a smooth elliptic curve.

Definition 2.5. Let E/S be a generalized elliptic curve, and let G be a 2-generated finite abelian group, say $G \cong \mathbb{Z}/(n_1) \times \mathbb{Z}/(n_2)$, with $n_2 | n_1$. A G-*structure* on E is a homomorphism $\phi : G \to E^{\mathrm{sm}}$ of group schemes over S such that:

- The relative effective Cartier divisor
$$D := \sum_{a \in G} [\phi(a)]$$
in E^{sm} is an n_1-torsion subgroup scheme.

- For every geometric point $\bar{p} \to S$, $D_{\bar{p}}$ meets every irreducible component of $E_{\bar{p}}$.

Theorem 2.6 [Conrad 2007, 3.1.7, 3.2.7, 3.3.1, 4.1.1]. *$\mathcal{X}_1(N)$ and $\mathcal{X}(N)$ are regular Deligne–Mumford stacks, proper and flat over* $\mathrm{Spec}(\mathbb{Z})$ *of pure relative dimension* 1.

In particular, it follows (see [Conrad 2007, 4.1.5]) that $\mathcal{X}_1(N)$ (resp. $\mathcal{X}(N)$) is canonically identified with the normalization of $\overline{\mathcal{M}}_{1,1}$ in the normal Deligne–Mumford stack $\mathcal{X}_1(N)|_{\mathbb{Z}[1/N]}$ (resp. $\mathcal{X}(N)|_{\mathbb{Z}[1/N]}$), as in [Deligne and Rapoport 1973, §VII.2; Katz and Mazur 1985, §8.6].

Reductions* mod *p of the moduli stacks. It will be useful for us to have a description of the "reduction mod p" of the stacks $\mathcal{X}_1(N)$ and $\mathcal{X}(N)$ for primes p dividing N. These reductions are described using Katz and Mazur's "crossings theorem", which we now recall.

We work over a fixed field k. Let Y/k be a smooth curve, and let $X \to Y$ be finite and flat. Suppose there is a nonempty finite set of k-rational points of Y (which we will call the *supersingular points* of Y) such that for each supersingular point y_0, X_{y_0} is a single k-rational point, and $\widehat{\mathcal{O}}_{X,x_0} \cong k[\![x, y]\!]/(f)$ for some f (depending on y_0).

Also suppose we have a finite collection of closed immersions $\{Z_i \hookrightarrow X\}_{i=1}^n$, where each Z_i is finite and flat over Y, with Z_i^{red} a smooth curve over k, such that the morphism $\sqcup Z_i \to X$ is an isomorphism over the nonsupersingular locus of Y, and such that for each i and each supersingular point $y_0 \in Y$, $(Z_i)_{y_0}$ is a single k-rational point.

Theorem 2.7 (crossings theorem, [Katz and Mazur 1985, 13.1.3]). *Under the above assumptions, let $y_0 \in Y$ be a supersingular point and $x_0 = X_{y_0}$. Then*

$$\widehat{\mathcal{O}}_{X,x_0} \cong k[\![x, y]\!] \Big/ \prod_{i=1}^m f_i^{e_i},$$

where each f_i is irreducible in $k[\![x, y]\!]$, each f_i and f_j $(i \neq j)$ are distinct in $k[\![x, y]\!]$ modulo multiplication by units, and for $z_{i,0} := (Z_i)_{y_0}$ we have

$$\widehat{\mathcal{O}}_{Z_i, z_{i,0}} \cong k[\![x, y]\!]/(f_i^{e_i}).$$

If Y is (geometrically) connected then each Z_i is (geometrically) connected, in which case $\{Z_i\}_{i=1}^n$ is the set of irreducible components of X.

Definition 2.8. Under the above assumptions, we say that X is the *disjoint union with crossings at the supersingular points* of the closed subschemes $\{Z_i\}_{i=1}^n$.

If $\mathcal{X} \to \mathcal{Y}$ is a finite (hence representable) flat morphism of algebraic stacks over k, with $\mathcal{Y}$ Deligne–Mumford (hence $\mathcal{X}$ also Deligne–Mumford), and $\{\mathcal{Z}_i \hookrightarrow \mathcal{X}\}_{i=1}^n$

is a finite collection of closed immersions of algebraic stacks, such that for some étale surjection $Y' \to \mathcal{Y}$ with Y' a scheme, the algebraic spaces Y', $\mathcal{X} \times_{\mathcal{Y}} Y'$, and $\{\mathcal{Z}_i \times_{\mathcal{Y}} Y'\}$ are schemes satisfying the assumptions of the crossings theorem, we say that $\mathcal{X}$ is the *disjoint union with crossings at the supersingular points* of the closed substacks $\{\mathcal{Z}_i\}$.

Remark 2.9. We will solely be applying the above theorem in the case $\mathcal{Y} = \overline{\mathcal{M}}_{1,1}$, with $\mathcal{X}$ an algebraic stack known to be quasifinite and proper over $\overline{\mathcal{M}}_{1,1}$. $\overline{\mathcal{M}}_{1,1}$ is a Deligne–Mumford stack with separated diagonal, so once we know that $\mathcal{X} \to \overline{\mathcal{M}}_{1,1}$ is representable, it follows from [Knutson 1971, II.6.16] that the algebraic spaces $\mathcal{X} \times_{\overline{\mathcal{M}}_{1,1}} Y'$ and $\{\mathcal{Z}_i \times_{\overline{\mathcal{M}}_{1,1}} Y'\}$ are schemes.

Definition 2.10. Let $S \in \mathrm{Sch}/\mathbb{F}_p$. Let E/S be an elliptic curve, with relative Frobenius $F : E \to E^{(p)}$. Let $G \subset E$ be a finite, locally free S-subgroup scheme of rank p^n, $n \geq 1$. For integers $a, b \geq 0$ with $a + b = n$, we say that G is an (a, b)-*subgroup* if $\ker(F^a) \subset G$, and if in the resulting factorization of $E \to E/G$ as

$$E \xrightarrow{F^a} E^{(p^a)} \xrightarrow{\pi} E/G$$

we have $\ker(\widehat{\pi}) = \ker(F^b_{E/G})$ (where $\widehat{\pi}$ denotes the dual isogeny and $F_{E/G} : E/G \to (E/G)^{(p)}$ is the relative Frobenius). In particular note that $E^{(p^a)} \cong (E/G)^{(p^b)}$.

We say that G is an (a, b)-*cyclic subgroup* if it is an (a, b)-subgroup, and either $a = 0$, $b = 0$, or there exists a closed subscheme $Z \subset S$ defined by a sheaf of ideals $\mathcal{I} \subset \mathcal{O}_S$, with $\mathcal{I}^{p-1} = 0$, such that the isomorphism $E^{(p^a)}|_Z \cong (E/G)^{(p^b)}|_Z$ is induced by an isomorphism $E^{(p^{a-1})}|_Z \cong (E/G)^{(p^{b-1})}|_Z$.

Finally, a $[\Gamma_1(p^n)]$-(a, b)-*cyclic structure* on E is a $[\Gamma_1(p^n)]$-structure $P \in E[p^n](S)$ such that the S-subgroup scheme

$$D := \sum_{m \in \mathbb{Z}/(p^n)} [m \cdot P]$$

in E is an (a, b)-cyclic subgroup of E. We write $\mathcal{Y}_1(p^n)^{(a,b)}_S$ for the substack of $\mathcal{Y}_1(p^n)_S$ associating to a scheme T/S the groupoid of pairs (E, P), where E/T is an elliptic curve and P is a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure on E.

If E is an ordinary elliptic curve over a scheme S of characteristic p, it is considerably easier to describe what is meant by a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure on E. Namely, it is just a $[\Gamma_1(p^n)]$-structure P on E such that the relative effective Cartier divisor

$$D_b := \sum_{m=1}^{p^b} [m \cdot P]$$

is a subgroup scheme of E which is étale over S. Note that for an arbitrary $[\Gamma_1(p^n)]$-structure P on E, D_b will not generally be a subgroup scheme of E if

$b < n$, and even if it is a subgroup scheme it might not be étale over S. Being a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure on E (for E ordinary) means that for any geometric fiber $E_{\bar{s}}$ (so $E_{\bar{s}}[p^n] \cong \mu_{p^n} \times \mathbb{Z}/(p^n)$ over $k(\bar{s})$) $P_{\bar{s}}$ has exact order p^b as an element of the group $E_{\bar{s}}[p^n](k(\bar{s})) \cong \mathbb{Z}/(p^n)$.

The following elementary result will be required when we study the interactions in characteristic p of different moduli stacks of elliptic curves contained in a moduli stack of twisted stable maps. We include a proof for lack of a suitable reference.

Lemma 2.11. *Let $E/S/\mathbb{F}_p$ be an elliptic curve, and let P be a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure on E. Then P is a $[\Gamma_1(p^{n+1})]$-structure on E and is $[\Gamma_1(p^{n+1})]$-$(a+1, b)$-cyclic.*

Proof. Consider the relative effective Cartier divisor

$$G := \sum_{m \in \mathbb{Z}/(p^n)} [m \cdot P]$$

in E. This is an S-subgroup scheme containing $\ker(F^a)$, and in the resulting factorization of the quotient map $E \to E/G$ as

$$E \xrightarrow{F^a} E^{(p^a)} \xrightarrow{\pi} E/G$$

we have $\ker(\widehat{\pi}) = \ker(F^b_{E/G})$. Consider also the relative effective Cartier divisor

$$G' := \sum_{m \in \mathbb{Z}/(p^{n+1})} [m \cdot P]$$

in E. G is a subgroup of G', and the image of G' in E/G is the relative effective Cartier divisor

$$\sum_{m \in \mathbb{Z}/(p)} [m \cdot 0_{E/G}].$$

This is simply the kernel of $F_{E/G} : E/G \to (E/G)^{(p)}$, so we may identify E/G' with $(E/G)^{(p)}$ and the quotient map $E/G \to E/G'$ with the relative Frobenius $F_{E/G} : E/G \to (E/G)^{(p)}$. In particular, the quotient map $E \to E/G'$ is a cyclic p^{n+1}-isogeny of elliptic curves with kernel generated by P, so we may already conclude that P is a $[\Gamma_1(p^{n+1})]$-structure on E.

Now, since $G' = p \cdot G$ as Cartier divisors, if $\ker(F^a) \subset G$ then $\ker(F^{a+1}) \subset G'$. Factor the quotient map $E \to E/G' = (E/G)^{(p)}$ as

$$E \xrightarrow{F^{a+1}} E^{(p^{a+1})} \xrightarrow{\pi'} (E/G)^{(p)}.$$

We have a diagram

$$\begin{tikzcd} E \arrow[r, "F^a"] \arrow[dr, "F^{a+1}"'] & E^{(p^a)} \arrow[r, "\pi"] \arrow[d, "F"] & E/G \arrow[r, "F_{E/G}"] & (E/G)^{(p)}, \\ & E^{(p^{a+1})} \arrow[urr, "\pi'"'] & & \end{tikzcd}$$

where the composite of the top arrows $E \to (E/G)^{(p)}$ is the natural quotient map $E \to E/G'$. The outer arrows and the left-hand triangle commute, hence the right-hand triangle commutes as well. Now consider the following diagram:

$$\begin{tikzcd} E^{(p^a)} \arrow[r, "\pi"] \arrow[d, "F"'] & E/G \arrow[r, "\widehat{\pi}"] \arrow[d] & E^{(p^a)} \arrow[d, "F"] \\ E^{(p^{a+1})} \arrow[r, "\pi'"'] & (E/G)^{(p)} \arrow[r, "\widehat{\pi}'"'] & E^{(p^{a+1})}. \end{tikzcd}$$

The outer rectangle commutes since the horizontal composites are the isogenies $[\deg(\pi)]$ and $[\deg(\pi')]$ (and $\deg(\pi) = \deg(\pi')$), and we have shown that the left-hand square commutes, so the right-hand square commutes as well. Since $\ker(\widehat{\pi}) = \ker(F^b_{E/G})$, we conclude that $\ker(\widehat{\pi}') = \ker(F^b_{(E/G)^{(p)}})$.

Finally, suppose $a, b \geq 1$. Since $E^{(p^a)} \cong (E/G)^{(p^b)}$ we also have

$$E^{(p^{a+1})} \cong (E/G)^{(p^{b+1})} = ((E/G)^{(p)})^{(p^b)},$$

and this is already induced by the isomorphism

$$E^{(p^a)} \cong (E/G)^{(p^b)} = ((E/G)^{(p)})^{(p^{b-1})}.$$

Therefore P is a $[\Gamma_1(p^{n+1})]$-$(a+1, b)$-cyclic structure on E. □

Remark 2.12. The reason the above argument fails when p is invertible on the base scheme S is that, preserving the notation of the above proof, the image of the relative Cartier divisor G' in E/G is not a subgroup of E/G unless S is an $\mathbb{F}_p$-scheme. Indeed, the image of G' in E/G is the relative effective Cartier divisor $p \cdot [0_{E/G}]$, which is finite flat of rank p over S. If p is invertible on S, any finite flat commutative group scheme of rank p over S is étale (see [Shatz 1986, Corollary 4.3]), but $p \cdot [0_{E/G}]$ is obviously not étale.

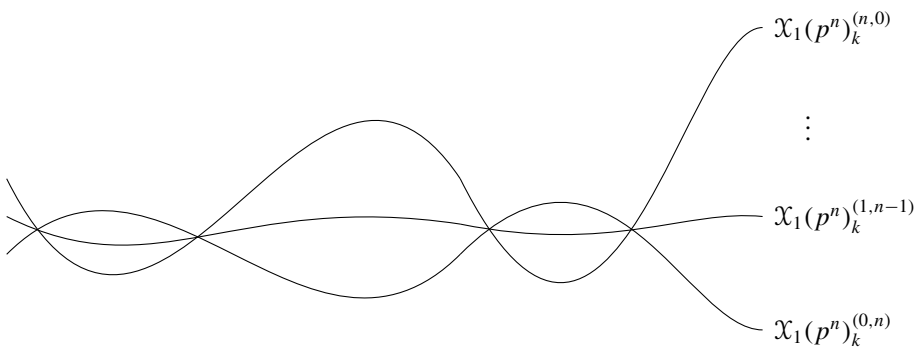
It is straightforward to extend the definition of a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure, and the result of the above lemma, to the case of generalized elliptic curves. Let $E/S/\mathbb{F}_p$ be a generalized elliptic curve; let $S_1 \subset S$ be the open locus where the morphism $E \to S$ is smooth, and let $S_2 \subset S$ be the complement of the closed locus in S where the geometric fibers are supersingular elliptic curves. We say that a $[\Gamma_1(p^n)]$-structure P on E/S is a $[\Gamma_1(p^n)]$-(a, b)-*cyclic structure* if:

- $P|_{S_1}$ is a $[\Gamma_1(p^n)]$-(a,b)-cyclic structure on the elliptic curve E_{S_1}/S_1 and
- the relative effective Cartier divisor
$$D_b := \sum_{m=1}^{p^b} [m \cdot P|_{S_2}]$$
in $E^{\mathrm{sm}}|_{S_2}$ is a subgroup scheme of $E^{\mathrm{sm}}|_{S_2}$ which is étale over S_2.

We write $\mathcal{X}_1(p^n)_S^{(a,b)}$ for the substack of $\mathcal{X}_1(p^n)_S$ associating to a scheme T/S the groupoid of pairs (E, P), where E/T is a generalized elliptic curve and P is a $[\Gamma_1(p^n)]$-(a,b)-cyclic structure on E.

Theorem 2.13 [Katz and Mazur 1985, 13.5.4]. *Let k be a perfect field of characteristic p. $\mathcal{Y}_1(p^n)_k$ (resp. $\mathcal{X}_1(p^n)_k$) is the disjoint union, with crossings at the supersingular points, of the $n+1$ substacks $\mathcal{Y}_1(p^n)_k^{(a,n-a)}$ (resp. $\mathcal{X}_1(p^n)_k^{(a,n-a)}$) for $0 \le a \le n$.*

It can be helpful to visualize $\mathcal{X}_1(p^n)_k$ as follows:



$\mathcal{X}_1(p^n)$ over a perfect field k of characteristic p

Definition 2.14. Let $S \in \mathrm{Sch}/\mathbb{F}_p$, and let $H \le (\mathbb{Z}/(p^n))^2$ such that $H \cong \mathbb{Z}/(p^n)$ (hence also $(\mathbb{Z}/(p^n))^2/H \cong \mathbb{Z}/(p^n)$). Let E/S be a generalized elliptic curve. Let (P, Q) be a $[\Gamma(p^n)]$-structure on E, corresponding to a group homomorphism $\phi : (\mathbb{Z}/(p^n))^2 \to E[p^n]$, $(1,0) \mapsto P$, $(0,1) \mapsto Q$. We say that (P, Q) has *component label* H if

- $\phi(H) \subseteq \ker(F^n)$, where $F : E \to E^{(p)}$ is the relative Frobenius, and
- the resulting group scheme homomorphism $\mathbb{Z}/(p^n) \cong (\mathbb{Z}/(p^n))^2/H \to E^{(p^n)} = E/\ker(F^n)$ is a $[\Gamma_1(p^n)]$-structure on $E^{(p^n)}$. (This is independent of the choice of isomorphism $(\mathbb{Z}/(p^n))^2/H \cong \mathbb{Z}/(p^n)$, although the $[\Gamma_1(p^n)]$-structure obtained depends on this choice.)

We define $\mathcal{Y}(p^n)_S^H$ (resp. $\mathcal{X}(p^n)_S^H$) to be the substack of $\mathcal{Y}(p^n)_S$ (resp. $\mathcal{X}(p^n)_S$) associating to a scheme T/S the groupoid of tuples $(E, (P, Q))$, where E/T is an

elliptic curve (resp. generalized elliptic curve) and (P, Q) is a $[\Gamma(p^n)]$-structure on E of component label H.

Theorem 2.15 [Katz and Mazur 1985, 13.7.6]. *Let k be a perfect field of characteristic p. $\mathcal{Y}(p^n)_k$ (resp. $\mathcal{X}(p^n)_k$) is the disjoint union, with crossings at the supersingular points, of the substacks $\mathcal{Y}(p^n)_k^H$ (resp. $\mathcal{X}(p^n)_k^H$) for $H \leq (\mathbb{Z}/(p^n))^2$ with $H \cong \mathbb{Z}/(p^n)$.*

The proof of the above theorem immediately generalizes to a slightly more general setting which we will find useful when studying compactified stacks of $[\Gamma(N)]$-structures below. Let $K \leq (\mathbb{Z}/(p^n))^2$, and write $G_K = (\mathbb{Z}/(p^n))^2/K$. Then there exist integers $m \geq l \geq 0$ with $G_K \cong \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$. Suppose that $l \geq 1$, so $G_K \cong \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$ with $m \geq l \geq 1$. Given a G_K-structure ϕ on an ordinary elliptic curve $E/T/\mathbb{F}_p$ (in the sense of Definition 2.5), étale locally on T we can consider the composite

$$G_K = \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l) \xrightarrow{\phi} E[p^n] \cong \mathbb{Z}/(p^n) \times \mu_{p^n} \xrightarrow{\pi_1} \mathbb{Z}/(p^n).$$

Since ϕ is a G_K-structure, the kernel and image of this composite are necessarily cyclic. The same argument used in [Katz and Mazur 1985, 13.7.6] to prove the above theorem shows that in characteristic p, $\mathcal{Y}_K$ breaks up into a union of substacks indexed in this way by the possible kernels of group homomorphisms $G_K \to \mathbb{Z}/(p^n)$, subject to the condition on the image just described. So we can describe $\mathcal{Y}_K$ as a union of closed substacks indexed by the set

$$L_K := \{H \leq G_K \mid H \text{ and } G/H \text{ are both cyclic}\}.$$

Since $G_K \cong \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$ with $m \geq l \geq 1$, we have that $H \in L_K$ if and only if $H \cong \mathbb{Z}/(p^m)$ or $H \cong \mathbb{Z}/(p^l)$.

The rigorous definition, accounting for elliptic curves that might not be ordinary and for the case where G_K is cyclic, is as follows:

Definition 2.16. Let $K \leq (\mathbb{Z}/(N))^2$ with corresponding quotient $G_K = (\mathbb{Z}/(N))^2/K$. We write $\mathcal{Y}_K$ (resp. $\mathcal{X}_K$) for the algebraic stack over $\mathrm{Spec}(\mathbb{Z})$ associating to a scheme S the groupoid of pairs (E, ϕ), where E/S is an elliptic curve (resp. generalized elliptic curve) and ϕ is a G_K-structure on E.

If S is an $\mathbb{F}_p$-scheme, $N = p^n$ and $G_K \cong \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$ with $m \geq l \geq 1$, then for any $H \in L_K$, we define $\mathcal{Y}_{K,S}^H \subset \mathcal{Y}_{K,S}$ (resp. $\mathcal{X}_{K,S}^H \subset \mathcal{X}_{K,S}$) to be the substack associating to a scheme T/S the groupoid of pairs (E, ϕ), where E/T is an elliptic curve (resp. generalized elliptic curve) and $\phi : G_K \to E[p^m]$ is a G_K-structure such that:

- $\phi(H) \subset \ker(F^m)$, where $F^m : E \to E^{(p^m)}$ is the m-fold relative Frobenius on E.
- The resulting group scheme homomorphism $G_K/H \to E/\ker(F^m) \cong E^{(p^m)}$ is a G_K/H-structure on $E^{(p^m)}$ in the sense of [Katz and Mazur 1985, §1.5].

In this case we say that the G_K-structure ϕ has *component label* H.

If $G_K \cong \mathbb{Z}/(p^m)$ (that is, $l = 0$), the stack $\mathcal{Y}_K$ (resp. $\mathcal{X}_K$) is isomorphic to $\mathcal{Y}_1(p^m)$ (resp. $\mathcal{X}_1(p^m)$), and for $H \cong \mathbb{Z}/(p^a) \in L_K$ and $S \in \mathrm{Sch}/\mathbb{F}_p$ we define $\mathcal{Y}^H_{K,S}$ (resp. $\mathcal{X}^H_{K,S}$) to be the substack $\mathcal{Y}_1(p^m)^{(a,m-a)}_S \subset \mathcal{Y}_1(p^m)_S$ (resp. $\mathcal{X}_1(p^m)^{(a,m-a)}_S \subset \mathcal{X}_1(p^m)_S$), as in Definition 2.10. We still say $\mathcal{Y}^H_{K,S}$ and $\mathcal{X}^H_{K,S}$ classify G_K-structures of *component label* H.

The result is:

Theorem 2.17. *Let k be a perfect field of characteristic p. $\mathcal{Y}_{K,k}$ (resp. $\mathcal{X}_{K,k}$) is the disjoint union, with crossings at the supersingular points, of the closed substacks $\mathcal{Y}^H_{K,k}$ (resp. $\mathcal{X}^H_{K,k}$) for $H \in L_K$.*

We also note that an analogue of Lemma 2.11 holds in this case, which we record for future use. It is clear that if $K' \leq K \leq (\mathbb{Z}/(p^n))^2$, giving a canonical quotient map $\pi : G_{K'} \to G_K$, and if $\phi : G_K \to E^{\mathrm{sm}}$ is a G_K-structure on a generalized elliptic curve $E/S/\mathbb{F}_p$, then a necessary condition for the composite $\phi \circ \pi : G_{K'} \to E^{\mathrm{sm}}$ to be a $G_{K'}$-structure is that $\ker(\pi)$ is cyclic. In fact, unwinding the definitions we immediately deduce:

Lemma 2.18. *Let $E/S/\mathbb{F}_p$ be a generalized elliptic curve with no supersingular fibers. Let $K' \leq K \leq (\mathbb{Z}/(p^n))^2$ such that the canonical quotient $\pi : G_{K'} \to G_K$ has cyclic kernel. Let $H \in L_K$ and suppose that $\phi : G_K \to E$ is a G_K-structure with component label H. Then $\phi \circ \pi$ is a $G_{K'}$ structure on E if and only if $\pi^{-1}(H) \subseteq G_{K'}$ is cyclic, in which case $\phi \circ \pi$ has component label $\pi^{-1}(H) \in L_{K'}$.*

3. Generalities/review of twisted stable maps

Twisted stable maps. We will be studying moduli stacks of elliptic curves embedded in moduli stacks of twisted stable maps to tame stacks. We now recall the relevant definitions and results relating to twisted stable maps.

Definition 3.1 [Abramovich et al. 2008b, 2.2]. Let G be a group scheme over a scheme S. Write $\mathrm{QCoh}^G(S)$ for the category of G-equivariant quasicoherent sheaves on S; writing $\mathcal{B}G$ for the classifying stack of G over S, this is equivalent to $\mathrm{QCoh}(\mathcal{B}G)$ [Abramovich et al. 2008b, §2.1]. We say that G is *linearly reductive* if the functor $\mathrm{QCoh}^G(S) \to \mathrm{QCoh}(S)$, $\mathcal{F} \mapsto \mathcal{F}^G$ is exact, or equivalently if the pushforward $\mathrm{QCoh}(\mathcal{B}G) \to \mathrm{QCoh}(S)$ is exact.

Linearly reductive group schemes are classified in [Abramovich et al. 2008b, §2.3]. The examples in which we are most interested for this paper are the finite flat commutative linearly reductive group schemes μ_N and μ_N^2 over S.

Definition 3.2 [Abramovich et al. 2008b, 3.1]. Let $\mathcal{X}$ be a locally finitely presented algebraic stack over a scheme S, with finite inertia. By [Keel and Mori 1997], $\mathcal{X}$

has a coarse moduli space $\rho : \mathscr{X} \to X$, with ρ proper. We say that $\mathscr{X}/S$ is *tame* if $\rho_* : \mathrm{QCoh}(\mathscr{X}) \to \mathrm{QCoh}(X)$ is exact.

As observed in [Abramovich et al. 2008b, §3], for any finite flat group scheme G over a scheme S, the classifying stack $\mathscr{B}G$ over S is tame if and only if G is linearly reductive. So in particular the classifying stacks $\mathscr{B}\mu_N$ and $\mathscr{B}\mu_N^2$ are always tame.

Definition 3.3 [Abramovich et al. 2011, §2]. An *n-marked twisted curve* over a scheme S is a proper tame stack $\mathscr{C}$ over S, with connected dimension-1 geometric fibers, and coarse space $f : C \to S$ a nodal curve over S; together with n closed substacks

$$\{\Sigma_i \subset \mathscr{C}\}_{i=1}^n$$

which are fppf gerbes over S mapping to n markings $\{p_i \in C^{\mathrm{sm}}(S)\}$, such that:

- the preimage in $\mathscr{C}$ of the complement $C' \subset C$ of the markings and singular locus of C/S maps isomorphically onto C',
- if $\bar{p} \to C$ is a geometric point mapping to the image in C of a marking $\Sigma_i \subset \mathscr{C}$, then
$$\mathrm{Spec}(\mathbb{O}_{C,\bar{p}}) \times_C \mathscr{C} \simeq [D^{\mathrm{sh}}/\mu_r]$$
for some $r \geq 1$, where D^{sh} is the strict Henselization at $(\mathfrak{m}_{S,f(\bar{p})}, z)$ of
$$D = \mathrm{Spec}(\mathbb{O}_{S,f(\bar{p})}[z])$$
and $\zeta \in \mu_r$ acts by $z \mapsto \zeta \cdot z$, and
- if $\bar{p} \to C$ is a geometric point mapping to a node of C, then
$$\mathrm{Spec}(\mathbb{O}_{C,\bar{p}}) \times_C \mathscr{C} \simeq [D^{\mathrm{sh}}/\mu_r]$$
for some $r \geq 1$, where D^{sh} is the strict Henselization at $(\mathfrak{m}_{S,f(\bar{p})}, x, y)$ of
$$D = \mathrm{Spec}(\mathbb{O}_{S,f(\bar{p})}[x, y]/(xy - t))$$
for some $t \in \mathfrak{m}_{S,f(\bar{p})}$, and $\zeta \in \mu_r$ acts by $x \mapsto \zeta \cdot x$ and $y \mapsto \zeta^{-1} \cdot y$.

We say a twisted curve $\mathscr{C}/S$ has *genus* g if the geometric fibers of its coarse space C/S have arithmetic genus g, and we say an n-marked genus-g twisted curve $\mathscr{C}/S$ is *stable* if the genus-g curve C/S with the markings $\{p_i\}$ is an n-marked genus-g stable curve over S.

Example 3.4. Over any base scheme S, consider a Néron 1-gon C/S as in Section 2. We have $C^{\mathrm{sm}} \cong \mathbb{G}_m$, and C admits the structure of a generalized elliptic curve with an action $C^{\mathrm{sm}} \times C \to C$ extending the group scheme structure of $\mathbb{G}_m$. For any positive integer N, the inclusion $\mu_N \subset \mathbb{G}_m$ determines an action of μ_N on C, and the stack quotient $\mathscr{C} := [C/\mu_N]$ is a twisted stable curve over S, with coarse space

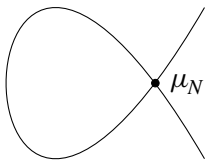
$f : C' \to S$ again a Néron 1-gon. If $\bar{p} \to C'$ is a geometric point mapping to a node of C', then

$$\mathrm{Spec}(\mathcal{O}_{C',\bar{p}}) \times_{C'} \mathscr{C} \simeq [D^{\mathrm{sh}}/\mu_N],$$

where D^{sh} denotes the strict Henselization of

$$D := \mathrm{Spec}(\mathcal{O}_{S,f(\bar{p})}[x, y]/(xy))$$

at the point $(\mathfrak{m}_{S,f(\bar{p})}, x, y)$ and $\zeta \in \mu_N$ acts by $x \mapsto \zeta \cdot x$ and $y \mapsto \zeta^{-1} \cdot y$. We will refer to this twisted curve as the *standard μ_N-stacky Néron 1-gon over S*.



Standard μ_N-stacky Néron 1-gon

Definition 3.5 [Abramovich et al. 2011, §4]. Let $\mathscr{X}$ be a finitely presented algebraic stack, proper over a scheme S and with finite inertia. A *twisted stable map* to $\mathscr{X}$ from an n-marked twisted curve $(\mathscr{C}/S, \{\Sigma_i\})$ over S with coarse space $(C/S, \{p_i\})$ is a morphism $\mathscr{C} \to \mathscr{X}$ of stacks over S such that.

- $\mathscr{C} \to \mathscr{X}$ is a representable morphism.
- The induced map $C \to X$ is a stable map from $(C, \{p_i\})$ to X.

Over any base scheme S, we write $\mathscr{K}_{g,n}(\mathscr{X})$ for the stack over S associating to a scheme T/S the groupoid of pairs $(\mathscr{C}, \alpha)$, where $\mathscr{C}/T$ is an n-marked twisted curve whose coarse space C/T is a genus-g nodal curve, and $\alpha : \mathscr{C} \to \mathscr{X} \times_S T$ is a twisted stable map.

Proposition 3.6 [Abramovich et al. 2011, 4.2]. *For $\mathscr{X}$ as above, $\mathscr{K}_{g,n}(\mathscr{X})$ is a locally finitely presented algebraic stack over S.*

There is a natural morphism of stacks $\mathscr{K}_{g,n}(\mathscr{X}) \to \overline{\mathscr{M}}_{g,n}(X)$ (where $\overline{\mathscr{M}}_{g,n}(X)$ is the usual Kontsevich stack of stable maps into X) defined by passing to coarse spaces. In particular, note that since a twisted stable map is required to be representable, if $\mathscr{X}$ is representable then this map is an equality.

Theorem 3.7 [Abramovich et al. 2011, 4.3]. *Let $\mathscr{X}$ be as above, and also assume that $\mathscr{X}$ is tame. Then $\mathscr{K}_{g,n}(\mathscr{X})$ is proper and quasifinite over $\overline{\mathscr{M}}_{g,n}(X)$.*

A word of caution: the twisted stable maps of [Abramovich et al. 2011] are a generalization of what are referred to as *balanced* twisted stable maps in [Abramovich and Vistoli 2002; Abramovich et al. 2003]; we are only interested in balanced maps for this paper, so hopefully our notation $\mathscr{K}_{g,n}(\mathscr{X})$ will not cause any confusion.

Let $(\mathscr{C}/S, \{\Sigma_i\})$ be an n-marked twisted curve with coarse space C/S. If $\bar{p} \to C$ is a geometric point mapping to the image in C of the gerbe Σ_i, such that $\mathrm{Spec}(\mathbb{O}_{C,\bar{p}}) \times_C \mathscr{C} \simeq [D^{\mathrm{sh}}/\mu_r]$ as in Definition 3.3, then the integer r is uniquely determined by $\bar{p}$, and we call r the *local index of* $\mathscr{C}$ *at* $\bar{p}$. In fact one verifies immediately (see [Abramovich and Vistoli 2002, 8.5.1]) that the local index only depends on i and on the image of $\bar{p}$ in S, and that the function $\epsilon_i : S \to \mathbb{Z}_{>0}$ sending $s = f(\bar{p}) \in S$ to the local index r is locally constant.

Notation 3.8. Let $\boldsymbol{e} = (e_1, \dots, e_n) \in \mathbb{Z}^n_{>0}$. We say an n-marked twisted curve $(\mathscr{C}/S, \{\Sigma_i\})$ has *global index* $\boldsymbol{e}$ if ϵ_i is constant of value e_i for all $i = 1, \dots, n$.

If $\mathscr{X}/S$ is a finitely presented algebraic stack, proper over S and with finite inertia, we write $\mathscr{K}^{\boldsymbol{e}}_{g,n}(\mathscr{X})$ for the substack of $\mathscr{K}_{g,n}(\mathscr{X})$ associating to a scheme T/S the groupoid of pairs $(\mathscr{C}, \alpha)$, where $\mathscr{C}/T$ is an n-marked genus-g twisted curve of global index $\boldsymbol{e}$, and $\alpha : \mathscr{C} \to \mathscr{X} \times_S T$ is a twisted stable map. Since the functions $\epsilon_i : S \to \mathbb{Z}_{>0}$ are locally constant, we see that $\mathscr{K}^{\boldsymbol{e}}_{g,n}(\mathscr{X})$ is an open and closed substack of $\mathscr{K}_{g,n}(\mathscr{X})$, and

$$\mathscr{K}_{g,n}(\mathscr{X}) = \coprod_{\boldsymbol{e} \in \mathbb{Z}^n_{>0}} \mathscr{K}^{\boldsymbol{e}}_{g,n}(\mathscr{X}).$$

In this paper we will generally only be interested in the case $\boldsymbol{e} = (1, \dots, 1)$, in which case $\mathscr{K}^{\boldsymbol{e}}_{g,n}(\mathscr{X})$ classifies twisted stable maps to $\mathscr{X}$ from twisted curves whose markings are honest sections. We will write $\mathscr{K}'_{g,n}(\mathscr{X})$ for $\mathscr{K}^{(1,\dots,1)}_{g,n}(\mathscr{X})$.

Notation 3.9. We write $\mathscr{K}^{\circ}_{g,n}(\mathscr{X})$ for the open substack of $\mathscr{K}'_{g,n}(\mathscr{X})$ associating to T/S the groupoid of pairs $(\mathscr{C}, \alpha)$, where $\mathscr{C}/T$ is a smooth n-marked genus-g curve (with no stacky structure) and $\alpha : \mathscr{C} \to \mathscr{X} \times_S T$ is a twisted stable map.

Twisted covers and Picard schemes of twisted curves. If G/S is a linearly reductive finite flat group scheme, then we have already observed that $\mathscr{B}G$ is tame. So in this case we can consider the algebraic stack $\mathscr{K}_{g,n}(\mathscr{B}G)$, which is proper and quasifinite over the Deligne–Mumford stack $\overline{\mathscr{M}}_{g,n}$ by Theorem 3.7, since the coarse space of $\mathscr{B}G$ is S.

Theorem 3.10 [Abramovich et al. 2011, 5.1]. *Let G/S be a linearly reductive finite flat group scheme. Then $\mathscr{K}_{g,n}(\mathscr{B}G)$ is flat over S, with local complete intersection fibers.*

Definition 3.11. Since a map $\mathscr{C} \to \mathscr{B}G$ is equivalent to a G-torsor $P \to \mathscr{C}$, it is often convenient to view $\mathscr{K}_{g,n}(\mathscr{B}G)$ as classifying such torsors. A G-torsor $P \to \mathscr{C}$ is a *twisted G-cover* of $\mathscr{C}$ if it arises in this way.

Note that since the coarse space of $\mathscr{B}G$ is S, the condition that the map $\mathscr{C} \to \mathscr{B}G$ is stable just says that C/S is stable in the usual sense. So for G as above, $\mathscr{K}_{g,n}(\mathscr{B}G)$

can and will be viewed as the stack classifying twisted G-covers of n-marked genus-g twisted stable curves.

Still writing G for a finite flat linearly reductive group scheme over S, suppose in addition that G is commutative. Then every object of $\mathcal{K}_{g,n}(\mathcal{B}G)$ canonically contains G in the center of its automorphism group scheme (determining a unique subgroup stack H in the center of the inertia stack of $\mathcal{K}_{g,n}(\mathcal{B}G)$ such that for any object $\xi \in \mathcal{K}_{g,n}(\mathcal{B}G)(T)$ the pullback of H to T is G). We can therefore apply the rigidification procedure described in [Abramovich et al. 2003, §5.1] and generalized in [Abramovich et al. 2008b, Appendix A], thereby "removing" G from the automorphism group of each object:

Definition 3.12. $\overline{\mathcal{K}}_{g,n}(\mathcal{B}G)$ is the rigidification of $\mathcal{K}_{g,n}(\mathcal{B}G)$ with respect to the copy of G naturally contained in the center of its inertia stack. We refer to the objects of $\overline{\mathcal{K}}_{g,n}(\mathcal{B}G)$ as *rigidified twisted stable maps* into $\mathcal{B}G$, or *rigidified twisted G-covers* of n-marked genus-g twisted stable curves. We write $\overline{\mathcal{K}}^{\circ}_{g,n}(\mathcal{B}G)$ for the open substack corresponding to smooth curves and $\overline{\mathcal{K}}'_{g,n}(\mathcal{B}G)$ for the open and closed substack corresponding to twisted stable curves whose markings are representable.

In [Abramovich et al. 2008a, §C.2] it is shown that the rigidification of an algebraic stack with respect to a group scheme admits a natural moduli interpretation. In this paper we are only interested in the case where G is diagonalizable, in which case $\overline{\mathcal{K}}_{g,n}(\mathcal{B}G)$ can be given a more concrete moduli interpretation which we will now describe.

Given a twisted curve $\mathscr{C}$ over a scheme S, let $\mathcal{P}ic_{\mathscr{C}/S}$ denote the stack associating to T/S the groupoid of line bundles on $\mathscr{C} \times_S T$.

Proposition 3.13 [Abramovich et al. 2011, 2.7]. *$\mathcal{P}ic_{\mathscr{C}/S}$ is a smooth locally finitely presented algebraic stack over S, and for any $\mathscr{L} \in \mathcal{P}ic_{\mathscr{C}/S}(T)$ the group scheme $\mathrm{Aut}_T(\mathscr{L})$ is canonically isomorphic to $\mathbb{G}_{m,T}$.*

Write $\mathrm{Pic}_{\mathscr{C}/S}$ for the rigidification of this stack with respect to $\mathbb{G}_m$; $\mathrm{Pic}_{\mathscr{C}/S}$ is none other than the relative Picard functor of $\mathscr{C}/S$. From the analysis of $\mathrm{Pic}_{\mathscr{C}/S}$ in [Abramovich et al. 2011, §2] we have:

Proposition 3.14. $\mathrm{Pic}_{\mathscr{C}/S}$ *is a smooth group scheme over S, and if $\pi : \mathscr{C} \to C$ is the coarse space of $\mathscr{C}/S$, then there is a short exact sequence of group schemes over S*

$$0 \to \mathrm{Pic}_{C/S} \xrightarrow{\pi^*} \mathrm{Pic}_{\mathscr{C}/S} \to W \to 0,$$

with W quasifinite and étale over S.

In fact, since $\pi_*\mathbb{G}_m = \mathbb{G}_m$ we can deduce from the exact sequence of low-degree terms of the fppf Leray spectral sequence

$$E_2^{p,q} = H^p(C, \boldsymbol{R}^q\pi_*\mathbb{G}_m) \Rightarrow H^{p+q}(\mathscr{C}, \mathbb{G}_m)$$

that W is the sheaf associated to the presheaf $T \mapsto H^0(C_T, \boldsymbol{R}^1\pi_*\mathbb{G}_m)$ (where we still write $\pi : \mathscr{C}_T \to C_T$ for the morphism induced by base change from $\pi : \mathscr{C} \to C$).

For any integer N annihilating W, we get a natural morphism

$$\times N : \mathrm{Pic}_{\mathscr{C}/S} \to \mathrm{Pic}_{C/S}\,.$$

Definition 3.15 [Abramovich et al. 2011, 2.11]. The *generalized Jacobian* of $\mathscr{C}$ is

$$\mathrm{Pic}^0_{\mathscr{C}/S} := \mathrm{Pic}_{\mathscr{C}/S} \times_{\times N, \mathrm{Pic}_{C/S}} \mathrm{Pic}^0_{C/S},$$

where $\mathrm{Pic}^0_{C/S}$ is the fiberwise connected component of the identity in the group scheme $\mathrm{Pic}_{C/S}$.

$\mathrm{Pic}^0_{\mathscr{C}/S}$ is independent of N and is viewed as classifying line bundles of fiberwise degree 0 on $\mathscr{C}/S$.

Remark 3.16. Unlike the case of $\mathrm{Pic}^0_{C/S}$ for C/S a (nonstacky) genus-g curve, the geometric fibers of $\mathrm{Pic}^0_{\mathscr{C}/S}$ need not be connected for $\mathscr{C}/S$ a twisted curve. In fact, when $\mathscr{C}/S$ is a 1-marked genus-1 twisted stable curve, $\mathrm{Pic}^0_{\mathscr{C}/S}$ behaves like the smooth part of a generalized elliptic curve over S. For instance, if $\mathscr{C} = [C/\mu_N]$ for C/S a Néron 1-gon as in Example 3.4, it is easily verified that $\mathrm{Pic}^0_{\mathscr{C}/S} \cong \mathbb{G}_m \times \mathbb{Z}/(N)$. Standard μ_N-stacky Néron 1-gons will play an analogous role in this paper to that of Néron polygons in [Deligne and Rapoport 1973; Conrad 2007]. In particular, we have:

Lemma 3.17. *Let k be an algebraically closed field, and $\mathscr{C}/k$ a 1-marked genus-1 twisted stable curve, with no stacky structure at its marking, such that the coarse space C/k is not smooth. Then $\mathscr{C}$ is a standard μ_N-stacky Néron 1-gon (Example 3.4) for some (unique) positive integer N.*

Proof. It follows from [Deligne and Rapoport 1973, II.1.15] that the 1-marked genus-1 stable curve C/k is a Néron 1-gon. Write $\pi : \mathscr{C} \to C$ for the map exhibiting C as the coarse space of $\mathscr{C}$. Write $\widetilde{C} = \mathbb{P}^1 \to C$ for the normalization of C, and write $\widetilde{\mathscr{C}}$ for the following fiber product:

$$\begin{tikzcd} \widetilde{\mathscr{C}} \arrow[r, "\tau"] \arrow[d, "\nu"'] & \mathbb{P}^1 \arrow[d] \\ \mathscr{C} \arrow[r, "\pi"'] & C. \end{tikzcd}$$

We have a short exact sequence of fppf sheaves on $\mathbb{P}^1$

$$0 \to \mathcal{O}^\times_{\mathbb{P}^1} \to \tau_*\mathcal{O}^\times_{\widetilde{\mathscr{C}}} \to \mathscr{F} \to 0,$$

where $\mathcal{F}$ is the pushforward to $\mathbb{P}^1$ of $\mathcal{O}_Z^\times$, for $Z \cong \mathcal{B}\mu_N \sqcup \mathcal{B}\mu_N$ the preimage in $\widetilde{\mathscr{C}}$ of the (stacky) node of $\mathscr{C}$. This induces an exact sequence

$$0 \to \mathrm{Pic}(\mathbb{P}^1) \to \mathrm{Pic}(\widetilde{\mathscr{C}}) \to \mathrm{Pic}(Z) \to H^2(\mathbb{P}^1, \mathcal{O}_{\mathbb{P}^1}^\times) = 0,$$

inducing an isomorphism $\mathrm{Pic}^0_{\widetilde{\mathscr{C}}/k} \cong \mathrm{Pic}(Z) \cong \mathbb{Z}/(N) \times \mathbb{Z}/(N)$ since $\mathrm{Pic}^0_{\mathbb{P}^1/k} = 0$.

Now consider the short exact sequence of fppf sheaves on $\mathscr{C}$

$$0 \to \mathcal{O}_{\mathscr{C}}^\times \to \nu_* \mathcal{O}_{\widetilde{\mathscr{C}}}^\times \to \mathcal{G} \to 0,$$

where $\mathcal{G}$ is the pushforward to $\mathscr{C}$ of $\mathcal{O}_p^\times$ for $p \cong \mathcal{B}\mu_N$ the (stacky) node of $\mathscr{C}$. This induces an exact sequence

$$0 \to k^\times \xrightarrow{\mathrm{id}} k^\times \xrightarrow{0} k^\times \to \mathrm{Pic}(\mathscr{C}) \to \mathrm{Pic}(\widetilde{\mathscr{C}}) \to \mathrm{Pic}(\mathcal{B}\mu_N).$$

We have $\mathrm{Pic}(\mathcal{B}\mu_N) \cong \mathbb{Z}/(N)$ and $\mathrm{Pic}^0_{\mathscr{C}/k} \cong \mathbb{Z}/(N) \times \mathbb{Z}/(N)$, and the map

$$\mathrm{Pic}^0_{\mathscr{C}/k} \cong \mathbb{Z}/(N) \times \mathbb{Z}/(N) \to \mathbb{Z}/(N) \cong \mathrm{Pic}(\mathcal{B}\mu_N)$$

is given by $(a, b) \mapsto a - b$. In particular the kernel is isomorphic to $\mathbb{Z}/(N)$, so we have a short exact sequence

$$0 \to k^\times \to \mathrm{Pic}^0_{\mathscr{C}/k} \xrightarrow{f} \mathbb{Z}/(N) \to 0.$$

This sequence splits (noncanonically) since $k^\times$ is divisible (as k is algebraically closed), so $\mathrm{Pic}^0_{\mathscr{C}/k} \cong \mathbb{G}_m \times \mathbb{Z}/(N)$.

The fppf exact sequence of sheaves on $\mathscr{C}$

$$0 \to \mu_N \to \mathbb{G}_m \to \mathbb{G}_m \to 0$$

then gives us an isomorphism

$$\begin{aligned} H^1(\mathscr{C}, \mu_N) &\cong \ker\big(\mathrm{Pic}_{\mathscr{C}/k} \xrightarrow{\times N} \mathrm{Pic}_{\mathscr{C}/k}\big) \\ &\cong \mu_N \times \mathbb{Z}/(N). \end{aligned}$$

Let $C' \to \mathscr{C}$ be the μ_N-torsor over $\mathscr{C}$ corresponding to the class

$$(1, 1) \in \mu_N \times \mathbb{Z}/(N) \cong H^1(\mathscr{C}, \mu_N).$$

In view of [Abramovich and Hassett 2011, 2.3.10], C' is representable, since $(1, 1)$ is the class in $H^1(\mathscr{C}, \mu_N)$ of a uniformizing line bundle on $\mathscr{C}$. Let

$$V = C^{\mathrm{sm}} \cong \mathscr{C} \times_C C^{\mathrm{sm}} \overset{\iota}{\hookrightarrow} \mathscr{C},$$

and consider the resulting μ_N-torsor $C'_V \to V \cong \mathbb{G}_m$. The pullback map

$$\iota^* : H^1(\mathscr{C}, \mu_N) \cong \mu_N \times \mathbb{Z}/(N) \to \mathbb{Z}/(N) \cong H^1(\mathbb{G}_m, \mu_N)$$

is given by $(\zeta, a) \mapsto a$. It therefore follows that $C'_V \cong \mathbb{G}_m$, with the μ_N-action given by the standard multiplication action on $\mathbb{G}_m$. The quotient map $\mathbb{G}_m \to [\mathbb{G}_m/\mu_N] \cong \mathbb{G}_m$ is the group scheme homomorphism $x \mapsto x^N$.

Fix an étale neighborhood W of the node of C of the form

$$W = \operatorname{Spec}(k[z, w]/(zw)),$$

such that

$$\mathscr{C} \times_C W \cong [D/\mu_N]$$

for $D = \operatorname{Spec}(k[x, y]/(xy))$, with $\zeta \in \mu_N$ acting by $x \mapsto \zeta x$ and $y \mapsto \zeta^{-1} y$. The composite

$$D \to [D/\mu_N] \to W$$

is given by the ring homomorphism $z \mapsto x^N$, $w \mapsto y^N$. Since C' is representable and $C' \times_C W \to \mathscr{C} \times_C W$ is a μ_N-torsor, it follows that $C' \times_C W \cong D$ with the above μ_N-action. In particular we see that C' is a nodal curve with one node.

Composing our original $C' \to \mathscr{C}$ with the coarse space map $\mathscr{C} \to C$ gives us a finite morphism of nodal curves $C' \to C$, which restricts to the μ_N-torsor $\mathbb{G}_m \to \mathbb{G}_m \cong C^{\mathrm{sm}}$ and which is totally ramified over the node of C. Riemann–Hurwitz for nodal curves implies that C' has arithmetic genus 1, so C' is a Néron 1-gon with smooth locus $C'^{\mathrm{sm}} = \mathbb{G}_m$. The multiplication action of μ_N on $\mathbb{G}_m$ extends uniquely to an action on C', and by assumption $\mathscr{C} = [C'/\mu_N]$. Thus $\mathscr{C}$ is a standard μ_N-stacky 1-gon. □

Relationship to moduli of elliptic curves with level structure.

Notation 3.18. For any finite flat commutative group scheme G over a base scheme S, $\mathcal{H}(G)$ is the stack over S associating to an S-scheme T the groupoid of pairs (E, ϕ), where E/T is an elliptic curve and $\phi : G^* \to E[N]$ is a homomorphism of group schemes over T (for G^* the Cartier dual of G). For the S-scheme $G = \mu_N$ we write $\mathcal{H}_1(N)$ for $\mathcal{H}(\mu_N)$, and for the S-scheme $G = \mu_N^2$ we write $\mathcal{H}(N)$ for $\mathcal{H}(\mu_N^2)$.

As in [Abramovich et al. 2008b, §2.3], we say a finite commutative group scheme is *diagonalizable* if its Cartier dual is constant, and *locally diagonalizable* if its Cartier dual is étale.

Lemma 3.19 [Abramovich et al. 2011, 5.7]. *Let G/S be a finite diagonalizable commutative group scheme, so $\mathcal{B}G$ is tame (since G is linearly reductive) and $X := G^*$ is constant. For any twisted curve $\mathscr{C}/S$ there is an equivalence of categories between the stack* $\mathrm{TORS}_{\mathscr{C}/S}(G)$ *classifying G-torsors on $\mathscr{C}/S$ and the Picard stack $\mathcal{P}ic_{\mathscr{C}/S}[X]$ of morphisms of Picard stacks $X \to \mathcal{P}ic_{\mathscr{C}/S}$.*

The construction for the above equivalence is as follows: Let $\pi : P \to \mathscr{C}$ be a G-torsor over $\mathscr{C}$. G acts on the sheaf $\pi_*\mathbb{O}_P$, yielding a decomposition

$$\pi_*\mathbb{O}_P = \bigoplus_{\chi \in X} \mathscr{L}_\chi.$$

Each $\mathscr{L}_\chi$ is invertible since P is a torsor over E, so this determines a morphism of Picard stacks

$$\phi_P : X \to \mathcal{P}ic_{\mathscr{C}/S}, \quad \chi \mapsto [\mathscr{L}_\chi].$$

Conversely, such a morphism $\phi : X \to \mathcal{P}ic_{\mathscr{C}/S}$ naturally determines an algebra structure on $\bigoplus_{\chi \in X} \phi(\chi)$, giving a G-torsor

$$\underline{\mathrm{Spec}}_{\mathscr{C}}\left(\bigoplus_{\chi \in X} \phi(\chi)\right) \to \mathscr{C}$$

with the G-action determined by the X-grading.

This defines an open immersion from $\mathscr{K}_{g,n}(\mathscr{B}G)$ into the algebraic stack over S associating to an S-scheme T the groupoid of pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/T$ is an n-marked genus-g twisted stable curve and $\phi \in \mathcal{P}ic_{\mathscr{C}/T}[X]$. Rigidifying $\mathscr{K}_{g,n}(\mathscr{B}G)$ and $\mathcal{P}ic_{\mathscr{C}/T}$ with respect to the group schemes G and $\mathbb{G}_m$, respectively, we have an open immersion from $\overline{\mathscr{K}}_{g,n}(\mathscr{B}G)$ into the stack classifying pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/T$ is an n-marked genus-g twisted stable curve and $\phi : X \to \mathrm{Pic}_{\mathscr{C}/T}$ is a homomorphism of group schemes over T.

Writing $f : \mathscr{C} \to S$ for the structural morphism, we have $\mathrm{Pic}_{\mathscr{C}/S} = \boldsymbol{R}^1 f_* \mathbb{G}_m$. Therefore fppf-locally on T, the morphism ϕ corresponds to the choice of an X-torsor $P \to \mathscr{C}$, with P representable if and only if $(\mathscr{C}, \phi)$ comes from an object of $\overline{\mathscr{K}}_{g,n}(\mathscr{B}G)$ (refer, for example, to [Abramovich and Hassett 2011, 2.3.10] to see that a morphism from a twisted curve $\mathscr{C}$ to $\mathscr{B}G$ is representable if and only if the corresponding G-torsor over $\mathscr{C}$ is representable). This gives us:

Corollary 3.20. *For a finite flat diagonalizable group scheme G/S, the above construction gives an equivalence between $\overline{\mathscr{K}}_{g,n}(\mathscr{B}G)$ and the stack over S associating to T/S the groupoid of pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/T$ is an n-marked genus-g twisted stable curve and $\phi : X \to \mathrm{Pic}_{\mathscr{C}/T}$ is a group scheme homomorphism such that, fppf locally on T, the G-torsor over $\mathscr{C}$ corresponding to ϕ as above is representable.*

In particular, we get an equivalence

$$\overline{\mathscr{K}}^{\circ}_{1,1}(\mathscr{B}G) \simeq \mathscr{H}(G),$$

since for an elliptic curve E/S every G-torsor over E is representable and we canonically have $E \cong \mathrm{Pic}^0_{E/S}$. This isomorphism sends $Q \in E(S)$ to the class of the line bundle $\mathscr{L}((Q) - (0_E))$, so as a special case we see that if $\phi : \mathbb{Z}/(N) \to E$

is a group scheme homomorphism with $\phi(1) = Q$, the corresponding μ_N-torsor over E is of the form

$$P = \underline{\mathrm{Spec}}_E\left(\bigoplus_{a=0}^{N-1} \mathscr{L}((a \cdot Q) - (0_E))\right).$$

Since $\mathcal{H}_1(N)$ naturally contains a closed substack isomorphic to the stack $\mathcal{Y}_1(N)$ over S classifying $[\Gamma_1(N)]$-structures on elliptic curves, we see that the algebraic stack $\overline{\mathcal{H}}_{1,1}(\mathcal{B}\mu_N)$ is a modular compactification of $\mathcal{Y}_1(N)$. Similarly, the algebraic stack $\overline{\mathcal{H}}_{1,1}(\mathcal{B}\mu_N^2)$ is a modular compactification of the stack $\mathcal{Y}(N)$ classifying (not necessarily symplectic) full level-N structures on elliptic curves. The task in both cases is to give a moduli interpretation of the closure of these classical moduli stacks in $\overline{\mathcal{H}}_{1,1}(\mathcal{B}G)$, and we address this in the following sections.

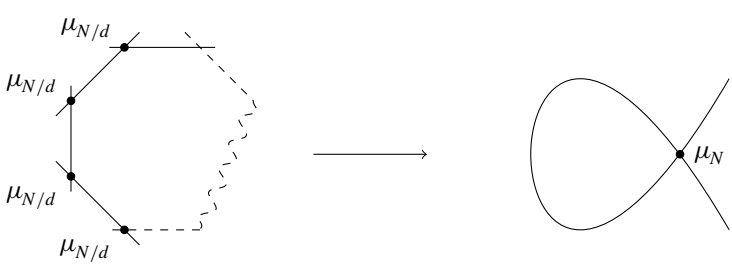
Lemma 3.21. *$\overline{\mathcal{H}}^{\circ}_{1,1}(\mathcal{B}\mu_N)$ is dense in $\overline{\mathcal{H}}'_{1,1}(\mathcal{B}\mu_N)$, and $\overline{\mathcal{H}}^{\circ}_{1,1}(\mathcal{B}\mu_N^2)$ is dense in $\overline{\mathcal{H}}'_{1,1}(\mathcal{B}\mu_N^2)$.*

Proof. Let $\mathscr{C}_0$ be a standard μ_d-stacky 1-gon over an algebraically closed field k, and let $\phi_0 : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_k} \cong \mathbb{G}_m \times \mathbb{Z}/(d)$ be a group scheme homomorphism such that $(\mathscr{C}_0, \phi_0) \in \overline{\mathcal{H}}'_{1,1}(\mathcal{B}\mu_N)(k)$ (so $d \mid N$ and ϕ_0 meets every component of $\mathrm{Pic}^0_{\mathscr{C}_0/k}$). We need to lift $(\mathscr{C}_0/k, \phi_0)$ to a pair $(\mathscr{C}, \phi) \in \overline{\mathcal{H}}'_{1,1}(\mathcal{B}\mu_N)(A)$ for a local ring A with residue field k, such that the generic fiber of $\mathscr{C}$ is a smooth elliptic curve.

Let $\mathscr{C}'_0/k$ be a standard μ_N-stacky 1-gon. Consider the morphism $\mathscr{C}'_0 \to \mathcal{B}\mu_N$ corresponding to the group scheme homomorphism

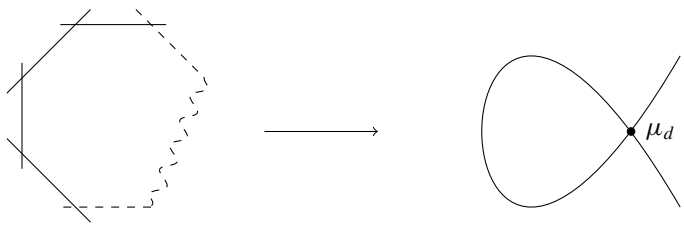
$$\mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}'_0/k} \cong \mathbb{G}_m \times \mathbb{Z}/(N)$$

sending 1 to $(1, N/d)$. This morphism is not representable; the corresponding μ_N-torsor over $\mathscr{C}'_0$ is as follows:



(Stacky N/d-gon) $\mathscr{C}'_0$

We may factor the morphism $\mathscr{C}'_0 \to \mathcal{B}\mu_N$ as $\mathscr{C}'_0 \to \mathscr{C}_0 \to \mathcal{B}\mu_N$, where $\mathscr{C}_0 \to \mathcal{B}\mu_N$ is the relative coarse moduli space [Abramovich et al. 2011, Theorem 3.1] of $\mathscr{C}'_0 \to \mathcal{B}\mu_N$:



(N/d-gon) $\mathscr{C}_0$

Since this $\mathscr{C}_0$ is a standard μ_d-stacky 1-gon, we may identify it with our original twisted curve $\mathscr{C}_0$. This gives us a morphism $\mathscr{C}_0' \to \mathscr{C}_0$, and the resulting pullback map $\mathrm{Pic}^0_{\mathscr{C}_0/k} \to \mathrm{Pic}^0_{\mathscr{C}_0'/k}$ is the monomorphism

$$\mathbb{G}_m \times \mathbb{Z}/(d) \to \mathbb{G}_m \times \mathbb{Z}/(N), \quad (\zeta, a) \mapsto \left(\zeta, \frac{N}{d} \cdot a\right).$$

Let $\mathscr{E}/k[\![q^{1/N}]\!]$ be an N-gon Tate curve, so the special fiber of $\mathscr{E}$ is a Néron N-gon, $\mathscr{E} \otimes k((q^{1/N}))$ is a smooth elliptic curve, and we have an isomorphism $\mathscr{E}^{\mathrm{sm}}[N] \cong \mu_N \times \mathbb{Z}/(N)$ of finite flat commutative group schemes over $k[\![q^{1/N}]\!]$. Let $Q = (1, 1) \in \mathscr{E}^{\mathrm{sm}}[N]$. The relative effective Cartier divisor

$$D := \sum_{a \in \mathbb{Z}/(N)} [a \cdot Q]$$

in $\mathscr{E}^{\mathrm{sm}}$ is a subgroup scheme, étale over $k[\![q^{1/N}]\!]$, and the quotient $\overline{\mathscr{E}} := \mathscr{E}/D$ is naturally a generalized elliptic curve whose special fiber is a 1-gon. The stack quotient $\mathscr{C} := [\mathscr{E}/\mathscr{E}^{\mathrm{sm}}[N]]$ is naturally a twisted curve, whose generic fiber is an elliptic curve and whose special fiber is $\mathscr{C}_0'$. Writing $\pi : \overline{\mathscr{E}} \to \mathscr{C}$ for the natural quotient map, we will see in Section 6 that for any line bundle $\mathscr{L}$ on $\overline{\mathscr{E}}$ there is a canonical decomposition

$$\pi_* \mathscr{L} \cong \bigoplus_{a \in \mathbb{Z}/(N)} \mathscr{L}_a,$$

where each $\mathscr{L}_a$ is a line bundle on $\mathscr{C}$.

For a section $R \in \mathscr{E}^{\mathrm{sm}}(k[\![q^{1/N}]\!])$, we write $\bar{R}$ for its image in $\overline{\mathscr{E}}^{\mathrm{sm}}(k[\![q^{1/N}]\!])$. Then we have the degree-0 line bundle $\mathscr{L}_R := \mathscr{L}((\bar{R}) - (0_{\overline{\mathscr{E}}}))$ on $\overline{\mathscr{E}}$, hence a canonical decomposition

$$\pi_* \mathscr{L}_R \cong \bigoplus_{a \in \mathbb{Z}/(N)} \mathscr{L}_{R,a}.$$

We will see in Section 6 that the map $\mathscr{E}^{\mathrm{sm}}[N] \to \mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}[N]$ sending

$$R = (\zeta, a) \in \mu_N \times \mathbb{Z}/(N) \cong \mathscr{E}^{\mathrm{sm}}[N]$$

to $\mathscr{L}_{R,a}$ is an isomorphism of group schemes over $k[\![q^{1/N}]\!]$.

Returning to our original pair $(\mathscr{C}_0, \phi_0) \in \overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N)(k)$, write

$$\phi(1) = (\zeta, a) \in \mu_N \times \mathbb{Z}/(d) \cong \mathrm{Pic}^0_{\mathscr{C}_0/k}[N].$$

Via the map $\mathscr{C}'_0 \to \mathscr{C}_0$ constructed above, this corresponds to $(\zeta, (N/d)\cdot a) \in \mu_N \times \mathbb{Z}/(N) \cong \mathrm{Pic}^0_{\mathscr{C}'_0/k}[N]$. $\mathscr{C}'_0$ is the special fiber of the twisted curve $\mathscr{C}$, and we have an isomorphism over $k[\![q^{1/N}]\!]$

$$\mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}[N] \cong \mathscr{E}^{\mathrm{sm}}[N] \cong \mu_N \times \mathbb{Z}/(N).$$

Now $(\zeta, (N/d)\cdot a)$ lifts to a section of $\mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}[N]$, corresponding to a group scheme homomorphism $\mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}$, hence to a morphism $\mathscr{C} \to \mathscr{B}\mu_N$. Writing $\overline{\mathscr{C}} \to \mathscr{B}\mu_N$ for the relative coarse moduli space and $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\overline{\mathscr{C}}/k[\![q^{1/N}]\!]}$ for the corresponding group scheme homomorphism, we see that $\overline{\mathscr{C}}/k[\![q^{1/N}]\!]$ is a twisted curve with special fiber $\mathscr{C}_0$ and generic fiber an elliptic curve. As desired, ϕ extends ϕ_0 and $(\overline{\mathscr{C}}, \phi) \in \overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N)(k[\![q^{1/N}]\!])$.

A similar argument of course applies to $\overline{\mathscr{K}}^{\circ}_{1,1}(\mathscr{B}\mu_N^2) \subset \overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N^2)$. □

We will require a concrete description of the μ_N-torsor corresponding to a particular sort of $[\Gamma_1(N)]$-structure on an elliptic curve:

Lemma 3.22. *Let K be a field and E/K an elliptic curve. Let $Q \in E(K)$ be a $[\Gamma_1(N)]$-structure on E such that the relative effective Cartier divisor*

$$D := \sum_{a=0}^{N-1} [a \cdot Q]$$

in E is étale over $\mathrm{Spec}(K)$. *Let $P \to E$ be the μ_N-torsor corresponding to Q as in Corollary 3.20*:

$$P = \underline{\mathrm{Spec}}_E\left(\bigoplus_{a=0}^{N-1} \mathscr{L}((a\cdot Q) - (0_E))\right).$$

Then P can be naturally identified with the quotient E/D, where D is viewed as a subgroup scheme of E, étale of rank N over $\mathrm{Spec}(K)$, *with the quotient map $P \to E$ corresponding to the natural map $E/D \to E/E[N] \cong E$.*

Proof. Consider the e_N-pairing on $E[N]$, a nondegenerate bilinear pairing of finite flat group schemes over $\mathrm{Spec}(K)$:

$$e_N : E[N] \times E[N] \to \mu_N.$$

Under our assumptions, the composite map

$$E[N]/D = \{Q\} \times E[N]/D \hookrightarrow D \times E[N]/D \xrightarrow{e_N} \mu_N$$

is an isomorphism of group schemes over $\mathrm{Spec}(K)$. Then via this isomorphism, μ_N acts on the quotient E/D through the natural action of the subgroup scheme

$E[N]/D \subset (E/D)[N]$, making E/D a μ_N-torsor over $E/E[N] \cong E$ with quotient map the obvious one induced from the factorization of $[N]$ as $E \to E/D \to E/E[N] \cong E$.

By Lemma 3.19, we may express the μ_N-torsor $E/D \to E$ as

$$E/D = \underline{\mathrm{Spec}}_E\left(\bigoplus_{a=0}^{N-1} \mathcal{L}_a\right)$$

for some line bundles $\mathcal{L}_a \in \mathrm{Pic}^0_{E/K}[N]$, with the algebra structure determined by isomorphisms $\mathcal{L}_a \otimes \mathcal{L}_b \cong \mathcal{L}_{a+b \bmod N}$ and the μ_N-action corresponding to the grading. We have a natural isomorphism of group schemes over K

$$E \to \mathrm{Pic}^0_{E/K}$$
$$R \in E(K) \mapsto \mathcal{L}((R) - (0_E)),$$

so we conclude that $\mathcal{L}_1 \cong \mathcal{L}((Q_0) - (0_E))$ for some $Q_0 \in E[N]$, and $\mathcal{L}_a \cong \mathcal{L}((a \cdot Q_0) - (0_E))$.

Let $\hat{\pi} : E \to E/D$ be the isogeny dual to $\pi : E/D \to E/E[N] \cong E$. Identifying $E \cong \mathrm{Pic}^0_{E/K}$ and $E/D \cong \mathrm{Pic}^0_{(E/D)/K}$, $\hat{\pi}$ is simply given by the pullback map $\pi^* : \mathrm{Pic}^0_{E/K} \to \mathrm{Pic}^0_{(E/D)/K}$. For any line bundle $\mathcal{L}$ on E/K we have

$$\pi^*(\mathcal{L}) = \bigoplus_{a=0}^{N-1} \mathcal{L} \otimes \mathcal{L}((a \cdot Q_0) - (0_E)),$$

viewing the direct sum of line bundles on E as a line bundle on

$$E/D = \underline{\mathrm{Spec}}_E\left(\bigoplus \mathcal{L}((a \cdot Q_0) - (0_E))\right).$$

In particular, for our original $[\Gamma_1(N)]$-structure Q,

$$\pi^*(\mathcal{L}((Q) - (0_E))) = \bigoplus_{a=0}^{N-1} \mathcal{L}((Q + a \cdot Q_0) - (0_E)).$$

But the dual isogeny to $\pi : E/D \to E/E[N] \cong E$ is the natural quotient map $E \to E/D$, and this maps Q to 0_E. Therefore the line bundle $\bigoplus \mathcal{L}((Q + a \cdot Q_0) - (0_E))$ on E/D is the trivial line bundle on E/D:

$$\bigoplus_{a=0}^{N-1} \mathcal{L}((Q + a \cdot Q_0) - (0_E)) \cong \bigoplus_{a=0}^{N-1} \mathcal{L}((a \cdot Q_0) - (0_E)).$$

Therefore Q is contained in the subgroup scheme of E generated by Q_0: $Q = b \cdot Q_0$ for some b. Since $Q_0 \in E[N]$ and N is the minimal positive integer with $N \cdot Q = 0_E$, this implies that $b \in (\mathbb{Z}/(N))^\times$, and in fact by the definition of the e_N-pairing we

have $b = 1$, that is, $Q_0 = Q$. Thus

$$\underline{\mathrm{Spec}}_E\left(\bigoplus_{a=0}^{N-1} \mathscr{L}((a \cdot Q_0) - (0_E))\right) = \underline{\mathrm{Spec}}_E\left(\bigoplus_{a=0}^{N-1} \mathscr{L}((a \cdot Q) - (0_E))\right),$$

that is, $E/D \cong P$ with the quotient map $P \to E$ of the given μ_N-action becoming identified with the natural quotient map $E/D \to E/E[N] \cong E$. □

4. Moduli of elliptic curves in $\overline{\mathscr{H}}_{1,1}(\mathscr{B}\boldsymbol{\mu}_N)$

***Reduction* mod *p of* $\mathcal{H}_1(N)$.** We first describe how the different components of $\overline{\mathscr{H}}^{\circ}_{1,1}(\mathscr{B}\mu_N) \cong \mathcal{H}_1(N)$ interact. These results are direct corollaries of [Katz and Mazur 1985, §13.5]. Continue working over an arbitrary base scheme S. First consider the case where $N = p^n$ is a prime power. We get a closed immersion

$$\iota^{(p^m)} : \mathcal{Y}_1(p^m) \hookrightarrow \mathcal{H}_1(p^n),$$

for each $0 \le m \le n$, sending a pair $(E, \phi) \in \mathcal{Y}_1(p^m)(T)$, where $\phi : \mathbb{Z}/(p^m) \to E[p^m]$ is a $[\Gamma_1(p^m)]$-structure on E/T, to the pair $(E, \widetilde{\phi}) \in \mathcal{H}_1(p^n)(T)$, where $\widetilde{\phi} : \mathbb{Z}/(p^n) \to E[p^n]$ is obtained from ϕ by precomposing with the canonical projection $\mathbb{Z}/(p^n) \to \mathbb{Z}/(p^m)$. These yield a proper surjection of algebraic stacks

$$\mathcal{Y}_1(p^n) \sqcup \mathcal{Y}_1(p^{n-1}) \sqcup \cdots \sqcup \mathcal{Y}_1(p) \sqcup \mathcal{Y}_1(1) \to \mathcal{H}_1(p^n)$$

which is an isomorphism over $S[1/p]$.

But this is not an isomorphism over $S \otimes \mathbb{F}_p$. Recall that by Theorem 2.13, for any perfect field k of characteristic p, $\mathcal{Y}_1(p^m)_k$ is the disjoint union, with crossings at the supersingular points, of components

$$\mathcal{Y}_1(p^m)_k^{(m-b,b)} \quad (0 \le b \le m),$$

where an object of $\mathcal{Y}_1(p^m)_k^{(m-b,b)}(T)$ is a pair (E, ϕ) where E/T is an elliptic curve and $\phi : \mathbb{Z}/(p^m) \to E[p^m]$ is a $[\Gamma_1(p^m)]$-$(m-b, b)$-cyclic structure on E/T (Definition 2.10). Such an object corresponds via $\iota^{(m)}$ to the pair $(E, \widetilde{\phi}) \in \mathcal{H}_1(p^n)(T)$ as described above.

The key observation is Lemma 2.11: if $\phi : \mathbb{Z}/(p^m) \to E$ is a $[\Gamma_1(p^m)]$-$(m-b, b)$-cyclic structure on an ordinary elliptic curve $E/T/\mathbb{F}_p$, then

$$\widetilde{\phi} := \phi \circ \pi : \mathbb{Z}/(p^n) \to E$$

is a $[\Gamma_1(p^n)]$-$(n-b, b)$-cyclic structure on E, where $\pi : \mathbb{Z}/(p^n) \to \mathbb{Z}/(p^m)$ is the natural projection. Étale locally on T such that $E[p^n] \cong \mu_{p^n} \times \mathbb{Z}/(p^n)$ and

$E[p^m] \cong \mu_{p^m} \times \mathbb{Z}/(p^m)$, the $[\Gamma_1(p^m)]$-$(m-b, b)$-cyclic structure ϕ corresponds to a section of

$$\begin{aligned} &\mu_{p^m}^{\times} \times (\mathbb{Z}/(p^b))^{\times} \quad \text{if } b < m, \\ &\mu_{p^m} \times (\mathbb{Z}/(p^b))^{\times} \quad \text{if } b = m, \end{aligned}$$

and the reason that such a section also gives a $[\Gamma_1(p^n)]$-structure is that in characteristic p, unlike in characteristic $\neq p$, if $c < n$ and $\mathbb{Z}/(p^c) \to \mu_{p^c}$ is a $\mathbb{Z}/(p^c)$-generator then the composite

$$\mathbb{Z}/(p^n) \twoheadrightarrow \mathbb{Z}/(p^c) \to \mu_{p^c} \hookrightarrow \mu_{p^n}$$

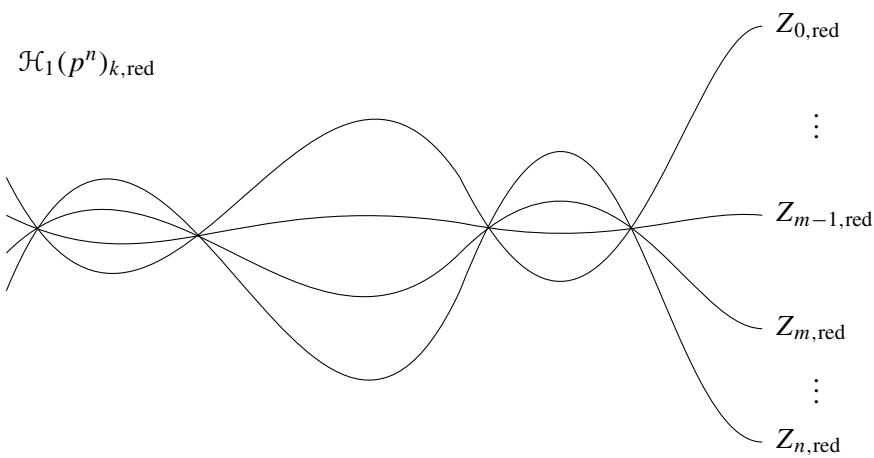
is a $\mathbb{Z}/(p^n)$-generator. This gives us:

Proposition 4.1. *Let k be a perfect field of characteristic p. $\mathcal{H}_1(p^n)_k$ is the disjoint union, with crossings at the supersingular points, of components Z_b for $0 \leq b \leq n$, where*

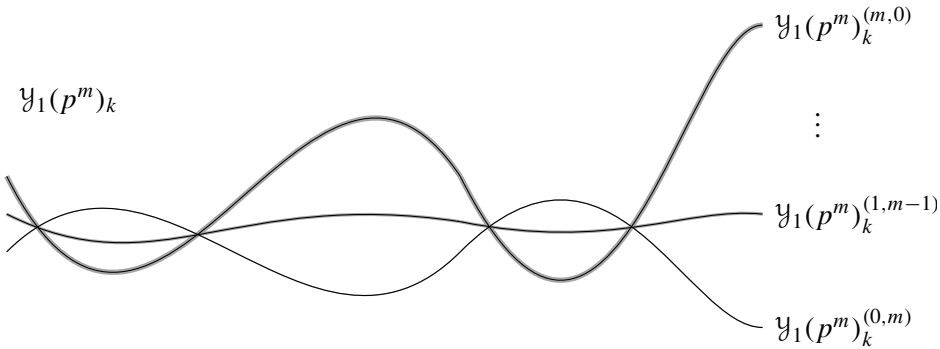
$$Z_b = \bigcup_{b \leq m \leq n} \mathcal{Y}_1(p^m)_k^{(m-b,b)},$$

identifying each $\mathcal{Y}_1(p^m)_k^{(m-b,b)}$ with a closed substack of $\mathcal{H}_1(p^n)_k$ via $\iota^{(p^m)}$. Each $\mathcal{Y}_1(p^m)_k^{(m-b,b)}$ is "set-theoretically identified with Z_b" in the sense that $(Z_b)_{\mathrm{red}} = \mathcal{Y}_1(p^m)_{k,\mathrm{red}}^{(m-b,b)}$ as substacks of $\mathcal{H}_1(p^n)_{k,\mathrm{red}}$ for all $b \leq m \leq n$.

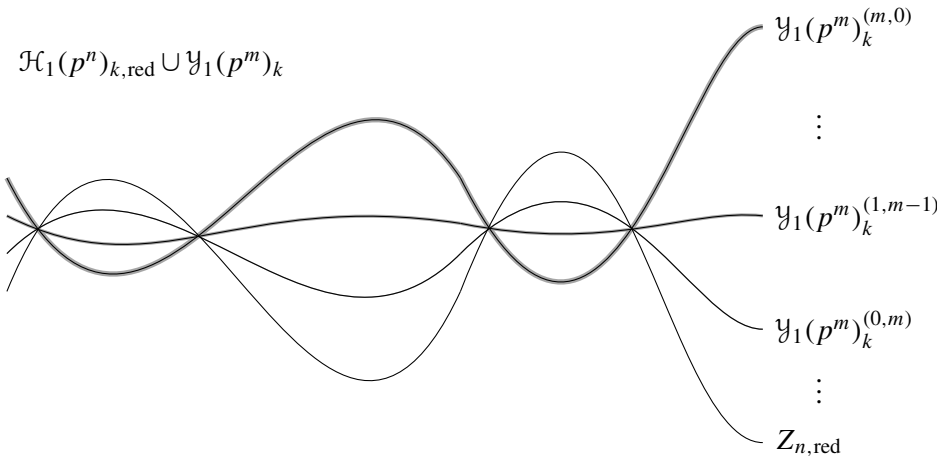
To illustrate, visualize $\mathcal{H}_1(p^n)_{k,\mathrm{red}}$ as follows:



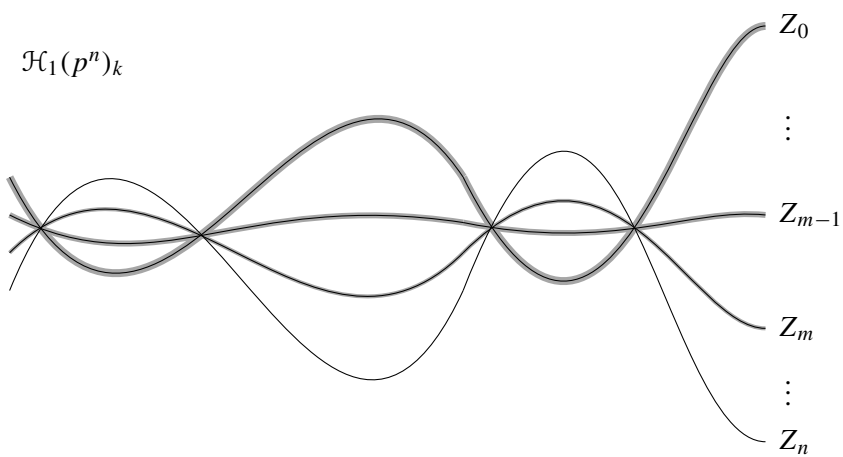
The closed immersion $\iota^{(p^m)}$ sends the following copy of $\mathcal{Y}_1(p^m)_k$ to the obvious closed substack of $\mathcal{H}_1(p^n)_k$, contributing nilpotent structure to the components $Z_0, \ldots, Z_m$:



The result is that $\mathcal{H}_1(p^n)_{k,\mathrm{red}} \cup \mathcal{Y}_1(p^m)_k \subseteq \mathcal{H}_1(p^n)_k$ looks something like this:



Each $\mathcal{Y}_1(p^m)_k$ (for $0 \leq m \leq n$) contributes additional nilpotent structure, giving us:



More generally, for arbitrary N we get a closed immersion

$$\iota^{(d)} : \mathcal{Y}_1(d) \hookrightarrow \mathcal{H}_1(N)$$

for each d dividing N, and the resulting map

$$\bigsqcup_{d \mid N} \mathcal{Y}_1(d) \to \mathcal{H}_1(N)$$

is an isomorphism over $S[1/N]$. It follows immediately from [Katz and Mazur 1985, §13.5] that if $(r, p) = 1$, then for a perfect field k of characteristic p, $\mathcal{Y}_1(p^m r)_k$ is the disjoint union, with crossings at the supersingular points, of $m+1$ components $\mathcal{Y}_1(p^m r)_k^{(m-b,b)}$ $(0 \le b \le m)$, where

$$\mathcal{Y}_1(p^m r)_k^{(m-b,b)} := \mathcal{Y}_1(r)_k \times_{\overline{\mathcal{M}}_{1,1,k}} \mathcal{Y}_1(p^m)_k^{(m-b,b)}.$$

Now let $N = p^n N'$, where $(N', p) = 1$, and for any $r \mid N'$ let $\mathcal{H}_1(N)_k^r \subset \mathcal{H}_1(N)_k$ denote the union of the components $\mathcal{Y}_1(p^m r)_k$ for $0 \le m \le n$. In summary:

Corollary 4.2. *Let k be a perfect field of characteristic p. For any r dividing N', $\mathcal{H}_1(N)_k^r$ is the disjoint union, with crossings at the supersingular points, of components Z_b^r for $0 \le b \le n$, where*

$$Z_b^r = \bigcup_{b \le m \le n} \mathcal{Y}_1(p^m r)_k^{(m-b,b)},$$

identifying each $\mathcal{Y}_1(p^m r)_k^{(m-b,b)}$ with a closed substack of $\mathcal{H}_1(N)_k$ via $\iota^{(p^m r)}$. Each $\mathcal{Y}_1(p^m r)_k^{(m-b,b)}$ is "set-theoretically identified with Z_b^r" in the sense that $(Z_b)_{\mathrm{red}} = \mathcal{Y}_1(p^m r)_{k,\mathrm{red}}^{(m-b,b)}$ as substacks of $\mathcal{H}_1(N)_{k,\mathrm{red}}$. $\mathcal{H}_1(N)_k$ is the disjoint union of the closed substacks

$$\{\mathcal{H}_1(N)_k^r\}_{r \mid N'}.$$

Closure of $\mathcal{Y}_1(N)$ in $\overline{\mathcal{H}}_{1,1}(\mathcal{B}\mu_N)$. We now want to describe the closure of $\mathcal{Y}_1(N)$ in $\overline{\mathcal{H}}_{1,1}(\mathcal{B}\mu_N)$, as a moduli stack classifying twisted curves with extra structure. This is accomplished in [Abramovich et al. 2003, §5.2] over $\mathbb{Z}[1/N]$; let us briefly recall how this is done.

Definition 4.3. Let G be a finite group, viewed as a finite étale group scheme over $\mathbb{Z}[1/|G|]$. Fix an isomorphism $G^* \cong G$, after adjoining the necessary roots of unity to $\mathbb{Z}[1/|G|]$. $\mathcal{B}_{g,n}^{\mathrm{tei}}(G)$ is defined as the substack of $\overline{\mathcal{H}}_{1,1}(\mathcal{B}G)$ over $\mathbb{Z}[1/|G|]$ whose objects are *twisted Teichmüller G-structures* on twisted curves. An object of $\overline{\mathcal{H}}_{g,n}(\mathcal{B}G)(T)$ is a pair $(\mathscr{C}, \phi)$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and ϕ is a group scheme homomorphism $G \to \mathrm{Pic}^0_{\mathscr{C}/T}$. By definition, $(\mathscr{C}, \phi) \in \mathcal{B}_{g,n}^{\mathrm{tei}}(G)(T)$ if and only if, whenever $P \to \mathscr{C}$ is a G-torsor corresponding to ϕ (fppf locally on T), the geometric fibers of P/T are connected.

$\mathcal{B}^{\mathrm{tei}}_{g,n}(G)$ is naturally a closed substack of $\overline{\mathcal{K}}_{g,n}(\mathcal{B}G)$. Working over $\mathbb{Z}[1/N]$, the choice of a primitive N-th root of unity ζ_N determines an isomorphism of group schemes $G := \mathbb{Z}/(N) \cong \mu_N$ over $\mathbb{Z}[\zeta_N, 1/N]$. Applying the resulting isomorphism $\overline{\mathcal{K}}_{1,1}(\mathcal{B}G) \cong \overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$, we may view $\mathcal{Y}_1(N)$ as a substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}G)$, and the closure of $\mathcal{Y}_1(N)$ in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}G)$ over $\mathbb{Z}[\zeta_N, 1/N]$ can be shown to be $\mathcal{B}^{\mathrm{tei}}_{1,1}(G)$ (indeed, this follows from Theorem 4.6). Thus:

Corollary 4.4. *The closure of $\mathcal{Y}_1(N)$ in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$ over $\mathbb{Z}[1/N]$ is the stack whose objects over a scheme $T/\mathbb{Z}[1/N]$ are pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ is a group scheme homomorphism, and whenever $P \to \mathscr{C}$ is a μ_N-torsor corresponding to ϕ (fppf locally on T) the geometric fibers of P/T are connected.*

However, in characteristics dividing N, simply requiring the μ_N-torsors to have connected geometric fibers will not give us a moduli stack isomorphic to the closure of $\mathcal{Y}_1(N)$ in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$. For example, the group scheme μ_{p^n} is itself connected over any field of characteristic p; so any μ_{p^n}-torsor over an n-marked genus-g twisted stable curve over a field of characteristic p will automatically be connected. So if the above result held over a base scheme $S \in \mathrm{Sch}/\mathbb{F}_p$, it would say the closure of $\mathcal{Y}_1(p^n)$ in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_{p^n})$ is all of $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_{p^n})$ (the substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_{p^n})$ where the marking is representable). We will see that this is not true; the closure of $\mathcal{Y}_1(p^n)$ turns out to be finite and flat of constant rank $p^{2n} - p^{2n-2}$ over $\overline{\mathcal{M}}_{1,1}$, while $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_{p^n})$ turns out to be finite and flat of constant rank p^{2n} over $\overline{\mathcal{M}}_{1,1}$.

A μ_N-torsor over $\mathscr{C}/S$ determines a group scheme homomorphism $\mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ (see Lemma 3.19), and over $\mathbb{Z}[1/N]$ a μ_N-torsor with connected geometric fibers corresponds to a group scheme homomorphism which is injective. We are therefore led to consider Drinfeld structures on our twisted curves:

Definition 4.5. Let $\mathscr{C}/S$ be a 1-marked genus-1 twisted stable curve with no stacky structure at its marking. A $[\Gamma_1(N)]$-*structure* on $\mathscr{C}$ is a group scheme homomorphism $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ such that:

- the relative effective Cartier divisor
$$D := \sum_{a \in \mathbb{Z}/(N)} [\phi(a)]$$
in $\mathrm{Pic}^0_{\mathscr{C}/S}$ is an S-subgroup scheme and
- for every geometric point $\bar{p} \to S$, $D_{\bar{p}}$ meets every irreducible component of $(\mathrm{Pic}^0_{\mathscr{C}/S})_{\bar{p}} = \mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})}$.

Over any scheme S, we define $\mathcal{X}^{\mathrm{tw}}_1(N)$ to be the substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$ whose objects over an S-scheme T are pairs $(\mathscr{C}, \phi) \in \overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)(T)$ such that ϕ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}$.

Note that if $\mathscr{C}/S$ is a 1-marked genus-1 stable curve (nontwisted), so that every geometric fiber of $\mathscr{C}/S$ is either a smooth elliptic curve or a Néron 1-gon, then this agrees with the definition of a $[\Gamma_1(N)]$-structure on $\mathscr{C}/S$ as given in [Conrad 2007, 2.4.1] and in Definition 2.3 above, identifying a $[\Gamma_1(N)]$-structure $P \in \mathscr{C}^{\mathrm{sm}}(S)$ with its corresponding group scheme homomorphism $\mathbb{Z}/(N) \to \mathscr{C}^{\mathrm{sm}}$, $1 \mapsto P$, since $\mathrm{Pic}^0_{\mathscr{C}/S} \cong \mathscr{C}^{\mathrm{sm}}$ in this case.

Theorem 4.6 (Restatement of Theorem 1.1). *Let S be a scheme and $\mathcal{X}_1^{\mathrm{tw}}(N)$ the stack over S classifying $[\Gamma_1(N)]$-structures on 1-marked genus-1 twisted stable curves with nonstacky marking. Then $\mathcal{X}_1^{\mathrm{tw}}(N)$ is a closed substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$, which contains $\mathcal{Y}_1(N)$ as an open dense substack.*

In particular $\mathcal{X}_1^{\mathrm{tw}}(N)$ is flat over S with local complete intersection fibers, and is proper and quasifinite over $\overline{\mathcal{M}}_{1,1}$.

Remark 4.7. Although this gives a new modular compactification of $\mathcal{Y}_1(N)$, it should be noted that the proof of the theorem relies in several places on the proof in [Conrad 2007] that the moduli stack classifying $[\Gamma_1(N)]$-structures on generalized elliptic curves is a proper algebraic stack over $\overline{\mathcal{M}}_{1,1}$.

Proof of Theorem 4.6. The main point is to verify the valuative criterion of properness for $\mathcal{X}_1^{\mathrm{tw}}(N)$, which implies $\mathcal{X}_1^{\mathrm{tw}}(N)$ is a closed substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$. It follows from Lemma 3.21 that $\mathcal{Y}_1(N)$ is dense.

Let R be a discrete valuation ring with $T := \mathrm{Spec}(R) \in \mathrm{Sch}/S$; write $\eta = \mathrm{Spec}(K)$ for the generic point of T and $s = \mathrm{Spec}(k)$ for the closed point. Let $(\mathscr{C}_\eta, \phi_\eta) \in \mathcal{X}_1^{\mathrm{tw}}(N)(\eta)$, so $\mathscr{C}_\eta/K$ is a 1-marked genus-1 twisted stable curve with nonstacky marking and $\phi_\eta : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_\eta/K}$ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}_\eta$. Since $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)$ is proper over S, there is a discrete valuation ring R_1 containing R as a local subring, with corresponding morphism of spectra $T_1 \to T$ over S, such that the pair $(\mathscr{C}_\eta \times_T T_1, (\phi_\eta)_{T_1})$ extends to a pair $(\mathscr{C}_{T_1}, \phi_{T_1}) \in \overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N)(T)$; whenever such an extension exists, it is unique. Therefore it suffices to show that for some such R_1 and T_1, there exists a $[\Gamma_1(N)]$-structure ϕ_{T_1} on a 1-marked genus-1 twisted stable curve $\mathscr{C}_{T_1}$ extending the $[\Gamma_1(N)]$-structure $(\phi_\eta)_{T_1}$ on $\mathscr{C}_\eta \times_T T_1$. This is accomplished in Lemmas 4.9–4.12.

Terminology 4.8. In the following lemmas and their proofs, "base change on R" will refer to replacing R with a discrete valuation ring R_1 as above.

By [Deligne and Rapoport 1973, IV.1.6], after a base change on R we may assume that the minimal proper regular model of the coarse space C_η of $\mathscr{C}_\eta$ is a generalized elliptic curve. We maintain this assumption for the rest of the proof.

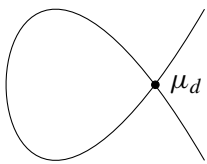
Lemma 4.9. *Suppose $\mathscr{C}_\eta/K$ is a smooth elliptic curve whose minimal proper regular model over R is smooth, and ϕ_η is a $[\Gamma_1(N)]$-structure on $\mathscr{C}_\eta$. Then after*

base change on R, $(\mathscr{C}_\eta, \phi_\eta)$ extends to a $[\Gamma_1(N)]$-structure ϕ on a smooth elliptic curve $\mathscr{C}/R$.

Proof. This follows immediately from the fact that the stack $\mathfrak{X}_1(N)$ classifying $[\Gamma_1(N)]$-structures on generalized elliptic curves is proper [Conrad 2007, 1.2.1]. □

Lemma 4.10. *Suppose $(\mathscr{C}_\eta, \phi_\eta) \in \mathfrak{X}_1^{\mathrm{tw}}(N)(\eta)$, such that the coarse space C_η of $\mathscr{C}_\eta$ is singular. Then after base change on R, $(\mathscr{C}_\eta, \phi_\eta)$ extends to $(\mathscr{C}, \phi) \in \mathfrak{X}_1^{\mathrm{tw}}(N)(R)$.*

Proof. After a base change on R, we may assume that $\mathscr{C}_\eta/K$ is the standard μ_d-stacky Néron 1-gon as in Example 3.4, for some $d \geq 1$; since $\mathscr{C}_\eta$ admits a $[\Gamma_1(N)]$-structure and $\mathrm{Pic}^0_{\mathscr{C}_\eta/K} \cong \mathbb{G}_{m,K} \times \mathbb{Z}/(d)$, we have $d \mid N$.



Standard μ_d-stacky Néron 1-gon

Let $\mathscr{C}/R$ be the standard μ_d-stacky Néron 1-gon over R, so $\mathscr{C}$ is a genus-1 twisted curve over R extending $\mathscr{C}_\eta$ (and of course the marking of $\mathscr{C}_\eta$ extends to $\mathscr{C}$). After further base change on R we may assume $\mu_N(K) = \mu_N(R)$. Let $\phi_\eta(1) = (\zeta, a) \in \mathbb{G}_m(K) \times \mathbb{Z}/(d)$; since ϕ_η is a $[\Gamma_1(N)]$-structure on $\mathscr{C}_\eta$, we have $\zeta \in \mu_N(K)$ and $a \in (\mathbb{Z}/(d))^\times$. Since $\mu_N(K) = \mu_N(R)$, this extends to a section $(\zeta, a) \in \mathbb{G}_m(R) \times \mathbb{Z}/(d)$, determining a group scheme homomorphism $\phi : \mathbb{Z}/(N) \to \mathbb{G}_{m,R} \times \mathbb{Z}/(d)$ with $\phi(1) = (\zeta, a)$. Since $a \in (\mathbb{Z}/(d))^\times$, ϕ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}$. □

Lemma 4.11. *Suppose R has pure characteristic $p \geq 0$, and $\mathscr{C}_\eta = E_\eta$ is an elliptic curve over K whose minimal proper regular model over R is not smooth; that is, E_η/K is an ordinary elliptic curve with bad reduction. If $\phi_\eta : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_\eta/K}$ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}_\eta$, then after base change on R, there exists a pair $(\mathscr{C}, \phi) \in \mathfrak{X}_1^{\mathrm{tw}}(N)(R)$ extending $(\mathscr{C}_\eta, \phi_\eta)$.*

Proof. Since for any coprime N and N' we obviously have

$$\overline{\mathscr{K}}_{1,1}(\mathscr{B}\mu_{NN'}) \cong \overline{\mathscr{K}}_{1,1}(\mathscr{B}\mu_N) \times_{\overline{\mathscr{M}}_{1,1}} \overline{\mathscr{K}}_{1,1}(\mathscr{B}\mu_{N'})$$

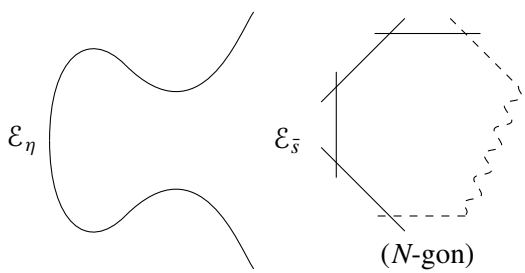
and

$$\mathfrak{X}_1^{\mathrm{tw}}(NN') \cong \mathfrak{X}_1^{\mathrm{tw}}(N) \times_{\overline{\mathscr{M}}_{1,1}} \mathfrak{X}_1^{\mathrm{tw}}(N'),$$

it suffices to consider the cases where (a) N is prime to p (including the case $p = 0$) and (b) $N = p^n$.

(a) First suppose $(N, p) = 1$ or $p = 0$. After base extension on R we may assume that the finite abelian group $E_\eta(K)[N]$ is isomorphic to $(\mathbb{Z}/(N))^2$. After further base extension on R, the map ϕ_η gives us a μ_N-torsor $P_\eta \to E_\eta$, corresponding to the

point $Q_\eta = \phi_\eta(1) \in \mathrm{Pic}^0_{E_\eta/K}(K) \cong E_\eta(K)$ of "exact order N" in the sense of [Katz and Mazur 1985, §1.4] (and since $(N, p) = 1$, this just says Q_η has exact order N as an element of the group $E_\eta(K)$). Let $\mathcal{E}/R$ be the minimal proper regular model; after further base extension on R and replacing $\mathcal{E}$ with the new minimal proper regular model of E_η, we may assume $\mathcal{E}$ has the structure of a generalized elliptic curve extending that of E_η (see [Deligne and Rapoport 1973, IV.1.6], and also [Conrad 2007, proof of 3.2.6]). Since $E_\eta(K) = \mathcal{E}^{\mathrm{sm}}(R)$, the finite flat R-group scheme $\mathcal{E}^{\mathrm{sm}}[N]$ has rank N^2, hence is étale over $\mathrm{Spec}(R)$ since $(N, p) = 1$. Therefore the special fiber $\mathcal{E}_s$ is geometrically a Néron mN-gon for some m. Replacing $\mathcal{E}$ with its contraction away from the subgroup scheme $\mathcal{E}^{\mathrm{sm}}[N] \subset \mathcal{E}^{\mathrm{sm}}$ (see [Deligne and Rapoport 1973, IV.1]), we get a generalized elliptic curve $\mathcal{E}/R$ extending E_η, whose special fiber is geometrically a Néron N-gon, with $E_\eta(K)[N] \cong \mathcal{E}^{\mathrm{sm}}(R)[N]$.



Note that Q_η extends uniquely to a section $Q \in \mathcal{E}^{\mathrm{sm}}(R)[N]$ of exact order N, which Conrad calls a "possibly nonample $[\Gamma_1(N)]$-structure on $\mathcal{E}$", meaning Q satisfies all the conditions of a $[\Gamma_1(N)]$-structure except that the relative effective Cartier divisor $\sum_{a \in \mathbb{Z}/(N)} [a \cdot Q]$ might not meet every irreducible component of a geometric closed fiber $\mathcal{E}_{\bar{s}}$.

Recall that by Corollary 3.20,

$$P_\eta = \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{N-1} \mathscr{L}((a \cdot Q_\eta) - (0_{E_\eta}))\right)$$

with the μ_N-action on $\bigoplus \mathscr{L}((a \cdot Q_\eta) - (0_{E_\eta}))$ induced by the $\mathbb{Z}/(N)$-grading. Since N is invertible on $\mathrm{Spec}(R)$, the assumptions of Lemma 3.22 are satisfied, so we may identify P_η with $E_\eta/\langle Q_\eta \rangle$, with the quotient map $P_\eta \to E_\eta$ identified with the quotient map $E_\eta/\langle Q_\eta \rangle \to E_\eta/E_\eta[N] \cong E_\eta$. Then the μ_N-action on $P_\eta = E_\eta/\langle Q_\eta \rangle$ is determined by the group scheme isomorphism $E_\eta[N]/\langle Q_\eta \rangle \cong \mu_N$ induced by the e_N-pairing and the choice of Q_η:

$$E_\eta[N]/\langle Q_\eta \rangle = \{Q_\eta\} \times E_\eta[N]/\langle Q_\eta \rangle \hookrightarrow \langle Q_\eta \rangle \times E_\eta[N]/\langle Q_\eta \rangle \xrightarrow{e_N} \mu_N.$$

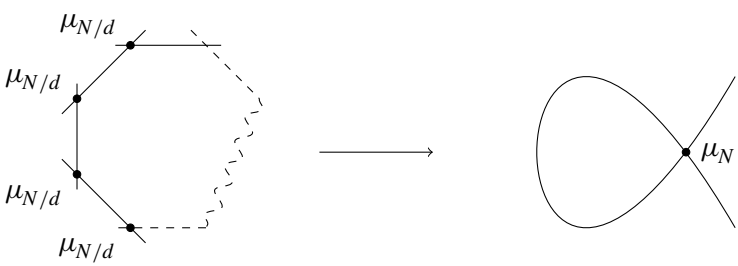
By [Conrad 2007, Theorem 4.1.1], the e_N-pairing on E_η/K extends (possibly after further base change on R) to a nondegenerate bilinear pairing of finite flat

commutative group schemes over R

$$e_N : \mathcal{E}^{\mathrm{sm}}[N] \times \mathcal{E}^{\mathrm{sm}}[N] \to \mu_N.$$

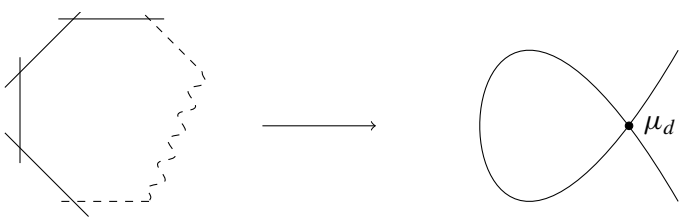
Therefore the isomorphism of group schemes $E_\eta[N]/\langle Q_\eta \rangle \cong \mu_N$ described above may be extended via the same formula to an isomorphism of group schemes $\mathcal{E}^{\mathrm{sm}}[N]/\langle Q \rangle \cong \mu_N$.

This isomorphism of group schemes makes $P' := [\mathcal{E}/\langle Q \rangle]$ a μ_N-torsor over the twisted curve $\mathscr{C}' := [\mathcal{E}/\mathcal{E}^{\mathrm{sm}}[N]]$, extending our original μ_N-torsor over E_η ($\mathscr{C}'$ is indeed a twisted curve: by [Abramovich et al. 2011, Proposition 2.3] we may detect this on the geometric fibers, where it is clear, since the geometric closed fiber of $\mathscr{C}'$ is a standard μ_N-stacky Néron 1-gon). But P' is not necessarily representable. Indeed, let d be the minimal positive integer such that $d \cdot Q_{\bar{s}}$ lies in the identity component of the geometric closed fiber $\mathcal{E}_{\bar{s}}$. Then the coarse space $\overline{P}'_{\bar{s}}$ of $P'_{\bar{s}}$ is a Néron N/d-gon, and for any geometric point $\bar{q} \to \overline{P}'_{\bar{s}}$ mapping to a node of $\overline{P}'_{\bar{s}}$, we have $P'_{\bar{s}} \times_{\overline{P}'_{\bar{s}}} \bar{q} \cong (\mathcal{B}\mu_{N/d})_{k(\bar{q})}$.



$P'_{\bar{s}} = [\mathcal{E}_{\bar{s}}/\langle Q_{\bar{s}} \rangle]$ (Stacky N/d-gon) $\qquad$ $\mathscr{C}'_{\bar{s}} = [\mathcal{E}_{\bar{s}}/\mathcal{E}^{\mathrm{sm}}_{\bar{s}}[N]]$

Let $P \to \mathscr{C}$ be the μ_N-torsor corresponding to the relative coarse moduli space [Abramovich et al. 2011, Theorem 3.1] of the map $\mathscr{C}' \to \mathcal{B}\mu_N$ coming from the μ_N-torsor $P' \to \mathscr{C}'$. The generic fibers are the same as those of $P' \to \mathscr{C}'$, $P_{\bar{s}}$ is a nonstacky Néron N/d-gon, and $\mathscr{C}_{\bar{s}}$ is a standard μ_d-stacky Néron 1-gon, say with coarse space $\pi : \mathscr{C}_{\bar{s}} \to C_{\bar{s}}$.



$P_{\bar{s}}$ (N/d-gon) $\qquad$ $\mathscr{C}_{\bar{s}}$

Let $\bar{q} \to C_{\bar{s}}$ be a geometric point mapping to the node of $C_{\bar{s}}$. Then $\mathscr{C}_{\bar{s}} \times_{C_{\bar{s}}} \bar{q} \cong (\mathscr{B}\mu_d)_{k(\bar{q})}$ and $P_{\bar{s}} \times_{C_{\bar{s}}} \bar{q} = \mu_{N/d} \times \bar{q}$ (which as a $k(\bar{q})$-scheme is just N/d disjoint copies of $\bar{q}$ since $(N, p) = 1$), so the resulting μ_N-torsor over $(\mathscr{B}\mu_d)_{k(\bar{q})}$ corresponds to a generator of $H^1(\mathscr{B}\mu_d, \mu_N) \cong \mathbb{Z}/(d)$. Therefore, with respect to the decomposition

$$\mathrm{Pic}^0_{\mathscr{C}_{\bar{s}}/k(\bar{s})} \cong \mathrm{Pic}^0_{C_{\bar{s}}/k(\bar{s})} \times H^0(C_{\bar{s}}, \boldsymbol{R}^1\pi_*\mathbb{G}_m) \cong \mathbb{G}_m \times \mathbb{Z}/(d),$$

the class of the μ_N-torsor $P_{\bar{s}} \to \mathscr{C}_{\bar{s}}$ projects in the second factor to a generator of $\mathbb{Z}/(d)$, so the map $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ induced by P is a $[\Gamma_1(N)]$-structure on the twisted curve $\mathscr{C}/R$ extending the $[\Gamma_1(N)]$-structure ϕ_η on $\mathscr{C}_\eta$.

(b) Now suppose $N = p^n$. After base change on R, the $[\Gamma_1(N)]$-structure ϕ_η gives us a μ_N-torsor $P_\eta \to E_\eta$ corresponding to the point $Q_\eta = \phi_\eta(1) \in \mathrm{Pic}^0_{E_\eta/K}(K) \cong E_\eta(K)$ of "exact order N" (in the sense of [Katz and Mazur 1985, §1.4], but not necessarily as an element of the group $E_\eta(K)$). Since E_η/K is ordinary, after base change on R we have an isomorphism of group schemes over K

$$E_\eta[N] \cong \mu_N \times \mathbb{Z}/(N),$$

so $E_\eta(K)[N] \cong \mu_N(K) \times \mathbb{Z}/(N) = \{1\} \times \mathbb{Z}/(N) = \mathbb{Z}/(N)$. The μ_N-torsor $P_\eta \to E_\eta$ is

$$P_\eta = \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{N-1} \mathscr{L}((a \cdot Q_\eta) - (0_{E_\eta}))\right)$$

with the μ_N-action on $\bigoplus \mathscr{L}((Q_\eta) - (0_{E_\eta}))$ induced by the $\mathbb{Z}/(N)$-grading.

Choose $m \geq 0$ minimal such that the image of $p^m \cdot Q_\eta$ in $\mathbb{Z}/(N)$ is zero. So $p^m \cdot Q_\eta = 0_{E_\eta}$ is viewed as a point of "exact order p^{n-m}" on E_η. Since $\mathscr{L}((p^m \cdot Q_\eta) - (0_{E_\eta})) \cong \mathbb{O}_{E_\eta}$, the corresponding $\mu_{p^{n-m}}$-torsor over E_η is trivial. So if $\mathscr{C}/R$ is a twisted curve extending E_η/K, the $\mu_{p^{n-m}}$-torsor corresponding to $p^m \cdot Q_\eta$ will automatically extend to the trivial $\mu_{p^{n-m}}$-torsor over $\mathscr{C}$.

Now view Q_η as a point of "exact order p^m" on E_η/K. The relative effective Cartier divisor

$$D_\eta := \sum_{a \in \mathbb{Z}/(p^m)} [a \cdot Q_\eta]$$

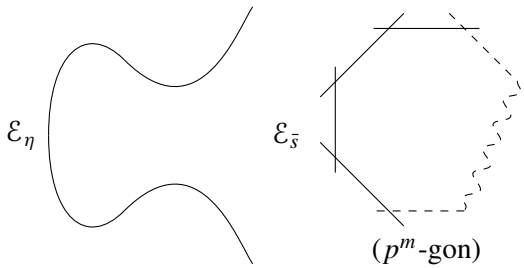
in E_η is a subgroup scheme which is étale over K. So by Lemma 3.22, E_η/D_η is the underlying scheme of the μ_{p^m}-torsor corresponding to Q_η. The μ_{p^m}-action on E_η/D_η is given by the group scheme isomorphism

$$E_\eta[p^m]/D_\eta \cong \mu_{p^m}$$

induced by the e_{p^m}-pairing and the choice of Q_η:

$$E_\eta[p^m]/D_\eta = \{Q_\eta\} \times E_\eta[p^m]/D_\eta \hookrightarrow D_\eta \times E_\eta[p^m]/D_\eta \xrightarrow{e_{p^m}} \mu_{p^m}.$$

After a base change on R if necessary, let $\mathcal{E}/R$ be a generalized elliptic curve extending E_η, with a $[\Gamma_1(p^m)]$-structure Q on $\mathcal{E}$ extending Q_η. The special fiber $\mathcal{E}_s/k$ is geometrically a Néron p^m-gon, and $\mathcal{E}^{\mathrm{sm}}[p^m] \cong \mu_{p^m} \times \mathbb{Z}/(p^m)$.



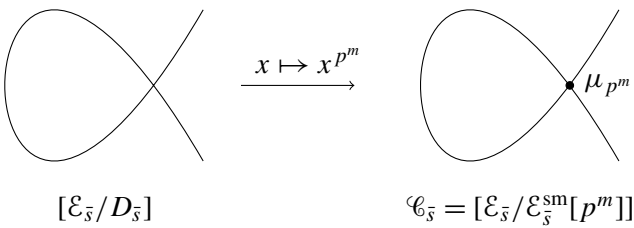
By [Conrad 2007, Theorem 4.1.1], the e_{p^m}-pairing on E_η/K extends (possibly after further base change on R) to a nondegenerate bilinear pairing of finite flat commutative group schemes over R

$$e_{p^m} : \mathcal{E}^{\mathrm{sm}}[p^m] \times \mathcal{E}^{\mathrm{sm}}[p^m] \to \mu_{p^m}.$$

Therefore the isomorphism of group schemes $E_\eta[p^m]/D_\eta \cong \mu_{p^m}$ described above may be extended via the same formula to an isomorphism of group schemes $\mathcal{E}^{\mathrm{sm}}[p^m]/D \cong \mu_{p^m}$, where D is the relative effective Cartier divisor

$$D = \sum_{a \in \mathbb{Z}/(p^m)} [a \cdot Q]$$

in $\mathcal{E}^{\mathrm{sm}}$. This makes the quotient $[\mathcal{E}/D] = \mathcal{E}/D$ a representable μ_{p^m}-torsor over the twisted curve $\mathscr{C} := [\mathcal{E}/\mathcal{E}^{\mathrm{sm}}[p^m]]$, which extends the given μ_{p^m}-torsor over E_η.



The geometric special fiber $\mathscr{C}_{\bar{s}}$ is a standard μ_{p^m}-stacky Néron 1-gon, say with coarse space $\pi : \mathscr{C}_{\bar{s}} \to C_{\bar{s}}$; $[\mathcal{E}_{\bar{s}}/D_{\bar{s}}] = \mathcal{E}_{\bar{s}}/D_{\bar{s}}$ is a (nonstacky) Néron 1-gon, and the quotient map $P_{\bar{s}} := \mathcal{E}_{\bar{s}}/D_{\bar{s}} \to \mathscr{C}_{\bar{s}}$ extends the map $P^{\mathrm{sm}}_{\bar{s}} = \mathbb{G}_m \to \mathscr{C}^{\mathrm{sm}}_{\bar{s}} = \mathbb{G}_m$, $x \mapsto x^{p^m}$. In particular, if $\bar{q} \to C_{\bar{s}}$ is a geometric point mapping to the node of $C_{\bar{s}}$, then $\mathscr{C} \times_C \bar{q} \cong (\mathcal{B}\mu_{p^m})_{k(\bar{q})}$ and $P_{\bar{s}} \times_C \bar{q} = \bar{q}$, so the resulting μ_{p^m}-torsor over $(\mathcal{B}\mu_{p^m})_{k(\bar{q})}$ corresponds to a generator of $H^1(\mathcal{B}\mu_{p^m}, \mu_{p^m}) \cong \mathbb{Z}/(p^m)$. Therefore, with respect to the decomposition

$$\mathrm{Pic}^0_{\mathscr{C}_{\bar{s}}/k(\bar{s})} \cong \mathrm{Pic}^0_{C_{\bar{s}}/k(\bar{s})} \times H^0(C_{\bar{s}}, \boldsymbol{R}^1\pi_*\mathbb{G}_m) \cong \mathbb{G}_m \times \mathbb{Z}/(p^m),$$

the class of the μ_{p^m}-torsor $P_{\bar{s}} \to \mathscr{C}_{\bar{s}}$ projects in the second factor to a generator of $\mathbb{Z}/(p^m)$, so the group scheme homomorphism $\phi : \mathbb{Z}/(p^m) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ corresponding to $P := [\mathcal{E}/D] = \mathcal{E}/D$ is a $[\Gamma_1(p^m)]$-structure on the twisted curve $\mathscr{C}/R$.

Finally, write

$$P = \underline{\mathrm{Spec}}_{\mathscr{C}}\left(\bigoplus_{a=0}^{p^m-1} \mathscr{L}_a\right),$$

with the grading determined by the μ_{p^m}-action on P. Since P extends the μ_{p^m}-torsor on E_η determined by Q_η, we have $\mathscr{L}_a|_\eta \cong \mathscr{L}((Q_\eta) - (0_{E_\eta}))$. Then

$$P \times \mu_{p^{n-m}} = \underline{\mathrm{Spec}}_{\mathscr{C}}\left(\bigoplus_{a=0}^{p^n-1} \mathscr{L}_{(a \bmod p^m)}\right)$$

is a μ_{p^n}-torsor over $\mathscr{C}$ with the μ_{p^n}-action determined by the grading, extending the original μ_{p^n}-torsor over E_η and representable since P is representable. Since the group scheme homomorphism $\phi : \mathbb{Z}/(p^m) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ corresponding to P is a $[\Gamma_1(p^m)]$-structure on $\mathscr{C}$, and the group scheme homomorphism $\phi' : \mathbb{Z}/(p^n) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ corresponding to $P \times \mu_{p^{n-m}}$ is $\phi \circ \pi$ for the canonical projection $\pi : \mathbb{Z}/(p^n) \to \mathbb{Z}/(p^m)$, it follows immediately that ϕ' is a $[\Gamma_1(p^n)]$-structure on $\mathscr{C}$. □

Lemma 4.12. *Suppose R has mixed characteristic $(0, p)$, and $\mathscr{C}_\eta = E_\eta$ is an elliptic curve over K whose minimal proper regular model over R is not smooth; that is, E_η/K is an ordinary elliptic curve with bad reduction. If $\phi_\eta : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_\eta/K}$ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}_\eta$, then after base change on R, there exists a pair $(\mathscr{C}, \phi) \in \mathcal{X}_1^{\mathrm{tw}}(N)(R)$ extending $(\mathscr{C}_\eta, \phi_\eta)$.*

Proof. As before, we can restrict to the two separate cases, one where $(N, p) = 1$ and another where $N = p^n$.

(a) If $(p, N) = 1$, the same argument as in Part (a) of the proof of Lemma 4.11 carries through.

(b) Suppose $N = p^n$. As in the proof of Lemma 4.11, after base extension on R we can extend E_η to a generalized elliptic curve $\mathcal{E}/R$ whose special fiber is geometrically a Néron N-gon, such that $\mathcal{E}^{\mathrm{sm}}(R)[N] \cong E_\eta(K)[N] \cong (\mathbb{Z}/(N))^2$ (the latter isomorphism being a noncanonical isomorphism of abelian groups). After further base change on R, the $[\Gamma_1(p^n)]$-structure ϕ_η gives us a μ_N-torsor $P_\eta \to E_\eta$, corresponding to $Q_\eta = \phi_\eta(1) \in \mathrm{Pic}^0_{E_\eta/K}(K) \cong E_\eta(K)$ of "exact order N" in the sense of [Katz and Mazur 1985, §1.4]. Q_η extends to a "possibly nonample $[\Gamma_1(N)]$-structure" on $\mathcal{E}/R$. Since the special fiber $\mathcal{E}_s$ is geometrically a Néron N-gon, after further base change on R we may assume $\mathcal{E}_s^{\mathrm{sm}}[N] \cong \mu_N \times \mathbb{Z}/(N)$, so

$$\mathcal{E}_s^{\mathrm{sm}}(k(s))[N] \cong \mu_N(k(s)) \times \mathbb{Z}/(N) = \{1\} \times \mathbb{Z}/(N) \cong \mathbb{Z}/(N).$$

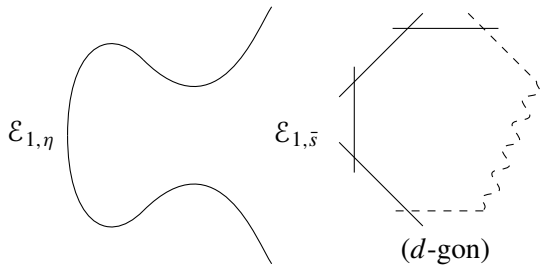
Suppose $d \geq 1$ is minimal such that $d \cdot Q_s$ maps to 0 in $\mathbb{Z}/(N)$. Then we can choose $Q^1, Q^2 \in \mathcal{E}^{\mathrm{sm}}(R)[N]$ such that:

- $Q = Q^1 + Q^2$ in $\mathcal{E}^{\mathrm{sm}}(R)$,
- Q^1 has exact order d in the abelian group $\mathcal{E}^{\mathrm{sm}}(R)[N]$, and the relative effective Cartier divisor
$$\sum_{a=0}^{d-1}[a \cdot Q]$$
in $\mathcal{E}^{\mathrm{sm}}$ is étale over $\mathrm{Spec}(R)$, and
- Q_s^2 maps to 0 in $\mathbb{Z}/(N)$ via the above isomorphism.

In the abelian group $\mathcal{E}^{\mathrm{sm}}(R)[N] \cong (\mathbb{Z}/(N))^2$, Q^2 has exact order e for some $e \mid N$. Therefore Q_η^1 is a $[\Gamma_1(d)]$-structure on E_η, and Q_η^2 is a $[\Gamma_1(e)]$-structure on E_η. They correspond via Lemma 3.19 to the μ_d-torsor
$$P_\eta^1 := \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{d-1} \mathcal{L}((a \cdot Q_\eta^1) - (0_{E_\eta}))\right)$$
and the μ_e-torsor
$$P_\eta^2 := \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{e-1} \mathcal{L}((a \cdot Q_\eta^2) - (0_{E_\eta}))\right),$$
respectively, with the gradings determining the actions of μ_d and μ_e. The μ_N-torsor corresponding to Q_η via Lemma 3.19 is
$$P_\eta := \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{N-1} \mathcal{L}((a \cdot Q_\eta) - (0_{E_\eta}))\right).$$
The group law on E_η tells us that
$$\big((Q_\eta^1) - (0_{E_\eta})\big) + \big((Q_\eta^2) - (0_{E_\eta})\big) \sim (Q_\eta^1 + Q_\eta^2) - (0_{E_\eta}) = (Q_\eta) - (0_{E_\eta}),$$
so we conclude that
$$P_\eta = \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{N-1} \mathcal{L}((a \cdot Q_\eta^1) - (0_{E_\eta})) \otimes \mathcal{L}((a \cdot Q_\eta^2) - (0_{E_\eta}))\right),$$
with the μ_N-action induced by the grading.

Consider the μ_d-torsor $P_\eta^1 \to E_\eta$. As in Lemma 4.11, factoring the isogeny $[d]$ on E_η as $E_\eta \to E_\eta/\langle Q_\eta^1\rangle \to E_\eta$, we have $P_\eta^1 = E_\eta/\langle Q_\eta^1\rangle$ with μ_d acting on P_η^1 through the isomorphism with the group scheme $E_\eta[d]/\langle Q_\eta^1\rangle$ induced via
$$E_\eta[d]/\langle Q_\eta^1\rangle \cong \{Q_\eta^1\} \times E_\eta[d]/\langle Q_\eta^1\rangle \hookrightarrow \langle Q_\eta^1\rangle \times E_\eta[d]/\langle Q_\eta^1\rangle \xrightarrow{e_d} \mu_d.$$

Let $\mathcal{E}_1/R$ (possibly after base change on R) be a generalized elliptic curve extending E_η, whose closed fiber is geometrically a Néron d-gon, with $\mathcal{E}_1^{\mathrm{sm}}(R)[d] \cong E_\eta(K)[d]$.



As in Lemma 4.11, after further base change on R we may assume that the e_d-pairing on E_η extends to a nondegenerate bilinear pairing of finite flat commutative group schemes over R

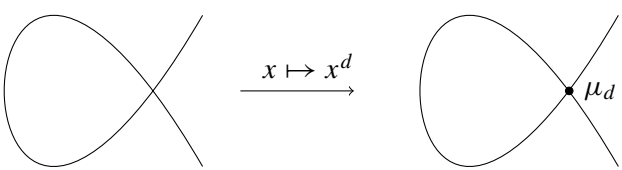
$$e_d : \mathcal{E}_1^{\mathrm{sm}}[d] \times \mathcal{E}_1^{\mathrm{sm}}[d] \to \mu_d.$$

Q_η^1 extends to a $[\Gamma_1(d)]$-structure Q^1 on $\mathcal{E}_1/R$, and the relative effective Cartier divisor

$$D := \sum_{a=0}^{d-1} [a \cdot Q^1]$$

in $\mathcal{E}_1^{\mathrm{sm}}$ is étale over R, so via the same formula as above we see that the isomorphism $E_\eta[d]/\langle Q_\eta^1 \rangle \cong \mu_d$ extends to a group scheme isomorphism $\mathcal{E}_1^{\mathrm{sm}}[d]/\langle Q^1 \rangle \cong \mu_d$. This makes $P^1 := [\mathcal{E}_1/\langle Q^1 \rangle] = \mathcal{E}_1/\langle Q_1 \rangle$ a representable μ_d-torsor over the twisted curve $\mathscr{C} := [\mathcal{E}/\mathcal{E}^{\mathrm{sm}}[d]]$, extending the given μ_d-torsor $P_\eta^1 \to E_\eta$.

The special fiber $\mathscr{C}_s$ is geometrically a standard μ_d-stacky Néron 1-gon, say with coarse space $\pi : \mathscr{C}_s \to C_s$; $[\mathcal{E}_{1,s}/D_s] = \mathcal{E}_{1,s}/D_s$ is geometrically a (nonstacky) Néron 1-gon, and the quotient map $P_s^1 := \mathcal{E}_{1,s}/D_s \to \mathscr{C}_s$ extends the map $P_s^{1,\mathrm{sm}} = \mathbb{G}_m \to \mathscr{C}_s^{\mathrm{sm}} = \mathbb{G}_m$, $x \mapsto x^d$.



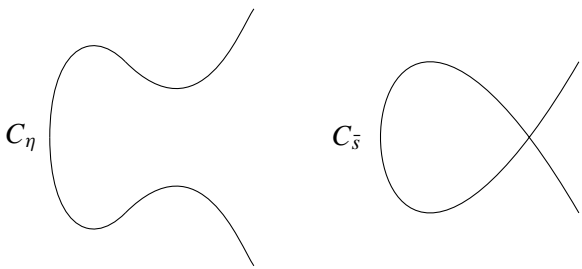
$P_{\bar{s}}^1 = [\mathcal{E}_{1,\bar{s}}/D_{\bar{s}}]$ $\qquad$ $\mathscr{C}_{\bar{s}} = [\mathcal{E}_{1,\bar{s}}/\mathcal{E}_{1,\bar{s}}^{\mathrm{sm}}[d]]$

In particular, if $\bar{q} \to C_s$ is a geometric point mapping to the node of C_s, then $\mathscr{C} \times_C \bar{q} \cong (\mathscr{B}\mu_d)_{k(\bar{q})}$ and $P_s^1 \times_C \bar{q} = \bar{q}$, so the resulting μ_d-torsor over $(\mathscr{B}\mu_d)_{k(\bar{q})}$ corresponds to a generator of $H^1(\mathscr{B}\mu_d, \mu_d) \cong \mathbb{Z}/(d)$. Therefore, with respect to the decomposition

$$\mathrm{Pic}^0_{\mathscr{C}_s/k(s)} \cong \mathrm{Pic}^0_{C_s/k(s)} \times H^0(C_s, \boldsymbol{R}^1\pi_*\mathbb{G}_m) \cong \mathbb{G}_m \times \mathbb{Z}/(d),$$

the class of the μ_d-torsor $P^1_s \to \mathscr{C}_s$ projects in the second factor to a generator of $\mathbb{Z}/(d)$. Therefore the group scheme homomorphism $\mathbb{Z}/(d) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ defined by the μ_d-torsor $P^1 := [\mathcal{E}_1/D] = \mathcal{E}_1/D$ over $\mathscr{C}$ is a $[\Gamma_1(d)]$-structure on the twisted curve $\mathscr{C}/R$.

Next consider the μ_e-torsor $P^2_\eta \to E_\eta$. Let $\pi : \mathscr{C} \to C$ be the coarse space of the twisted curve $\mathscr{C}/R$ described above. By [Deligne and Rapoport 1973, IV.1.6], after further base change on R we may assume that C/R is a generalized elliptic curve, with structure extending that of E_η; note that $C_\eta = E_\eta$ and that C_s is geometrically a Néron 1-gon.



We may take the scheme-theoretic closure of the section $Q^2_\eta \in E_\eta(K)$ to get a unique section $Q^2 \in C^{\mathrm{sm}}(R)$; necessarily $e \cdot Q^2 = 0_C$ since $e \cdot Q^2_\eta = 0_{E_\eta}$. The isomorphisms

$$\mathscr{L}((a \cdot Q^2_\eta) - (0_{E_\eta})) \otimes \mathscr{L}((b \cdot Q^2_\eta) - (0_{E_\eta})) \cong \mathscr{L}(((a+b) \cdot Q^2_\eta) - (0_{E_\eta}))$$

of line bundles on E_η extend uniquely to isomorphisms

$$\mathscr{L}((a \cdot Q^2) - (0_C)) \otimes \mathscr{L}((b \cdot Q^2) - (0_C)) \cong \mathscr{L}(((a+b) \cdot Q^2) - (0_C))$$

of line bundles on C; therefore the μ_e-torsor

$$P^2_\eta = \underline{\mathrm{Spec}}_{E_\eta}\left(\bigoplus_{a=0}^{e-1} \mathscr{L}((a \cdot Q^2_\eta) - (0_{E_\eta}))\right)$$

extends uniquely to a μ_e-torsor

$$P^2 := \underline{\mathrm{Spec}}_C\left(\bigoplus_{a=0}^{e-1} \mathscr{L}((a \cdot Q^2) - (0_C))\right)$$

over C, with the μ_e-action induced by the grading. Since $\mathrm{Pic}^0_{C/R}$ has irreducible geometric fibers, this is a $[\Gamma_1(e)]$-structure on the generalized elliptic curve C/R. Pulling this back to $\mathscr{C}$ via the coarse moduli space map $\pi : \mathscr{C} \to C$, we get a μ_e-torsor $\mathscr{P}^2 \to \mathscr{C}$ extending $P^2_\eta \to E_\eta$, such that the corresponding map $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ lands in the identity component of every geometric fiber.

Finally, we return to the μ_N-torsor $P_\eta \to E_\eta$. Write the μ_d-torsor $P^1 \to \mathscr{C}$ as

$$P^1 = \underline{\mathrm{Spec}}_{\mathscr{C}}\left(\bigoplus_{a=0}^{d-1} \mathscr{L}_a\right),$$

so $\mathscr{L}_a|_\eta = \mathscr{L}((a \cdot Q_\eta^1) - (0_{E_\eta}))$; write the μ_e-torsor $P^2 \to \mathscr{C}$ as

$$P^2 = \underline{\mathrm{Spec}}_{\mathscr{C}}\left(\bigoplus_{a=0}^{e-1} \mathscr{L}'_a\right),$$

so $\mathscr{L}'_a|_\eta = \mathscr{L}((a \cdot Q_\eta^2) - (0_{E_\eta}))$. Consider the μ_N-torsor

$$P := \underline{\mathrm{Spec}}_{\mathscr{C}}\left(\bigoplus_{a=0}^{N-1} \mathscr{L}_{a \bmod d} \otimes \mathscr{L}'_{a \bmod e}\right)$$

over $\mathscr{C}$, with the μ_N-action induced by the grading. Since

$$\mathscr{L}((a \cdot Q_\eta^1) - (0_{E_\eta})) \otimes \mathscr{L}((a \cdot Q_\eta^2) - (0_{E_\eta})) \cong \mathscr{L}((a \cdot Q_\eta) - (0_{E_\eta})),$$

we conclude that $P \to \mathscr{C}$ extends the original μ_N-torsor $P_\eta \to E_\eta$. Furthermore, with respect to the decomposition

$$\mathrm{Pic}^0_{\mathscr{C}_s/k(s)} \cong \mathrm{Pic}^0_{C_s/k(s)} \times H^0(C_s, \boldsymbol{R}^1\pi_*\mathbb{G}_m) \cong \mathbb{G}_m \times \mathbb{Z}/(d),$$

the line bundle $\mathscr{L}_1|_s$ projects to a generator of $\mathbb{Z}/(d)$, and the line bundle $\mathscr{L}'_1|_s$ projects to $0 \in \mathbb{Z}/(d)$; therefore the line bundle $(\mathscr{L}_1 \otimes \mathscr{L}'_1)|_s$ projects to a generator of $\mathbb{Z}/(d)$, so the group scheme homomorphism $\mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/R}$ corresponding to the μ_N-torsor $P \to \mathscr{C}$ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}/R$, extending our original $[\Gamma_1(N)]$-structure $\phi_\eta : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_\eta/K}$. □

This concludes the proof of our final lemma and thus of Theorem 4.6. □

***Reduction* mod *p of* $\overline{\mathscr{K}}'_{1,1}(\mathscr{B}\boldsymbol{\mu}_N)$.** The analysis of $\mathscr{H}_1(N)$ above immediately generalizes to the compactified case. Recall from Notation 3.8 that $\overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N) \subset \overline{\mathscr{K}}_{1,1}(\mathscr{B}\mu_N)$ denotes the closed substack classifying rigidified twisted stable μ_N-covers of twisted curves with nonstacky marking; $\overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N)$ is the closure of $\overline{\mathscr{K}}^\circ_{1,1}(\mathscr{B}\mu_N) \simeq \mathscr{H}_1(N)$ in $\overline{\mathscr{K}}_{1,1}(\mathscr{B}\mu_N)$ by Lemma 3.21.

We have a natural closed immersion

$$\iota^{(d)} : \mathscr{X}_1^{\mathrm{tw}}(d) \hookrightarrow \overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N)$$

for each d dividing N, precomposing a map $\mathbb{Z}/(d) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ with the canonical projection $\mathbb{Z}/(N) \to \mathbb{Z}/(d)$. The resulting map

$$\bigsqcup_{d \mid N} \mathscr{X}_1^{\mathrm{tw}}(d) \to \overline{\mathscr{K}}'_{1,1}(\mathscr{B}\mu_N)$$

is an isomorphism over $S[1/N]$.

Definition 4.13. Let p be prime and $S \in \mathrm{Sch}/\mathbb{F}_p$. Let $\mathscr{C}/S$ be a 1-marked genus-1 twisted stable curve with no stacky structure at its marking. Let $n \geq 1$ and $a, b \geq 0$ with $a + b = n$. A $[\Gamma_1(p^n)]$-(a, b)-*cyclic structure* on $\mathscr{C}$ is a $[\Gamma_1(p^n)]$-structure $\phi : \mathbb{Z}/(p^n) \to \mathrm{Pic}^0_{\mathscr{C}/S}$, such that:

- if $S_1 \subset S$ is the maximal Zariski open subset such that $\mathscr{C}_{S_1} \to S_1$ is smooth, $\phi_{S_1} : \mathbb{Z}/(p^n) \to \mathscr{C}_{S_1}$ is a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure in the sense of [Katz and Mazur 1985], and
- if $S_2 \subset S$ is the complement of the supersingular locus of $\mathscr{C} \to S$, then the relative effective Cartier divisor

$$D := \sum_{m=1}^{p^b} [\phi(m)]$$

in $\mathrm{Pic}^0_{\mathscr{C}_{S_2}/S_2}$ is a subgroup scheme of $\mathrm{Pic}^0_{\mathscr{C}_{S_2}/S_2}$ which is étale over S_2.

Over the base scheme $S \in \mathrm{Sch}/\mathbb{F}_p$, we define $\mathcal{X}_1^{\mathrm{tw}}(p^n)^{(a,b)} \subset \mathcal{X}_1^{\mathrm{tw}}(p^n)$ to be the closed substack associating to T/S the groupoid of pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/S$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and ϕ is a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure on $\mathscr{C}$.

If $N = p^n N'$ with $(N', p) = 1$, we define

$$\mathcal{X}_1^{\mathrm{tw}}(N)^{(a,b)} := \mathcal{X}_1^{\mathrm{tw}}(N') \times_{\overline{\mathcal{M}}_{1,1}} \mathcal{X}_1^{\mathrm{tw}}(p^n)^{(a,b)}.$$

The same argument as that used to prove Lemma 2.11 immediately gives us:

Lemma 4.14. *Let $\mathscr{C}/S/\mathbb{F}_p$ be a 1-marked genus-1 twisted stable curve with nonstacky marking, and let $\phi : \mathbb{Z}/(p^n) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ be a $[\Gamma_1(p^n)]$-(a, b)-cyclic structure on $\mathscr{C}$. Then for the canonical projection $\pi : \mathbb{Z}/(p^{n+1}) \twoheadrightarrow \mathbb{Z}/(p^n)$, the composite $\phi \circ \pi : \mathbb{Z}/(p^{n+1}) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ is a $[\Gamma_1(p^{n+1})]$-structure on $\mathscr{C}$, and is $[\Gamma_1(p^{n+1})]$-$(a + 1, b)$-cyclic.*

If $N = p^n N'$ with $(N', p) = 1$, for any $r \mid N'$ write $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)^r \subset \overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)$ for the union of the components $\mathcal{X}_1^{\mathrm{tw}}(p^m r)$ over $0 \leq m \leq n$.

Corollary 4.15. *Let k be a perfect field of characteristic p, and let $N = p^n N'$ where $(N', p) = 1$. For any $r \mid N'$, $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)^r_k$ is the disjoint union, with crossings at the supersingular points, of components $\mathcal{Z}^r_b$ for $0 \leq b \leq n$, where*

$$\mathcal{Z}^r_b = \bigcup_{b \leq m \leq n} \mathcal{X}_1^{\mathrm{tw}}(p^m r)^{(m-b,b)}_k,$$

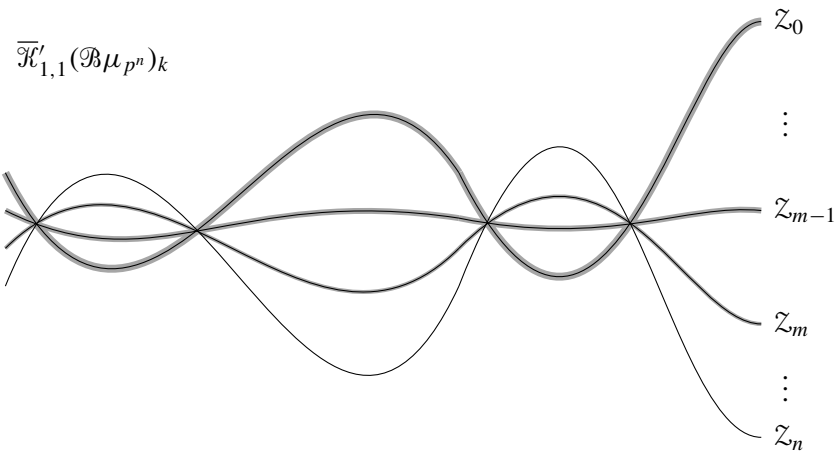
identifying each $\mathcal{X}_1^{\mathrm{tw}}(p^m r)^{(m-b,b)}_k$ with a closed substack of $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)_k$ via $\iota^{(p^m r)}$. Each substack $\mathcal{X}_1^{\mathrm{tw}}(p^m r)^{(m-b,b)}_k$ is "set-theoretically identified with $\mathcal{Z}^r_b$" in the sense

that $(\mathcal{Z}_b^r)_{\mathrm{red}} = \mathcal{X}_1(p^m r)_{k,\mathrm{red}}^{(m-b,b)}$ *as substacks of* $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)_{k,\mathrm{red}}$. $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)_k$ *is the disjoint union of the open and closed substacks*

$$\{\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N)_k^r\}_{r\,|\,N'}.$$

Strictly speaking, to apply the crossings theorem (Theorem 2.7) to get the above corollary, we need to know that the morphism $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N) \to \overline{\mathcal{M}}_{1,1}$ is finite. This follows from Corollary 6.3.

The picture in the case $N = p^n$ is essentially the same as the picture for $\overline{\mathcal{K}}^{\circ}_{1,1}(\mathcal{B}\mu_{p^n})_k$ (as discussed after Proposition 4.1), except now each component is proper:



5. Moduli of elliptic curves in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$

***Reduction* mod *p of* $\mathcal{H}(N)$.** Next we turn our attention to $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$, working as before over an arbitrary base scheme S. Recall that by Corollary 3.20, the open substack $\overline{\mathcal{K}}^{\circ}_{1,1}(\mathcal{B}\mu_N^2)$ classifying rigidified twisted μ_N^2-covers of smooth elliptic curves is naturally equivalent to the stack $\mathcal{H}(N)$ associating to a scheme T/S the groupoid of pairs (E, ϕ) where E/T is an elliptic curve and $\phi : (\mathbb{Z}/(N))^2 \to E[N]$ is a homomorphism of group schemes over T.

For any subgroup $K \leq (\mathbb{Z}/(N))^2$ with corresponding quotient $G_K = (\mathbb{Z}/(N))^2/K$ of $(\mathbb{Z}/(N))^2$, recall (Definition 2.16) that $\mathcal{Y}_K$ denotes the moduli stack associating to a scheme T/S the groupoid of pairs (E, ψ), where E/T is an elliptic curve and $\psi : G_K \to E$ is a G_K-structure (in the sense of [Katz and Mazur 1985, §1.5]). So, for example, if $G_K \cong \mathbb{Z}/(d)$ for some $d\,|\,N$, $\mathcal{Y}_K$ is isomorphic to the stack $\mathcal{Y}_1(d)$, and if $G_K \cong (\mathbb{Z}/(d))^2$ then $\mathcal{Y}_K$ is isomorphic to the stack $\mathcal{Y}(d)$ classifying (not necessarily symplectic) $[\Gamma(d)]$-structures on elliptic curves. For every such K, we

have a closed immersion

$$\iota^K : \mathcal{Y}_K \hookrightarrow \mathcal{H}(N),$$

given by precomposing a G_K-structure $\phi : G_K \to E[N]$ with the canonical projection $(\mathbb{Z}/(N))^2 \to G_K$. Together, these give a proper surjection

$$\bigsqcup_{K \leq (\mathbb{Z}/(N))^2} \mathcal{Y}_K \to \mathcal{H}(N)$$

which is an isomorphism over $S[1/N]$.

But in characteristics dividing N this is not an isomorphism. First we consider the case where $N = p^n$ for some prime p. Any quotient $G_K = (\mathbb{Z}/(p^n))^2/K$ is isomorphic as an abelian group to $\mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$ for some $l \leq m \leq n$. The corresponding moduli stack $\mathcal{Y}_K$ classifies G_K-structures on elliptic curves, and we saw in Theorem 2.17 that over a perfect field k of characteristic p, $\mathcal{Y}_{K,k}$ is the disjoint union, with crossings at the supersingular points, of substacks $\mathcal{Y}^H_{K,k}$ indexed by the set

$$L_K := \{H \leq G_K \mid H \text{ and } G_K/H \text{ are both cyclic}\}.$$

The component $\mathcal{Y}^H_{K,k}$ classifies G_K-structures of component label H.

Now consider two subgroups $K' \leq K \leq (\mathbb{Z}/(p^n))^2$, and write $\pi : G_{K'} \to G_K$ for the canonical surjection. If $\phi : G_K \to E[p^n]$ is a G_K-structure on an ordinary elliptic curve $E/T/k$, then $\phi \circ \pi$ may or may not be a $G_{K'}$-structure on E. Indeed, we saw in Lemma 2.18 that $\phi \circ \pi$ is a $G_{K'}$ structure if and only if $\pi^{-1}(H) \in L_{K'}$, that is, if and only if $\pi^{-1}(H) \subseteq G_{K'}$ is cyclic.

Consider the set $\{(K, H) \mid K \leq (\mathbb{Z}/(p^n))^2, H \in L_K\}$. Let $\sim$ be the equivalence relation on this set generated by requiring that $(K, H) \sim (K', H')$ if $K' \leq K$ and H and H' are as above, and let $\Lambda = \{(K, H)\}/\sim$.

Proposition 5.1. *Let k be a perfect field of characteristic p. $\mathcal{H}(p^n)_k$ is the disjoint union, with crossings at the supersingular points, of components $\mathcal{H}(p^n)^\lambda_k$ for $\lambda \in \Lambda$, where*

$$\mathcal{H}(p^n)^\lambda_k := \bigcup_{[(K,H)]=\lambda} \mathcal{Y}^H_{K,k},$$

identifying each $\mathcal{Y}^H_{K,k}$ with a closed substack of $\mathcal{H}(p^n)^\lambda_k$ via ι^K. Each $\mathcal{Y}^H_{K,k}$ is "set-theoretically identified with $\mathcal{H}(p^n)^\lambda_k$" in the sense that $(\mathcal{H}(p^n)^\lambda_k)_{\mathrm{red}} = \mathcal{Y}^H_{K,k,\mathrm{red}}$ as substacks of $\mathcal{H}(p^n)_{k,\mathrm{red}}$ for all (K, H) with $[(K, H)] = \lambda \in \Lambda$.

If $N = p^n N'$ with $(p, N') = 1$, for any $A \leq (\mathbb{Z}/(N))^2$ of order prime to p, let $\mathcal{H}(N)^A_k \subset \mathcal{H}(N)_k$ be the union of the substacks $\mathcal{Y}_{K,k}$ for $A \leq K \leq (\mathbb{Z}/(N))^2$ with $(K : A)$ a power of p. Then similarly $\mathcal{H}(N)^A_k$ is the disjoint union, with crossings at the supersingular points, of components $\mathcal{H}(N)^{A,\lambda}_k$ for $\lambda \in \Lambda$, and $\mathcal{H}(N)_k$ is the

disjoint union of the open and closed substacks $\mathcal{H}(N)_k^A$ *for* $A \leq (\mathbb{Z}/(N))^2$ *of order prime to* p.

As in the case of $\mathcal{H}_1(p^n)$, the reduction mod p of $\mathcal{H}(p^n)$ has an appealing geometric description. To keep our pictures from getting unreasonably large, we restrict our attention to the case $n = 1$. The group $K := (\mathbb{Z}/(p))^2$ has $p+3$ subgroups, namely the entire group K, $K_0 := 0$, and $p+1$ subgroups $K_1, \ldots, K_{p+1}$ isomorphic to $\mathbb{Z}/(p)$. The corresponding moduli stacks are

$$\begin{aligned} \mathcal{Y}_K &= \mathcal{Y}(1), \\ \mathcal{Y}_{K_0} &= \mathcal{Y}(p), \\ \mathcal{Y}_{K_i} &\cong \mathcal{Y}_1(p) \quad \text{for } i = 1, \ldots, p+1, \end{aligned}$$

so we see that over $\mathbb{Z}[1/p]$, $\mathcal{H}(p)$ is the disjoint union of $\mathcal{Y}(1)$, $\mathcal{Y}(p)$, and $p+1$ copies of $\mathcal{Y}_1(p)$.

By definition we have

$$\begin{aligned} L_K &= \{0\}, \\ L_{K_0} &= \{K_1, \ldots, K_{p+1}\}, \\ L_{K_i} &= \{G_{K_i}, 0\} \quad \text{for } i = 1, \ldots, p+1, \end{aligned}$$

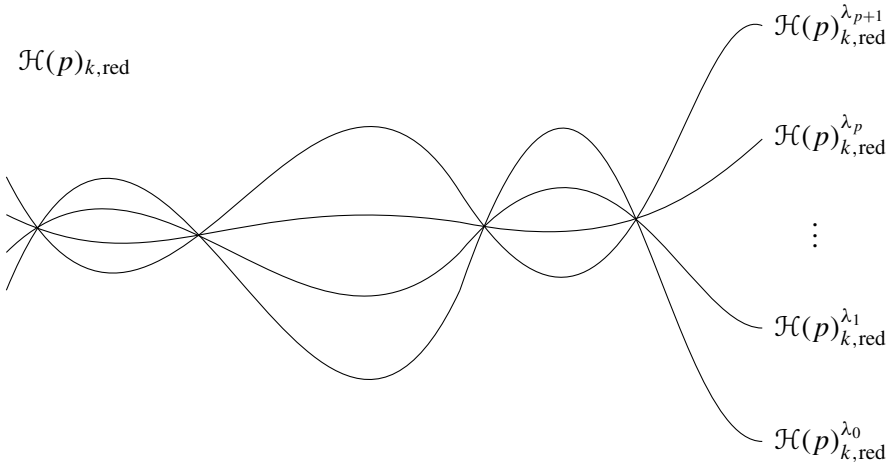
where as usual $G_{K_i} = (\mathbb{Z}/(p))^2/K_i$. The set of labels Λ is built by putting an equivalence relation on the set consisting of the following pairs:

$$\begin{aligned} &(K, 0), \\ &(K_0, K_i) \quad \text{for } i = 1, \ldots, p+1, \\ &(K_i, G_{K_i}) \quad \text{for } i = 1, \ldots, p+1, \\ &(K_i, 0) \quad \text{for } i = 1, \ldots, p+1. \end{aligned}$$

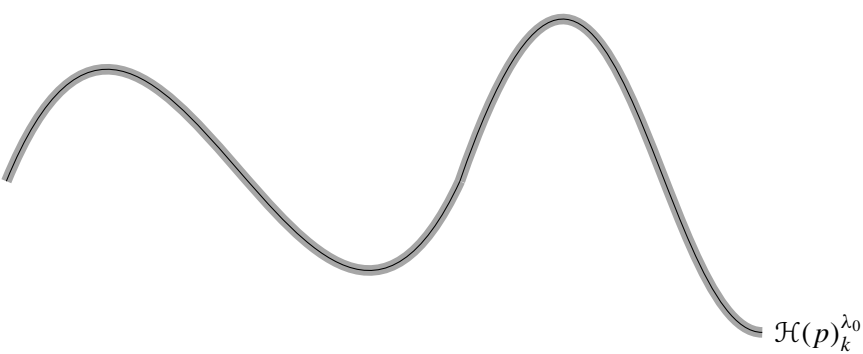
By definition, working over a perfect field k of characteristic p, the pair $(K, 0)$ corresponds to $\mathcal{Y}(1)_k$; the pair (K_0, K_i) corresponds to the component $\mathcal{Y}(p)_k^{K_i}$ of $\mathcal{Y}(p)_k$; the pair (K_i, G_{K_i}) corresponds to the component $\mathcal{Y}_1(p)_k^{(1,0)}$ in $\mathcal{Y}_{K_i,k} \cong \mathcal{Y}_1(p)_k$; and the pair $(K_i, 0)$ corresponds to the component $\mathcal{Y}_1(p)_k^{(0,1)}$ in $\mathcal{Y}_{K_i,k} \cong \mathcal{Y}_1(p)_k$. Unwinding the definition, we see that the equivalence relation defining Λ just says that $(K, 0) \sim (K_i, G_{K_i})$ for $i = 1, \ldots, p+1$, and that $(K_0, K_i) \sim (K_i, 0)$ for $i = 1, \ldots, p+1$. Thus the set of labels is $\Lambda = \{\lambda_0, \lambda_1, \ldots, \lambda_{p+1}\}$, where

$$\begin{aligned} \lambda_0 &= [(K, 0)] = [(K_i, G_{K_i})] \quad \text{for } i = 1, \ldots, p+1, \\ \lambda_i &= [(K_0, K_i)] = [(K_i, 0)] \quad \text{for } i = 1, \ldots, p+1. \end{aligned}$$

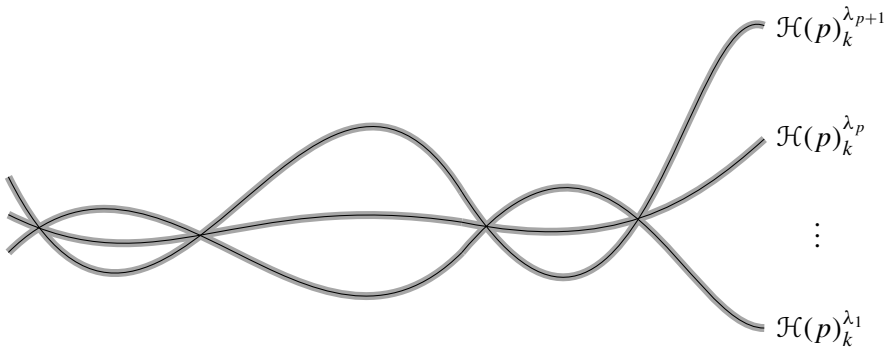
Visualize $\mathcal{H}(p)_{k,\mathrm{red}}$ as follows:



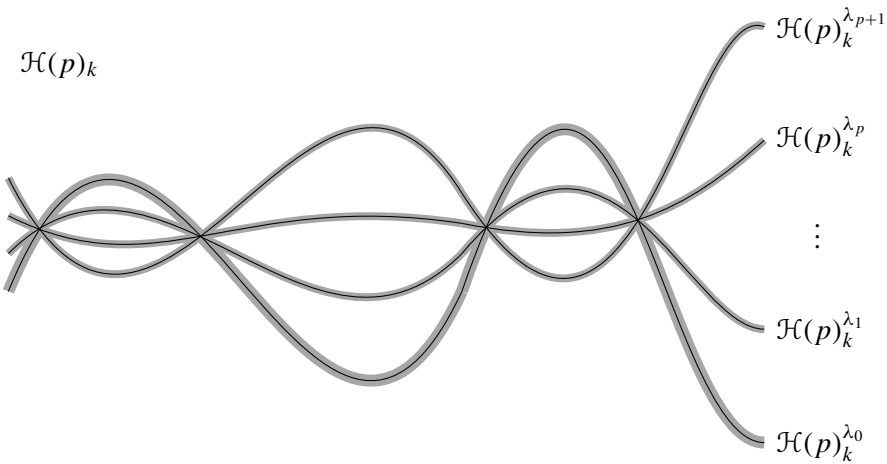
The component $\mathcal{H}(p)_k^{\lambda_0}$ is "set-theoretically identified" with the component $\mathcal{Y}(1)_k$ and with the component $\mathcal{Y}_1(p)_k^{(1,0)}$ in each copy of $\mathcal{Y}_1(p)_k$; each of these contributes additional nilpotent structure to the component $\mathcal{H}(p)_k^{\lambda_0}$ (each $\mathcal{Y}_1(p)_k^{(1,0)}$ has length $p-1$ over $\mathcal{H}(p)_{k,\mathrm{red}}^{\lambda_0} = \mathcal{Y}(1)_k$, so $\mathcal{H}(p)_k^{\lambda_0}$ has length $(p+1)(p-1)+1 = p^2$ over $\mathcal{H}(p)_{k,\mathrm{red}}^{\lambda_0} = \mathcal{Y}(1)_k$).



For $i = 1, \dots, p+1$, the component $\mathcal{H}(p)_k^{\lambda_i}$ is "set-theoretically identified" with the component $\mathcal{Y}(p)_k^{K_i}$ of $\mathcal{Y}(p)_k$, contributing nilpotent structure to $\mathcal{H}(p)_k^{\lambda_i}$ (each $\mathcal{Y}(p)_k^{K_i}$ has length $p-1$ over $\mathcal{H}(p)_{k,\mathrm{red}}^{\lambda_i}$). The component $\mathcal{H}(p)_k^{\lambda_i}$ is also "set-theoretically identified" with the component $\mathcal{Y}_1(p)_k^{(0,1)}$ of $\mathcal{Y}_{K_i,k} \cong \mathcal{Y}_1(p)_k$; each of these is reduced, adding 1 to the length of the component $\mathcal{H}(p)_k^{\lambda_i}$ over $\mathcal{H}(p)_{k,\mathrm{red}}^{\lambda_i}$. Thus $\mathcal{H}(p)_k^{\lambda_i}$ has length p over $\mathcal{H}(p)_{k,\mathrm{red}}^{\lambda_i}$, which is isomorphic to $\mathcal{Y}_1(p)_k^{(0,1)}$ and hence has degree $p^2 - p$ over $\mathcal{Y}(1)_k$ (see [Katz and Mazur 1985, 13.5.6]). The result is that the component $\mathcal{H}(p)_k^{\lambda_i}$ has length p over the underlying reduced stack $\mathcal{Y}_1(p)_k^{(0,1)}$, which has degree $p^2 - p$ over $\mathcal{Y}(1)_k$:



This gives us the following picture of $\mathcal{H}(p)_k$:



Note that adding up the lengths calculated in the course of the above construction, we recover the fact that the stack $\mathcal{H}(p)$ has length p^4 over $\mathcal{Y}(1) = \mathcal{M}_{1,1}$.

Closure of $\mathcal{Y}(N)$ in $\overline{\mathcal{H}}_{1,1}(\mathcal{B}\boldsymbol{\mu}_N^2)$.

Definition 5.2. Let $\mathscr{C}/S$ be a 1-marked genus-1 twisted stable curve over a scheme S, with no stacky structure at its marking. A $[\Gamma(N)]$-*structure* on $\mathscr{C}$ is a group scheme homomorphism $\phi : (\mathbb{Z}/(N))^2 \to \mathrm{Pic}^0_{\mathscr{C}/S}$ such that:

- the relative effective Cartier divisor

$$D := \sum_{a \in (\mathbb{Z}/(N))^2} [\phi(a)]$$

  in $\mathrm{Pic}^0_{\mathscr{C}/S}$ is an N-torsion subgroup scheme, hence $D = \mathrm{Pic}^0_{\mathscr{C}/S}[N]$, and
- for every geometric point $\bar{p} \to S$, $D_{\bar{p}}$ meets every irreducible component of $(\mathrm{Pic}^0_{\mathscr{C}/S})_{\bar{p}} = \mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})}$.

We write $\mathcal{X}^{\mathrm{tw}}(N)$ for the substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$ associating to T/S the groupoid of pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/S$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and ϕ is a $[\Gamma(N)]$-structure on $\mathscr{C}$.

If $\mathscr{C}/S$ is a twisted curve admitting a $[\Gamma(N)]$-structure and $\bar{p} \to S$ is a geometric point such that $C_{\bar{p}}$ is singular, then necessarily

$$\mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})} \cong \mathbb{G}_m \times \mathbb{Z}/(N),$$

so by Lemma 3.17 $\mathscr{C}_{\bar{p}}$ is a standard μ_N-stacky Néron 1-gon over $k(\bar{p})$, as in Example 3.4.

Applying the methods of our study of $\mathcal{X}_1^{\mathrm{tw}}(N)$ to the stack $\mathcal{X}^{\mathrm{tw}}(N)$, we have:

Theorem 5.3. *Let S be a scheme and let $\mathcal{X}^{\mathrm{tw}}(N)$ be the stack over S classifying $[\Gamma(N)]$-structures on 1-marked genus-1 twisted stable curves with nonstacky marking. Then $\mathcal{X}^{\mathrm{tw}}(N)$ is a closed substack of $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$, which contains $\mathcal{Y}(N)$ as an open dense substack.*

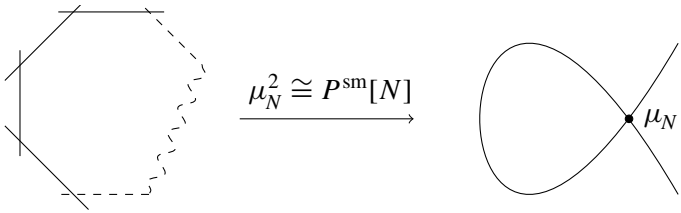
In particular $\mathcal{X}^{\mathrm{tw}}(N)$ is flat over S with local complete intersection fibers, and is proper and quasifinite over $\overline{\mathcal{M}}_{1,1}$.

Remark 5.4. In [Petersen 2012] a direct proof is given that over $\mathbb{Z}[1/N]$, $\mathcal{X}^{\mathrm{tw}}(N)$ agrees with the stack $\mathcal{X}(N)$ classifying $[\Gamma(N)]$-structures on generalized elliptic curves; in this case a μ_N^2-torsor over a 1-marked genus-1 twisted stable curve with nonstacky marking is in fact a generalized elliptic curve, and it is this observation that gives the desired equivalence. This argument does not generalize to characteristics dividing N, because, for example, if $N = p^n$ then the Néron N-gon in characteristic p (which is a generalized elliptic curve admitting various $[\Gamma(N)]$-structures) cannot be realized as a μ_N^2-torsor over a 1-marked genus-1 twisted stable curve.

Proof of Theorem 5.3. This is proved in exactly the same manner as Theorem 4.6. An immediate consequence of Lemma 3.22 is that if E/K is an elliptic curve over a field K in which N is invertible, and (Q_1, Q_2) is a $[\Gamma(N)]$-structure on E, then the μ_N^2-torsor

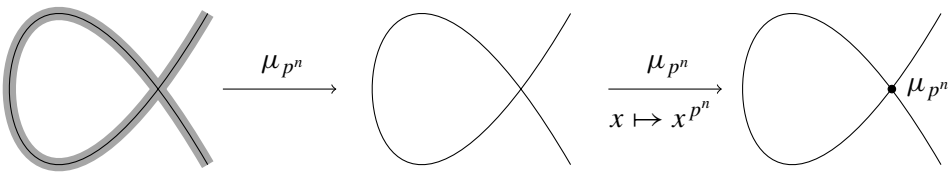
$$P = \underline{\mathrm{Spec}}_E\Big(\bigoplus_{a,b\in\mathbb{Z}/(N)} \mathscr{L}((a\cdot Q_1 + b\cdot Q_2) - (0_E))\Big)$$

over E as in Corollary 3.20 may be identified with E itself, with the quotient map $P \to E$ corresponding to the isogeny $[N] : E \to E$. We immediately deduce via the methods of Theorem 4.6 that over an algebraically closed field k in which N is invertible, the μ_N^2-torsor obtained when we pass to the cusp of $\overline{\mathcal{M}}_{1,1}$ is a Néron N-gon P over a standard μ_N-stacky Néron 1-gon $\mathscr{C}$, with the μ_N^2-action on P induced by some choice of isomorphism $P^{\mathrm{sm}}[N] \cong \mu_N^2$:



P (N-gon) $\qquad$ $\mathscr{C} = [P/P^{\mathrm{sm}}[N]]$

And in the case of $N = p^n$, over an algebraically closed k field of characteristic p, the $\mu_{p^n}^2$-torsor P obtained in passing to the cusp of $\overline{\mathcal{M}}_{1,1}$ may be realized as a trivial μ_{p^n}-torsor over a standard Néron 1-gon C', which in turn is a μ_{p^n}-torsor over a standard μ_{p^n}-stacky Néron 1-gon $\mathscr{C} = [C'/\mu_{p^n}]$ via the choice of an isomorphism $(C')^{\mathrm{sm}}[p^n] \cong \mu_{p^n}$:



$P = C' \times \mu_{p^n}$ $\qquad$ C' $\qquad$ $\mathscr{C}$

In both of the above cases, it is immediately verified that for each of these μ_N^2-torsors, the corresponding group scheme homomorphism $(\mathbb{Z}/(N))^2 \to \mathrm{Pic}^0_{\mathscr{C}/k} \cong \mathbb{G}_m \times \mathbb{Z}/(N)$ is a $[\Gamma(N)]$-structure on the standard μ_N-stacky Néron 1-gon $\mathscr{C}$ in the sense of Definition 5.2, giving the valuative criterion of properness for $\mathcal{X}^{\mathrm{tw}}(N)$, hence Theorem 5.3. $\square$

***Reduction* mod *p* of $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\boldsymbol{\mu}_N^2)$.** Recall that $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N^2) \subset \overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$ is the closed substack classifying rigidified twisted stable μ_N^2-covers of twisted curves with nonstacky marking; so $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_N^2)$ is the closure of

$$\overline{\mathcal{K}}^{\circ}_{1,1}(\mathcal{B}\mu_N^2) \simeq \mathcal{H}(N)$$

in $\overline{\mathcal{K}}_{1,1}(\mathcal{B}\mu_N^2)$.

Definition 5.5. Let $\mathscr{C}/S$ be a 1-marked genus-1 twisted stable curve with nonstacky marking, and let G be a 2-generated finite abelian group, say $G \cong \mathbb{Z}/(n_1) \times \mathbb{Z}/(n_2)$, $n_1 \geq n_2$. A *G-structure* on $\mathscr{C}$ is a homomorphism $\phi : G \to \mathrm{Pic}^0_{\mathscr{C}/S}$ of group schemes over S such that:

- the relative effective Cartier divisor

$$D := \sum_{a \in G} [\phi(a)]$$

in $\mathrm{Pic}^0_{\mathscr{C}/S}$ is an n_1-torsion subgroup scheme and

- for every geometric point $\bar{p} \to S$, $D_{\bar{p}}$ meets every irreducible component of $(\mathrm{Pic}^0_{\mathscr{C}/S})_{\bar{p}} = \mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})}$.

For any subgroup $K \leq (\mathbb{Z}/(N))^2$ with corresponding quotient $G_K = (\mathbb{Z}/(N))^2/K$ of $(\mathbb{Z}/(N))^2$, we write $\mathfrak{X}^{\mathrm{tw}}_K$ for the moduli stack over S associating to a scheme T/S the groupoid of pairs $(\mathscr{C}, \psi)$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and $\psi : G_K \to \mathrm{Pic}^0_{\mathscr{C}/T}$ is a G_K-structure on $\mathrm{Pic}^0_{\mathscr{C}/T}$.

For a twisted curve $\mathscr{C}/S/\mathbb{F}_p$, if $N = p^n$ and $G_K \cong \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$ with $m \geq l \geq 1$, we set

$$L_K := \{H \leq G_K \mid H \text{ and } G_K/H \text{ are both cyclic}\}.$$

For any $H \in L_K$ we say a G_K-structure $\phi : G_K \to \mathrm{Pic}^0_{\mathscr{C}/S}$ has *component label* H if H maps to the kernel of the n-fold relative Frobenius F^n on the group scheme $\mathrm{Pic}^0_{\mathscr{C}/S}$ over S, and the resulting group scheme homomorphism $G_K/H \to \mathrm{Pic}^0_{\mathscr{C}/S}[p^n]/\ker(F^n)$ is a G_K/H-structure in the sense of [Katz and Mazur 1985, §1.5].

If $G_K \cong \mathbb{Z}/(p^m)$ (that is, $l = 0$), then $\mathfrak{X}^{\mathrm{tw}}_K \cong \mathfrak{X}^{\mathrm{tw}}_1(p^m)$, and for $H \cong \mathbb{Z}/(p^a) \in L_K$ we define $\mathfrak{X}^{\mathrm{tw},H}_K \subset \mathfrak{X}^{\mathrm{tw}}_K$ to be the substack $\mathfrak{X}^{\mathrm{tw}}_1(p^m)^{(a,m-a)} \subset \mathfrak{X}^{\mathrm{tw}}_1(p^m)$ as in Definition 4.13. We still say that $\mathfrak{X}^{\mathrm{tw},H}_K$ classifies G_K-structures of *component label* H.

So, for example, if $G_K \cong \mathbb{Z}/(d)$ for some $d \mid N$, $\mathfrak{X}^{\mathrm{tw}}_K$ is isomorphic to the stack $\mathfrak{X}^{\mathrm{tw}}_1(d)$, and if $G_K \cong (\mathbb{Z}/(d))^2$ then $\mathfrak{X}^{\mathrm{tw}}_K$ is isomorphic to the stack $\mathfrak{X}^{\mathrm{tw}}(d)$. For every such K, we have a closed immersion

$$\iota^K : \mathfrak{X}^{\mathrm{tw}}_K \hookrightarrow \overline{\mathfrak{K}}'_{1,1}(\mathcal{B}\mu^2_N),$$

given by precomposing a G_K-structure $\phi : G_K \to \mathrm{Pic}^0_{\mathscr{C}/T}$ with the canonical projection $(\mathbb{Z}/(N))^2 \to G_K$. Together, these give a proper surjection

$$\coprod_{K \leq (\mathbb{Z}/(N))^2} \mathfrak{X}^{\mathrm{tw}}_K \to \overline{\mathfrak{K}}'_{1,1}(\mathcal{B}\mu^2_N)$$

which is an isomorphism over $S[1/N]$.

Let k be a perfect field of characteristic p. The same argument that proves Lemma 2.18 immediately gives us:

Corollary 5.6. *If $K \leq (\mathbb{Z}/(p^n))^2$ such that $G_K := (\mathbb{Z}/(p^n))^2/K \cong \mathbb{Z}/(p^m) \times \mathbb{Z}/(p^l)$ with $l \leq m \leq n$, then $\mathfrak{X}^{\mathrm{tw}}_{K,k}$ is the disjoint union, with crossings at the supersingular points, of closed substacks $\mathfrak{X}^{\mathrm{tw},H}_{K,k}$ for $H \in L_K$. $\mathfrak{X}^{\mathrm{tw},H}_{K,k}$ classifies G_K-structures with component label H.*

As before, we let Λ denote the set $\{(K, H) \mid K \leq (\mathbb{Z}/(p^n))^2, H \in L_K\}$, modulo the equivalence relation generated by declaring $(K, H) \sim (K', \pi^{-1}(H))$ whenever

$K' \leq K$ with corresponding quotient map $\pi : G_{K'} \to G_K$ such that $\pi^{-1}(H) \in L_{K'}$. We conclude:

Corollary 5.7. *Let k be a perfect field of characteristic p. $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_{p^n})_k$ is the disjoint union, with crossings at the supersingular points, of components $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_{p^n})^\lambda_k$ for $\lambda \in \Lambda$, where*

$$\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_{p^n})^\lambda_k := \bigcup_{[(K,H)]=\lambda} \mathcal{X}^{\mathrm{tw},H}_{K,k},$$

identifying each $\mathcal{X}^{\mathrm{tw},H}_{K,k}$ with a closed substack of $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_{p^n})^\lambda_k$ via ι^K. Each $\mathcal{X}^{\mathrm{tw},H}_{K,k}$ is "set-theoretically identified with $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_{p^n})^\lambda_k$" in the sense that

$$\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_{p^n})^\lambda_{k,\mathrm{red}} = \mathcal{X}^{\mathrm{tw},H}_{K,k,\mathrm{red}}$$

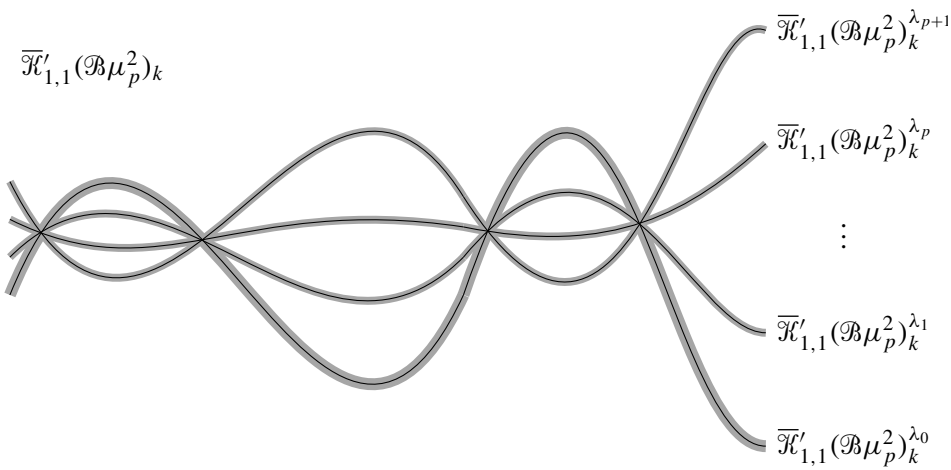
as substacks of $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu_{p^n})_{k,\mathrm{red}}$ for all (K, H) with $[(K, H)] = \lambda \in \Lambda$.

If $N = p^n N'$ with $(p, N') = 1$, for any $A \leq (\mathbb{Z}/(N))^2$ of order prime to p, let

$$\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_N)^A_k \subset \overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_N)_k$$

be the union of the substacks $\mathcal{X}^{\mathrm{tw}}_{K,k}$ for $A \leq K \leq (\mathbb{Z}/(N))^2$ with $(K : A)$ a power of p. Then similarly $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_N)^A_k$ is the disjoint union, with crossings at the supersingular points, of components $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_N)^{A,\lambda}_k$ for $\lambda \in \Lambda$, and $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_N)_k$ is the disjoint union of the open and closed substacks $\overline{\mathcal{K}}'_{1,1}(\mathcal{B}\mu^2_N)^A_k$ for $A \leq (\mathbb{Z}/(N))^2$ of order prime to p.

The picture in the case $N = p$ is essentially the same as the picture for $\overline{\mathcal{K}}^\circ_{1,1}(\mathcal{B}\mu^2_p)_k \simeq \mathcal{H}(p)_k$ (as discussed after Proposition 5.1), except now each component is proper:



6. Comparison with the classical moduli stacks

As promised, we verify that the moduli stacks $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are isomorphic to the corresponding classical moduli stacks, justifying the claim in [Abramovich et al. 2011] that we have recovered the Katz–Mazur regular models:

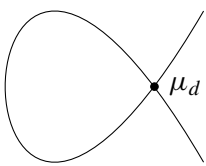
Theorem 6.1 (Restatement of Theorem 1.2). *Over the base $S = \mathrm{Spec}(\mathbb{Z})$, there is a canonical isomorphism of algebraic stacks $\mathcal{X}_1^{\mathrm{tw}}(N) \cong \mathcal{X}_1(N)$ extending the identity map on $\mathcal{Y}_1(N)$, and a canonical isomorphism of algebraic stacks $\mathcal{X}^{\mathrm{tw}}(N) \cong \mathcal{X}(N)$ extending the identity map on $\mathcal{Y}(N)$.*

We prove this after some preliminary results; the main point is to demonstrate that $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are normal.

Proposition 6.2. *The morphism $\pi : \mathcal{X}_1^{\mathrm{tw}}(N) \to \overline{\mathcal{M}}_{1,1}$ sending $(\mathscr{C}, \phi)$ to the coarse space C of $\mathscr{C}$ is a representable morphism of stacks. In particular, $\mathcal{X}_1^{\mathrm{tw}}(N)$ is Deligne–Mumford.*

Similarly the natural morphism $\mathcal{X}^{\mathrm{tw}}(N) \to \overline{\mathcal{M}}_{1,1}$ is representable, hence $\mathcal{X}^{\mathrm{tw}}(N)$ is Deligne–Mumford.

Proof. We have already seen that $\mathcal{X}_1^{\mathrm{tw}}(N)$ is an algebraic stack, so it suffices to show that for any object $(\mathscr{C}, \phi) \in \mathcal{X}_1^{\mathrm{tw}}(N)(k)$ with k an algebraically closed field, the natural map $\mathrm{Aut}(\mathscr{C}, \phi) \to \mathrm{Aut}(C)$ is a monomorphism of group schemes. Here C/k is the coarse space of $\mathscr{C}$, and automorphisms are required to preserve the marking. It is obvious that $\mathrm{Aut}(\mathscr{C}, \phi) \to \mathrm{Aut}(C)$ is a monomorphism if $\mathscr{C} = C$ is a smooth elliptic curve over k, so by Lemma 3.17 we reduce to the case where $\mathscr{C}/k$ is a standard μ_d-stacky Néron 1-gon for some $d \mid N$.



In this case an automorphism of $\mathscr{C}$ is an automorphism of the coarse space C, together with an automorphism of the μ_d-gerbe in $\mathscr{C}$ lying over the node of C. Thus

$$\mathrm{Aut}(\mathscr{C}) \cong \mathrm{Aut}(C) \times \mathrm{Aut}(\mathscr{B}\mu_{d,k}).$$

The only nontrivial automorphism of C preserving the marked point $1 \in C$ is the automorphism $\iota : C \to C$ induced by the inversion automorphism of $\mathbb{G}_m$. We have $\mathrm{Aut}(\mathscr{B}\mu_d) \cong \mu_d$, and the automorphism of $\mathrm{Pic}^0_{\mathscr{C}/k} \cong \mathbb{G}_m \times \mathbb{Z}/(d)$ induced by $(0, \zeta) \in \mathrm{Aut}(\mathscr{C}) \cong \langle \iota \rangle \times \mu_d$ sends (η, a) to $(\zeta^a \eta, a)$. Since $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/k}$ meets every component, the only automorphisms of $\mathscr{C}$ than can possibly preserve ϕ are the automorphisms $\langle \iota \rangle \times \{0\} \subset \mathrm{Aut}(\mathscr{C})$ (see [Conrad 2007, proof of 3.1.8]). Thus $\mathrm{Aut}(\mathscr{C}, \phi) \subset \langle \iota \rangle \times \{0\} \cong \mathrm{Aut}(C)$.

The same argument applies to $\mathcal{X}^{\mathrm{tw}}(N)$. □

Corollary 6.3. *$\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are finite over $\overline{\mathcal{M}}_{1,1}$.*

Proof. By Theorems 4.6 and 5.3, the natural maps $\mathcal{X}_1^{\mathrm{tw}}(N) \to \overline{\mathcal{M}}_{1,1}$ and $\mathcal{X}^{\mathrm{tw}}(N) \to \overline{\mathcal{M}}_{1,1}$ are proper and quasifinite; since they are also representable, they are finite. □

Let $\infty \hookrightarrow \overline{\mathcal{M}}_{1,1}$ denote the closed substack classifying 1-marked genus-1 stable curves whose geometric fibers are singular. Let $\mathcal{X}_1^{\mathrm{tw}}(N)^\infty = \mathcal{X}_1^{\mathrm{tw}}(N) \times_{\overline{\mathcal{M}}_{1,1}} \infty$ and $\mathcal{X}^{\mathrm{tw}}(N)^\infty = \mathcal{X}^{\mathrm{tw}}(N) \times_{\overline{\mathcal{M}}_{1,1}} \infty$. Exactly analogously to [Conrad 2007, 2.1.12], formation of these closed substacks is compatible with arbitrary base change.

Proposition 6.4. *The proper flat morphisms $\mathcal{X}_1^{\mathrm{tw}}(N) \to \mathrm{Spec}(\mathbb{Z})$ and $\mathcal{X}^{\mathrm{tw}}(N) \to \mathrm{Spec}(\mathbb{Z})$ are Cohen–Macaulay (of pure relative dimension 1).*

Proof. Let $\mathcal{X}$ denote $\mathcal{X}_1^{\mathrm{tw}}(N)$ or $\mathcal{X}^{\mathrm{tw}}(N)$. The canonical morphism $\mathcal{X} \to \overline{\mathcal{M}}_{1,1}$ is finite (by Corollary 6.3) and flat (by Theorems 4.6 and 5.3), and the structural morphism $\overline{\mathcal{M}}_{1,1} \to \mathrm{Spec}(\mathbb{Z})$ is Cohen–Macaulay (see [Conrad 2007, 3.3.1]), so by [Bourbaki 1998, 2.7.9, Corollary 3], the composite $\mathcal{X} \to \mathrm{Spec}(\mathbb{Z})$ is Cohen–Macaulay. □

Lemma 6.5. *$\mathcal{X}_1^{\mathrm{tw}}(N)^\infty$ and $\mathcal{X}^{\mathrm{tw}}(N)^\infty$ are relative effective Cartier divisors over $\mathrm{Spec}(\mathbb{Z})$ in $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$, respectively.*

Proof. Here we are using the notion of a Cartier divisor on a Deligne–Mumford stack; see [Arbarello et al. 2011, Chapter XIII]. For $\mathcal{X} = \mathcal{X}_1^{\mathrm{tw}}(N)$ or $\mathcal{X} = \mathcal{X}^{\mathrm{tw}}(N)$, the closed substack $\mathcal{X}^\infty$ is the pullback $\mathcal{X} \times_{\overline{\mathcal{M}}_{1,1}} \infty$. We know $\infty \subset \overline{\mathcal{M}}_{1,1}$ is a relative effective Cartier divisor over $\mathrm{Spec}(\mathbb{Z})$ (meaning an effective Cartier divisor which is flat over $\mathrm{Spec}(\mathbb{Z})$), and by Theorems 4.6 and 5.3 the morphism $\mathcal{X} \to \overline{\mathcal{M}}_{1,1}$ is flat. Cartier divisors are preserved by flat morphisms (see [Fulton 1998, §1.7]), so $\mathcal{X}^\infty$ is an effective Cartier divisor in $\mathcal{X}$. Since $\mathcal{X} \to \overline{\mathcal{M}}_{1,1}$ is flat, so is $\mathcal{X}^\infty \to \infty$, so $\mathcal{X}^\infty$ is flat over $\mathrm{Spec}(\mathbb{Z})$, that is, $\mathcal{X}^\infty$ is a relative effective Cartier divisor in $\mathcal{X}$ over $\mathrm{Spec}(\mathbb{Z})$. □

Corollary 6.6. *$\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are normal.*

Proof. This is proven in an identical manner to [Conrad 2007, 4.1.4]. The stacks $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are Deligne–Mumford, and from [Abramovich et al. 2003, 3.0.2] we know $\mathcal{X}_1^{\mathrm{tw}}(N) \otimes_{\mathbb{Z}} \mathbb{Z}[1/N]$ and $\mathcal{X}^{\mathrm{tw}}(N) \otimes_{\mathbb{Z}} \mathbb{Z}[1/N]$ are smooth over $\mathrm{Spec}(\mathbb{Z}[1/N])$. In particular, they are regular at any characteristic-0 points. Furthermore, by Proposition 6.4 they are Cohen–Macaulay over $\mathrm{Spec}(\mathbb{Z})$ of pure relative dimension 1. As in [Conrad 2007, 4.1.4], we can conclude from Serre's criterion for normality that it suffices to prove that these stacks are regular away from some relative effective Cartier divisor, since such a divisor cannot contain any codimension-1 points of positive residue characteristic. Use the divisors $\mathcal{X}_1^{\mathrm{tw}}(N)^\infty$ and $\mathcal{X}^{\mathrm{tw}}(N)^\infty$; their complements are $\mathcal{Y}_1(N)$ and $\mathcal{Y}(N)$, which are regular by [Katz and Mazur 1985, 5.1.1]. □

Proof of Theorem 6.1. $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ are finite, flat, and normal over $\overline{\mathcal{M}}_{1,1}$, so they are naturally identified with the normalizations (in the sense of [Deligne and Rapoport 1973, IV.3.3]) of $\overline{\mathcal{M}}_{1,1}$ in $\mathcal{X}_1^{\mathrm{tw}}(N)|_{\mathcal{M}_{1,1}} = \mathcal{Y}_1(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)|_{\mathcal{M}_{1,1}} = \mathcal{Y}(N)$, respectively; see [Conrad 2007, 4.1.5]. □

We now give a moduli interpretation of the equivalence $\mathcal{X}_1(N) \simeq \mathcal{X}_1^{\mathrm{tw}}(N)$. Let S be a scheme, E/S be a generalized elliptic curve, and $P \in E^{\mathrm{sm}}(S)[N]$ be a $[\Gamma_1(N)]$-structure on E. From this data we want to construct a pair $(\mathscr{C}_P, \phi_P)$, where $\mathscr{C}_P/S$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and $\phi_P : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_P/S}$ is a $[\Gamma_1(N)]$-structure on $\mathscr{C}_P$.

If E/S is a smooth elliptic curve, there is nothing to show: we simply take $(\mathscr{C}_P, \phi_P) = (E, \phi_P)$ where $\phi_P : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{E/S} \cong E$ sends $1 \mapsto P$. Therefore to construct $(\mathscr{C}_P, \phi_P)$ in general, we may restrict to the open subscheme of S where E/S has no supersingular geometric fibers; once we have constructed $(\mathscr{C}_P, \phi_P)$ in this case, we only need to check that it agrees with our previous construction for ordinary elliptic curves.

For the rest of the construction, assume that E/S has no supersingular geometric fibers.

Note that by [Conrad 2007, 4.2.3], fppf locally on S there exists a generalized elliptic curve E'/S, whose singular geometric fibers are N-gons, together with an open S-immersion $\iota : E^{\mathrm{sm}} \hookrightarrow E'^{\mathrm{sm}}$ of group schemes over S. In particular, by [Deligne and Rapoport 1973, II.1.20] the group scheme $E'^{\mathrm{sm}}[N]/S$ is finite and flat of constant rank N^2.

Since E'/S has no supersingular geometric fibers and all its singular geometric fibers are N-gons, it follows that fppf locally on S, there exists a $[\Gamma(N)]$-structure (Q, R) on E' such that:

- the relative effective Cartier divisor
$$D := \sum_{a \in \mathbb{Z}/(N)} [a \cdot Q]$$
in E'^{sm} is étale over S and
- R meets the identity component of every geometric fiber of E'/S.

The choice of Q and the pairing
$$e_N : E'^{\mathrm{sm}}[N] \times E'^{\mathrm{sm}}[N] \to \mu_N$$
induce a canonical isomorphism $E'^{\mathrm{sm}}[N]/D \cong E'^{\mathrm{sm}}[N]/D \times \{Q\} \cong \mu_N$. Identifying $E'^{\mathrm{sm}}[N]/D$ with its image in the N-torsion of the generalized elliptic curve $C := E'/D$ (a generalized elliptic curve whose singular fibers are 1-gons), the group law of C and the above isomorphism give us an action of μ_N on C, making C a

μ_N-torsor over the twisted curve $\mathscr{C} := [C/\mu_N] = [E'/E'^{\mathrm{sm}}[N]] = [E/E^{\mathrm{sm}}[N]]$. Write

$$C = \underline{\mathrm{Spec}}_{\mathscr{C}}\Big(\bigoplus_{a\in\mathbb{Z}/(N)} \mathscr{G}_a\Big) \xrightarrow{\pi} \mathscr{C},$$

where each $\mathscr{G}_a$ is an invertible $\mathbb{O}_{\mathscr{C}}$-module, with the grading and algebra structure corresponding to the structure of C as a μ_N-torsor over $\mathscr{C}$.

The image $\bar{R}$ of R in C is a $[\Gamma_1(N)]$-structure on C, so we get a μ_N-torsor

$$T := \underline{\mathrm{Spec}}_C\Big(\bigoplus_{b\in\mathbb{Z}/(N)} \mathscr{L}((b\cdot\bar{R}) - (0_C))\Big)$$

over C; the μ_N-action on T corresponds to the $\mathbb{Z}/(N)$-grading and the algebra structure on

$$\bigoplus_{b\in\mathbb{Z}/(N)} \mathscr{L}((b\cdot\bar{R}) - (0_C))$$

comes from the group law on C^{sm} and the canonical isomorphism $C^{\mathrm{sm}} \cong \mathrm{Pic}^0_{C/S}$.

Since C is a μ_N-torsor over $\mathscr{C}$, if $\mathscr{L} \in \mathrm{Pic}(C)$ we have a canonical decomposition

$$\pi_*\mathscr{L} = \bigoplus_{a\in\mathbb{Z}/(N)} \mathscr{L}_a,$$

where each $\mathscr{L}_a$ is an invertible sheaf on $\mathscr{C}$ and $\zeta \in \mu_N$ acts on $\mathscr{L}_a$ via multiplication by ζ^a. In particular this applies to the invertible sheaf $\mathscr{L} = \mathscr{L}((b\cdot\bar{R}) - (0_C))$, giving us a canonical decomposition

$$\pi_*\mathscr{L}((b\cdot\bar{R}) - (0_C)) = \bigoplus_{a\in\mathbb{Z}/(N)} \mathscr{L}_{(a,b)}.$$

We have $\mathscr{L}_{0,0} = \mathscr{G}_0 = \mathbb{O}_{\mathscr{C}}$, and the isomorphisms

$$\mathscr{L}((b_0\cdot\bar{R}) - (0_C)) \otimes_{\mathbb{O}_C} \mathscr{L}((b_1\cdot\bar{R}) - (0_C)) \cong \mathscr{L}(((b_0+b_1)\cdot\bar{R}) - (0_C))$$

(coming from the algebra structure of $\bigoplus_b \mathscr{L}((b\cdot\bar{R}) - (0_C))$) induce isomorphisms

$$\mathscr{L}_{(a_0,b_0)} \otimes_{\mathbb{O}_{\mathscr{C}}} \mathscr{L}_{(a_1,b_1)} \cong \mathscr{L}_{(a_0+a_1,b_0+b_1)}$$

for all $(a_0, b_0), (a_1, b_1) \in (\mathbb{Z}/(N))^2$, giving us a canonical algebra structure on the direct sum

$$\bigoplus_{(a,b)\in(\mathbb{Z}/(N))^2} \mathscr{L}_{(a,b)}.$$

Identifying our original $[\Gamma_1(N)]$-structure P with its image in $E'^{\mathrm{sm}}(S)[N]$, there exists some $(a_0, b_0) \in (\mathbb{Z}/(N))^2$ with $P = a_0\cdot Q + b_0\cdot R \in E'^{\mathrm{sm}}[N]$. This determines

a μ_N-torsor

$$\mathcal{T} := \underline{\mathrm{Spec}}_{\mathscr{C}}\Big(\bigoplus_{c\in\mathbb{Z}/(N)} \mathscr{L}_{(ca_0,cb_0)}\Big)$$

over $\mathscr{C}$, corresponding to a morphism $\mathscr{C} \to \mathcal{B}\mu_N$. Here $\bigoplus \mathscr{L}_{(ca_0,cb_0)}$ is viewed as a sub-$\mathcal{O}_{\mathscr{C}}$-algebra of the algebra $\bigoplus \mathscr{L}_{(a,b)}$.

Definition 6.7. We define $\mathscr{C}_P \to \mathcal{B}\mu_N$ to be the relative coarse moduli space of the above morphism $\mathscr{C} \to \mathcal{B}\mu_N$, and we write $\phi_P : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}_P/S}$ for the corresponding group scheme homomorphism.

It is immediate that ϕ_P is a $[\Gamma_1(N)]$-structure on the twisted curve $\mathscr{C}_P$.

Lemma 6.8. *$(\mathscr{C}_P, \phi_P)$ is independent of the choice of (a_0, b_0) with $P = a_0 \cdot Q + b_0 \cdot R$, and of the choice of generalized elliptic curve E' and $[\Gamma(N)]$-structure (Q, R) on E' such that $D = \sum [a \cdot Q]$ is étale over S and R meets the fiberwise identity components of E'/S.*

Proof. First of all, if E/S is an ordinary elliptic curve, then $E' = E$. Our construction defines a map $E[N] \to \mathrm{Pic}^0_{E/S}[N] \cong E[N]$, which in fact is simply the identity map (which in particular is independent of the choice of (a_0, b_0) and the $[\Gamma(N)]$-structure (Q, R)). To see this, recall that by Lemma 3.22 we may identify the generalized elliptic curve $C = E/\langle Q\rangle$ (viewed as a μ_N-torsor over $E \cong [E/E[N]]$ as discussed above) with the μ_N-torsor

$$\underline{\mathrm{Spec}}_E\Big(\bigoplus_{a\in\mathbb{Z}/(N)} \mathscr{L}((a \cdot Q) - (0_E))\Big)$$

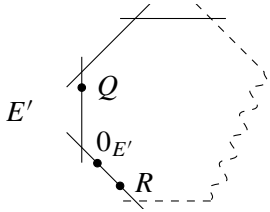
over E. So in the notation of the above construction, $\mathscr{C} = E$ and $\mathcal{G}_a = \mathscr{L}_{(a,0)} = \mathscr{L}((a \cdot Q) - (0_E))$. We have

$$\begin{aligned}\pi_*\mathscr{L}((\bar{R}) - (0_C)) &\cong \bigoplus_{a\in\mathbb{Z}/(N)} \big(\mathcal{G}_a \otimes \mathscr{L}((R) - (0_E))\big)\\ &\cong \bigoplus_{a\in\mathbb{Z}/(N)} \mathscr{L}((a \cdot Q + R) - (0_E)).\end{aligned}$$

So the map defined in the above construction sends Q to $\mathscr{L}_{(1,0)} = \mathscr{L}((Q) - (0_E))$ and R to $\mathscr{L}_{(0,1)} = \mathscr{L}((R) - (0_E))$. Composing with the usual isomorphism $\mathrm{Pic}^0_{E/S} \cong E$ yields the identity map on $E[N]$.

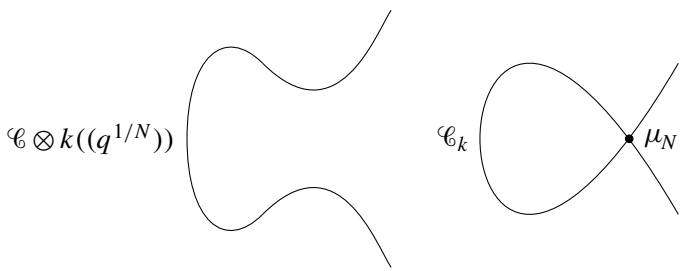
To complete the proof in the case where E/S is not necessarily smooth, it suffices to consider the case where S is the spectrum of an algebraically closed field k and E/k is a Néron d-gon (for some $d \mid N$). E'/k is then a Néron N-gon, and our Drinfeld basis (Q, R) was chosen so that $\langle Q\rangle$ meets every irreducible component of E' and $\langle R\rangle$ lies on the identity component. We may therefore choose

an isomorphism $E'^{\mathrm{sm}} \cong \mathbb{Z}/(N) \times \mu_N$ such that $Q = (1, 1)$ and $R = (0, \zeta)$ for some $\zeta \in \mu_N^\times(k)$.



Then if $P = a_0 \cdot Q + b_0 \cdot R = a_1 \cdot Q + b_1 \cdot R$, it follows that $a_0 = a_1$ and $\zeta^{b_0} = \zeta^{b_1}$, the latter of which implies that $b_0 \cdot \bar{R} = b_1 \cdot \bar{R}$. Thus $\mathscr{L}((b_0 \cdot \bar{R}) - (0_C)) = \mathscr{L}((b_1 \cdot \bar{R}) - (0_C))$, so $\mathscr{L}_{(a_0, b_0)} = \mathscr{L}_{(a_1, b_1)}$, hence $(\mathscr{C}_P, \phi_P)$ is independent of the choice of (a_0, b_0).

Now we must check that $(\mathscr{C}_P, \phi_P)$ is independent of the choice of $(E', (Q, R))$. E'/k is a Néron N-gon, so the choice of E' is unique up to composition with an automorphism of E' fixing $E^{\mathrm{sm}} \subset E'^{\mathrm{sm}}$. E' is the special fiber of an N-gon Tate curve $\mathscr{E}'/k[\![q^{1/N}]\!]$. Let $\mathscr{C} = [\mathscr{E}'/\mathscr{E}'^{\mathrm{sm}}[N]]$, so $\mathscr{C}_k = [E'/E'^{\mathrm{sm}}[N]] = [E/E^{\mathrm{sm}}[N]]$.



We may choose an isomorphism $\mathscr{E}'^{\mathrm{sm}}[N] \cong \mathbb{Z}/(N) \times \mu_N$ of finite flat group schemes over $k[\![q^{1/N}]\!]$. (Q, R) extends to a $[\Gamma(N)]$-structure $(\mathfrak{Q}, \mathfrak{R})$ on $\mathscr{E}'$ with $\mathfrak{D} = \sum (a \cdot \mathfrak{Q})$ étale over $k[\![q^{1/N}]\!]$ (and of course $\mathfrak{R}$ meets the identity component of every geometric fiber of $\mathscr{E}'/k[\![q^{1/N}]\!]$). Given such a $[\Gamma(N)]$-structure on $\mathscr{E}'$, our construction defines a group scheme homomorphism

$$\mathscr{E}'^{\mathrm{sm}}[N] \to \mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}[N].$$

Both of these are finite flat group schemes over $k[\![q^{1/N}]\!]$ which are isomorphic to $\mathbb{Z}/(N) \times \mu_N$, and $\underline{\mathrm{End}}(\mathbb{Z}/(N) \times \mu_N)$ is finite (hence proper) over $k[\![q^{1/N}]\!]$. Since

$$\mathscr{E}'^{\mathrm{sm}}[N] \otimes k((q^{1/N})) \to \mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}[N] \otimes k((q^{1/N}))$$

is independent of the choice of $[\Gamma(N)]$-structure over $k((q^{1/N}))$ (as $\mathscr{E}' \otimes k((q^{1/N}))$ is an elliptic curve), we conclude that $\mathscr{E}'^{\mathrm{sm}}[N] \to \mathrm{Pic}^0_{\mathscr{C}/k[\![q^{1/N}]\!]}[N]$ is independent of the choice of $(\mathfrak{Q}, \mathfrak{R})$. Thus in particular $E'^{\mathrm{sm}}[N] \to \mathrm{Pic}^0_{\mathscr{C}_k/k}[N]$ is independent of the choice of (Q, R) and the resulting homomorphism $E^{\mathrm{sm}}[N] \to \mathrm{Pic}^0_{\mathscr{C}_k/k}[N]$ is independent of the choice of E'. $\square$

Thus by descent, $(\mathscr{C}_P, \phi_P) \in \mathcal{X}_1^{\mathrm{tw}}(N)(S)$ is well defined globally over our initial base scheme S (even allowing supersingular fibers) and depends only on the pair $(E, P) \in \mathcal{X}_1(N)(S)$. We define our map $\mathcal{X}_1(N) \to \mathcal{X}_1^{\mathrm{tw}}(N)$ by sending (E, P) to $(\mathscr{C}_P, \phi_P)$.

Corollary 6.9. *Over any base scheme S, the morphism $\mathcal{X}_1(N) \to \mathcal{X}_1^{\mathrm{tw}}(N)$ sending (E, P) to $(\mathscr{C}_P, \phi_P)$ is an isomorphism of algebraic stacks.*

Similarly, given a generalized elliptic curve E/S which is equipped with a $[\Gamma(N)]$-structure (P_1, P_2), the above procedure produces a $[\Gamma(N)]$-structure $\phi_{(P_1,P_2)}$ on the twisted curve $\mathscr{C}_E := [E/E^{\mathrm{sm}}[N]]$.

Corollary 6.10. *Over any base scheme S, the morphism $\mathcal{X}(N) \to \mathcal{X}^{\mathrm{tw}}(N)$ sending $(E, (P_1, P_2))$ to $(\mathscr{C}_E, \phi_{(P_1,P_2)})$ is an isomorphism of algebraic stacks.*

7. Other compactified moduli stacks of elliptic curves

It is worth noting that the techniques in the proof of Theorem 4.6 are easily adapted to prove properness of the natural analogues in our current setting of well-known modular compactifications of other various moduli stacks of elliptic curves with extra structure, even when these moduli stacks do not naturally lie in a moduli stack of twisted stable maps:

Definition 7.1. Let $\mathscr{C}/S$ be a 1-marked genus-1 twisted stable curve with nonstacky marking.

(i) A $[\Gamma_0(N)]$-*structure* on $\mathscr{C}$ is a finite locally free S-subgroup scheme G of $\mathrm{Pic}^0_{\mathscr{C}/S}$ of rank N over S which is cyclic (fppf locally admits a $\mathbb{Z}/(N)$-generator), such that for every geometric point $\bar{p} \to S$, $G_{\bar{p}}$ meets every irreducible component of $\mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})}$. We write $\mathcal{X}_0^{\mathrm{tw}}(N)$ for the stack over S associating to T/S the groupoid of pairs $(\mathscr{C}, G)$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and G is a $[\Gamma_0(N)]$-structure on $\mathscr{C}$.

(ii) A *balanced* $[\Gamma_1(N)]$-*structure* (see [Katz and Mazur 1985, §3.3]) on $\mathscr{C}$ is an fppf short exact sequence of commutative group schemes over S

$$0 \to K \to \mathrm{Pic}^0_{\mathscr{C}/S}[N] \to K' \to 0, \qquad (\dagger)$$

where K and K' are locally free of rank N over S, together with sections $P \in K(S)$ and $P' \in K'(S)$ which are $\mathbb{Z}/(N)$-generators of K and K' in the sense of [Katz and Mazur 1985, §1.4]. We write $\mathcal{X}_1^{\mathrm{bal,tw}}(N)$ for the stack over S associating to T/S the groupoid of pairs $(\mathscr{C}, \dagger)$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and $\dagger$ is a balanced $[\Gamma_1(N)]$-structure on $\mathscr{C}$.

(iii) An $[N\text{-Isog}]$-*structure* (see [Katz and Mazur 1985, §6.5]) on $\mathscr{C}$ is a finite locally free commutative S-subgroup scheme $G \subset \mathrm{Pic}^0_{\mathscr{C}/S}[N]$ of rank N over S,

such that for every geometric point $\bar{p} \to S$, $G_{\bar{s}}$ meets every irreducible component of $\mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})}$. We write $\mathcal{X}^{\mathrm{tw}}(N\text{-Isog})$ for the stack over S associating to T/S the groupoid of pairs $(\mathscr{C}, G)$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and G is an $[N\text{-Isog}]$-structure on $\mathscr{C}$.

(iv) If N and n are positive integers such that $\mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ for all primes p dividing both N and n, a $[\Gamma_1(N; n)]$-*structure* (see [Conrad 2007, 2.4.3]) on $\mathscr{C}$ is a pair (ϕ, G), where

- $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{\mathscr{C}/S}$ is a $\mathbb{Z}/(N)$-structure in the sense of [Katz and Mazur 1985, §1.5],
- $G \subset \mathrm{Pic}^0_{\mathscr{C}/S}$ is a finite locally free S-subgroup scheme which is cyclic of order n,
- the degree-Nn relative effective Cartier divisor
$$\sum_{a=0}^{N-1} (\phi(a) + G)$$
in $\mathrm{Pic}^0_{\mathscr{C}/S}$ meets every irreducible component of each geometric fiber of $\mathrm{Pic}^0_{\mathscr{C}/S}$ over S, and
- for all primes p dividing both N and n, for $e = \mathrm{ord}_p(n)$ we have an equality of closed subschemes of $\mathrm{Pic}^0_{\mathscr{C}/S}$
$$\sum_{a=0}^{p^e-1} ((N/p^e) \cdot \phi(a) + G[p^e]) = \mathrm{Pic}^0_{\mathscr{C}/S}[p^e].$$

We write $\mathcal{X}_1^{\mathrm{tw}}(N; n)$ for the stack over S associating to T/S the groupoid of tuples $(\mathscr{C}, (\phi, G))$, where $\mathscr{C}/T$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, and (ϕ, G) is a $[\Gamma_1(N; n)]$-structure on $\mathscr{C}$.

Corollary 7.2. *The stacks $\mathcal{X}_0^{\mathrm{tw}}(N)$, $\mathcal{X}_1^{\mathrm{bal,tw}}(N)$, $\mathcal{X}^{\mathrm{tw}}(N\text{-Isog})$, and $\mathcal{X}_1^{\mathrm{tw}}(N; n)$ are algebraic stacks which are flat and locally finitely presented over S, with local complete intersection fibers. They are proper and quasifinite over $\overline{\mathcal{M}}_{1,1}$, and each is isomorphic to the corresponding moduli stack for generalized elliptic curves.*

As shown explicitly for the stacks $\mathcal{X}_1^{\mathrm{tw}}(N)$ and $\mathcal{X}^{\mathrm{tw}}(N)$ earlier in this paper, one may study the reductions modulo p of these moduli stacks, and one finds that over a perfect field of characteristic p, each stack is a disjoint union with crossings at the supersingular points of various closed substacks, which come naturally equipped with moduli interpretations extending the definitions given in [Katz and Mazur 1985] (except for $\mathcal{X}_1^{\mathrm{tw}}(N; n)$, which is not studied there).

Question 7.3. If $\mathscr{C}/S$ is a 1-marked genus-1 twisted stable curve with nonstacky marking, the group scheme $\mathrm{Pic}^0_{\mathscr{C}/S}$ behaves just like the smooth part of a generalized

elliptic curve. Over the Zariski open set of S where $\mathscr{C} \to S$ is smooth, it agrees with C (which is in this case a smooth elliptic curve); if $\bar{p} \to S$ is a geometric point such that $C_{\bar{p}}$ is singular, then $\mathrm{Pic}^0_{\mathscr{C}_{\bar{p}}/k(\bar{p})} \cong \mathbb{G}_m \times \mathbb{Z}/(N)$ for some N, and this is just the smooth part of a Néron N-gon over $k(\bar{p})$. Is there a natural way to exhibit $\mathrm{Pic}^0_{\mathscr{C}/S}$ as the smooth part of a generalized elliptic curve, giving an equivalence between the stack of generalized elliptic curves over S and the stack of 1-marked genus-1 twisted stable curves over S with nonstacky marking? More precisely, is there a natural way of defining compactified Jacobians of twisted curves, such that the degree-0 compactified Jacobian of a standard μ_d-stacky Néron 1-gon is a Néron d-gon?

Appendix: On moduli of curves of higher genus

As in the case of elliptic curves, stacks of twisted stable maps allow for natural compactifications of stacks of genus-g curves equipped with certain extra structure. For example,

$$\overline{\mathscr{K}}^{\circ}_{g,0}(\mathscr{B}\mu_N) := \overline{\mathscr{K}}_{g,0}(\mathscr{B}\mu_N) \times_{\overline{\mathscr{M}}_g} \mathscr{M}_g$$

classifies pairs $(C/S, \phi)$, where C/S is a smooth genus-g curve and $\phi : \mathbb{Z}/(N) \to \mathrm{Pic}^0_{C/S}$, which we view as an N-torsion point in $\mathrm{Pic}^0_{C/S} = \mathrm{Jac}(C/S)$. This stack is naturally contained in the proper algebraic stack $\overline{\mathscr{K}}_{g,0}(\mathscr{B}\mu_N)$ as an open dense substack. However, the situation becomes considerably more complicated when we attempt to use this to obtain proper moduli stacks of curves with level structure, for example, replacing "N-torsion points" with "points of exact order N".

Over $\mathbb{Z}[1/N]$, we have a stack $\overline{\mathscr{M}}^{(N)}_g$ of twisted curves with level-N structure, studied in [Abramovich et al. 2003, §6]; this is a smooth proper modular compactification of the stack classifying (not necessarily symplectic) Jacobi level-N structures on smooth genus-g curves, as in [Deligne and Mumford 1969, 5.4]. One may be tempted to proceed as follows:

Definition A.1. Let $\mathscr{C}/S$ be an unmarked genus-g ($g > 1$) twisted stable curve over a scheme S. A *full level-N structure* on $\mathscr{C}$ is a group scheme homomorphism $\phi : (\mathbb{Z}/(N))^{2g} \to \mathrm{Pic}^0_{\mathscr{C}/S}$ such that $\{\phi(a) \mid a \in (\mathbb{Z}/(N))^2\}$ is a full set of sections for the finite flat group scheme $\mathrm{Pic}^0_{\mathscr{C}/S}[N]$ over S in the sense of [Katz and Mazur 1985, §1.8].

We write $\overline{\mathscr{M}}^{(N),\mathrm{tw}}_g$ for the substack of $\overline{\mathscr{K}}_{g,0}(\mathscr{B}\mu_N^{2g})$ associating to T/S the groupoid of pairs $(\mathscr{C}, \phi)$, where $\mathscr{C}/T$ is an unmarked genus-g twisted stable curve and ϕ is a full level-N structure on $\mathscr{C}$.

Unfortunately, this is not the "right" definition. By this we mean that we would like the stack $\overline{\mathscr{M}}^{(N),\mathrm{tw}}_g$ to be a closed substack of $\overline{\mathscr{K}}_{g,0}(\mathscr{B}\mu_N^{2g})$, flat over S; but it follows immediately from the example [Chai and Norman 1990, Appendix] that flatness of $\overline{\mathscr{M}}^{(N),\mathrm{tw}}_g$ fails in mixed characteristic, even over the ordinary locus of

$\mathcal{M}_g$. Of course, if N is invertible on S then this definition is the correct one. More precisely, the choice of an isomorphism $(\mathbb{Z}/(N))^{2g} \cong \mu_N^{2g}$ identifies $\overline{\mathcal{M}}_g^{(N),\mathrm{tw}}$ with the stack $\overline{\mathcal{M}}_g^{(N)}$ of [Abramovich et al. 2003, §6], which is shown in [loc. cit.] to be smooth over $\mathbb{Z}[1/N]$ and proper over $\overline{\mathcal{M}}_g$. One would hope to be able to change the above definition to get a closed substack $\overline{\mathcal{M}}_g^{(N),\mathrm{tw}}$ of $\overline{\mathcal{H}}_{g,0}(\mathcal{B}\mu_N^{2g})$, flat over S, agreeing with $\overline{\mathcal{M}}_g^{(N)}$ over $S[1/N]$ and with a natural moduli interpretation in terms of the maps from $(\mathbb{Z}/(N))^{2g}$ to the group schemes $\mathrm{Pic}^0_{\mathscr{C}/S}$.

More generally we have good properties for the moduli stack $\overline{\mathcal{H}}_{g,0}(\mathcal{B}G)$ whenever G is a finite diagonalizable or locally diagonalizable group scheme over S, so the Cartier dual G^* is commutative and constant or locally constant. Recall that for a finite group G there is a stack ${}_G\mathcal{M}_g$ over $\mathbb{Z}[1/|G|]$ of Teichmüller structures of level G on smooth curves (see [Deligne and Mumford 1969, 5.6; Pikaart and de Jong 1995]). Now if G is a diagonalizable group scheme with $|G|$ invertible on the base scheme S, then after adjoining appropriate roots of unity we may fix an isomorphism $G \cong G^*$. In [Abramovich et al. 2003, 5.2.3] it is observed that this determines an isomorphism between ${}_{G^*}\mathcal{M}_g$ and a substack $\mathcal{B}_g^{\mathrm{tei}}(G)^\circ$ of $\overline{\mathcal{H}}^\circ_{g,0}(\mathcal{B}G)$ whose closure $\mathcal{B}_g^{\mathrm{tei}}(G)$ in $\overline{\mathcal{H}}_{g,0}(\mathcal{B}G)$ is a moduli stack whose geometric objects correspond precisely to G-torsors $P \to \mathscr{C}$ which are connected (where $\mathscr{C}$ is a genus-g twisted stable curve); these are called *twisted Teichmüller G-structures*. One would hope that $\mathcal{B}_g^{\mathrm{tei}}(G)$ can be defined in arbitrary characteristic, with a natural moduli interpretation, but it is not clear to the author how to proceed with this for genus $g > 1$; as discussed above, it does not suffice to simply consider the substack of $\overline{\mathcal{H}}_{g,0}(\mathcal{B}G)$ whose geometric objects correspond to G-torsors which are connected, since μ_{p^n} is connected in characteristic p, and the definition in terms of "full sets of sections" does not give a stack flat over the base scheme in mixed characteristic.

Acknowledgements

The author is grateful to his advisor, Martin Olsson, for suggesting this research topic and for many helpful conversations, as well as to the reviewer for several helpful comments and corrections. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1106400.

References

[Abramovich 2012] D. Abramovich, "Raynaud's group-scheme and reduction of coverings", pp. 1–18 in *Number theory, analysis and geometry*, edited by D. Goldfeld et al., Springer, New York, 2012. MR 2867909 Zbl 06085939

[Abramovich and Hassett 2011] D. Abramovich and B. Hassett, "Stable varieties with a twist", pp. 1–38 in *Classification of algebraic varieties* (Schiermonnikoog, 2009), edited by C. Faber et al., Eur. Math. Soc., Zürich, 2011. MR 2012c:14023 Zbl 1223.14039

[Abramovich and Romagny 2012] D. Abramovich and M. Romagny, "Moduli of Galois p-covers in mixed characteristics", *Algebra Number Theory* **6**:4 (2012), 757–780. MR 2966718 Zbl 06064729

[Abramovich and Vistoli 2002] D. Abramovich and A. Vistoli, "Compactifying the space of stable maps", *J. Amer. Math. Soc.* **15**:1 (2002), 27–75. MR 2002i:14030 Zbl 0991.14007

[Abramovich et al. 2003] D. Abramovich, A. Corti, and A. Vistoli, "Twisted bundles and admissible covers", *Comm. Algebra* **31**:8 (2003), 3547–3618. MR 2005b:14049 Zbl 1077.14034

[Abramovich et al. 2008a] D. Abramovich, T. Graber, and A. Vistoli, "Gromov–Witten theory of Deligne–Mumford stacks", *Amer. J. Math.* **130**:5 (2008), 1337–1398. MR 2009k:14108 Zbl 1193.14070

[Abramovich et al. 2008b] D. Abramovich, M. Olsson, and A. Vistoli, "Tame stacks in positive characteristic", *Ann. Inst. Fourier* (*Grenoble*) **58**:4 (2008), 1057–1091. MR 2009c:14002 Zbl 1222.14004

[Abramovich et al. 2011] D. Abramovich, M. Olsson, and A. Vistoli, "Twisted stable maps to tame Artin stacks", *J. Algebraic Geom.* **20**:3 (2011), 399–477. MR 2012c:14024 Zbl 1225.14020

[Arbarello et al. 2011] E. Arbarello, M. Cornalba, and P. A. Griffiths, *Geometry of algebraic curves*, vol. 2, Grundlehren Math. Wiss. **268**, Springer, Heidelberg, 2011. MR 2012e:14059 Zbl 1235.14002

[Bourbaki 1998] N. Bourbaki, *Éléments de mathématique, algèbre commutative, chapitre 10: Profondeur, régularité, dualité*, Masson, Paris, 1998. MR 2007f:13001

[Chai and Norman 1990] C.-L. Chai and P. Norman, "Bad reduction of the Siegel moduli scheme of genus two with $\Gamma_0(p)$-level structure", *Amer. J. Math.* **112**:6 (1990), 1003–1071. MR 91i:14033 Zbl 0734.14010

[Conrad 2007] B. Conrad, "Arithmetic moduli of generalized elliptic curves", *J. Inst. Math. Jussieu* **6**:2 (2007), 209–278. MR 2008e:11073 Zbl 1140.14018

[Deligne and Mumford 1969] P. Deligne and D. Mumford, "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* **36**:1 (1969), 75–109. MR 41 #6850 Zbl 0181.48803

[Deligne and Rapoport 1973] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques", pp. 143–316 in *Modular functions of one variable* (Antwerp, 1972), vol. 2, edited by P. Deligne and W. Kuyk, Lecture Notes in Mathematics **349**, Springer, Berlin, 1973. MR 49 #2762 Zbl 0281.14010

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergeb. Math. Grenzgeb. (3) **2**, Springer, Berlin, 1998. MR 99d:14003 Zbl 0885.14002

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies **108**, Princeton University Press, 1985. MR 86i:11024 Zbl 0576.14026

[Keel and Mori 1997] S. Keel and S. Mori, "Quotients by groupoids", *Ann. of Math.* (2) **145**:1 (1997), 193–213. MR 97m:14014 Zbl 0881.14018

[Knutson 1971] D. Knutson, *Algebraic spaces*, Lecture Notes in Mathematics **203**, Springer, Berlin, 1971. MR 46 #1791 Zbl 0221.14001

[Petersen 2012] D. Petersen, "Cusp form motives and admissible G-covers", *Algebra Number Theory* **6**:6 (2012), 1199–1221. MR 2968638 Zbl 06126973

[Pikaart and de Jong 1995] M. Pikaart and A. J. de Jong, "Moduli of curves with non-abelian level structure", pp. 483–509 in *The moduli space of curves* (Texel Island, 1994), edited by R. Dijkgraaf et al., Progress in Mathematics **129**, Birkhäuser, Boston, MA, 1995. MR 97e:14037 Zbl 0860.14024

[Shatz 1986] S. S. Shatz, "Group schemes, formal groups, and p-divisible groups", pp. 29–78 in *Arithmetic geometry* (Storrs, CT, 1984), edited by G. Cornell and J. H. Silverman, Springer, New York, 1986. MR 861972 Zbl 0603.14033

Communicated by Brian Conrad
Received 2012-08-01 Revised 2013-01-04 Accepted 2013-02-09

andrew@math.berkeley.edu *Department of Mathematics, University of California, Berkeley, Berkeley, CA 94720, United States*
http://math.berkeley.edu/~andrew/

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2203

Regular permutation groups of order *mp* and Hopf Galois structures

Timothy Kohl

Let Γ be a group of order mp where p is prime and $p > m$. We give a strategy to enumerate the regular subgroups of $\mathrm{Perm}(\Gamma)$ normalized by the left representation $\lambda(\Gamma)$ of Γ. These regular subgroups are in one-to-one correspondence with the Hopf Galois structures on Galois field extensions L/K with $\Gamma = \mathrm{Gal}(L/K)$. We prove that every such regular subgroup is contained in the normalizer in $\mathrm{Perm}(\Gamma)$ of the p-Sylow subgroup of $\lambda(\Gamma)$. This normalizer has an affine representation that makes feasible the explicit determination of regular subgroups in many cases. We illustrate our approach with a number of examples, including the cases of groups whose order is the product of two distinct primes and groups of order $p(p-1)$, where p is a "safe prime". These cases were previously studied by N. Byott and L. Childs, respectively.

Introduction

Let L/K be a finite Galois extension of fields with Galois group $\Gamma = \mathrm{Gal}(L/K)$. Then the action of the group ring $K[\Gamma]$ of the Galois group Γ makes L/K into a Hopf Galois extension, in the sense of Chase and Sweedler [1969]. However, the classical Hopf Galois structure on L/K may not be the only Hopf Galois structure. For many Galois groups Γ, every Γ-Galois extension L/K has Hopf Galois structures by cocommutative K-Hopf algebras other than the classical Hopf Galois structure by the group ring $K[\Gamma]$ of the Galois group. Greither and Pareigis [1987] demonstrated this lack of uniqueness, by showing that the Hopf Galois structures on L/K are in direct correspondence with the regular subgroups $N \leq \mathrm{Perm}(\Gamma)$ normalized by $\lambda(\Gamma)$, where λ is the left action of Γ on Γ.

Subsequently Byott [2000] showed that nonclassical Hopf Galois structures are of interest in local Galois module theory settings, involving wildly ramified Galois extensions of local fields. Byott showed that a nonclassical Hopf Galois structure can yield freeness of the valuation ring of the extension over the corresponding

MSC2010: primary 20B35; secondary 12F10, 20E22, 16W30.
Keywords: regular permutation group, Hopf–Galois extension, holomorph.

associated order, whereas freeness fails over the associated order for the classical Galois structure given by the Galois group.

The Greither–Pareigis correspondence is via Galois descent: if H is a cocommutative K-Hopf algebra and L is an H-module algebra via some Galois structure map $H \otimes_K L \to L$, then base changing to L yields a Galois structure map $(L \otimes_K H) \otimes_L (L \otimes_K L) \to (L \otimes_K L)$. But then $L \otimes_K L \cong \mathrm{Hom}_L(L[\Gamma], L) = L[\Gamma]^* \cong \sum_{\gamma \in \Gamma} L\varphi_\gamma$ and $L \otimes_K H \cong L[N]$, where N is a group that acts on $L \otimes_K L$ via acting as a regular group of permutations on the subscripts of the dual basis $\{\varphi_\gamma : \gamma \in \Gamma\}$ of $L[\Gamma]^*$. Then N is normalized by $\lambda(\Gamma)$. Conversely, given a regular subgroup N of $\mathrm{Perm}\,\Gamma$, then $L[N]$ yields a Hopf Galois structure on $L[\Gamma]^*$. If N is normalized by $\lambda(\Gamma)$, then Galois descent yields a K-Hopf algebra structure by $H = (L[N])^G$ on L/K.

Thus determining Hopf Galois structures on Galois extensions L/K of fields with Galois group Γ is translated into the purely group-theoretic problem of determining regular subgroups of $B = \mathrm{Perm}(\Gamma)$ normalized by $\lambda(\Gamma)$.

Nearly all of the work since [Greither and Pareigis 1987] on determining the Hopf Galois structures on a Galois extension L/K of fields with Galois group Γ, or on counting or estimating the number of Hopf Galois structures on such field extensions, has involved a further translation of the problem. The idea of the translation, as formulated by Byott [1996], is to stratify the problem into a set of problems, one for each isomorphism type of group of the same cardinality as Γ. For each such group M, one seeks regular embeddings (modulo a certain equivalence) of Γ into the holomorph $\mathrm{Hol}(M) \subset \mathrm{Perm}(M)$ of M, where $\mathrm{Hol}(M) \cong M \rtimes \mathrm{Aut}(M)$. The number of such regular embeddings is equal to the number of Hopf Galois structures on L/K via K-Hopf algebras H such that $L \otimes_K H \cong L[M]$: then the Hopf Galois structure is said to have *type* M. This translation turns the problem of classifying Hopf Galois structures into a collection of somewhat easier problems, easier because it has seemed more tractable to identify regular subgroups in $\mathrm{Hol}\,M$ than in the usually much larger group $\mathrm{Perm}\,\Gamma$.

On the other hand, once one has a regular embedding β of Γ in $\mathrm{Hol}\,M$, two translations are required to actually describe the corresponding Hopf Galois structure on L/K. It is typically not easy to identify the regular subgroup N of $\mathrm{Perm}\,\Gamma$ isomorphic to M that corresponds to the embedding β and the action of N on $L[G]^*$ on which one may apply Galois descent. For this reason, it is of interest to find groups Γ where regular subgroups of $\mathrm{Perm}\,\Gamma$ normalized by $\lambda(\Gamma)$ may be determined directly.

The aim of this paper is to do exactly that for a special class of groups. We consider groups Γ of order mp where p is prime and $p > m$. Then $\lambda(\Gamma)$ has a unique p-Sylow subgroup $\mathcal{P}$ of order p. Our main result is that every regular subgroup of $\mathrm{Perm}\,\Gamma$ normalized by Γ is contained in $\mathrm{Norm}_B(\mathcal{P})$, the normalizer in

$B = \mathrm{Perm}(\Gamma)$ of $\mathcal{P}$. The group $\mathrm{Norm}_B(\mathcal{P})$ may be identified as the subgroup of the affine group $\mathrm{AGL}_m(\mathbb{F}_p) \subset \mathrm{GL}_{m+1}(\mathbb{F}_p)$ consisting of $(m+1) \times (m+1)$ matrices of the form

$$\begin{pmatrix} A & v \\ 0 & 1 \end{pmatrix},$$

where A is a scalar multiple of an $m \times m$ permutation matrix and v is in $\mathbb{F}_p^m$. For $m < p$, $\mathrm{Norm}_B(\mathcal{P})$ is far smaller and much more amenable than the symmetric group $\mathrm{Perm}(\Gamma) \cong S_{mp}$. (For example, for $p = 7$ and $m = 4$, $\mathrm{Norm}_B(\mathcal{P})$ has order $7^4 \cdot 6 \cdot 4! = 345779$, while S_{28} has order $28! \sim 3 \cdot 10^{29}$.)

The first application of our main result is to determine all regular subgroups of $\mathrm{Perm}\,\Gamma$ normalized by $\lambda(\Gamma)$ where Γ has order pq, distinct primes. N. Byott [2004] determined the Hopf Galois structures on a field extension L/K with Galois group Γ of order pq by looking at the holomorph $\mathrm{Hol}\,M$ of M for M a group of order pq and determining the regular embeddings of Γ whose intersection with $\mathrm{Aut}\,M$ has a given cardinality. The method of this paper is quite different; the reader may judge the relative efficiency of the two methods.

For our second application we consider the Hopf Galois structures on a Galois extension L/K where the Galois group Γ has order mp with $m = 2q$, q prime, and $p = 2q + 1$ prime: thus p is a safe prime and q is a Sophie Germain prime. L. Childs [2003] determined all of the Hopf Galois structures on a Galois extension L/K of fields with Galois group $\Gamma \cong \mathrm{Hol}(C_p)$ by determining embeddings of Γ into $\mathrm{Hol}\,M$ for each of the six isomorphism types of groups of order mp. We extend [Childs 2003] by determining the number of Hopf Galois structures for Γ and M running through all 36 pairs (Γ, M). Since the computations are in many cases similar to those in the pq case, we provide only a few sample cases to illustrate the variety of approaches needed.

This paper generalizes the results for $m = 4$ in [Kohl 2007]. Some of the ideas here are similar to those in that paper, but for the benefit of the reader we have made this paper independent of [Kohl 2007] and reasonably self-contained.

1. Preliminaries

Groups of order mp. We begin with some observations about abstract groups G of order mp, where $m < p$.

First, G has a p-Sylow subgroup P that is unique, and hence a characteristic subgroup of G. Also, by the Schur–Zassenhaus lemma, there exists a subgroup $Q \leq G$ of order m, and $G \cong P \rtimes_\tau Q$ with $\tau : Q \to \mathrm{Aut}(P)$ induced by conjugation within G.

Lemma 1.1. *Let G have order mp with p prime and $p > m$, with $G \cong P \rtimes_\tau Q$ as above.*

(a) *If τ is trivial, that is, $G \cong P \times Q$, then p does not divide the order of* $\operatorname{Aut} G$.

(b) *If τ is not-trivial, then* $\operatorname{Aut} G$ *has a unique p-Sylow subgroup, consisting of inner automorphisms given by conjugation by elements of P.*

Proof. Since $P \leq G$ is unique and thus characteristic, if $\psi \in \operatorname{Aut}(G)$ then ψ induces $\bar{\psi} \in \operatorname{Aut}(G/P)$. Our claim is that $|\psi|$ cannot be p^k for any $k > 1$. Since $|G/P| = m < p$ then $p \nmid |\operatorname{Aut}(G/P)|$ so if ψ has order p^k then $\bar{\psi} = \mathrm{id}_{G/P}$. Therefore, for any $g \in G$ one has $\psi(gP) = gP$ and so $g^{-1}\psi(g) \in P$ and likewise $g^{-1}\psi^r(g) \in P$ for any power r. If $|\psi| = p^k$ for $k > 1$ then there exists $g \in G$ such that

$$g, \psi(g), \ldots, \psi^{p^k-1}(g)$$

are distinct elements of G, but then

$$1, g^{-1}\psi(g), \ldots, g^{-1}\psi^{p^k-1}(g)$$

are p^k distinct elements of P, which is impossible since $|P| = p$. Therefore the p torsion of $\operatorname{Aut} G$ cannot be larger than p. If τ is trivial then $G \cong P \times Q$ for Q of order m. As such, $\operatorname{Aut}(G) \cong \operatorname{Aut}(P) \times \operatorname{Aut}(Q)$ and neither $\operatorname{Aut} P$ nor $\operatorname{Aut} Q$ can have elements of order p so $p \nmid |\operatorname{Aut}(G)|$. If $\tau : Q \to \operatorname{Aut}(P)$ is nontrivial then one can show that $|P \cap Z(G)| = 1$, so that if $P = \langle x \rangle$ then conjugation by x provides an element of order p in $\operatorname{Aut} G$ which therefore generates the p-Sylow subgroup of $\operatorname{Aut} G$. □

Regular subgroups.

Definition. Let $\mathcal{P} \leq \lambda(\Gamma)$ be the unique p-Sylow subgroup of $\lambda(\Gamma)$.

Definition. A subgroup $N \leq B = \operatorname{Perm}(\Gamma)$ is *semiregular* [Wielandt 1955] if $\operatorname{Stab}_N(\gamma) = \{\eta \in N \mid \eta(\gamma) = \gamma\}$ is the trivial group for all $\gamma \in \Gamma$.

A subgroup $N \leq B$ is *regular* if N is semiregular and either $|N| = |\Gamma|$ or N acts transitively on Γ.

If N is semiregular and $\eta \neq e$ (the identity) of N, then η acts on Γ without fixed points. Thus for η in N, if η has order h, then for each γ in Γ,

$$(\gamma, \eta(\gamma), \ldots, \eta^{h-1}(\gamma))$$

is the cycle containing γ in the cycle decomposition of η in $B = \operatorname{Perm}(\Gamma)$. Hence η is a product of k cycles of length h, where $hk = |\Gamma|$.

Definition. For η in $B = \operatorname{Perm}(\Gamma)$,

$$\operatorname{Supp}(\eta) = \{\gamma \in \Gamma \mid \eta(\gamma) \neq \gamma\}.$$

Thus if N is semiregular and $\eta \in N$ is not the identity, then $\mathrm{Supp}(\eta) = \Gamma$.

Because of the connection to Hopf Galois structures, in this paper we are not interested in all the regular subgroups of B, but only in those normalized by $\lambda(\Gamma)$, the image of the left regular representation of Γ in B.

Definition. Let $R(\Gamma)$ denote the set of regular subgroups N of $B = \mathrm{Perm}(\Gamma)$ such that $\lambda(\Gamma) \leq \mathrm{Norm}_B(N)$, the normalizer in B of N.

We partition $R(\Gamma)$ as follows:

Definition. For M a group of order $|\Gamma|$, let $[M]$ denote the isomorphism type of M, and let $R(\Gamma, [M])$ denote the subset of $R(\Gamma)$ consisting of the regular subgroups N in $R(\Gamma)$ that are isomorphic to M.

Then $R(\Gamma)$ is the disjoint union of the sets $R(\Gamma, [M])$ where $[M]$ runs through the isomorphism types of groups of order equal to $|\Gamma|$.

To enumerate $R(\Gamma)$, we enumerate $R(\Gamma, [M])$ for each isomorphism type $[M]$. As noted in the introduction, the Hopf Galois structures on a Galois extension L/K with Galois group $\Gamma = \mathrm{Gal}(L/K)$ correspond in a one-to-one fashion to the elements of $R(\Gamma)$; if a Hopf Galois structure corresponds to N in $R(\Gamma, [M])$, then the K-Hopf algebra acting on L has *type* M (because $L \otimes_K H \cong L[M]$).

Our goal in this paper is to develop a new method to enumerate $R(\Gamma)$ for $|\Gamma| = mp$.

Cycle structures. Let N be a regular subgroup of $B = \mathrm{Perm}(\Gamma)$ normalized by $\lambda(\Gamma)$, and let $P(N)$ be the unique order-p subgroup of N. Then we can relate the cycle structure of a generator of $\mathcal{P} = P(\lambda(\Gamma))$ to the cycle structure of a generator of $P(N)$:

Proposition 1.2. *Let $\mathcal{P}$ be the unique subgroup of $\lambda(\Gamma)$ of order p, and let $\mathcal{P} = \langle \phi \rangle$, where $\phi = \pi_1 \pi_2 \cdots \pi_m$ with $\pi_1, \ldots, \pi_m$ disjoint p-cycles in* $\mathrm{Perm}(\Gamma) \cong S_{pm}$. *Let N be a regular subgroup of* $\mathrm{Perm}\,\Gamma$ *normalized by $\lambda(\Gamma)$ and let $P(N)$ be the p-Sylow subgroup of N. Then $P(N)$ is generated by $\theta = \pi_i^{a_1} \pi_2^{a_2} \cdots \pi_m^{a_m}$ where $a_i \in U_p = \mathbb{F}_p^\times$ for each i.*

Proof. N is normalized by $\lambda(\Gamma)$ and $P(N)$ is characteristic in N. Hence $\lambda(\Gamma)$, and therefore also $\mathcal{P}$, normalizes $P(N)$. But $\gcd(|\mathrm{Aut}(P(N))|, p) = 1$, so $\mathcal{P}$ centralizes $P(N)$, hence $P(N)$ centralizes $\mathcal{P}$.

Let θ be a generator of $P(N)$. Then

$$\pi_1 \pi_2 \cdots \pi_m = \phi = \theta \phi \theta^{-1} = \theta(\pi_1 \pi_2 \cdots \pi_m)\theta^{-1} = \pi_1 \pi_2 \cdots \pi_m,$$

and so θ permutes the cycles $\pi_1, \ldots, \pi_m$. But conjugation by θ has order dividing p, and $\mathrm{Perm}(\{\pi_1, \ldots, \pi_m\})$ has order $m!$ coprime to p, so for all i, $\theta \pi_i \theta^{-1} = \pi_i$.

For each i and for any c in $\operatorname{Supp}\pi_i$, π_i is the cycle

$$\pi_i = (c, \pi_i(c), \pi_i^2(c), \ldots, \pi_i^{p-1}(c)),$$

and $\theta\pi_i\theta^{-1}$ is the cycle

$$\theta\pi_i\theta^{-1} = (\theta(c), \theta\pi_i(c), \theta\pi_i^2(c), \ldots, \theta\pi_i^{p-1}(c)).$$

If $\theta(c) = \pi_i^a(c)$, then comparing the two cycles, we see that $\theta\pi_i^r(c) = \pi_i^{a+r}(c)$ for all r. Thus for each i, on $\operatorname{Supp}\pi_i$, $\theta = \pi_i^a$. Hence $\theta = \pi_i^{a_1}\pi_2^{a_2}\cdots\pi_m^{a_m}$ in B. No a_i can equal 0 modulo p; if it did, c_i would be fixed under θ, and θ is an element of the semiregular subgroup $P(N)$ of B. □

Let N be a regular subgroup of $B = \operatorname{Perm}(\Gamma)$, let $P(N)$ be the p-Sylow subgroup of N, and let $N = P(N)Q(N)$, where $Q(N)$ is a complementary subgroup of order m to $P(N)$ in N. Then $Q(N)$ normalizes $P(N) = \langle \pi_1^{a_1}\cdots\pi_m^{a_m}\rangle$. Let $Q(N) = \{q_1 = e, q_2, \ldots, q_m\}$. Since N is a regular subgroup of $\operatorname{Perm}\Gamma$,

$$\Gamma = Ne_\Gamma = \bigcup_{i=1}^{m} P(N)q_ie_\Gamma,$$

and $P(N) = \langle\theta\rangle$ acts on $P(N)q_ie_\Gamma$ via the left regular representation. After renumbering the elements of $Q(N)$ as needed, we have $\Pi_i = \operatorname{Supp}(\pi_i) = P(N)q_ie_\Gamma$.

Proposition 1.3. *$Q(N)$ is a regular group of permutations of $\{\Pi_1, \ldots, \Pi_m\}$.*

Proof. For q in $Q(N)$,

$$q\Pi_i = qP(N)q_ie_\Gamma = qP(N)q^{-1}q_ie_\Gamma = P(N)qq_ie_\Gamma,$$

since $P(N)$ is a normal subgroup of N. So the action of $Q(N)$ on $\{\Pi_1, \ldots, \Pi_m\}$ is the same as the left regular representation $\lambda(Q(N))$ on $Q(N)$.

The partition $\{\Pi_1, \ldots, \Pi_m\}$ arising from $P(N)$ is the same as that from $\mathcal{P}$. So we conclude that each regular subgroup N of $\operatorname{Perm}\Gamma$ normalized by $\lambda(\Gamma)$ has the form $P(N)Q(N)$ where $P(N) = \langle\pi_1^{a_1}\cdots\pi_m^{a_m}\rangle$ and $Q(N)$ is a regular subgroup of $\operatorname{Perm}(\{\Pi_1, \ldots, \Pi_m\})$ with $\Pi_i = \operatorname{Supp}(\pi_i)$. □

2. Characters and generators of $P(N)$

In this section we determine the semiregular order-p subgroups of $B = \operatorname{Perm}(\Gamma)$ normalized by $\lambda(\Gamma)$.

Recall that $\lambda(\Gamma) = \mathcal{P}\mathcal{Q}$ where $\mathcal{P}$ is the unique p-Sylow subgroup of $\lambda(\Gamma)$ and $\mathcal{Q}$ is a complement of $\mathcal{P}$ in $\lambda(\Gamma)$. Then $\mathcal{P} = \langle\phi\rangle$ where $\phi = \pi_1\cdots\pi_m$, a product of p-cycles, $\Pi_i = \operatorname{Supp}(\pi_i)$ for $i = 1, \ldots, m$, and $\mathcal{Q}$ is a regular group of permutations of $\{\Pi_1, \ldots, \Pi_m\}$, hence may be viewed as a regular subgroup of S_m. From the last result of the previous section, every semiregular order-p subgroup P of B

normalized by $\lambda(\Gamma)$ has the form $P = \langle \pi_1^{a_1} \cdots \pi_m^{a_m} \rangle$ for $a_1, \ldots, a_m$ in $\mathbb{F}_p^\times$. Here we describe the possible P more precisely.

There is an isomorphism from $V = \langle \pi_1, \ldots, \pi_m \rangle$ to $\mathbb{F}_p^m$ by

$$\pi_1^{i_1} \cdots \pi_m^{i_m} \mapsto (i_1, \ldots, i_m).$$

Denote $\pi_1^{i_1} \cdots \pi_m^{i_m}$ by $[i_1, \ldots, i_m]$. Then $\hat{v}_i = (0, \ldots, 1, \ldots, 0)$ in $\mathbb{F}_p^m$ corresponds to π_i. By abuse of notation, we will identify $\hat{v}_i$ in $\mathbb{F}_p^m$ with π_i in V.

Let $\chi : \mathfrak{Q} \to \mathbb{F}_p^\times$ be a homomorphism, that is, a degree-one representation or linear character of $\mathfrak{Q}$ in $\mathbb{F}_p$ [Isaacs 1976].

Let $\hat{p}_\chi = \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma) \hat{v}_{\gamma(1)}$. As with $\hat{v}_i$, we will identify $\hat{p}_\chi$ with the corresponding element of V, as in the statement of the following result:

Theorem 2.1. *For each linear character $\chi : \mathfrak{Q} \to \mathbb{F}_p^\times$, $\hat{p}_\chi$ is a generator of a semiregular order-p subgroup of V normalized by $\lambda(\Gamma)$. Conversely, let P be an order-p semiregular subgroup of V that is normalized by $\lambda(\Gamma)$. Then $P = \langle \hat{p}_\chi \rangle$ for some linear character $\chi : \mathfrak{Q} \to \mathbb{F}_p^\times$.*

Proof. For the first part, we begin by observing that $\mathfrak{Q}$ normalizes $\mathcal{P} = \langle \pi \rangle$, so for all μ in $\mathfrak{Q}$, $\mu(\pi) = \mu \pi \mu^{-1} = \pi^{\tau(\mu)}$ for some $\tau(\mu)$ in $\mathbb{F}_p^\times$. Now

$$\hat{p}_\chi = \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma) \hat{v}_{\gamma(1)} = \sum_{\gamma \in \mathfrak{Q}} \chi(\mu\gamma) \hat{v}_{\mu\gamma(1)} = \chi(\mu) \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma) \hat{v}_{\mu\gamma(1)},$$

and so

$$\begin{aligned} \mu \hat{p}_\chi \mu^{-1} &= \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma)(\mu \hat{v}_{\gamma(1)} \mu^{-1}) = \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma) \tau(\mu) \hat{v}_{\mu\gamma(1)} \\ &= \tau(\mu) \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma) \hat{v}_{\mu\gamma(1)} = \tau(\mu) \chi(\mu)^{-1} \hat{p}_\chi. \end{aligned}$$

Hence $\langle \hat{p}_\chi \rangle$ is normalized by $\mathfrak{Q}$. Since $\langle \hat{p}_\chi \rangle$ is a subgroup of V, $\langle \hat{p}_\chi \rangle$ is centralized by $\mathcal{P}$, hence $\langle \hat{p}_\chi \rangle$ is normalized by $\lambda(\Gamma)$.

Now we show the converse.

Let $[a_1, \ldots, a_m]$ be in V with all $a_i \neq 0$ in $\mathbb{F}_p$, such that $\langle [a_1, \ldots, a_m] \rangle$ is normalized by $\lambda(\Gamma)$. Then for γ in $\mathfrak{Q}$,

$$\gamma [a_1, \ldots, a_m] \gamma^{-1} = [a_1, \ldots, a_m]^{d_\gamma} = [d_\gamma a_1, \ldots, d_\gamma a_m].$$

The map from $\mathfrak{Q}$ to $\mathbb{F}_p^\times$ given by $\gamma \mapsto d_\gamma$ is a homomorphism, hence a linear character. Also, for every γ in $\mathfrak{Q}$,

$$\gamma \pi_i \gamma^{-1} = \pi_{\gamma(i)}^{c_\gamma},$$

where $\mathfrak{Q}$ acts as a regular subgroup of $\mathrm{Perm}(1, \ldots, m)$ as noted above, and c_γ is in $\mathbb{F}_p^\times$. Then $c_{\gamma'\gamma} = c_{\gamma'} c_\gamma$, so the map $\gamma \mapsto c_\gamma$ is a linear character from $\mathfrak{Q}$ to $\mathbb{F}_p^\times$.

Since all $a_i \neq 0$, in the subgroup $\langle [a_1, \ldots, a_m] \rangle$ we may replace the generator by a suitable power so that $a_1 = 1$, so we assume henceforth that $a_1 = 1$. Now for γ in $\mathfrak{Q}$,

$$\gamma [a_1, \ldots, a_m] \gamma^{-1} = [c_\gamma a_{\gamma^{-1}(1)}, \ldots, c_\gamma a_{\gamma^{-1}(m)}],$$

and so

$$c_\gamma a_{\gamma^{-1}(i)} = d_\gamma a_i,$$

for every i. Setting $i = \gamma(j)$, this becomes

$$c_\gamma a_j = d_\gamma a_{\gamma(j)},$$

or

$$a_{\gamma(j)} = \frac{c_\gamma}{d_\gamma} a_j.$$

In particular,

$$a_{\gamma(1)} = \frac{c_\gamma}{d_\gamma} a_1 = \frac{c_\gamma}{d_\gamma}.$$

Since $\mathfrak{Q}$ acts as a regular subgroup of permutations of $1, \ldots, m$, this last formula determines a_i for all $i = 1, \ldots, m$.

The mapping $\chi : \mathfrak{Q} \to \mathbb{F}_p^\times$ defined by $\chi(\gamma) = c_\gamma / d_\gamma$ is a homomorphism, hence a linear character of $\mathfrak{Q}$ in $\mathbb{F}_p^\times$, and we have:

$$[a_1, \ldots, a_m] = \prod_{\gamma \in \mathfrak{Q}} \pi_{\gamma(1)}^{a_\gamma(1)} = \prod_{\gamma \in \mathfrak{Q}} \pi_{\gamma(1)}^{\chi(\gamma)} = \sum_{\gamma \in \mathfrak{Q}} \chi(\gamma) \hat{v}_{\gamma(1)} = \hat{p}_\chi. \qquad \square$$

Example 2.1. In [Kohl 2007] we examined groups of order $4p$. There were two cases. If $\mathfrak{Q} = C_p \times C_p = \langle x, y \rangle$, then there are four linear characters, defined by the following table:

	1	x	y	xy
χ_1	1	1	1	1
χ_2	1	1	-1	-1
χ_3	1	-1	1	-1
χ_4	1	-1	-1	1

For $\mathfrak{Q} = C_4 = \langle x \rangle$, we have two or four linear characters:

	1	x	x^2	x^3
ψ_1	1	1	1	1
ψ_2	1	-1	1	-1
ψ_3	1	ζ	-1	ζ^3
ψ_4	1	ζ^3	-1	ζ

with the last two characters occurring only when $p \equiv 1 \pmod 4$. These linear characters corresponded to the possible groups $P_1, \dots, P_6$ found in [Kohl 2007] by other methods.

The following lemma is critical for the results in the next section. Let $\iota : \mathfrak{Q} \to \mathbb{F}_p^\times$ be the trivial linear character, $\iota(\gamma) = 1$ for all γ in $\mathfrak{Q}$. Then $\hat{p}_\iota = [1, \dots, 1] = \pi$, the generator of $\mathcal{P}$.

Lemma 2.2. *Let χ_1 and χ_2 be distinct nontrivial linear characters of $\mathfrak{Q}$. Then $\langle \hat{p}_{\chi_1}, \hat{p}_{\chi_2} \rangle$ cannot contain $\hat{p}_\iota$.*

Proof. If $\hat{p}_\iota = r\hat{p}_{\chi_1} + s\hat{p}_{\chi_2}$, then for all γ in $\mathfrak{Q}$ we have

$$1 = r\chi_1(\gamma) + s\chi_2(\gamma).$$

Hence

$$m = r\sum_{\gamma\in\mathfrak{Q}} \chi_1(\gamma) + s\sum_{\gamma\in\mathfrak{Q}} \chi_2(\gamma). \tag{1}$$

But for $i = 1, 2$, if $T_i = \chi_i(\mathfrak{Q}) \subset \mathbb{F}_p^\times$, then

$$\sum_{\gamma\in\mathfrak{Q}} \chi_i(\gamma)$$

is $[\mathbb{F}_p^\times : T_i]$ times the sum of the elements of T_i. Since $\mathbb{F}_p^\times$ is a cyclic group, T_i is a cyclic subgroup of $\mathbb{F}_p^\times$, hence elements of T_i sum to $0 \pmod p$. So (1) becomes $m = 0 \pmod p$. Thus it is impossible for $\hat{p}_\iota = r\hat{p}_{\chi_1} + s\hat{p}_{\chi_2}$. □

3. The main theorem

Let N be a regular subgroup of $B = \mathrm{Perm}(\Gamma)$. Let $\lambda(\Gamma) = \mathcal{P}\cdot\mathfrak{Q}$ where $\mathcal{P}$ is the p-Sylow subgroup of $\lambda(\Gamma)$. Our main theorem, Theorem 3.5, is

$$N \text{ is a subgroup of } \mathrm{Norm}_B(\mathcal{P}).$$

As we'll see in Theorem 3.7, $\mathrm{Norm}_B(\mathcal{P})$ can be viewed as a subgroup of the affine group of $\mathbb{F}_p^m$ generated by scalar matrices, permutation matrices, and $\mathbb{F}_p^m$. So this result reduces the question of determining regular subgroups of $\mathrm{Perm}(\Gamma) \cong S_{mp}$ to a question about subgroups of a much smaller group, a semidirect product of S_m with a metabelian group.

We begin by studying $\mathrm{Norm}_B(N)$, for N a regular subgroup of $B = \mathrm{Perm}(\Gamma)$.

Recall that the normalizer $\mathrm{Norm}_B(\lambda(\Gamma))$ in $\mathrm{Perm}\,\Gamma$ of $\lambda(\Gamma)$ is denoted by $\mathrm{Hol}\,\Gamma$ and is the group $\mathrm{Hol}(\Gamma) = \rho(\Gamma) \rtimes \mathrm{Aut}(\Gamma) \cong \Gamma \rtimes \mathrm{Aut}(\Gamma)$, where ρ is the right regular representation of Γ in $\mathrm{Perm}\,\Gamma$ and $\mathrm{Aut}\,\Gamma$ is embedded inside $\mathrm{Perm}\,\Gamma$ in the natural way. Since $\mathrm{Perm}(\Gamma) \cong \mathrm{Perm}(N)$ if N is a regular subgroup of $\mathrm{Perm}\,\Gamma$, we have:

Proposition 3.1. *Let N be a regular subgroup of $B = \mathrm{Perm}(\Gamma)$. Then*

$$\mathrm{Norm}_B(N) \cong \mathrm{Hol}(N).$$

Proof. Since N is regular in $\mathrm{Perm}\,\Gamma$, the map $b : N \to \Gamma$ by $b(\eta) = \eta(1)$ is a bijection. So $C(b^{-1}) : \mathrm{Perm}(\Gamma) \to \mathrm{Perm}(N)$, given by $C(b^{-1})(\pi) = b^{-1}\pi b$, is an isomorphism. Under this map, η in $N \subset \mathrm{Perm}(\Gamma)$ maps to $b^{-1}\eta b$, where for μ in N,

$$b^{-1}\eta b(\mu) = b^{-1}\eta(\mu(1)) = b^{-1}(\eta\mu(1)) = \eta\mu.$$

Thus inside $\mathrm{Perm}\,N$, the image $C(b^{-1})(N) = \lambda(N)$, and so

$$C(b^{-1})(\mathrm{Norm}_B(N)) = \mathrm{Norm}_{\mathrm{Perm}(N)}(\lambda(N)) \cong N \rtimes \mathrm{Aut}(N).$$

Since $C(b^{-1})$ is an isomorphism from $\mathrm{Perm}\,\Gamma$ to $\mathrm{Perm}\,N$, $C(b^{-1})$ is an isomorphism from $\mathrm{Norm}_B(N)$ to $\mathrm{Hol}(N) \cong N \rtimes \mathrm{Aut}(N)$. □

In order to obtain Theorem 3.5, we need to introduce the *opposite group*, $N^{\mathrm{opp}} = \mathrm{Cent}_B(N)$, the centralizer of N in $B = \mathrm{Perm}(\Gamma)$. We denote by 1 the identity element of the set Γ on which B acts. The following is a recapitulation of [Greither and Pareigis 1987, Lemma 2.4.2].

Lemma 3.2. *For N a regular subgroup of $B = \mathrm{Perm}(\Gamma)$, let ϕ be in $\mathrm{Cent}_B(N)$. Then $\phi(\gamma) = \eta_\gamma\phi(1)$, where η_γ is the unique element η of N such that $\eta(1) = \gamma$. Conversely, if ϕ is in B and $\phi(\gamma) = \eta_\gamma\phi(1)$ for all γ, then ϕ is in $\mathrm{Cent}_B(N)$.*

Proof. For ϕ in $\mathrm{Cent}_B(N)$, $\phi(\gamma) = \phi(\eta_\gamma(1)) = \eta_\gamma\phi(1)$. Let $\phi(1) = \sigma(1)$ for unique σ in N. Then ϕ is uniquely determined by σ: denote that ϕ by ϕ_σ. Thus $\phi_\sigma(\gamma) = \eta_\gamma\sigma(1)$.

Conversely, suppose ϕ is in B and there is some σ in N such that $\phi(\gamma) = \eta_\gamma\sigma(1)$ for all γ, so that $\phi = \phi_\sigma$. Then ϕ_σ is in $\mathrm{Cent}_B(N)$. Indeed,

$$\phi_\sigma\eta_\epsilon(\gamma) = \phi_\sigma\eta_{\eta_\epsilon}(\gamma) = \eta_{\eta_\epsilon}(\gamma)\sigma(1),$$

while

$$\eta_\epsilon\phi_\sigma(\gamma) = \eta_\epsilon\eta_\gamma\sigma(1).$$

We claim that $\eta_{\eta_\epsilon(\gamma)} = \eta_\epsilon\eta_\gamma$. Since elements η of N bijectively correspond with their images $\eta(1)$ in Γ, it suffices to observe that

$$\eta_{\eta_\epsilon(\gamma)}(1) = \eta_\epsilon(\gamma) = \eta_\epsilon(\eta_\gamma(1)) = (\eta_\epsilon\eta_\gamma)(1).$$

Thus $\mathrm{Cent}_B(N) = \{\phi_\sigma : \sigma \in N\}$. □

Corollary 3.3. *Let N be a regular subgroup of* $\mathrm{Perm}\,\Gamma$. *Then*:

(a) *N^{opp} is also a regular subgroup of* $\mathrm{Perm}\,\Gamma$.

(b) *$N \cap N^{\mathrm{opp}} = Z(N)$, the center of N.*

(c) *If N is abelian, then $N = N^{\mathrm{opp}}$.*

(d) $(N^{\mathrm{opp}})^{\mathrm{opp}} = N$.

Proof. (a) Observe that for σ in N, $\phi_\sigma(1) = \eta_1\sigma(1)$. But η_1 is the unique element of N that maps 1 to 1 in Γ, hence η_1 is the identity element of N. Thus $\phi_\sigma(1) = \sigma(1)$. Thus if N is regular, then so is N^{opp}.

(b), and hence (c), are clear since $N^{\mathrm{opp}} = \mathrm{Cent}_B(N)$.

(d) Clearly N is contained in the centralizer of $\mathrm{Cent}_B(N)$, so is in $(N^{\mathrm{opp}})^{\mathrm{opp}}$. But by (a), this last group is regular; hence it has the same cardinality as N. So $N = (N^{\mathrm{opp}})^{\mathrm{opp}}$. □

Proposition 3.4. $\mathrm{Norm}_B(N) = \mathrm{Norm}_B(N^{\mathrm{opp}})$. *Hence N is normalized by $\lambda(\Gamma)$ if and only if N^{opp} is normalized by $\lambda(\Gamma)$.*

Proof. We show that $N^{\mathrm{opp}} = \mathrm{Cent}_B(N)$ is a normal subgroup of $\mathrm{Norm}_B(N)$. Let α be in $\mathrm{Cent}_B(N)$, δ in $\mathrm{Norm}_B(N)$. We show $\delta\alpha\delta^{-1}$ is in $\mathrm{Cent}_B(N)$. Since every element η of N has the form $\delta\sigma\delta^{-1}$ for some σ in N and $\alpha\sigma = \sigma\alpha$, we have

$$\begin{aligned}\delta\alpha\delta^{-1}\eta &= \delta\alpha\delta^{-1}(\delta\sigma\delta^{-1}) = \delta\alpha\sigma\delta^{-1}\\ &= \delta\sigma\alpha\delta^{-1} = \delta\sigma\delta^{-1}\delta\alpha\delta^{-1} = \eta\delta\alpha\delta^{-1}.\end{aligned}$$

Thus $\delta\alpha\delta^{-1}$ is in $\mathrm{Cent}_B(N)$, and so N^{opp} is a normal subgroup of $\mathrm{Norm}_B(N)$. Hence

$$\mathrm{Norm}_B(N) \subset \mathrm{Norm}_B(N^{\mathrm{opp}}).$$

The same is true replacing N by N^{opp}. Equality then follows by part (d) of Corollary 3.3. The last sentence follows easily from the equality $\mathrm{Norm}_B(N) = \mathrm{Norm}_B(N^{\mathrm{opp}})$. □

Now we can prove the main theorem.

Theorem 3.5. *Let N be a regular subgroup of $B = \mathrm{Perm}(\Gamma)$ normalized by $\lambda(\Gamma) = \mathcal{P}\cdot\mathcal{Q}$, with $\mathcal{P}$ the p-Sylow subgroup of $\lambda(\Gamma)$. Then N is a subgroup of* $\mathrm{Norm}_B(\mathcal{P})$.

Proof. Since $\lambda(\Gamma)$ is contained in $\mathrm{Norm}_B(N)$, we have $\mathcal{P}$ inside $\mathrm{Norm}_B(N) = \mathrm{Norm}_B(N^{\mathrm{opp}})$.

Since $\mathrm{Norm}_B(N) \cong \mathrm{Hol}(N) = N \rtimes \mathrm{Aut}(N)$, we know by Proposition 1.2 what the p-Sylow subgroup of $\mathrm{Norm}_B(N)$ is:

• If $N = P(N) \times Q(N)$, then the p-Sylow subgroup of $\mathrm{Norm}_B(N)$ is $P(N)$, which is unique and has order p. Hence $\mathcal{P} = P(N) = P(N^{\mathrm{opp}})$.

• If $N = P(N) \rtimes_\tau Q(N)$ where τ is nontrivial, then $\mathrm{Norm}_B(N) \cong \mathrm{Hol}(N) \cong N \rtimes \mathrm{Aut}(N)$ has a p-Sylow subgroup isomorphic to $C_p \times C_p$, where one copy of C_p is $P(N)$ and the other copy is the group $C(P(N))$ of inner automorphisms of N obtained by conjugation by the elements of $P(N)$ (see Lemma 1.1). We check that

the subgroup $P(N)\cdot C(P(N))$ is normal in $\mathrm{Hol}(N) = N \rtimes \mathrm{Aut}(N)$. Take $\sigma, \tau \in P$, $h \in G$, $\alpha \in \mathrm{Aut}\, G$. Then

$$(\alpha(h)^{-1}\alpha)(h\alpha^{-1}) = 1,$$

so conjugating an element $\sigma C(\tau)$ of $P(N)\cdot C(P(N))$ by $(h\alpha^{-1})^{-1}$ yields:

$$\begin{aligned}(\alpha(h)^{-1}\alpha)(\sigma C(\tau))(h\alpha^{-1}) &= \alpha(h)^{-1}\alpha(\sigma)\alpha(\tau h\tau^{-1})\cdot \alpha C(\tau)\alpha^{-1}\\ &= \alpha(h)^{-1}\alpha(\sigma)\alpha(\tau)\alpha(h)\alpha(\tau^{-1})\cdot C(\alpha(\tau))\\ &= C(\alpha(h)^{-1})(\alpha(\sigma\tau))\alpha(\tau^{-1})\cdot C(\alpha(\tau)).\end{aligned}$$

Since P is a characteristic subgroup of G, $C(\alpha(h)^{-1})(\alpha(\sigma\tau))$ is in P, as are $\alpha(\tau^{-1})$ and $\alpha(\tau)$. Hence $P(N)\cdot C(P(N))$ is a normal subgroup of $\mathrm{Hol}\, N$, hence is the unique p-Sylow subgroup of $\mathrm{Hol}\, N$.

Since N in this case is nonabelian, $Z(N)$ has no p-torsion, and so since $N \cap N^{\mathrm{opp}} = Z(N)$, $P(N)\cap P(N^{\mathrm{opp}}) = (1)$. Since $P(N)$ and $P(N^{\mathrm{opp}})$ centralize each other, $P(N)\cdot P(N^{\mathrm{opp}}) \cong C_p \times C_p$, and hence $P(N)\cdot P(N^{\mathrm{opp}})$ is the p-Sylow subgroup of $\mathrm{Hol}(N) = \mathrm{Norm}_B(N)$.

Now we identify $\mathcal{P}$, the p-Sylow subgroup of $\lambda(\Gamma)$, inside $\mathrm{Norm}_B(N)$. Clearly, $\mathcal{P} \subset P(N)\cdot P(N^{\mathrm{opp}})$. The groups $\mathcal{P}$, $P(N)$, and $P(N^{\mathrm{opp}})$ are order-p semiregular subgroups of $\mathrm{Perm}\,\Gamma$ normalized by $\lambda(\Gamma)$; hence they have generators $\hat{p}_\iota$, $\hat{p}_{\chi_1}$, and $\hat{p}_{\chi_2}$ that correspond to linear characters ι, χ_1, and χ_2 from $\mathfrak{Q} = Q(\lambda(\Gamma))$ to $\mathbb{F}_p^\times$, where ι, corresponding to $\mathcal{P}$, is the trivial character. Since $P(N)$ and $P(N^{\mathrm{opp}})$ are distinct subgroups, χ_1 and χ_2 are distinct characters. Since $\mathcal{P}$ is contained in $P(N)\cdot P(N^{\mathrm{opp}})$, we have

$$\iota = r\chi_1 + s\chi_2,$$

for some integers r and s. But by Lemma 2.2, this can only occur if χ_1 or χ_2 is the trivial character, that is, $\mathcal{P} = P(N)$ or $\mathcal{P} = P(N^{\mathrm{opp}})$.

If $\mathcal{P} = P(N^{\mathrm{opp}})$, then N centralizes $\mathcal{P}$, so N is contained in $\mathrm{Norm}_B(\mathcal{P})$.

If $\mathcal{P} = P(N)$, then N normalizes $P(N) = \mathcal{P}$, so N is contained in $\mathrm{Norm}_B(\mathcal{P})$. □

Definition. For groups Γ and M of order mp and P an order-p semiregular subgroup of $\mathrm{Norm}_B(\mathcal{P})$ that is normalized by $\mathrm{Norm}_B(\mathcal{P})$ (see Theorem 2.1), let $R(\Gamma, [M]; P)$ be the set of regular subgroups N of $\mathrm{Norm}_B(\mathcal{P})$ isomorphic to M and normalized by $\lambda(\Gamma)$ such that $P(N) = P$.

Then $R(G, [M])$ is the disjoint union of $R(\Gamma, [M]; P)$ for P running through all order-p semiregular subgroups of $\mathrm{Norm}_B(\mathcal{P})$.

To count $R(G, [M])$, we combine Proposition 3.4 with the proof of Theorem 3.5:

Corollary 3.6. *With Γ and M as above, let $\mathcal{P} = P(\lambda(\Gamma))$, the p-Sylow subgroup of $\lambda(\Gamma)$.*

If $M = P(N) \times Q(N)$, then $R(\Gamma, [M]) = R(\Gamma, [M]; \mathcal{P})$.
If M is a nontrivial semidirect product of $P(N)$ and $Q(N)$, then

$$|R(G, [M])| = 2|R(G, [M]; \mathcal{P})|.$$

Proof. Lemma 1.1 showed that if N is the direct product of $P(N)$ and $Q(N)$, then $\mathcal{P}$ is the unique order-p subgroup of $\mathrm{Norm}_B(\mathcal{P})$, hence $P(N) = \mathcal{P}$ for all regular subgroups of $\mathrm{Norm}_B(\mathcal{P})$ normalized by $\lambda(\Gamma)$. Otherwise, N and N^{opp} are regular subgroups of $\mathrm{Perm}\,\Gamma$ normalized by $\lambda(\Gamma)$ such that $P(N)$ and $P(N^{\mathrm{opp}})$ are distinct subgroups of $\mathrm{Norm}_B(\mathcal{P})$, and as observed at the end of the proof of Theorem 3.5, exactly one of $P(N)$ and $P(N^{\mathrm{opp}})$ is equal to $\mathcal{P}$. Thus when M is a nontrivial semidirect product, counting $R(\Gamma, [M]; \mathcal{P})$ counts half of the set $R(\Gamma, [M])$. □

Now we identify $\mathrm{Norm}_B(\mathcal{P})$ as a semidirect product and as a subgroup of the affine group of $\mathbb{F}_p^m$. The first description makes computing regular subgroups of $\mathrm{Norm}_B(\mathcal{P})$ feasible in many cases.

Theorem 3.7. *Let $\lambda(\Gamma) = \mathcal{P}\mathcal{Q}$, where $\mathcal{P} = \langle \pi \rangle$, $\pi = \pi_1\pi_2 \cdots \pi_m$, a product of disjoint p-cycles in $B = \mathrm{Perm}(\Gamma)$. Let $V = \langle \pi_1, \ldots, \pi_m \rangle \cong \mathbb{F}_p^m$, as before. Then $\mathrm{Norm}_B(\mathcal{P}) \cong \mathbb{F}_p^m \rtimes (\mathbb{F}_p^\times \cdot S_m)$ and embeds in*

$$\mathrm{AGL}_m(\mathbb{F}_p) = \left\{ \begin{pmatrix} A & \hat{v} \\ 0 & 1 \end{pmatrix} : A \in \mathrm{GL}_m(\mathbb{F}_p),\ \hat{v} \in \mathbb{F}_p^m \right\},$$

the affine group of $\mathbb{F}_p^m$.

Proof. We first show that $\mathrm{Norm}_B(\mathcal{P}) \cong \mathbb{F}_p^m \rtimes (\mathbb{F}_p^\times \cdot S_m)$.

Given an element τ of $\mathrm{Norm}_B(\mathcal{P})$, $\tau\pi\tau^{-1} = \pi^{c(\tau)}$, and so τ induces a permutation, denoted by t_τ, of the set $\{1, 2, \ldots, m\}$ by

$$\tau\pi_j\tau^{-1} = \pi_{t_\tau(i)}^{c(\tau)}.$$

This defines homomorphisms $c : \mathrm{Norm}_B(\mathcal{P}) \to \mathbb{F}_p^\times$, $t : \mathrm{Norm}_B(\mathcal{P}) \to S_m$, and $\phi : \mathrm{Norm}_B(\mathcal{P}) \to \mathbb{F}_p^\times \cdot S_m$ by $\phi(\tau) = (c(\tau), t(\tau))$. The kernel $\ker\phi$ of ϕ is the set of elements τ in $\mathrm{Norm}_B(\mathcal{P})$ such that $\tau\pi_j\tau^{-1} = \pi_j$ for all j, that is, the centralizer of V. We show that $\ker\phi = V$.

For $i = 1, \ldots, m$, choose γ_i in $\Pi_i = \mathrm{Supp}(\pi_i)$. Then π_i is the p-cycle

$$\pi = (\gamma_i, \pi(\gamma_i), \ldots, \pi^{p-1}(\gamma_i)),$$

hence

$$\Gamma = \{\pi_i^k(\gamma_i) \mid i = 1, \ldots, m, k = 0, \ldots, p-1\}.$$

If τ in $\mathrm{Perm}\,\Gamma$ centralizes π_i, then since

$$\tau\pi_i\tau^{-1} = \big(\tau(\gamma_i), \tau(\pi(\gamma_i)), \ldots, \tau(\pi^{p-1}(\gamma_i))\big) = \pi_i,$$

τ conjugates $\mathrm{Supp}(\pi_i) = \Pi_i$ to itself, and hence yields a permutation of the set Π_i. But the only permutations in $S_p = \mathrm{Perm}(\Pi_i)$ that centralize the p-cycle π_i are the powers of π_i. Thus τ commutes with π_i for all $i = 1, \dots, m$ if and only if τ is in V. Therefore $V = \ker \phi$ and we have a short exact sequence:

$$1 \to V \to \mathrm{Norm}_B(\mathcal{P}) \to \mathbb{F}_p^\times \cdot S_m \to 1.$$

The sequence splits. For inside $\mathrm{Norm}_B(\mathcal{P})$ are the permutations σ_c for c in $\mathbb{F}_p^\times$ induced by the c-th power map $\pi \mapsto \pi^c$, for $(c, p) = 1$, that take $\pi_i^k(\gamma_i)$ to $\pi_i^{ck}(\gamma_i)$ for all $i = 1, \dots, m$ and $k = 0, \dots, p-1$. The σ_c generate a subgroup $\mathcal{U}$ of $\mathrm{Norm}_B(\mathcal{P})$ isomorphic to $\mathbb{F}_p^\times$. Also, a permutation $\bar{\alpha}$ of S_m defines a permutation α of $\mathrm{Perm}\,\Gamma$ by

$$\alpha(\pi_i^k(\gamma_i)) = \pi_{\bar{\alpha}(i)}^k(\gamma_{\bar{\alpha}(i)}).$$

Then $\{\alpha \in \mathrm{Perm}(\Gamma) : \bar{\alpha} \in S_m\}$ is a subgroup $\mathcal{S}$ of $\mathrm{Norm}_B(\mathcal{P})$ isomorphic to S_m. Clearly $\mathcal{S}$ and $\mathcal{U}$ centralize each other, so the group $\mathcal{S}\mathcal{U} \subset \mathrm{Norm}_B(\mathcal{P})$ is a preimage of $\mathbb{F}_p^\times \cdot S_m$ under ϕ. So ϕ splits, and $\mathrm{Norm}_B(\mathcal{P}) = V \cdot (\mathcal{U}\mathcal{S}) \cong \mathbb{F}_p^m \rtimes (\mathbb{F}_p^\times \cdot S_m)$.

A convenient way to view $\mathbb{F}_p^m \rtimes (\mathbb{F}_p^\times \cdot S_m)$ is as the subgroup of $\mathrm{AGL}_m(\mathbb{F}_p)$ consisting of matrices

$$\begin{pmatrix} A & v \\ 0 & 1 \end{pmatrix},$$

where $v \in V = \mathbb{F}_p^m$, and A in $\mathrm{GL}_m(\mathbb{F}_p)$ is a nonzero scalar multiple of a permutation matrix. In other words, we view S_m as $m \times m$ permutation matrices of the components of $\mathbb{F}_p^m$ and $\mathbb{F}_p^\times$ as nonzero scalar multiples (in $\mathbb{F}_p$) of the $m \times m$ identity matrix. Such matrices are examples of monomial matrices, whose properties in general are explored by various authors such as Ore [1942]. □

In the sequel we will need to understand $\mathrm{Norm}_B(\mathcal{P})$ as a subgroup of $B = \mathrm{Perm}(\Gamma)$. Writing the elements of $\mathrm{Norm}_B(\mathcal{P}) = V \cdot (\mathcal{U}\mathcal{S})$ as $(\hat{a}, u^r, \alpha)$, the explicit action of elements of $\mathrm{Norm}_B(\mathcal{P})$ on $\Gamma = \{\pi_i^k(\gamma_i) \mid i = 1, \dots, m, k = 0, \dots, p-1\}$ is given by

$$(\hat{a}, u^r, \alpha)(\pi_i^k(\gamma_i)) = \pi_1^{a_1} \cdots \pi_m^{a_m}(\pi_{\alpha(i)}^{ku^r}(\gamma_{\alpha(i)})) = \pi_{\alpha(i)}^{ku^r + a_{\alpha(i)}}(\gamma_{\alpha(i)}).$$

Then we have the following easily verified formulas:

$$(\hat{a}, u^r, \alpha)^k = \left(\sum_{i=0}^{k-1} u^{ir} \alpha^r(\hat{a}),\ u^{rk},\ \alpha^k \right). \tag{2}$$

The inverse of $(\hat{b}, u^s, \beta)$ is $(-u^{-s}\beta^{-1}(\hat{b}), u^{-s}, \beta^{-1})$, so

$$(\hat{b}, u^s, \beta)(\hat{a}, u^r, \alpha)(\hat{b}, u^s, \beta)^{-1} = (\hat{b} + u^s \beta(\hat{a}) - u^r(\beta\alpha\beta^{-1})(\hat{b}), u^r, \beta\alpha\beta^{-1}).$$

In particular, elements of $\mathrm{Norm}_B(\mathcal{P})$ act on $\mathcal{P}$ by:

$$(\hat{b}, u^s, \beta)(\hat{p}_\iota, 1, I)(\hat{b}, u^s, \beta)^{-1} = (u^s \hat{p}_\iota, 1, I).$$

Let N be a regular subgroup of $\mathrm{Perm}\,\Gamma$ normalized by $\lambda(\Gamma)$ and recall that $N = P(N)Q(N)$ where $P(N)$ is the p-Sylow subgroup of N and $Q(N)$ is a group of order m. We know that $N \subset \mathrm{Norm}_B(\mathcal{P})$ and that $P(N) = \langle(\hat{p}_\chi, 1, I)\rangle$ for some linear character from $\mathfrak{Q} = Q(\lambda(\Gamma))$ to $\mathbb{F}_p^\times$. We need to examine $Q(N)$.

Now N is a regular subgroup of $\mathrm{Perm}\,\Gamma$, so $Q(N)$ acts fixed-point-freely on Γ. We need to identify fixed-point-free elements of $\mathrm{Norm}_B(\mathcal{P})$.

Proposition 3.8. *If the order of* $(\hat{a}, u^r, \alpha) \neq 1$ *in* $\mathrm{Norm}_B(\mathcal{P})$ *is coprime to* p, *then* $(\hat{a}, u^r, \alpha)$ *is fixed-point free on* Γ *if and only if* α *is fixed-point free in* S_m.

Proof. Suppose α is fixed-point free in S_m. Then for all i, $1 \le i \le m$, $\alpha(i) \neq i$, so $(\hat{a}, u^r, \alpha)(\pi_i^k(\gamma_i))$ is in $\Pi_{\alpha(i)} \neq \Pi_i$. So $(\hat{a}, u^r, \alpha)$ is fixed-point free.

Suppose $\alpha(i) = i$ for some i. Then

$$(\hat{a}, u^r, \alpha)(\pi_i^k(\gamma_i)) = \pi_i^{u^r k + a_i} = \pi_i^k,$$

for k satisfying $(1-u^r)k \equiv a_i \pmod p$. If $u^r \neq 1$, then such a k exists, so $(\hat{a}, u^r, \alpha)$ has a fixed point whenever α has a fixed point and $u^r \neq 1$.

If $\alpha(i) = i$ and $u^r = 1$, then

$$(\hat{a}, 1, \alpha)^s(\pi_i^k(\gamma_i)) = \pi_i^{k + a_i s}(\gamma_i),$$

for all s. If s is the order of $(\hat{a}, 1, \alpha)$, then $\pi_i^{k+a_i s}(\gamma_i) = \pi_i^k(\gamma_i)$, so $a_i s \equiv 0 \pmod p$. If s and p are coprime, then $a_i = 0$. But then $\pi_i^k(\gamma_i)$ is a fixed point for $(\hat{a}, 1, \alpha)$. □

Let $t : \mathrm{Norm}_B(\mathcal{P}) \to S_m$ be the map sending $(\hat{a}, u^r, \alpha)$ to $\bar{\alpha}$ in S_m defined by $\alpha(\pi_i^k(\gamma_i)) = \pi_{\bar{\alpha}(i)}^k(\gamma_{\bar{\alpha}(i)})$. Proposition 3.8 implies immediately:

Corollary 3.9. *Let* Q *be a subgroup of* $\mathrm{Norm}_B(\mathcal{P})$ *of order* m, *and suppose* $t : \mathrm{Norm}_B(\mathcal{P}) \to S_m$ *is one-to-one on* Q. *Then* Q *is fixed-point free on* Γ, *hence a semiregular subgroup of* $\mathrm{Norm}_B(\mathcal{P})$, *if and only if* $t(Q)$ *is a regular subgroup of* S_m.

Corollary 3.10. *If* N *is a regular subgroup of* $\mathrm{Norm}_B(\mathcal{P})$, *then* $t(Q(N^{\mathrm{opp}})) = (t(Q(N)))^{\mathrm{opp}}$, *where the right-hand group is viewed within* $\mathcal{S} \cong S_m$.

Proof. For $(\hat{a}, u^r, \alpha)$ in $Q(N)$ and $(\hat{c}, u^s, \delta)$ in $Q(N^{\mathrm{opp}})$, we have $\alpha\delta = \delta\alpha$, so $t(\hat{a}, u^r, \alpha) = \bar{\alpha}$ and $t(\hat{c}, u^s, \delta) = \bar{\delta}$ commute in S_m. So $t(Q(N^{\mathrm{opp}})) \subset (t(Q(N)))^{\mathrm{opp}}$. But because $Q(N)$ is regular in S_m, both sides have cardinality m. Hence the two groups are equal. □

It is interesting to observe that $\mathrm{Cent}_B(\mathcal{P})$ consists precisely of those elements of the form $(\hat{b}, 1, \beta)$, which is consistent with the classical fact (due to Burnside [1911, §170]) that $\mathrm{Cent}_B(\mathcal{P})$ is isomorphic to the wreath product $C_p \wr S_m$. This

wreath product is isomorphic to the semidirect product $(C_p \times \cdots \times C_p) \rtimes S_m$ where the action of S_m on the m-fold product of the C_p's is given by the natural action on the coordinates. The group $\mathrm{Norm}_B(\mathcal{P})$ is also not unknown. It is an example of a *twisted* wreath product whose precise definition (which may be found in [Neumann 1963]) is not so important here since we have the semidirect product description given above. The appearance of wreath products, by the way, is a natural consequence of the action of $\mathrm{Norm}_B(\mathcal{P})$ (as well as any other subgroups thereof, such as $\mathrm{Cent}_B(\mathcal{P})$) on the blocks $\{\Pi_1, \ldots, \Pi_m\}$. We may, in fact, frame part of Theorem 3.5 in terms of one of the important consequences of the so-called universal embedding theorem of Krasner and Kaloujnine [1951]. Specifically, if one has an exact sequence $1 \to P \to N \to Q \to 1$, expressing N as an extension of P by Q, then $P \wr Q$ contains a subgroup isomorphic to N. In the setting of this work, where $|N| = |P| \cdot |Q| = pm$ our group Q may, of course, be embedded as a subgroup of S_m. As such we have an embedding of N into $P \wr S_m$. This dovetails with the above observation that $\mathrm{Cent}_B(\mathcal{P}) \cong C_p \wr S_m$ since, for a given $N \in R(\Gamma, [M])$, either N or N^{opp} centralizes $\mathcal{P}$ and $N \cong N^{\mathrm{opp}}$ so that indeed $\mathrm{Cent}_B(\mathcal{P})$ contains a subgroup isomorphic to N. One of the upshots of Corollary 3.6, in fact, is that either all $N \in R(\Gamma, [M])$ are subgroups of $\mathrm{Cent}_B(\mathcal{P})$ (when $P(N)$ is a direct factor) or (when $P(N)$ is not a direct factor) exactly half of the elements centralize $\mathcal{P}$, indeed all those for which $P(N) \neq \mathcal{P}$. As such, one *could* enumerate only those N that lie in $\mathrm{Cent}_B(\mathcal{P})$ and then apply Corollary 3.6 in order to determine $|R(\Gamma, [M])|$.

What the affine representation above yields for us is a very concrete way of performing the enumeration of these subgroups of $\mathrm{Norm}_B(\mathcal{P})$.

In order to apply Theorem 3.5 to deal with all possible Γ and all possible N of a given order mp, it is convenient to apply the following (in the author's opinion quite important) observation:

Proposition 3.11 [Dixon 1971, Lemma 1]. *If N and N' are regular subgroups of S_n that are isomorphic as abstract groups, they are conjugate as subgroups of S_n.*

Proof. Identify $S_n = \mathrm{Perm}(Z/nZ) = \mathrm{Perm}(C_n)$. Let $\phi : N \to N'$ be an isomorphism. Then the conjugation map $C(\phi) : \mathrm{Perm}(N) \to \mathrm{Perm}(N')$ is an isomorphism, under which $\lambda(N)$ maps to $\lambda(N')$, as is easily verified. If $b : N \to C_n$ and $c : N' \to C_n$ are bijections, then $C(b^{-1}) : \mathrm{Perm}(C_n) \to \mathrm{Perm}(N)$ maps N in $\mathrm{Perm}\, C_n$ to $\lambda(N)$ in $\mathrm{Perm}\, N$, and $C(c^{-1}) : \mathrm{Perm}(C_n) \to \mathrm{Perm}(N')$ maps N' in $\mathrm{Perm}\, C_n$ to $\lambda(N')$. The composition $C(c^{-1})C(\phi)C(b) = C(c^{-1} \circ \phi \circ b)$ maps N in $\mathrm{Perm}\, C_n$ to N' in $\mathrm{Perm}\, C_n$. $\square$

This result allows us to determine $R(\Gamma, [M])$, for all pairings of groups of order mp, while working entirely within the single group $B = S_{mp}$.

Here is an outline of the strategy.

Let $B = S_{mp}$.

Suppose that $\mathcal{P} = \langle \pi \rangle$ is a cyclic semiregular subgroup of B of order p and that $\pi = \pi_1 \cdot \pi_2 \cdot \cdots \cdot \pi_m$, where $\pi_1, \ldots, \pi_m$ are disjoint p-cycles. We may choose $\mathcal{P}$ at our convenience.

Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_s$ be subgroups of $\mathrm{Norm}_B(\mathcal{P})$ that act regularly on the set $\{\Pi_1, \ldots, \Pi_m\}$, where $\Pi_i = \mathrm{Supp}(\pi_i)$, and represent all isomorphism classes of groups of order m.

For each $\mathcal{Q}_i$, find the $\mathbb{F}_p$-linear characters χ_{ij} of $\mathcal{Q}_i$. Then $\langle \hat{p}_{\chi_{ij}} \rangle$ is normalized by $\mathcal{Q}_i$, so, as we shall show below, $\langle \hat{p}_{\chi_{ij}} \rangle \mathcal{Q}_i$ is a regular subgroup of S_{mp} and is contained in $\mathrm{Norm}_B(\mathcal{P})$. If $\langle \hat{p}_{\chi_{ij}} \rangle \mathcal{Q}_i$ is a direct product or χ_{ij} is not the trivial character, we find $(\langle \hat{p}_{\chi_{ij}} \rangle \mathcal{Q}_i)^{\mathrm{opp}}$ in S_{mp}. Then $(\langle \hat{p}_{\chi_{ij}} \rangle \mathcal{Q}_i)^{\mathrm{opp}}$ is contained in $\mathrm{Norm}_B(\mathcal{P})$ and its p-Sylow subgroup is $\mathcal{P}$. We represent the isomorphism types of groups Γ by suitable groups $(\langle \hat{p}_{\chi_{ij}} \rangle \mathcal{Q}_i)^{\mathrm{opp}}$.

Having done so, we then seek to construct regular subgroups N normalized by Γ by looking for fixed-point-free elements in $\mathrm{Norm}_B(\mathcal{P})$ of suitable orders that are normalized by Γ.

In the next sections we demonstrate this program.

4. Groups of order pq

N. Byott [2004] determined the number of Hopf Galois structures on a Galois extension of fields L/K with Galois group Γ of order pq where p and q are primes and $p \equiv 1 \pmod{q}$. As Byott notes, the case where $p \not\equiv 1 \pmod{q}$ is of little interest because then pq and $\phi(pq)$ are coprime, in which case Byott [1996] shows that the only Hopf Galois structure on LK is the classical structure given by the Galois group Γ.

Let G_1 and G_2 be the two isomorphism types of groups of order pq. Byott's [2004] approach for counting Hopf Galois structures is to apply the strategy, suggested in [Childs 1989] and codified in [Byott 1996], of looking for regular subgroups isomorphic to G_i inside $\mathrm{Hol}(G_j) \cong G_j \rtimes \mathrm{Aut}(G_j)$ for $i, j = 1, 2$. Equivalence classes of such regular subgroups correspond to Hopf Galois structures on field extensions with Galois group G_i whose Hopf algebra has type G_j.

In this section we count the number of Hopf Galois structures on L/K with Galois group G_i whose Hopf algebra has type G_j by looking for regular subgroups G_j inside $\mathrm{Norm}_{\mathrm{Perm}(G_i)}(\mathcal{P}) \subset \mathrm{Perm}(G_i)$. Thus we obtain Byott's count by a refinement of the direct Greither–Pareigis approach. As may be observed, the two methods are rather different.

Let $\mathbb{F}_p^\times = \langle u \rangle$. The two groups of order pq are the cyclic group $C_{pq} \cong \mathbb{F}_p \times \langle u^d \rangle$ and the group $C_p \rtimes_\tau C_q = \mathbb{F}_p \rtimes \langle u^d \rangle$, where in $C_p \rtimes_\tau C_q$ we have $(0, u^d)(x, 1) = (u^d x, 1)(0, u^d)$ and $qd = p - 1$; hence u^d is an element of $\mathbb{F}_p^\times$ of order q.

The result is:

Theorem 4.1. *Let $R(\Gamma, [G])$ be the regular subgroups of* $\operatorname{Perm}\Gamma$ *isomorphic to G and normalized by $\lambda(\Gamma)$. Then*

$$\begin{aligned} |R(C_{pq}, [C_{pq}])| &= 1, \\ |R(C_{pq}, [C_p \rtimes_\tau C_q])| &= 2(q-1), \\ |R(C_p \rtimes_\tau C_q, [C_{pq}])| &= p, \\ |R(C_p \rtimes_\tau C_q, [C_p \rtimes_\tau C_q])| &= 2(1+p(q-2)). \end{aligned}$$

By [Greither and Pareigis 1987], in each case the right-hand side equals the number of Hopf Galois structures on a Galois extension of fields with Galois group Γ with Hopf algebra of type $[M]$.

Before doing the particular cases, we obtain some preliminary information that applies in all four cases. Also, some notational conventions will be used throughout the rest of the paper. In $\mathbb{F}_p^m$ we shall denote the vectors $[0, 0, \dots, 0]$ and $[1, 1, \dots, 1] = \hat{p}_\iota = \langle \pi \rangle$ (both of which are fixed by any $\alpha \in S_m$) by $\hat{0}$ and $\hat{1}$, respectively, and any scalar multiple $[c, c, \dots, c]$ of $\hat{1}$ shall be expressed as $c\hat{1}$. Also, an arbitrary $\hat{a} \in \mathbb{F}_p^m$ has the form $[a_1, a_2, \dots, a_m]$ for $a_i \in \mathbb{F}_p$.

Lemma 4.2. *Suppose*

$$G = \langle (\hat{1}, 1, I), (\hat{a}, u^r, \sigma) \rangle \subset \operatorname{Norm}_B(\mathcal{P}),$$

where $x = (\hat{1}, 1, I)$ and $y = (\hat{a}, u^r, \sigma)$ satisfy $x^p = y^q = 1$ and $yx = x^{u^d} y$ and σ is a nontrivial permutation of S_q. Then σ is a q-cycle in S_q and $u^r = u^d$.

Proof. If $(\hat{a}, u^r, \sigma)^q = (\hat{1}, 1, I)$, then $\sigma^q = 1$. Since σ is nontrivial, it must have order q, hence be a q-cycle since q is prime. From the defining relation

$$(\hat{a}, u^r, \sigma)(\hat{1}, 1, I) = (\hat{1}, 1, I)^{u^d} (\hat{a}, u^r, \sigma),$$

we have $\hat{a} + u^r \hat{1} = u^d \hat{1} + \hat{a}$, hence $u^r = u^d$. □

Lemma 4.3. *Suppose G is as in Lemma 4.2 and*

$$H = \langle (\hat{1}, 1, I), (\hat{b}, u^s, \alpha) \rangle \subset \operatorname{Norm}_B(\mathcal{P}),$$

with α a q-cycle. If H is normalized by G, then $\alpha = \sigma^t$ for some $t \in \mathbb{F}_p^\times$.

Proof. Since G normalizes H, G must conjugate the generator of H of order q to an element of H. Thus

$$(\hat{a}, u^r, \sigma)(\hat{b}, u^s, \alpha)(\hat{a}, u^r, \sigma)^{-1} = (\hat{1}, 1, I)^f (\hat{b}, u^s, \alpha)^e,$$

for some $f \in \mathbb{F}_p$ and $e \in \mathbb{F}_p^\times$. Looking at the rightmost components, we have

$$\sigma \alpha \sigma^{-1} = \alpha^e.$$

Since conjugation by the order-q element σ is an automorphism of the cyclic q group $\langle\alpha\rangle$, whose automorphism group has order $q-1$, conjugation by σ must be trivial on $\langle\alpha\rangle$. Hence $\alpha\sigma=\sigma\alpha$. Now α is the q-cycle

$$\alpha=(1,\alpha(1),\dots,\alpha^r(1),\dots).$$

So

$$\alpha=\sigma\alpha\sigma^{-1}=(\sigma(1),\sigma\alpha(1),\dots,\sigma\alpha^r(1),\dots).$$

If $\sigma(1)=\alpha^k(1)$ for $k\neq 0$, then for all $s>0$,

$$\sigma(\alpha^s(1))=\alpha^s\sigma(1)=\alpha^s\alpha^k(1)=\alpha^k(\alpha^s(1)).$$

Hence $\sigma=\alpha^k$. □

We outline the strategy of the proof of Theorem 4.1.

Given that

$$\Gamma=\langle(\hat{1},1,I),(\hat{0},u^r,\sigma)\rangle,\qquad N=\langle(\hat{1},1,I),(\hat{a},u^s,\sigma^t)\rangle,$$

we know that $N\subset\mathrm{Norm}_B(\mathcal{P})$. The constraints on N arise from the requirements that, first, Γ normalizes N, and, second, $(\hat{a},u^s,\sigma^t)$ has order q. Regarding the first constraint, conjugating $(\hat{a},s^s,\sigma^t)$ by $(\hat{1},1,I)$ poses no constraint on N since

$$(\hat{1},1,I)(\hat{a},u^s,\sigma^t)=((1-u^s)\hat{1},1,I)(\hat{a},u^s,\sigma^t)\in N.$$

But the condition

$$(\hat{0},u^r,\sigma)(\hat{a},u^s,\sigma^t)(\hat{0},u^r,\sigma)\text{ is in }N \tag{3}$$

typically yields conditions on $\hat{a}$.

Now we do each case in turn.

$|R(C_{pq},[C_{pq}])|=1$. We identify $\Gamma=C_p\times C_q$ inside $\mathrm{Norm}_B(\mathcal{P})$ as

$$\Gamma=\langle(\hat{1},1,I),(0,1,\sigma)\rangle,$$

where σ is a fixed q-cycle in S_q. Then, since $N\cong C_p\times C_q$, N must have the form

$$N=\langle(\hat{1},1,I),(\hat{a},1,\sigma^t)\rangle,$$

for some integer t modulo $p-1$ by Lemmas 4.2 and 4.3.

Since $Q(N)$ is characteristic in N, condition (3) becomes the condition that $(0,1,\sigma)$ conjugates the generator $(\hat{a},1,\alpha)$ of $Q(N)$ to a power of itself:

$$(\hat{0},1,\sigma)(\hat{a},1,\sigma^t)(\hat{0},1,\sigma^{-1})=(\hat{a},1,\sigma)^e,$$

for some integer e. Looking at the rightmost components shows that $e=1$. Thus

$$N=\langle(\hat{1},1,I),(\hat{a},1,\sigma^t)\rangle,$$

and looking at the leftmost components yields that $\sigma(\hat{a}) = \hat{a}$, hence $\hat{a} = k\hat{1}$. Then

$$(k\hat{1}, 1, \sigma^t) = (\hat{1}, 1, I)^k(\hat{0}, 1, \sigma)^t$$

is in Γ. Hence $N = \Gamma$.

$|R(C_{pq}, [C_p \rtimes_\tau C_q])| = 2(q-1)$. Since $C_p \rtimes_\tau C_q$ is a nontrivial semidirect product, to count the regular subgroups N, by Corollary 3.6 we may restrict to those N such that $P(N) = \mathcal{P}$, hence $P(N) = \langle(\hat{1}, 1, I)\rangle$. Again, $\Gamma = \langle(\hat{1}, 1, I), (0, 1, \sigma)\rangle$. By Lemmas 4.2 and 4.3,

$$N = \langle(\hat{1}, 1, I), (\hat{a}, u^d, \sigma^t)\rangle,$$

where $(t, q) = 1$. We claim that $\hat{a} = \hat{0}$.

We first observe that we may replace the generator $(\hat{a}, u^d, \sigma^t)$ by $(\hat{a}, u^d, \sigma^t)(l\hat{1}, 1, I)$ for any l, and choose l so that $a_1 = 0$, where a_1 is the first component of $\hat{a} \in \mathbb{F}_p^q$. The normalization condition (3) becomes

$$(\hat{0}, 1, \sigma)(\hat{a}, u^d, \sigma^t)(\hat{0}, 1, \sigma^{-1}) = (f\hat{1}, 1, I)(\hat{a}, u^d, \sigma^t),$$

for some f. Looking at the leftmost components yields

$$\sigma(\hat{a}) = \hat{a} + f\hat{1}. \tag{4}$$

This equation implies that

$$a_{\sigma^{-1}(k)} = a_k + f,$$

for all k. In particular, since $a_1 = 0$, we have

$$a_{\sigma^{-n}(1)} = nf,$$

for all n.

Now we consider the condition that $(\hat{a}, u^d, \sigma^t)$ have order q. Looking at the leftmost components in $(\hat{0}, 1, I) = (\hat{a}, u^d, \sigma^t)^q$ yields

$$\hat{0} = \sum_{j=1}^{q-1} u^{dj}\sigma^{tj}(\hat{a}). \tag{5}$$

Since σ is a q-cycle, we may write

$$\hat{a} = [a_1, a_{\sigma(1)}, \ldots, a_{\sigma^r(1)}, \ldots, a_{\sigma^{q-1}(1)}]. \tag{6}$$

Now σ cyclically permutes the components of $\hat{a}$, so

$$\sigma(\hat{a}) = [a_{\sigma^{-1}(1)}, a_1, \ldots, a_{\sigma^{r-1}(1)}, \ldots, a_{\sigma^{q-2}(1)}]. \tag{7}$$

Thus looking at the first components of (5), we obtain

$$0 = \sum_{j=1}^{q-1} u^{dj} a_{\sigma^{-tj}(1)} = \sum_{j=1}^{q-1} u^{dj} tjf = tf \sum_{j=1}^{q-1} ju^{dj}. \tag{8}$$

Now for any indeterminate x, we have

$$\sum_{j=0}^{q-1} jx^j = x\frac{d}{dx}(1 + x + \cdots + x^q) = x\frac{d}{dx}\left(\frac{x^q - 1}{x - 1}\right) = x\left(\frac{qx^{q-1}}{x-1} - \frac{x^q - 1}{(x-1)^2}\right).$$

Setting $x = u^d$, the second term is $(u^{dq} - 1)/(u^d - 1)^2 = 0$, and so (8) becomes

$$0 = tfu^d \frac{qu^{d(q-1)}}{u^d - 1}. \tag{9}$$

Since $u^d \neq 1$ is a unit modulo p and $0 < t < q$, this equation only holds when $f = 0$. Hence $\hat{a} = \hat{0}$ and

$$N = \langle (\hat{1}, 1, I), (\hat{0}, u^d, \sigma^t) \rangle.$$

We have a distinct group N for each t coprime to q. Hence there are $q - 1$ regular subgroups of $\mathrm{Norm}_B(\mathcal{P})$ normalized by Γ such that $P(N) = \mathcal{P}$. By Corollary 3.6, $R(C_{pq}, [C_p \rtimes_\tau C_q]) = 2(q - 1)$.

$|R(C_p \rtimes_\tau C_q, [C_{pq}])| = p$. Let

$$\Gamma = C_p \rtimes_\tau C_q = \langle (\hat{1}, 0, I), (\hat{0}, u^d, \sigma) \rangle$$

and assume $P(N) = \mathcal{P}$. Then

$$N = \langle (\hat{1}, 1, I), (\hat{a}, 1, \sigma^t) \rangle,$$

for some $\hat{a}$ and some t coprime to q. Now Γ normalizes N, and $Q(N)$ is characteristic in N, so the normalization equation (3) becomes

$$(\hat{0}, u^d, \sigma)(\hat{a}, 1, \sigma^t)(\hat{0}, u^{-d}, \sigma^{-1}) = (\hat{a}, 1, \sigma^t).$$

Looking at the leftmost components gives

$$\sigma(\hat{a}) = u^{-d}\hat{a}.$$

Then

$$\sigma^k(\hat{a}) = u^{-dk}\hat{a},$$

hence

$$a_{\sigma^{-k}(1)} = u^{-dk} a_1,$$

for all k.

Thus $\hat{a}$ is uniquely determined by a_1, and, in fact, $\hat{a} = a_1 \hat{p}_{\psi_d}$. So

$$N = \langle (\hat{1}, 1, I), (a_1 \hat{p}_{\psi_d}, 1, \sigma^t) \rangle.$$

Now $\sigma(\hat{p}_{\psi_d}) = u^{-d} \hat{p}_{\psi_d}$ (see Lemma 5.2). So if $st \equiv 1 \pmod q$, then we may replace the generator $(a_1 \hat{p}_{\psi_d}, 1, \sigma^t)$ by its s-th power:

$$(a_1 \hat{p}_{\psi_d}, 1, \sigma^t)^s = \left(a_1 \left(\frac{u^{-dst} - 1}{u^{-dt} - 1} \right) \hat{p}_{\psi_d}, 1, \sigma \right).$$

Since d and t are coprime to q, $((u^{-dst} - 1)/(u^{-dt} - 1))$ is a unit modulo q. The constraint that $(b_1 \hat{p}_{\psi_d}, 1, \sigma)^q = (\hat{1}, 1, I)$ poses no further constraint, for the first component of $(b_1 \hat{p}_{\psi_d}, 1, \sigma)^q$ is

$$\sum_{i=0}^{q-1} \sigma^i (b_1 \hat{p}_{\psi_d}) = b_1 \left(\sum_{i=0}^{q-1} u^{-di} \right) \hat{p}_{\psi_d} = b_1 \left(\frac{u^{-dq} - 1}{u^d - 1} \right) \hat{p}_{\psi_d} = \hat{0}.$$

Thus we may choose a generator of $Q(N)$ to be $(b_1 \hat{p}_{\psi_d}, 1, \sigma)$ for any b_1 modulo p, and the p choices for b_1 yield different N. Thus $R(C_p \rtimes_\tau C_q, [C_{pq}]) = p$.

$|R(C_p \rtimes_\tau C_q, [C_p \rtimes_\tau C_q])| = 2(1 + p(q-2))$. Let

$$\Gamma = C_p \rtimes_\tau C_q = \langle (\hat{1}, 0, I), (\hat{0}, u^d, \sigma) \rangle$$

and assume $P(N) = \mathcal{P}$. Then we may assume that

$$N = \langle (\hat{1}, 1, I), (\hat{a}, u^d, \sigma^t) \rangle,$$

with $(t, q) = 1$. Constraint (3) is that conjugation by $(\hat{0}, u^d, \sigma)$ sends $(\hat{a}, u^d, \alpha)$ to an element of order q in N:

$$(\hat{0}, u^d, \sigma)(\hat{a}, u^d, \sigma^t)(\hat{0}, u^{-d}, \sigma^{-1}) = (\hat{a}, u^d, \sigma^t)^e (f\hat{1}, i, I), \tag{10}$$

for some e and f, where e is necessarily equal to 1 since σ commutes with σ^t. Looking at the left components of (10), we obtain $u^d \sigma(\hat{a}) = \hat{a} + u^d f \hat{1}$, since $\sigma(\hat{1}) = \hat{1}$. Thus

$$\sigma(\hat{a}) = u^{-d} \hat{a} + f\hat{1}.$$

Recalling (6) and (7), the action

$$\sigma(\hat{a}) = u^{-d} \hat{a} + f\hat{1}$$

translates at the component level to

$$a_{\sigma^{r-1}(1)} = u^{-d} a_{\sigma^r(1)} + f,$$

for all r. This implies that $\hat{a}$ is determined by a_1 and f, and so N is determined by (a_1, f, t).

From $a_{\sigma^{r-1}(1)} = u^{-d} a_{\sigma^r(1)} + f$, we obtain

$$a_{\sigma^{-r}(1)} = u^{-rd} a_1 + (1 + u^{-d} + \cdots + u^{-(r-1)d}) f,$$

for all r. Letting $u^{-d} = w$, we have

$$a_{\sigma^{-r}(1)} = w^r a_1 + \left(\frac{w^r - 1}{w - 1}\right) f,$$

for all r, where $w^q \equiv 1 \pmod{p}$.

The condition that $(\hat{a}, u^d, \sigma^t)^q = 1$ places potential constraints on (a_1, f, t). We have

$$(\hat{a}, u^d, \sigma^t)^q = (\hat{a} + u^d \sigma^t \hat{a} + \cdots + u^{d(q-1)} \sigma^{t(q-1)} \hat{a}, u^{dq}, \sigma^{tq}),$$

which equals $(\hat{0}, 1, I)$ provided that

$$\hat{a} + u^d \sigma^t \hat{a} + \cdots + u^{d(q-1)} \sigma^{t(q-1)} \hat{a} = \hat{0}.$$

Looking at the leftmost component of this last equation gives

$$a_1 + u^d a_{\sigma^{-t}(1)} + u^{2d} a_{\sigma^{-2t}(1)} + \cdots + u^{(q-1)d} a_{\sigma^{-(q-1)t}(1)} = 0.$$

Setting $u^{-d} = w$, this is

$$\begin{aligned} 0 &= \sum_{r=0}^{q-1} w^{-r} a_{\sigma^{-rt}(1)} = \sum_{r=0}^{q-1} w^{-r} \left(w^{rt} a_1 + \frac{w^{rt} - 1}{w - 1} f\right) \\ &= \sum_{r=0}^{q-1} w^{r(t-1)} a_1 + \frac{f}{w - 1} \sum_{r=0}^{q-1} (w^{r(t-1)} - w^{-r}). \end{aligned}$$

If $t \neq 1$, then this is equal to

$$a_1 \left(\frac{w^{(t-1)q} - 1}{w^{t-1} - 1}\right) + \frac{f}{w - 1} \left(\frac{w^{(t-1)q} - 1}{w^{t-1} - 1} - \frac{w^{-q} - 1}{w^{-1} - 1}\right).$$

Since $w^q \equiv 1 \pmod{p}$, this is congruent to $0 \pmod{p}$.

If $t = 1$, then this yields

$$f = (1 - w) a_1 = (1 - u^{-d}) a_1. \tag{11}$$

For $t \neq 1$, every pair (a, f) yields a group N. But if we vary the generator $(\hat{a}, u^d, \sigma^t)$ of N of order q by multiplying it by $(k\hat{1}, 1, I)$, we obtain a new generator

$$(k\hat{1}, 1, I)(\hat{a}, u^d, \sigma^t) = (\hat{a} + k\hat{1}, u^d, \sigma^t) = (\hat{b}, u^d, \sigma^t),$$

where $\hat{b} = \hat{a} + k\hat{1}$. Then, since $\sigma(\hat{a}) = u^{-d}\hat{a} + f\hat{1}$, we have

$$\sigma(\hat{b}) = \sigma(\hat{a}) + k\hat{1} = (u^{-d}\hat{a} + f\hat{1}) + k\hat{1} = u^{-d}\hat{b} + (f + (1 - u^{-d})k)\hat{1}.$$

So changing the generator of order q changes (a_1, f, t) to $(a_1+k, f+(1-u^{-d})k, t)$. Since $1-u^{-d}$ is a unit modulo p, the p^2 pairs (a, f) for each $t \neq 1$ yield p different groups N. Thus there are $(q-2)p$ different regular subgroups N isomorphic to $C_p \rtimes_\tau C_q$ with $t \neq 1$.

For $t = 1$,

$$N = \langle (\hat{1}, 1, I), (\hat{a}, u^d, \sigma) \rangle$$

and for the second generator to have order q, we must have (11):

$$(1-u^{-d})a_1 = f,$$

where $\sigma(\hat{a}) = u^{-d}\hat{a} + f\hat{1}$. Replacing $(\hat{a}, u^d, \sigma)$ by $(k\hat{1}, 1, I)(\hat{a}, u^d, \sigma)$ gives an order-q generator $(\hat{b}, u^d, \sigma)$ for N where

$$\hat{b} = \hat{a} + k\hat{1}.$$

Then

$$\begin{aligned}\sigma(\hat{b}) &= \sigma(\hat{a}) + k\hat{1} = (u^{-d}\hat{a} + f\hat{1}) + k\hat{1} \\ &= u^{-d}(\hat{b} - k\hat{1}) + (f+k)\hat{1} = u^{-d}b + f'\hat{1},\end{aligned}$$

where

$$f' = f + k(1-u^{-d}).$$

By choosing k so that $f' = 0$, then $\sigma(\hat{b}) = u^{-d}\hat{b}$, and the condition on the order-q generator becomes

$$(1-u^{-d})b_1 = 0.$$

Hence $b_1 = 0$ and since

$$b_{\sigma^{-r}(1)} = u^{-rd}b_1,$$

we have $\hat{b} = \hat{0}$ and $N = \Gamma$. Thus we obtain $1 + (q-1)p$ regular subgroups N of $\mathrm{Norm}_B(\mathcal{P})$ isomorphic to $C_p \rtimes_\tau C_q$ with $P(N) = \mathcal{P}$ that are normalized by $\Gamma \cong C_p \rtimes_\tau C_q$. By Corollary 3.6, we conclude $R(C_p \rtimes_\tau C_q, [C_p \rtimes_\tau C_q]) = 2(1+(q-1)p)$.

That completes the proof of Theorem 4.1.

5. Groups of order $(2q+1)2q$

In this section we consider $R(\Gamma)$ for groups of order mp where $p = 2q+1$ with q an odd prime and $m = 2q = \phi(p)$; p is then a safe prime. Such groups were explored in some detail in [Childs 2003] (and in [Moody 1994, Example 8.7, p. 133 ff.] for $q = 3$). There are six isomorphism classes of groups of order $p(p-1)$ where $p - 1 = 2q$ with q prime:

$$\begin{aligned}
C_{mp} &= C_p \times C_m = \langle x, y \mid x^p = y^m = 1\rangle,\\
F \times C_2 &= (C_p \rtimes C_q) \times C_2\\
&= \langle x, y \mid x^p = y^m = 1;\ yxy^{-1} = x^{u^2}\rangle,\\
C_p \times D_q &= C_p \times (C_q \rtimes C_q)\\
&= \langle x, a, b \mid x^p = a^q = b^2 = 1;\ bx = xb;\ ax = xa,\ bab^{-1} = a^{-1}\rangle,\\
D_{pq} &= C_p \rtimes (C_q \rtimes C_2)\\
&= \langle x, a, b \mid x^p - a^q = b^2 = 1;\ bab^{-1} = x^{-1};\ ax = xa;\ bab^{-1} = a^{-1}\rangle,\\
D_p \times C_q &= (C_p \rtimes C_m = \langle x, y \mid x^p = y^m = 1;\ yxy^{-1} = x^{-1}\rangle,\\
\mathrm{Hol}(C_p) &= C_p \rtimes C_m = \langle x, y \mid x^p = y^m = 1;\ yxy^{-1} = x^u\rangle.
\end{aligned}$$

Here u is a primitive root modulo p: $\langle u\rangle = \mathbb{F}_p^\times = U_p = \mathrm{Aut}(C_p)$.

The main result in this section is:

Theorem 5.1. *Let $R(\Gamma, [M])$ be the set of regular subgroups N isomorphic to M in* $\mathrm{Perm}\,\Gamma_i$ *that are normalized by $\lambda(\Gamma)$. Then the cardinality of $R(\Gamma, [M])$ is given by the following table*:

$\Gamma\downarrow \quad M\rightarrow$	C_{mp}	$C_p\times D_q$	$F\times C_2$	$C_q\times D_p$	D_{pq}	$\mathrm{Hol}\,C_p$
C_{mp}	1	2	$2(q-1)$	2	4	$2(q-1)$
$C_p\times D_q$	q	2	0	$2q$	4	0
$F\times C_2$	p	$2p$	$2(p(q-2)+1)$	$2p$	$4p$	$2p(q-1)$
$C_q\times D_p$	p	$2p$	$2p(q-1)$	2	4	$2p(q-1)$
D_{pq}	qp	$2p$	0	$2q$	4	0
$\mathrm{Hol}\,C_p$	p	$2p$	$2p(q-1)$	$2p$	$4p$	$2(p(q-2)+1)$

For each pair (Γ, M), the table shows $|R(\Gamma, [M])|$, the number of Hopf Galois structures of type M on a Galois extension L/K with Galois group Γ. Thus the row sum for that Γ is the number of Hopf Galois structures on L/K. Observe that whenever M is not a direct product of the p-Sylow subgroup of M with a group of order m, the entries in the M-column are even: that is a consequence of Corollary 3.6.

We now construct subgroups $\mathcal{P}\mathcal{Q}$ of S_{mp} isomorphic to Γ for each isomorphism type of groups Γ of order mp. We will work within $B = S_{mp}$ and set $\mathcal{P} = \langle \pi_1\pi_2\cdots\pi_m\rangle$, where π_i is the p-cycle

$$\pi_i = \big((i-1)p+1 \ \ (i-1)p+2 \ \ \ldots \ \ ip\big).$$

Then $\mathrm{Norm}_B(\mathcal{P})$ is isomorphic to the group of 3-tuples $(\hat{a}, u^s, \alpha)$, where $\hat{a} = [a_1, \ldots, a_m]$ with a_i in $\mathbb{F}_p$, $\langle u\rangle = U_p$, and $\alpha \in S_m$. We set

$$\Pi_i = \mathrm{Supp}(\pi_i) = \{(i-1)p+1, (i-1)p+2, \ldots, ip\}.$$

Then we choose regular subgroups $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ of $\mathrm{Perm}(\{\Pi_1, \ldots, \Pi_m\}) \cong S_m$ representing the isomorphism types of groups of order $m = 2q$, namely $\mathfrak{Q}_1 \cong C_m$ and $\mathfrak{Q}_2 \cong D_q$, and embed them in $\mathrm{Norm}_B(\mathcal{P})$ by

$$\alpha \in \mathfrak{Q} \mapsto (\hat{0}, 1, \alpha) \in \mathrm{Norm}_B(\mathcal{P}).$$

By slight abuse of notation, we denote the image of $\mathfrak{Q}_i$ in $\mathrm{Norm}_B(\mathcal{P})$ also by $\mathfrak{Q}_i$.

We choose $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ as follows: let $\mathfrak{Q}_1 = \langle \sigma \rangle \cong C_m$ and $\mathfrak{Q}_2 = \langle \sigma^2, \delta \rangle \cong D_q$, where

$$\begin{aligned}
\sigma &= (1, 4, 5, 8, 9, \ldots, 2q-1, 2, 3, 6, \ldots, 2q),\\
\sigma^2 &= (1, 5, 9, \ldots, 2q-3)(2, 6, 10, \ldots, 2q-2), \text{ which we denote by } \sigma_L\sigma_R,\\
\delta &= (1, 2)(3, 2q)(4, 2q-1)(5, 2q-2)\cdots(q, q+3)(q+1, q+2).
\end{aligned}$$

Then $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ are regular subgroups of S_m. We observe that $(\mathfrak{Q}_1)^{\mathrm{opp}} = \mathfrak{Q}_1$ (since $\mathfrak{Q}_1$ is abelian), and that $\mathfrak{Q}_2^{\mathrm{opp}} = \langle \sigma_L\sigma_r^{-1}, \sigma^q \rangle$, where

$$\sigma^q = (1, 2)(3, 4)\cdots(2q-1, 2q).$$

To find the possible order-p subgroups of $N \in R(\Gamma)$, we follow Theorem 2.1 and consider linear characters $\psi_i : C_m \to \mathbb{F}_p^\times$,

$$\psi_i(\sigma) = u^i, \text{ for } i = 0, \ldots, m-1,$$

and $\chi_i : D_q \to \mathbb{F}_p^\times$,

$$\chi_i(\sigma^2) = 1,\ \chi_i(\delta) = u^{qi} = (-1)^i, \text{ for } i = 0, 1.$$

Since $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ centralize $\langle \hat{p}_\iota \rangle = \mathcal{P}$ (since the elements of $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ act as permutations of $\{\pi_1, \ldots, \pi_m\}$), the proof of Theorem 2.1 shows that $\mathfrak{Q}_i$ normalizes $\langle \hat{p}_\chi \rangle$ for each linear character χ of $\mathfrak{Q}_i$. In fact, from Theorem 2.1, if $\mathfrak{Q}$ is a regular subgroup of $\mathrm{Perm}(\pi_1, \ldots, \pi_m)$ and χ is a character of $\mathfrak{Q}$, then for all μ in $\mathfrak{Q}$, $\mu\pi\mu^{-1} = \pi$, so

$$\mu\hat{p}_\chi\mu^{-1} = \chi(\mu)^{-1}\hat{p}_\chi.$$

Hence $\hat{p}_\chi$ is an eigenvector under the action of $\mathfrak{Q}$.

More precisely, we have

Lemma 5.2. *For σ the generator of $\mathfrak{Q}_1 \cong C_m$ and σ^2 and δ the generators of $\mathfrak{Q}_2 \cong D_q$, we have*:

$$\sigma(\hat{p}_{\chi_0}) = \delta(\hat{p}_{\chi_0}) = \hat{p}_{\chi_0}, \quad \sigma(\hat{p}_{\psi_i}) = u^{-1}\hat{p}_{\psi_i}, \quad \sigma^2(\hat{p}_{\chi_1}) = \hat{p}_{\chi_1},$$
$$\delta(\hat{p}_{\chi_1}) = u^q\hat{p}_{\chi_1}, \quad \delta(\hat{p}_{\psi_i}) = \hat{p}_{\psi_{-i}}.$$

Proof. All of these follow from

$$\mu\hat{p}_\chi\mu^{-1} = \mu(\hat{p}_\chi) = \chi(\mu)^{-1}\hat{p}_\chi$$

except the last, in which ψ_i is not a character of $\mathcal{Q}_2$. For the last, we have

$$\hat{p}_{\psi_i} = \sum_{\gamma \in \mathcal{Q}_1} \psi_i(\gamma)\hat{v}_{\gamma(1)} = \sum_{j=0}^{m-1} \psi_i(\sigma^j)\hat{v}_{\sigma^j(1)}.$$

Now $\delta(\sigma) = \sigma^{-1}$, so

$$\begin{aligned}\delta(\hat{p}_{\psi_i}) &= \sum_{j=0}^{m-1} \psi_i(\sigma^j)\hat{v}_{\delta(\sigma^j)(1)} = \sum_{j=0}^{m-1} \psi_i(\sigma^j)\hat{v}_{\sigma^{-j}(1)} = \sum_{j=0}^{m-1} u^{ij}\hat{v}_{\sigma^{-j}(1)} \\ &= \sum_{j=0}^{m-1} u^{-ij}\hat{v}_{\sigma^j(1)} = \sum_{j=0}^{m-1} \psi_{-i}(\sigma^j)\hat{v}_{\sigma^j(1)} = \hat{p}_{\psi_{-i}}. \qquad \square\end{aligned}$$

We set $P_i = \langle \hat{p}_{\psi_i} \rangle$ for $i = 0, \ldots, m-1$. In particular, $P_0 = \langle \hat{p}_{\chi_0} \rangle = \langle \hat{p}_{\psi_0} \rangle = \langle [1, 1, \ldots, 1] \rangle = \langle \hat{1} \rangle = \mathcal{P}$. We also have that

$$\hat{p}_{\chi_1} = \sum_{\gamma \in \mathcal{Q}_2} \chi_1(\gamma)\hat{v}_{\gamma(1)} = \sum_{i=0}^{m-1} (-1)^i \hat{v}_{\delta^i\sigma^{2i}(1)}$$

while

$$\hat{p}_{\psi_q} = \sum_{\gamma \in \mathcal{Q}_1} \psi_q(\gamma)\hat{v}_{\gamma(1)} = \sum_{i=0}^{m-1} (-1)^i \hat{v}_{\sigma(1)}.$$

Both are equal to $\langle [1, -1, 1, -1, \ldots, 1, -1] \rangle$.

We thus have subgroups of $\mathrm{Norm}_B(\mathcal{P})$ of the form $P_i\mathcal{Q}_j$ for certain pairs (i, j). We identify their isomorphism types as follows:

Proposition 5.3. *With P_i and $\mathcal{Q}_j$ as defined above, we have*

$$\begin{aligned}
&P_0\mathcal{Q}_1 \cong C_p \times C_m, \\
&P_i\mathcal{Q}_1 \cong F \times C_2 \text{ for } i \text{ even}, i \neq 0, \\
&P_i\mathcal{Q}_1 \cong \mathrm{Hol}(C_p) \text{ for } i \text{ odd}, i \neq q, \\
&P_q\mathcal{Q}_1 \cong D_p \times C_q, \\
&P_0\mathcal{Q}_2 \cong D_q \times C_p, \\
&P_q\mathcal{Q}_2 \cong D_{pq}.
\end{aligned}$$

Proof. This follows from Lemma 5.2 and the definitions for the P_i. $\square$

Each group $\mathcal{P}_i\mathcal{Q}_j$ above centralizes $\mathcal{P} = P_0 = \langle [1, 1, \ldots, 1] \rangle = \langle \hat{1} \rangle$, so each opposite group $(\mathcal{P}_i\mathcal{Q}_j)^{\mathrm{opp}}$ will contain $\mathcal{P}$. We will use those opposite groups for the groups Γ in the computations.

We need to observe:

Proposition 5.4. *Each group $P_i\mathcal{Q}_j$ is a regular subgroup of* $\mathrm{Norm}_B(\mathcal{P})$.

Proof. Each $P_i\mathcal{Q}_j$ is a subgroup of order mp by Proposition 5.3. To show regularity we show that each nonidentity element of $P_i\mathcal{Q}_j$ acts fixed-point-freely. Now each element of $P_i\mathcal{Q}_j$ has the form $(\hat{a}, 1, \alpha)$ for $\hat{a}$ in $\mathbb{F}_p^n$ and α in S_m. Since $\mathcal{Q}_j$ is a regular subgroup of S_m acting on $\{\Pi_1, \dots, \Pi_m\}$, $(\hat{a}, 1, \alpha)$ is fixed-point free for $\alpha \neq 1$ by Proposition 3.8. If an element $(\hat{a}, 1, I)$ is not the identity, then $\hat{a} = [a_1, a_2, \dots, a_m]$ with all $a_i \neq 0$ (since $\hat{a}$ is a power of $\hat{p}_\chi$ for some linear character with values in $\mathbb{F}_p^\times$). Hence for t in Π_i, $(\hat{a}, 1, I)(t) = a_i + t \neq t$; hence $(\hat{a}, 1, I)$ has no fixed points. □

For each isomorphism type of Γ, we have the following (recall that $P(\Gamma) = P_0 = \mathcal{P} = \langle [1, 1, \dots, 1] \rangle = \langle \hat{1} \rangle$):

$$\begin{aligned}
\Gamma = C_p \times C_m &= (P_0\mathcal{Q}_1)^{\mathrm{opp}} = P_0\mathcal{Q}_1 \\
&= P_0\langle (\hat{0}, 1, \sigma) \rangle, \\
\Gamma = C_p \times D_q &= (P_0\mathcal{Q}_2)^{\mathrm{opp}} \\
&= P_0\langle (\hat{0}, 1, \sigma^q)(\hat{0}, 1, \sigma_L\sigma_R^{-1}) \rangle, \\
\Gamma = D_p \times C_q &= (P_q\mathcal{Q}_1)^{\mathrm{opp}} \\
&= P_0\langle (\hat{0}, u^q, \sigma) \rangle, \\
\Gamma = D_{pq} &= (P_q\mathcal{Q}_2)^{\mathrm{opp}} \\
&= P_0\langle (\hat{0}, u^q, \sigma^q) \rangle, \\
\Gamma = F \times C_2 &= (P_2\mathcal{Q}_1)^{\mathrm{opp}} \\
&= P_0\langle (\hat{0}, u^2, \sigma) \rangle, \\
\Gamma = \mathrm{Hol}(C_p) &= (P_1\mathcal{Q}_1)^{\mathrm{opp}} \\
&= P_0\langle (\hat{0}, u, \sigma) \rangle.
\end{aligned}$$

There is a certain arbitrariness concerning these last two choices.

Recall from Proposition 3.8 that if $(\hat{a}, 1, \alpha)$ in $\mathrm{Norm}_B(\mathcal{P})$ has order coprime to p, then $(\hat{a}, 1, \alpha)$ is fixed-point free in $\mathrm{Norm}_B(\mathcal{P})$ if and only if α is fixed-point free in S_m.

Lemma 5.5. *Let $\alpha = [a_1, \dots, a_m] \in \mathbb{F}_p^m$ and $\alpha \in S_m$.*

If the element $(\hat{a}, 1, \alpha)$ has order 2, then $\alpha = x_1 \cdots x_q$, a product of q disjoint 2-cycles such that for each $x_i = (r, s)$, $a_r + a_s = 0$.

If the element $(\hat{a}, 1, \alpha)$ has order q, then $\alpha = x_1x_2$, disjoint q-cycles, and $\sum_{i \in \mathrm{Supp}(x_j)} a_i = 0$ for $i = 1, 2$.

If the element $(\hat{a}, 1, \alpha)$ has order $m = 2q$, then α is an m-cycle and $\sum_{i=0}^{m-1} a_i = 0$.

Proof. Let $d = |(\hat{a}, 1, \alpha)|$. If d is coprime to p, then $|\alpha| = d$; for otherwise $|\alpha| = e < d$, in which case $(\hat{a}, 1, \alpha)^e = (\hat{b}, 1, I)$, with $\hat{b} \neq 0$. But then $(\hat{b}, 1, I)$ has order p, and so p divides $|(\hat{a}, 1, \alpha)|$, a contradiction.

So if d is coprime to p, then α has order d. Since α is fixed-point free, if $d = 2$, then α is a product of q disjoint 2-cycles; if $d = q$ then α is a product of two disjoint

q-cycles, and if α has order $m = 2q$ then α is an m-cycle. Now

$$(\hat{a}, 1, \alpha)^n = \left(\sum_{k=0}^{n-1} \alpha^k(\hat{a}),\ 1,\ \alpha^n\right).$$

If n is the order of $(\hat{a}, 1, \alpha)$, hence also the order of α, then by what was just observed,

$$\sum_{k=0}^{n-1} \alpha^k(\hat{a}) = \hat{0},$$

and hence for each a_i,

$$\sum_{k=0}^{n-1} a_{\alpha^{-k}(i)} = \sum_{k=0}^{n-1} a_{\alpha^k(i)} = 0.$$

The conclusions of the lemma follow. □

Using that

$$(\hat{a}, u^r, \alpha)^n = \left(\sum_{k=0}^{n-1} u^{rk}\alpha^k(\hat{a}),\ u^{rn},\ \alpha^n\right),$$

the same argument gives:

Lemma 5.6. *Let $\hat{a} = [a_1, \ldots, a_m] \in \mathbb{F}_p^m$, $r \neq 0$ in $\mathbb{F}_p^\times$, and $\alpha \in S_m$.*

If the element $(\hat{a}, u^r, \alpha)$ has order 2, then $r = q$ and $u^r = u^q = -1$, and $\alpha = (x_1, \ldots, x_q)$, a product of q disjoint 2-cycles such that for each $x_i = (r, \alpha(r))$, $a_r - a_{\alpha(r)} = 0$.

If the element $(\hat{a}, u^r, \alpha)$ has order q, then $\alpha = x_1x_2$, where x_1 and x_2 are disjoint q-cycles, and for t_i in $\operatorname{Supp} x_i$,

$$\sum_{k=0}^{q-1} u^{kr} a_{\alpha^{-k}(t_i)} = 0,$$

for $i = 1, 2$.

If the element $(\hat{a}, u^r, \alpha)$ has order $m = 2q$, then α is an m-cycle and

$$\sum_{i=0}^{m-1} u^{ri}\alpha^{-i}(a_1) = 0.$$

Enumeration of the $R(\Gamma, [M])$ for each of the 36 pairs (Γ, M) in Theorem 5.1 breaks up into subcases. Recall that $R(\Gamma, [M]; P_i)$ is the set of regular subgroups N of $\mathrm{Norm}_B(\mathcal{P}) \subset S_{mp}$ such that the p-Sylow subgroup of N is $P(N) = P_i$. By Corollary 3.6, if $M \cong C_{mp}$ or $C_p \times D_q$, $R(\Gamma, [M]) = R(\Gamma, [M]; P_0)$. For other M, Corollary 3.6 shows that to count $R(\Gamma, [M])$ we need only count $R(\Gamma, [M]; P_0)$ (where $P_0 = \mathcal{P}$). But given that regular subgroups N yield Hopf Galois structures

on Galois extensions of fields with Galois group Γ, it is useful to explicitly consider $R(\Gamma, [M]; P_i)$ for $i \neq 0$.

Thus, rather than just the 36 cases described in Theorem 5.1, a more complete story would involve 57 cases: 36 of the form $R(\Gamma, [M]; P_0)$, and 21 of the form $R(\Gamma, [M]; P_i)$ with $i \neq 0$ where for each $[M]$, the possible P_i with $i \neq 0$, where $P(N) = P_i$ and $N \cong M$, are as listed in Proposition 5.3. The counts in those cases are as follows.

For $N \cong M = D_p \times C_q$ or D_{pq}, we have $P(N) = P_0$ or P_q and Corollary 3.6 shows that $|R(\Gamma, [M]; P_q)| = |R(\Gamma, [M]; P_0)|$.

For $N \cong M = F \times C_2$ or $\operatorname{Hol} C_p$, there are $\phi(2q)$ possible i, and $|R(\Gamma, [M]; P_i)| = |R(\Gamma, [M]; P_j)|$ for all possible $i \neq j$ and $i, j \neq 0$, except when $\Gamma \cong M$.

For $\Gamma = M = F \times C_2$ we have

$$|R(F \times C_2, [F \times C_2]; P_2)| = 1,$$
$$|R(F \times C_2, [F \times C_2]; P_i)| = p \quad \text{for } i = 4, 6, \ldots, 2q-2.$$

The case $\Gamma = M = \operatorname{Hol}(C_p)$ is similar and will be described below.

Since most of the computations are very similar in outline and details to those in Section 4, we will limit ourselves to just three cases. Before we begin, we pause to give the reader some perspective, with a view toward dealing with other classes of groups of order mp, beyond those considered here. There are some common themes that arise in the enumeration of $N \in R(\Gamma, [M])$, in particular in the determination of the 3-tuples $(\hat{a}, v, \alpha)$ that generate $Q(N)$, some of which have been seen already in the work in Section 4.

- The given generator of $Q(N)$ must, of course, normalize (and possibly even centralize) $P(N)$.
- Any $Q(N)$ is semiregular so any generator of $Q(N)$ must act without fixed points, which imposes restrictions on its components as seen above. And if one is dealing with several generators of $Q(N)$, the products of these generators also cannot have fixed points.
- The order of a given generator of $Q(N)$ imposes restrictions on its components.
- Any N is normalized by Γ, so when a given generator of $Q(N)$ is conjugated by an element of Γ it is mapped to another element of N and the form of this conjugate is determined by whether $Q(N)$ is a direct factor of N or not.
- The restrictions imposed by order, semiregularity, and being normalized by Γ will frequently imply that $\hat{a}$ is the solution to a particular set of linear equations and so linear algebra techniques may be applied.
- The number of free variables that determine the solution sets for the aforementioned linear systems determines whether or not the resulting generators

$(\hat{a}, v, \alpha)$ lie in $Q(N)$ for a *single* N or, in fact, multiple N. As such, the count of $|R(\Gamma, [M])|$ may vary linearly with p (as when we showed that $|R(C_p \rtimes_\tau C_q, [C_{pq}])| = p$ earlier) or be "combinatorially" determined, that is, in terms of some intrinsic property of regular subgroups of S_m, as will be seen at the end of the determination of $|R(C_{mp}, [C_p \times D_q])|$ later on.

$R(C_p \times D_q, [F \times C_2])$.

Proposition 5.7. *With $p > q$ primes, $|R(C_p \times D_q, [F \times C_2])| = 0$.*

Proof. We have

$$\Gamma = \mathcal{P}\langle(\hat{0}, 1, \sigma^2), (\hat{0}, q, \delta)\rangle.$$

Since $N \cong F \times C_2 \cong (C_p \rtimes C_q) \times C_2$, it has the form

$$N = \langle(\hat{0}, 1, I), (\hat{a}, u^r, \alpha)\rangle,$$

where $(\hat{a}, u^r, \alpha)$ has order $m = 2q$, and therefore α is an m-cycle in S_m. Now $(\hat{a}, u^r, \alpha)$ conjugates the order-p generator of N to its u^2 power

$$(\hat{a}, u^r, \alpha)(\hat{1}, 1, I)(\hat{a}, u^r, \alpha)^{-1} = (u^2\hat{1}, 1, I),$$

so $r = 2$.

Also α has order $m = 2q$, and being fixed-point free, must be an m-cycle.

If Γ normalizes N, then conjugation by $(\hat{0}, 1, \sigma^2)$ and $(\hat{0}, 1, \delta)$ are automorphisms of N. Every automorphism of $F \times C_2$ sends the order-m element y to xy for some element x of order p. Thus conjugating the order-m generator $(\hat{a}, u^2, \alpha)$ of N by $(\hat{0}, 1, \sigma^2)$ and $(\hat{0}, 1, \delta)$, and looking at the rightmost S_m components of the result, we have that $\sigma^2\alpha\sigma^{-2} = \alpha$ and $\delta\alpha\delta^{-1} = \alpha$. Thus σ^2 and δ commute with α. But since α is an m-cycle in S_m, the centralizer in S_m of α is $\langle\alpha\rangle$. So σ^2 and δ are powers of α in S_m, and hence commute. But that's impossible. Thus no α exists, and hence there is no N isomorphic to $F \times C_2$ that is normalized by $\Gamma \cong C_p \times D_q$.

By Corollary 3.6, $R(C_p \times D_q, [F \times C_2]; P_0) = 0$. □

Essentially the same argument shows that $|R(C_p \times D_q), [\mathrm{Hol}(C_p)])|$, $|R(D_{pq}, [F \times C_2])|$, and $|R(D_{pq}, [\mathrm{Hol}(C_p)])|$ are all zero.

$R(C_{mp}, [C_p \times D_q]) = R(C_{mp}, [C_p \times D_q]; P_0)$. We will need the following technical information.

Lemma 5.8. *If $x = (a_1, a_2, \ldots, a_q)$ and $y = (b_1, b_2, \ldots, b_q)$ are elements with disjoint support in $S_{2q} = \mathrm{Perm}(\{1, \ldots, 2q\})$ then $\mathrm{Norm}_{S_{2q}}(\langle xy\rangle)$ contains $2q(q-1)$ elements z of order $2q$ with no fixed points (which are therefore $2q$-cycles), half of which centralize xy and are such that $\langle z^2\rangle = \langle(xy)^2\rangle$ and the other half invert xy and satisfy $\langle z^2\rangle = \langle(xy^{-1})^2\rangle$. Also, $\mathrm{Norm}_{S_{2q}}(\langle xy\rangle)$ contains two subgroups isomorphic to D_q, which are opposites of each other, one of which is contained in $\mathrm{Cent}_{S_{2q}}(xy)$.*

Proof. First we observe that $\mathrm{Norm}_{S_{2q}}(\langle xy\rangle)$ is isomorphic to $\mathbb{F}_q^2 \rtimes (\langle u\rangle \times S_2)$ where $\langle u\rangle = \mathbb{F}_q^\times$. As such, one may readily count how many elements have order $2q$. In particular, since a typical element is a 3-tuple $(\hat{v}, u^r, \alpha)$ with $\hat{v} = (v_1, v_2) \in \mathbb{F}_q^2$, $\langle u\rangle = \mathbb{F}_q^*$, and $\alpha \in S_2$, then, using (2), one may show that $|(\hat{v}, u^r, \alpha)| = 2q$ provided that $\alpha = (1, 2)$, and either $v_1 \neq v_2$ and $u^r = -1$ or $v_1 \neq -v_2$ and $u^r = 1$. This yields precisely $2(q^2 - q) = 2q(q-1)$ elements as claimed. We can exhibit the particular elements of order $2q$ (as elements in S_{2q}) as follows. First, let

$$\begin{aligned}
t_0 &= (a_1, b_1)(a_2, b_2)\cdots(a_q, b_q),\\
t_1 &= (a_1, b_2)(a_2, b_3)\cdots(a_q, b_1),\\
&\vdots\\
t_{q-1} &= (a_1, b_q)(a_2, b_1)\cdots(a_q, b_{q-1}),\\
\tau_0 &= (a_1, b_1)(a_2, b_q)\cdots(a_q, b_2),\\
\tau_1 &= (a_1, b_2)(a_2, b_1)\cdots(a_q, b_3),\\
&\vdots\\
\tau_{q-1} &= (a_1, b_q)(a_2, b_{q-1})\cdots(a_q, b_1),
\end{aligned}$$

and consider the elements xyt_i and $xy^{-1}\tau_i$. One may verify that each t_i interchanges x and y, so that xyt_i centralizes xy and that $\tau_i x \tau_i^{-1} = y^{-1}$ and $\tau_i y \tau_i^{-1} = x^{-1}$; therefore $xy^{-1}\tau_i$ inverts xy. Each of the elements xyt_i and $xy^{-1}\tau_i$ are $2q$-cycles and each generates a distinct subgroup. Moreover $(xyt_i)^2 = (xy)^2 \in \langle xy\rangle$ while $(xy^{-1}\tau_i)^2 = (xy^{-1})^2 \in \langle xy^{-1}\rangle$. The conclusion we get is that if a $2q$-cycle z inverts or centralizes xy then $z^2 \in \langle xy^{-1}\rangle$ or $\langle xy\rangle$. The groups $\langle xy^{-1}, t_i\rangle$ for each i are all equal and isomorphic to D_q (and are contained in $\mathrm{Cent}_{S_{2q}}(xy)$), and the groups $\langle xy, \tau_i\rangle$ are all equal and isomorphic to D_q but are not subgroups of $\mathrm{Cent}_{S_{2q}}(xy)$. Moreover $\langle xy^{-1}, t_i\rangle^{\mathrm{opp}} = \langle xy, \tau_i\rangle$ since each clearly centralizes the other. One may also observe that each of the $2q$-cycles above clearly normalize each of these two copies of D_q. □

If C is a cyclic regular subgroup of S_{2q} and $\langle xy\rangle = Q(C)$, then C must be generated by one of the $2q$-cycles given in Lemma 5.8. If $N \cong D_q \subset S_{2q}$ is normalized by C, then $Q(N) = \langle xy\rangle$, and so $N = \langle xy, \tau_i\rangle$. Thus $|R(C_{2q}, [D_q]; P_0)| = 1$. This is in agreement with Theorem 4.1 (if in Theorem 4.1 we set $p = 2$ and exchange the roles of p and q).

Proposition 5.9. $|R(C_{mp}, [C_p \times D_q])| = 2$.

Proof. Here $P(N) = \mathcal{P}$, since $Q(N)$ is a direct factor of N. In this case $Q(N)$ is generated by $(\hat{a}, 1, \alpha)$ of order q and $(\hat{b}, 1, \beta)$ of order 2. Note that both $Q(N)$ and

$\langle(\hat{a}, 1, \alpha)\rangle$ are characteristic subgroups of N. So

$$(\hat{0}, 1, \sigma)(\hat{a}, 1, \alpha)(\hat{0}, 1, \sigma^{-1}) = (\sigma(\hat{a}), 1, \sigma\alpha, \sigma^{-1})$$

must equal $(\hat{a}, 1, \alpha)^k$ for some k. By Lemma 5.8, σ must either centralize or invert α, so $k = 1$ or -1.

First, we look at the case where σ *centralizes* α. Then

$$(\sigma(\hat{a}), 1, \sigma\alpha, \sigma^{-1}) = (\hat{a}, 1, \alpha),$$

so $\sigma(\hat{a}) = \hat{a}$, and therefore $\hat{a} = a\hat{1}$ for some a in $\mathbb{F}_p$. Consequently, $\alpha(\hat{a}) = \hat{a}$. Since $(\hat{a}, 1, \alpha)$ has order q, we have that $q\hat{a} = qa\hat{1} = \hat{0}$, and so $a = 0$ and $\hat{a} = \hat{0}$.

Now, since $(\hat{b}, 1, \beta)$ normalizes $\langle(\hat{a}, 1, \alpha)\rangle$ then

$$\begin{aligned}(\hat{b}, 1, \beta)(\hat{a}, 1, \alpha)(-\beta^{-1}(\hat{b}), 1, \beta^{-1}) &= (\hat{b}, 1, \beta)(\hat{0}, 1, \alpha)(-\beta^{-1}(\hat{b}), 1, \beta^{-1}) \\ &= (\hat{b} - (\beta\alpha\beta^{-1})(\hat{b}), 1, \beta\alpha\beta^{-1}),\end{aligned}$$

which must equal

$$(\hat{0}, 1, \alpha)^{-1} = (\hat{0}, 1, \alpha^{-1}).$$

As $\beta\alpha\beta^{-1} = \alpha^{-1}$ we have $\hat{b} - \alpha^{-1}(\hat{b}) = \hat{0}$, so that $\alpha(\hat{b}) = \hat{b}$. Now, we must have that $(\hat{0}, 1, \sigma)$ conjugates $(\hat{b}, 1, \beta)$ to another order-2 element of $Q(N)$, ergo

$$\begin{aligned}(\hat{0}, 1, \sigma)(\hat{b}, 1, \beta)(\hat{0}, 1, \sigma^{-1}) &= (\hat{0}, 1, \alpha)^k(\hat{b}, 1, \beta) \\ &= (\alpha^k(\hat{b}), 1, \alpha^k\beta) \\ &= (\hat{b}, 1, \alpha^k\beta), \quad \text{since } \alpha(\hat{b}) = \hat{b}.\end{aligned}$$

So we must have $\sigma(\hat{b}) = \hat{b}$, which means $\hat{b} = b\hat{1}$ for some b in $\mathbb{F}_p$. But $\beta(\hat{b}) = -\hat{b}$ since $(\hat{b}, 1, \beta)$ has order 2. Thus $b = 0$. We conclude that

$$Q(N) = \langle(\hat{0}, 1, \alpha), (\hat{0}, 1, \beta)\rangle,$$

where $\langle\alpha, \beta\rangle \cong D_q$ and is centralized by σ.

Letting $\alpha = xy$ in Lemma 5.8, σ is an element of $\mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ of order $2q$ that centralizes α, hence by Lemma 5.8 $\sigma^2 \in \langle\alpha\rangle$, hence $\langle\sigma^2\rangle = \langle\alpha\rangle$. Now $\mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ contains a unique copy of D_q that does not centralize α. That copy must be $\langle\alpha, \beta\rangle$, since clearly $\langle\alpha, \beta\rangle$ does not centralize α,

We show that $\mathfrak{Q}_2$ is also in $\mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ and does not centralize α. Recall (from Lemma 5.2) that $\mathfrak{Q}_2 = \langle\sigma^2, \delta\rangle \cong D_q$, hence $\delta\sigma^2 = \sigma^{-2}\delta$. Since $\langle\sigma^2\rangle = \langle\alpha\rangle$, δ normalizes but does not centralize $\langle\alpha\rangle$. Hence $\mathfrak{Q}_2$ is contained in $\mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ and does not centralize α. By the uniqueness, $\mathfrak{Q}_2 = \langle\alpha, \beta\rangle$. We conclude that the group N above is the unique regular subgroup of $\mathrm{Norm}_B(\mathcal{P})$ such that $Q(N)$ maps to $\mathfrak{Q}_2$ in S_{2m}.

Now assume that σ *inverts* α. We show that $\mathfrak{Q}_2$ is in $\mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$. We have that σ is in $\mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ and σ^2 is in $\langle xy^{-1}\rangle$. So σ^2 centralizes α by the proof of Lemma 5.8. Now δ inverts σ^2, hence inverts xy^{-1}. Since $\delta(xy^{-1})\delta^{-1} = x^{-1}y$, either $\delta x\delta^{-1} = x^{-1}$ or $\delta x\delta^{-1} = y$. But δ is a fixed-point-free product of transpositions in S_{2m}. If $\delta x\delta^{-1} = x^{-1}$ then δ restricts to a fixed-point-free product of transpositions of $\mathrm{Supp}\, x$, a set with an odd number of elements. That is not possible. So $\delta x\delta^{-1} = y$ and $\delta y\delta^{-1} = x$, so δ centralizes $\alpha = xy$. Thus $\mathfrak{Q}_2 = \langle\sigma^2, \delta\rangle \in \mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ and centralizes α. Since $\langle\alpha, \beta\rangle \in \mathrm{Norm}_{S_{2q}}(\langle\alpha\rangle)$ and does not centralize α, therefore $\langle\alpha, \beta\rangle = \mathfrak{Q}_2^{\mathrm{opp}}$ by Lemma 5.8.

Now

$$Q(N) = \langle(\hat{a}, 1, \alpha), (\hat{b}, 1, \beta)\rangle.$$

Since $(\hat{0}, 1, \sigma)$ normalizes $\langle(\hat{a}, 1, \alpha)\rangle$, which is characteristic in N, and $\sigma\alpha\sigma = \alpha^{-1}$, we have

$$(\hat{0}, 1, \sigma)(\hat{a}, 1, \alpha)(\hat{0}, 1, \sigma^{-1}) = (\hat{a}, 1, \alpha)^{-1},$$

hence $\sigma(\hat{a}) = -\alpha(\hat{a})$, and so

$$\alpha\sigma(\hat{a}) = -\hat{a}.$$

Since σ inverts α, σ has order $2q$, and α has order q, one sees easily that $\alpha\sigma$ has order $2q$. Hence

$$\hat{a} = [a_1, a_{\alpha\sigma(1)}, \ldots, a_{(\alpha\sigma)^{2q-1}(1)}],$$

while

$$\alpha\sigma(\hat{a}) = [a_{(\alpha\sigma)^{-1}(1)}, a_1, a_{\alpha\sigma(1)}, \ldots, a_{(\alpha\sigma)^{2q-2}(1)}].$$

We have $\alpha\sigma(\hat{a}) = -\hat{a}$, while $(\alpha\sigma)^2(\hat{a}) = \hat{a}$. Thus

$$a_{(\alpha\sigma)^r(1)} = \begin{cases} a_1 & \text{if } r \text{ is even,} \\ -a_1 & \text{if } r \text{ is odd.} \end{cases}$$

Now $(\hat{a}, 1, \alpha)$ has order q, so

$$\sum_{i=0}^{q-1} \alpha^i(\hat{a}) = \hat{0};$$

hence

$$\sum_{i=0}^{q-1} a_{\alpha^{-i}(1)} = 0.$$

But the sum of an odd number of elements of $\mathbb{F}_p$ from a set consisting of copies of a and $-a$ can equal 0 only when $a = 0$.

Thus $\hat{a} = \hat{0}$. Since $(\hat{b}, 1, \beta)$ normalizes $(\hat{0}, 1, \alpha)$, the same argument as in the first case of this proof shows that $\hat{b} = \hat{0}$. Thus

$$N = \mathcal{P} \cdot \langle(\hat{0}, 1, \alpha), (\hat{0}, 1, \beta)\rangle,$$

where $\langle \alpha, \beta \rangle = \mathbb{Q}_2^{\mathrm{opp}}$, hence N is the unique regular subgroup of $\mathrm{Norm}_B(\mathcal{P})$ with $Q(N)$ mapping to $\mathbb{Q}_2^{\mathrm{opp}}$ in S_{2q}. □

***R*(Hol(*C*$_p$), [Hol(*C*$_p$)]).**

Proposition 5.10. $|R(\mathrm{Hol}(C_p), [\mathrm{Hol}(C_p)])| = 2(1 + p(q-2))$.

Proof. $\mathrm{Hol}\, C_p$ is not a direct product of a group of order p and a group of order $m = 2q$, so it suffices to show that $|R(\mathrm{Hol}(C_p), [\mathrm{Hol}(C_p)]; P_0)| = 1 + p(q-2)$. This case is essentially similar to the computation for $R(C_p \rtimes_\tau C_q, [C_p \rtimes_\tau C_q])$ in Section 5, and yields the same cardinality. So instead, we focus here on the case where $P(N) \neq P_0$.

Let $\Gamma = \langle (\hat{1}, 1, I), (\hat{0}, u, \sigma) \rangle$ and let $N = \langle (\hat{p}_{\psi_i}, 1, I), (\hat{b}, u^s, \beta) \rangle$, where $(\hat{b}, u^s, \beta)$ has order m. Since N is regular, β is fixed-point free of order $m = 2q$, so must be an m-cycle, and by the argument of Lemma 4.3 using that $(\hat{b}, u^s, \beta)$ is normalized by Γ, we find that $\beta = \sigma^t$ for some t coprime to m.

Since $N \cong \mathrm{Hol}(C_p)$, the two generators of N, x of order p and y of order m, must satisfy the defining relation $yx = x^u y$, so we must have

$$(\hat{b}, u^s, \sigma^t)(\hat{p}_{\psi_i}, 1, I)(\hat{b}, u^s, \sigma^t)^{-1} = (u\hat{p}_{\psi_i}, 1, I),$$

and hence $u^s \sigma^t \hat{p}_{\psi_i} = u\hat{p}_{\psi_i}$. Since $\sigma(\hat{p}_{\psi_i}) = u^{-i}\hat{p}_{\psi_i}$, this becomes

$$u^{s-it}\hat{p}_{\psi_i} = u\hat{p}_{\psi_i},$$

hence

$$s - it \equiv 1 \pmod{m}. \tag{12}$$

Also, Γ normalizes N. Thus we require that

$$(\hat{1}, 1, I)(\hat{b}, u^s, \sigma^t)(-\hat{1}, 1, I) \in N,$$

hence

$$\hat{b} + (1 - u^s)\hat{1} = f\hat{p}_{\psi_i} + \hat{b}.$$

Thus $(1 - u^s)\hat{1} = f\hat{p}_{\psi_i}$, which for $i \neq 0$ can only occur when both sides equal zero. Thus $s = 0$ and $f = 0$. From (12) we obtain

$$-it \equiv 1 \pmod{m}, \tag{13}$$

hence t is odd and coprime to m.

Since Γ normalizes N, conjugation by $(\hat{0}, u, \sigma^t)$ is an automorphism of N. Every automorphism of N must take the generator y of order m to $x^k y$ for some power x^k of the generator of order p. Thus (noting that $u^s = 1$),

$$(\hat{0}, u, \sigma)(\hat{b}, 1, \sigma^t)(\hat{0}, u^{-1}, \sigma^{-1}) = (k\hat{p}_{\psi_i}, 1, I)(\hat{b}, 1, \sigma^t),$$

for some k, so

$$u\sigma(\hat{b}) = \hat{b} + k\sigma^t(\hat{p}_{\psi_i}) = \hat{b} + ku^{-it}\hat{p}_{\psi_i},$$

which, in view of (13), yields

$$\sigma(\hat{b}) = u^{-1}\hat{b} + k\hat{p}_{\psi_i}.$$

Setting $u^{-1} = w$, we have

$$\sigma(\hat{b}) = w\hat{b} + k\hat{p}_{\psi_i}.$$

For $(\hat{b}, 1, \sigma^t)^m = (\hat{0}, 1, I)$, we need that

$$\hat{b} + \sigma^t(\hat{b}) + \cdots + \sigma^{(m-1)t}(\hat{b}) = 0.$$

This holds if the first elements of the terms on the left side sum to 0:

$$b_1 + b_{\sigma^{-t}(1)} + \cdots + b_{\sigma^{-tj}(1)} + \cdots + b_{\sigma^{-t(m-1)}(1)} = 0. \tag{14}$$

First assume $i \neq 1$. Then for all r, we have

$$\sigma^r(\hat{b}) = w^r\hat{b} + \frac{w^r - w^{ri}}{w - w^i}k\hat{p}_{\psi_i}.$$

Thus, since $(\hat{p}_{\psi_i})_1 = 1$, the first component of $\sigma^r(\hat{b})$ is

$$b_{\sigma^{-r}(1)} = (\sigma^r(\hat{b}))_1 = w^r b_1 + \frac{w^r - w^{ri}}{w - w^i}k.$$

Thus (14) is

$$\begin{aligned}\sum_{l=0}^{m-1} b_{\sigma^{-tl}(1)} &= \sum_{l=0}^{m-1}\left(w^{tl}b_1 + k\left(\frac{w^{tl} - w^{tli}}{w - w^i}\right)\right)\\ &= b_1\left(\frac{w^{tm} - 1}{w^t - 1}\right) + k\sum_{l=0}^{m-1}\left(\frac{w^{tl} - w^{tli}}{w - w^i}\right).\end{aligned}$$

Since $w^m = 1$, the first sum is 0; so this becomes

$$\begin{aligned}&= \frac{k}{w - w^i}\sum_{l=0}^{m-1} w^{tl} - \sum_{l=0}^{m-1} w^{tli}\\ &= \frac{k}{w - w^i}\,\frac{w^{tm} - 1}{w^t - 1} - \frac{w^{tim} - 1}{w^{ti} - 1}.\end{aligned}$$

Now $ti \equiv -1 \pmod{m}$, so $w^{ti} = w^{-1}$ and so both terms in this last equation equal zero. Thus (14) holds if $i \neq 1$.

If $i = 1$, then $t = -1$ and $\sigma^r(\hat{b}) = w^r\hat{b} + rw^{r-1}k\hat{p}_{\psi_i}$ for all r. Thus (14) becomes

$$\sum_{l=0}^{m-1} b_{\sigma^{-tl}(1)} = \sum_{l=0}^{m-1} w^{tl}b_1 + k\sum_{l=0}^{m-1}(tlw^{tl} - 1).$$

The first sum on the right is equal to zero. By the same observation as with (9), the second sum on the right equals zero if and only if $k = 0$. Thus when $i = 1$ and $t = -1$, the generator $(\hat{b}, 1, \sigma^t)$ has order m if and only if $\sigma(\hat{b}) = u^{-1}\hat{b}$. In that case, $\hat{b} = b_1 \hat{p}_{\psi_1}$, and so replacing the generator $(\hat{b}, 1, \sigma^{-1})$ of N by $(-b_1 \hat{p}_{\psi_1}, 1, I)(\hat{b}, 1, \sigma^{-1}) = (0, 1, \sigma^{-1})$ yields

$$N = \langle (\hat{p}_{\psi_1}, 1, I), (\hat{0}, 1, \sigma^{-1}) \rangle.$$

Thus there is a unique regular subgroup N when $t = -1$. For $t \neq -1$, each b_1 yields a different N, hence we have a total of $1 + (q-2)p$ regular subgroups N with $P(N) \neq \mathcal{P}$. By Corollary 3.6, this implies that $|R(\mathrm{Hol}(C_p), [\mathrm{Hol}(C_p)])| = 2(1 + (q-2)p)$. □

The enumeration of $R(\mathrm{Hol}(C_p), [\mathrm{Hol}(C_p)])$ is in agreement with that in [Childs 2003].

6. Conclusion

The program developed here to enumerate $R(\Gamma, [M])$ may be readily applied to any class of groups of order mp with $p > m$. The primary requirement is to start with the groups of order m and for the particular p determine the set of linear characters for each group of order m. One may find that, depending on congruence conditions between m and p the number of possible characters may vary greatly. Nonetheless, one is presented with a very interesting set of calculations, wherein one may apply many different techniques. What is most interesting is the interplay between the linear and combinatorial information in the different cases. For small m and p these computations may be readily implemented in a computer algebra system such as GAP [2002]. This was done by the author in the development of this work, especially in gathering empirical information about some specific cases, for example, with $mp = 42$. Lastly, and this is mildly conjectural, it seems that the theory developed here applies to certain cases where actually $p < m$. Specifically, one might consider those cases where $p \nmid m$ and the order-p subgroup is automatically characteristic due to basic Sylow theory, for example, $p = 5$ and $m = 8$.

Acknowledgements

The author wishes to thank the referee, Lindsay Childs, for the time and attention spent on reading this article. His insights greatly improved the exposition of the results and the concision of many of the arguments.

References

[Burnside 1911] W. Burnside, *Theory of groups of finite order*, 2nd ed., Cambridge University Press, Cambridge, 1911. Reprinted by Dover in 1955. MR 16,1086c JFM 42.0151.02

[Byott 1996] N. P. Byott, "Uniqueness of Hopf Galois structure for separable field extensions", *Comm. Algebra* **24**:10 (1996), 3217–3228. MR 97j:16051a Zbl 0878.12001

[Byott 2000] N. P. Byott, "Galois module theory and Kummer theory for Lubin–Tate formal groups", pp. 55–67 in *Algebraic number theory and Diophantine analysis* (Graz, 1998), edited by F. Halter-Koch and R. F. Tichy, de Gruyter, Berlin, 2000. Zbl 0958.11076

[Byott 2004] N. P. Byott, "Hopf–Galois structures on Galois field extensions of degree pq", *J. Pure Appl. Algebra* **188**:1-3 (2004), 45–57. MR 2004j:16041 Zbl 1047.16022

[Chase and Sweedler 1969] S. U. Chase and M. E. Sweedler, *Hopf algebras and Galois theory*, Lecture Notes in Mathematics **97**, Springer, Berlin, 1969. MR 41 #5348 Zbl 0197.01403

[Childs 1989] L. N. Childs, "On the Hopf Galois theory for separable field extensions", *Comm. Algebra* **17**:4 (1989), 809–825. MR 90g:12003 Zbl 0692.12007

[Childs 2003] L. N. Childs, "On Hopf Galois structures and complete groups", *New York J. Math.* **9** (2003), 99–115. MR 2004k:16097 Zbl 1038.12003

[Dixon 1971] J. D. Dixon, "Maximal abelian subgroups of the symmetric groups", *Canad. J. Math.* **23** (1971), 426–438. MR 43 #7496 Zbl 0213.03301

[GAP Group 2002] GAP Group, "GAP: groups, algorithms, and programming", 2002, Available at http://www.gap-system.org. Version 4.3.

[Greither and Pareigis 1987] C. Greither and B. Pareigis, "Hopf Galois theory for separable field extensions", *J. Algebra* **106**:1 (1987), 239–258. MR 88i:12006 Zbl 0615.12026

[Isaacs 1976] I. M. Isaacs, *Character theory of finite groups*, Pure and Applied Mathematics **69**, Academic Press, New York, 1976. MR 57 #417 Zbl 0337.20005

[Kohl 2007] T. Kohl, "Groups of order $4p$, twisted wreath products and Hopf–Galois theory", *J. Algebra* **314**:1 (2007), 42–74. MR 2008e:12001 Zbl 1129.16031

[Krasner and Kaloujnine 1951] M. Krasner and L. Kaloujnine, "Produit complet des groupes de permutations et problème d'extension de groupes, III", *Acta Sci. Math. Szeged* **14** (1951), 69–82. MR 14,242d Zbl 0045.30301

[Moody 1994] J. A. Moody, *Groups for undergraduates*, World Scientific, River Edge, NJ, 1994. MR 96e:20001 Zbl 0832.20001

[Neumann 1963] B. H. Neumann, "Twisted wreath products of groups", *Arch. Math.* (*Basel*) **14** (1963), 1–6. MR 26 #5040 Zbl 0108.02602

[Ore 1942] O. Ore, "Theory of monomial groups", *Trans. Amer. Math. Soc.* **51**:1 (1942), 15–64. MR 3,197e Zbl 0028.00304

[Wielandt 1955] H. Wielandt, *Permutationsgruppen*, Mathematische Institut, Tübingen, 1955. Translated as *Finite permutation groups*, Academic Press, New York, 1964. MR 32 #1252 Zbl 0138.02501

Communicated by Susan Montgomery
Received 2012-09-08 Revised 2013-02-02 Accepted 2013-03-11

tkohl@math.bu.edu *Department of Mathematics and Statistics, Boston University, 111 Cummington Mall, Boston, MA 02215, United States*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2241

Further evidence for conjectures in block theory

Benjamin Sambale

We prove new inequalities concerning Brauer's $k(B)$-conjecture and Olsson's conjecture by generalizing old results. After that, we obtain the invariants for 2-blocks of finite groups with certain bicyclic defect groups. Here, a bicyclic group is a product of two cyclic subgroups. This provides an application for the classification of the corresponding fusion systems in a previous paper. To some extent, this generalizes previously known cases with defect groups of types $D_{2^n} \times C_{2^m}$, $Q_{2^n} \times C_{2^m}$ and $D_{2^n} * C_{2^m}$. As a consequence, we prove Alperin's weight conjecture and other conjectures for several new infinite families of nonnilpotent blocks. We also prove Brauer's $k(B)$-conjecture and Olsson's conjecture for the 2-blocks of defect at most 5. This completes results from a previous paper. The $k(B)$-conjecture is also verified for defect groups with a cyclic subgroup of index at most 4. Finally, we consider Olsson's conjecture for certain 3-blocks.

1. Introduction

Let B be a p-block of a finite group G. One aim of this paper is to establish new inequalities on the number of irreducible characters of B in terms of subsections. We outline the idea behind these things.

Olsson [1981] proved the following:

$$l(B) \leq 2 \implies k(B) \leq p^d,$$

where d is the defect of B. In particular, this gives an example for Brauer's $k(B)$-conjecture. However, in practice this implication is not so useful because usually the knowledge of $l(B)$ already implies the exact value of $k(B)$. Since the proofs in [Olsson 1981] only rely on computations with the contributions of the trivial subsection $(1, B)$, it seems likely that one can extend this result to major subsections. Then we would be able to apply induction on d (see Theorem 4.9). Hence, let (z, b_z) be a major subsection such that $l(b_z) \leq 2$. In case $l(b_z) = 1$, we have

MSC2010: primary 20C15; secondary 20C20.
Keywords: 2-blocks, bicyclic defect groups, Brauer's $k(B)$-conjecture, Alperin's weight conjecture.

$\sum k_i(B)p^{2i} \leq p^d$ by [Robinson 1991, Theorem 3.4] (a stronger bound can be found in [Héthelyi et al. 2013]).

In Section 2, we show

$$l(b_z) \leq 2 \implies k(B) \leq p^d.$$

In contrast to Olsson's paper, we use methods from [Robinson 1991] and [Sambale 2011b]. For $p = 2$, Olsson proved the stronger statement $l(B) \leq 3 \Longrightarrow k(B) \leq p^d$. Using his ideas, we generalize this to major subsections as well. The underlying properties of the contribution matrices were first discovered by Brauer [1968], but we will refer to [Feit 1982] for a more accessible account. Using Galois theory, we overcome the difficulty that the contributions are not necessarily integers in this general setting.

More generally, we consider arbitrary subsections for the prime 2 in order to give bounds on the number of characters of height 0. Here it is known by [Broué 1980] (more recent accounts can be found in [Robinson 1992; Murai 2000]) that the corresponding contributions for characters of height 0 do not vanish. Using exactly the same method, we show that $k_0(B) \leq 2^q$ if there is a subsection (u, b_u) such that b_u has defect q and $l(b_u) \leq 3$.

In Section 3, we present new infinite families of defect groups for which the block invariants can be calculated. These defect groups are examples of bicyclic 2-groups (i.e., $D = \langle x\rangle\langle y\rangle$ for some $x, y \in D$). The proofs make use of the classification of the corresponding fusion systems in [Sambale 2012b]. However, we cannot handle all bicyclic 2-groups. We also remark that these defect groups are in a sense noncommutative versions of the groups $D_{2^n} \times C_{2^m}$, $Q_{2^n} \times C_{2^m}$ and $D_{2^n} * C_{2^m}$ covered in [Sambale 2012a; 2013b; 2013a]. As a consequence, we verify numerous conjectures including Alperin's weight conjecture for these blocks.

In Section 4, we collect some more or less related examples for block invariants. In particular, we discuss some defect groups of order 32. One of the main results here is the verification of Brauer's $k(B)$-conjecture for the 2-blocks of defect at most 5. This completes [Sambale 2011c, Theorem 3]. The new ingredient here is in fact an old result of Brauer that uses the inverse of the Cartan matrix of a major subsection.

In Section 5, we obtain new cases for Olsson's conjecture. In particular, we handle the 2-blocks of defect at most 5 and some 3-blocks with defect group of 3-rank 2 that were left over in [Héthelyi et al. 2013].

2. New inequalities

Let B be a p-block of a finite group G with defect group D. We define the *height* $h(\chi)$ of a character $\chi \in \mathrm{Irr}(B)$ by $\chi(1)_p = p^{h(\chi)}|G : D|_p$. Moreover,

$\mathrm{Irr}_i(B) := \{\chi \in \mathrm{Irr}(B) : h(\chi) = i\}$, $k(B) := |\mathrm{Irr}(B)|$ and $k_i(B) := |\mathrm{Irr}_i(B)|$ for $i \geq 0$. As usual, we denote the set of irreducible Brauer characters of B by $\mathrm{IBr}(B)$ and its cardinality by $l(B) := |\mathrm{IBr}(B)|$.

In the following, we choose an element $z \in \mathrm{Z}(D)$. Then there exists a Brauer correspondent b_z of B in $\mathrm{C}_G(z)$. The pair (z, b_z) is called a *major subsection*.

Theorem 2.1. *Let B be a p-block of a finite group with defect d, and let (z, b_z) be a major subsection such that $l(b_z) \leq 2$. Then one of the following holds*:

$$\sum_{i=0}^{\infty} k_i(B) p^{2i} \leq p^d. \tag{1}$$

$$k(B) \leq \begin{cases} ((p+3)/2)p^{d-1} & \text{if } p > 2, \\ \frac{2}{3}2^d & \text{if } p = 2. \end{cases} \tag{2}$$

In particular, Brauer's $k(B)$-conjecture holds for B.

Proof. In case $l(b_z) = 1$, equation (1) holds. Hence, let $l(b_z) = 2$, and let $C_z = (c_{ij})$ be the Cartan matrix of b_z up to basic sets. We consider the number

$$q(b_z) := \min\{x p^d C_z^{-1} x^{\mathrm{T}} : 0 \neq x \in \mathbb{Z}^{l(b_z)}\} \in \mathbb{N}.$$

If $q(b_z) = 1$, equation (1) follows from [Robinson 1991, Theorem 3.4.1]. Therefore, we may assume $q(b_z) \geq 2$. Then Brauer's $k(B)$-conjecture already holds by [Feit 1982, Theorem V.9.17], but we want to obtain the stronger bound (2). Since p^d is always an elementary divisor of C_z, we see that C_z is not a diagonal matrix. This allows us to apply [Héthelyi et al. 2013, Theorem 2.4]. All entries of C_z are divisible by the smallest elementary divisor $\gamma := p^{-d} \det C_z$. Hence, we may consider the integral matrix $\widetilde{C}_z = (\tilde{c}_{ij}) := \gamma^{-1} C_z$. After changing the basic set, we may assume that $0 < 2\tilde{c}_{12} \leq \tilde{c}_{11} \leq \tilde{c}_{22}$. Then

$$\tilde{c}_{11}\tilde{c}_{22} - \frac{\tilde{c}_{11}^2}{4} \leq \tilde{c}_{11}\tilde{c}_{22} - \tilde{c}_{12}^2 = \det \widetilde{C}_z = \frac{p^d}{\gamma}$$

and

$$\tilde{c}_{11} + \tilde{c}_{22} \leq \tfrac{5}{4}\tilde{c}_{11} + \frac{\det \widetilde{C}_z}{\tilde{c}_{11}} =: f(\tilde{c}_{11}).$$

A discussion of the convex function $f(\tilde{c}_{11})$ as in [Sambale 2011b, Theorem 1] shows that $\tilde{c}_{11} + \tilde{c}_{22} \leq f(2)$. Now [Héthelyi et al. 2013, Theorem 2.4] leads to

$$k(B) \leq \gamma(\tilde{c}_{11} + \tilde{c}_{22} - \tilde{c}_{12}) \leq \gamma(f(2) - 1) \leq \frac{p^d + 3\gamma}{2}.$$

Since $\gamma \leq p^{d-1}$, we get (2) for p odd. In order to deduce the $k(B)$-conjecture, we need to consider the case $p = 2$. If $\tilde{c}_{11} = 2$, we must have $\tilde{c}_{12} = 1$. Hence, under

these circumstances $p > 2$ since otherwise $\det \widetilde{C}_z$ is not a p-power. Now assume $\tilde{c}_{11} \geq 3$ and $p = 2$. Since

$$p^d C_z^{-1} = \frac{p^d}{\gamma} \widetilde{C}_z^{-1} = \begin{pmatrix} \tilde{c}_{22} & -\tilde{c}_{12} \\ -\tilde{c}_{12} & \tilde{c}_{11} \end{pmatrix},$$

we have $q(b_z) \geq 3$. Now [Feit 1982, Theorem V.9.17] implies (2). We will derive another estimation for $p = 2$ in Theorem 2.2 below. □

It is conjectured that the matrix C_z for $l(b_z) \geq 2$ in the proof of Theorem 2.1 cannot have diagonal shape (this holds for p-solvable groups by [Sambale 2011b, Lemma 1]). Hence, for $l(b_z) = 2$ Theorem 2.1(2) might always apply. Then $k(B) < p^d$ unless $p = 3$.

In order to improve Theorem 2.1 for $p = 2$, we need more notation. Suppose as before that (z, b_z) is a major subsection. We denote the corresponding part of the generalized decomposition matrix by $D_z := (d^z_{\chi\varphi} : \chi \in \mathrm{Irr}(B),\ \varphi \in \mathrm{IBr}(b_z))$. Then the Cartan matrix of b_z is given by $C_z := D_z^{\mathrm{T}} D_z$. Moreover, the contribution matrix of b_z is defined as

$$M_z := (m^z_{\chi\psi})_{\chi,\psi \in \mathrm{Irr}(B)} = |D| D_z C_z^{-1} \overline{D_z}^{\mathrm{T}}.$$

In case $|\langle z \rangle| \leq 2$, it can be seen easily that M_z is an integral matrix. Then most proofs of [Olsson 1981] remain true without any changes. This was more or less done in [Robinson 2008] (compare also Corollary 3.5 in [Robinson 1991]). In the general case, we have to put a bit more effort into the proof.

Theorem 2.2. *Let B be a 2-block of a finite group with defect d, and let (z, b_z) be a major B-subsection such that $l(b_z) \leq 3$. Then*

$$k(B) \leq k_0(B) + \frac{2}{3} \sum_{i=1}^{\infty} 2^i k_i(B) \leq 2^d.$$

In particular, Brauer's $k(B)$-conjecture is satisfied for B.

Proof. Observe that by construction $m^z_{\chi\chi}$ is a positive real number for every $\chi \in \mathrm{Irr}(B)$ since C_z is positive definite. Since all elementary divisors of C_z are divisors of 2^d, the matrix $2^d C_z^{-1}$ is integral. In particular, the numbers $m^z_{\chi\psi}$ are also algebraic integers. Let $\chi \in \mathrm{Irr}(B)$ be a character of height 0. Let $|\langle z \rangle| = 2^n$. In case $n \leq 1$, the proof is much easier. For this reason, we assume $n \geq 2$. We write

$$m^z_{\chi\chi} = \sum_{j=0}^{2^{n-1}-1} a_j(\chi) \zeta^j$$

with $\zeta := e^{2\pi i/2^n}$ and $a_j(\chi) \in \mathbb{Z}$ for $j = 0, \ldots, 2^{n-1} - 1$. As usual, the Galois group $\mathcal{G}$ of the 2^n-th cyclotomic field acts on $\mathrm{Irr}(B)$, on the rows of D_z and thus

also on M_z in an obvious manner. Let Γ be the orbit of χ under $\mathcal{G}$. Set $m := |\Gamma|$. Then we have

$$ma_0(\chi) = \sum_{\psi\in\Gamma} m^z_{\psi\psi} > 0.$$

Assume first that $a_0(\chi) = 1$. Since $M_z^2 = M_z\overline{M}_z{}^{\mathrm{T}} = 2^d M_z$ (see [Feit 1982, Theorem V.9.4]), it follows that

$$m2^d = \sum_{\substack{\psi\in\Gamma\\ \tau\in\mathrm{Irr}(B)}} |m^z_{\psi\tau}|^2.$$

Applying Galois theory gives

$$\prod_{\substack{\psi\in\Gamma\\ \tau\in\mathrm{Irr}_i(B)}} |m^z_{\psi\tau}|^2 \in \mathbb{Q}$$

for all $i \geq 0$. By [Feit 1982, Theorem V.9.4], we also know $\nu(m^z_{\psi\tau}) = h(\tau)$, where ν is the 2-adic valuation and $\psi \in \Gamma$. Hence, also the numbers $m^z_{\psi\tau}2^{-h(\tau)}$ are algebraic integers. This implies

$$\mathbb{Z} \ni \prod_{\substack{\psi\in\Gamma\\ \tau\in\mathrm{Irr}_i(B)}} p^{-2i}|m^z_{\psi\tau}|^2 \geq 1.$$

Now using the inequality of arithmetic and geometric means, we obtain

$$\sum_{\substack{\psi\in\Gamma\\ \tau\in\mathrm{Irr}_i(B)}} |m^z_{\psi\tau}|^2 \geq m2^{2i}k_i(B)$$

for all $i \geq 0$. Summing over i gives

$$m2^d = \sum_{\substack{\psi\in\Gamma\\ \tau\in\mathrm{Irr}(B)}} |m^z_{\psi\tau}|^2 \geq m\sum_{i=0}^{\infty} 2^{2i}k_i(B),$$

which is even more than we wanted to prove.

Hence, we can assume that $a_0(\chi) \geq 2$ for all $\chi \in \mathrm{Irr}(B)$ such that $h(\chi) = 0$. It is well-known that the ring of integers of $\mathbb{Q}(\zeta) \cap \mathbb{R}$ has basis $\{1, \zeta^j + \zeta^{-j} = \zeta^j - \zeta^{2^{n-1}-j} : j = 1, \ldots, 2^{n-2} - 1\}$. In particular, the numbers $a_j(\chi)$ for $j \geq 1$ come in pairs modulo 2. Since $\nu(m^z_{\chi\chi}) = 0$, we even have $a_0(\chi) \geq 3$. For an arbitrary character $\psi \in \mathrm{Irr}(B)$ of positive height, we already know that $m^z_{\psi\psi}2^{-h(\psi)}$ is a positive algebraic integer. Hence, $2^{h(\psi)} \mid a_j(\psi)$ for all $j \geq 0$. By [Feit 1982, Theorem V.9.4], we have $\nu(m^z_{\psi\psi}) > h(\psi)$. Thus, we even have $2^{h(\psi)+1} \mid a_0(\psi)$. As above, we also have $a_0(\psi) > 0$. This implies $\sum_{\psi\in\mathrm{Irr}_i(B)} m^z_{\psi\psi} \geq 2^{i+1}k_i(B)$ for $i \geq 1$

via Galois theory. Using $\operatorname{tr} M_z = 2^d l(b_z)$, it follows that

$$3 \cdot 2^d \geq \sum_{\psi \in \operatorname{Irr}(B)} m^z_{\psi\psi} \geq 3k_0(B) + \sum_{i=1}^{\infty} 2^{i+1} k_i(B).$$

This proves the claim. □

We remark that Theorem 6(ii) in [Olsson 1981] should read $l(B) \leq p^2 - 1$ (compare with Theorem 6*(ii)).

It is easy to see that the proof of Theorem 2.2 can be generalized to the following:

Proposition 2.3. *Let B be a 2-block of a finite group with defect d, and let (z, b_z) be a major B-subsection. Then for every odd number α one of the following holds*:

$$\sum_{i=0}^{\infty} 2^{2i} k_i(B) \leq 2^d \alpha. \tag{1}$$

$$(\alpha + 2)k_0(B) + \sum_{i=1}^{\infty} 2^{i+1} k_i(B) \leq 2^d l(b_z). \tag{2}$$

Proof. As in Theorem 2.2, let $\chi \in \operatorname{Irr}_0(B)$ and define $a_0(\chi)$ similarly. In case $a_0(\chi) \leq \alpha$, the first inequality applies. Otherwise, the second inequality applies. □

Observe that Proposition 2.3 also covers (a generalization of) [Olsson 1981, Theorem 8] for $p = 2$.

Going over to arbitrary subsections (i.e., the element does not necessarily belong to $\mathrm{Z}(D)$), we can prove the following result concerning Olsson's conjecture. This improves [Robinson 1992, Theorem 3.1] for $p = 2$.

Theorem 2.4. *Let B be a 2-block of a finite group, and let (u, b_u) be a B-subsection such that b_u has defect q. Set $\alpha := \lfloor \sqrt{l(b_u)} \rfloor$ if $\lfloor \sqrt{l(b_u)} \rfloor$ is odd and $\alpha := l(b_u)/(\lfloor \sqrt{l(b_u)} \rfloor + 1)$ otherwise. Then $k_0(B) \leq \alpha 2^q$. In particular, $k_0(B) \leq 2^q$ if $l(b_u) \leq 3$.*

Proof. The contributions for (u, b_u) are defined by

$$M_u := (m^u_{\chi\psi})_{\chi,\psi \in \operatorname{Irr}(B)} = p^q D_u C_u^{-1} \overline{D_u}^{\mathrm{T}}.$$

By [Murai 2000, Corollary 1.15], we still have $m^u_{\chi\psi} \neq 0$ as long as $h(\chi) = h(\psi) = 0$. However, in all other cases it is possible that $m^u_{\chi\psi} = 0$. So we can copy the proof of Theorem 2.2 by leaving out the characters of positive height. This gives $k_0(B) \leq \alpha 2^q$ or $k_0(B) \leq 2^q l(b_u)/(\alpha + 2)$ for every odd number α. If $\lfloor \sqrt{l(b_u)} \rfloor$ is odd, we choose $\alpha := \lfloor \sqrt{l(b_u)} \rfloor$. Otherwise, we take $\alpha := \lfloor \sqrt{l(b_u)} \rfloor - 1$. The result follows. □

Finally, we generalize the "dual" inequalities in [Olsson 1981]. For this, let $M'_z := (m'_{\chi\psi}) = 2^d 1_{k(B)} - M_z$.

Proposition 2.5. *Let B be a 2-block of a finite group with defect d, and let (z, b_z) be a major B-subsection. Then for every odd number α one of the following holds*:

$$\sum_{i=0}^{\infty} 2^{2i} k_i(B) \leq 2^d \alpha. \tag{1}$$

$$(\alpha+2)k_0(B) + \sum_{i=1}^{\infty} 2^{i+1} k_i(B) \leq 2^d (k(B) - l(b_z)). \tag{2}$$

In particular, Brauer's $k(B)$-conjecture holds if $k(B) - l(b_z) \leq 3$.

Proof. By [Feit 1982, Lemma V.9.3], the numbers $m'_{\chi\chi}$ for $\chi \in \mathrm{Irr}(B)$ are still real, positive algebraic integers. As in Theorem 2.2, we may assume $|\langle z \rangle| = 2^n \geq 4$. Let us write

$$m'_{\chi\chi} = \sum_{j=0}^{2^{n-1}-1} a_j(\chi)\zeta^j$$

with $\chi \in \mathrm{Irr}_0(B)$, $\zeta := e^{2\pi i/2^n}$ and $a_j(\chi) \in \mathbb{Z}$ for $j = 0, \ldots, 2^{n-1}-1$. The Galois group still acts on M'_z. Also the equation $(M'_z)^2 = M'_z \overline{M}'^{\mathrm{T}}_z = 2^d M'_z$ remains true. For $\tau \in \mathrm{Irr}(B)$, we have $\nu(m'_{\chi\tau}) = \nu(2^d - m^z_{\chi\tau}) = \nu(m^z_{\chi\tau}) = h(\tau)$. Hence, in case $a_0(\chi) \leq \alpha$ we can carry over the arguments in Theorem 2.2.

Now assume that $a_0(\chi) > \alpha$ for all characters $\chi \in \mathrm{Irr}_0(B)$. Here too the proof works much as in Theorem 2.2. In fact, for a character $\psi \in \mathrm{Irr}(B)$ of positive height we have $\nu(m'_{\psi\psi}) = \nu(2^d - m^z_{\psi\psi}) \geq \min\{\nu(2^d), \nu(m^z_{\psi\psi})\} > h(\psi)$ by [Feit 1982, Theorem V.9.4]. Moreover, $\mathrm{tr}\, M'_z = 2^d(k(B) - l(B))$. The claim follows. □

It should be pointed out that usually $k(B) - l(B) = k(B) - l(b_1) \leq k(B) - l(b_z)$ for a major subsection (z, b_z) (this holds for example if z lies in the center of the fusion system of B; see [Külshammer and Okuyama ∼ 2000]). However, this is not true in general as we see in [Külshammer and Sambale 2013, Proposition 2.1(vii)]. Another problem is that $k(B) - l(b_z)$ for $z \neq 1$ is not locally determined (in contrast to $k(B) - l(B)$). By combining with Proposition 2.3, we can replace Proposition 2.5(2) by

$$(\alpha+2)k_0(B) + \sum_{i=1}^{\infty} 2^{i+1} k_i(B) \leq 2^d \min\{l(b_z), k(B) - l(b_z)\}.$$

3. Bicyclic defect groups

As mentioned in the introduction, we consider in this section blocks with defect groups coming from [Sambale 2012b, Theorem 4.19]. A key feature of the groups in the next three theorems is that all their irreducible characters have degree 1 or 2. We also remark that Olsson's conjecture was verified for all blocks with bicyclic defect groups in [Sambale 2012b].

Theorem 3.1. *Let B be a nonnilpotent 2-block of a finite group with defect group*

$$D \cong \langle v, x, a \mid v^{2^n} = x^2 = a^{2^m} = 1,\ {}^x v = {}^a v = v^{-1},\ {}^a x = vx \rangle \cong D_{2^{n+1}} \rtimes C_{2^m}$$

for some $n, m \geq 2$. Then $k(B) = 2^{m-1}(2^n+3)$, $k_0(B) = 2^{m+1}$, $k_1(B) = 2^{m-1}(2^n-1)$ and $l(B) = 2$. In particular, Brauer's $k(B)$-conjecture and Alperin's weight conjecture are satisfied.

Proof. Let $\mathcal{F}$ be the fusion system of B, and let $z := v^{2^{n-1}}$. Then by [Sambale 2012b, Theorem 4.19], $Q := \langle z, x, a^2 \rangle$ is the only $\mathcal{F}$-essential subgroup up to conjugation. In order to calculate $k(B)$, we use Brauer's formula [Nagao and Tsushima 1989, Theorem 5.9.4]. We will see that it is not necessary to obtain a complete set of representatives for the $\mathcal{F}$-conjugacy classes. Since $\langle v, ax \rangle$ is an abelian maximal subgroup of D, all characters in $\mathrm{Irr}(D)$ have degree 1 or 2. In particular, $k(D) := |\mathrm{Irr}(D)| = |D/D'| + (|D| - |D/D'|)/4 = 2^{m-1}(2^n+3)$. Now we have to count how many conjugacy classes of D are fused under $\mathrm{Aut}_{\mathcal{F}}(Q)$. According to [Sambale 2012b, Theorem 4.19], there are two possibilities $\mathrm{C}_Q(\mathrm{Aut}_{\mathcal{F}}(Q)) = \mathrm{Z}(\mathcal{F}) \in \{\langle a^2 \rangle, \langle a^2 z \rangle\}$. In the first case, the elements of the form xa^{2j} are conjugate to corresponding elements za^{2j} under $\mathrm{Aut}_{\mathcal{F}}(Q)$. In the second case, a similar statement is true for a^{2j}. Observe that the elements xa^{2^j} and xza^{2j} are already conjugate in D. Since $\langle a^2, z \rangle \subseteq \mathrm{Z}(D)$, no more fusion can occur. Hence, the number of $\mathcal{F}$-conjugacy classes is $2^{m-1}(2^n+3) - 2^{m-1} = 2^m(2^{n-1}+1)$.

Now we have to determine at least some of the numbers $l(b_u)$ where $u \in D$. The groups $\overline{D}_1 := D/\langle a^2 \rangle$ and $\overline{D}_2 := D/\langle a^2 z \rangle$ have commutator subgroups $D'\langle a^2 \rangle/\langle a^2 \rangle$ and $D'\langle a^2 z \rangle/\langle a^2 z \rangle$ of index 4, respectively. Hence, $\overline{D}_1$ and $\overline{D}_2$ have maximal class. The blocks b_{a^2} and $b_{a^2 z}$ dominate blocks $\overline{b_{a^2}}$ and $\overline{b_{a^2 z}}$, respectively, with defect group $\overline{D}_1$. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be the fusion systems of $\overline{b_{a^2}}$ and $\overline{b_{a^2 z}}$, respectively. Then in case $\mathrm{Z}(\mathcal{F}) = \langle a^2 \rangle$ or $\mathrm{Z}(\mathcal{F}) = \langle a^2 z \rangle$, $\overline{Q}$ is the only $\mathcal{F}_1$-essential or $\mathcal{F}_2$-essential subgroup of $\overline{D}_1$ or $\overline{D}_2$, respectively, up to conjugation. Thus, [Brauer 1974; Olsson 1975] imply $l(b_{a^2}) = l(\overline{b_{a^2}}) = 2$ and $l(b_{a^2 z}) = l(\overline{b_{a^2 z}}) = 2$. The same holds for all odd powers of a^2 and $a^2 z$. Next we consider the elements $u := a^{2^j}$ for $2 \leq j \leq m-1$. It can be seen that the isomorphism type of $D/\langle u \rangle$ is the same as for D except that we have to replace m by j. Also the essential subgroup Q carries over to the block $\overline{b_u}$. Hence, induction on m gives $l(b_u) = 2$ as well. For all other nontrivial subsections (u, b_u), we only know $l(b_u) \geq 1$. Finally, $l(B) \geq 2$ since B is centrally controlled, by Theorem 1.1 in [Külshammer and Okuyama ~ 2000]. Applying Brauer's formula gives

$$k(B) \geq 2^m + 2^m(2^{n-1}+1) - 2^{m-1} = 2^{m-1}(2^n+3) = k(D).$$

We already know from [Sambale 2012b, Theorem 5.3] that Olsson's conjecture holds for B, i.e., $k_0(B) \leq |D : D'| = 2^{m+1}$. Now we apply [Robinson 1991, Theorem 3.4]

to the subsection (z, b_z), which gives

$$|D| = 2^{m+1} + 2^{m+1}(2^n - 1) \leq k_0(B) + 4(k(B) - k_0(B)) \leq \sum_{i=0}^{\infty} 2^{2i} k_i(B) \leq |D|.$$

This implies $k(B) = k(D) = 2^{m-1}(2^n + 3)$, $k_0(B) = 2^{m+1}$, $k_1(B) = 2^{m-1}(2^n - 1)$ and $l(B) = 2$. Brauer's $k(B)$-conjecture follows immediately. In order to prove Alerin's weight conjecture (see [Kessar 2007, Proposition 5.4]), it suffices to show that Q and D are the only $\mathcal{F}$-radical, $\mathcal{F}$-centric subgroups of D. Thus, assume by way of contradiction that Q_1 is another $\mathcal{F}$-radical, $\mathcal{F}$-centric subgroup. Since Q_1 is $\mathcal{F}$-centric, it cannot lie inside Q. Moreover, $\mathrm{Out}_{\mathcal{F}}(Q_1)$ must provide a morphism of odd order because $Q_1 < D$. However, by Alperin's fusion theorem $\mathcal{F}$ is generated by $\mathrm{Aut}_{\mathcal{F}}(Q)$ and $\mathrm{Aut}_{\mathcal{F}}(D)$. This gives the desired contradiction. □

We add some remarks. First, the direct products of similar type $D_{2^{n+1}} \times C_{2^m}$ were already handled in [Sambale 2012a]. Also if $n = 1$, we obtain the minimal nonabelian group $C_2^2 \rtimes C_{2^m}$ for which the block invariants are also known by [Sambale 2011a]. Moreover, it is an easy exercise to check that various other conjectures (for example [Eaton and Moretó 2013; Eaton 2003; Malle and Navarro 2006]) are also true in the situation of Theorem 3.1. We will not go into the details here.

The next theorem concerns defect groups that have a similar structure as the central products $Q_{2^{n+1}} * C_{2^m}$ discussed in [Sambale 2013a]. Also, this result is needed for the induction step in the theorem after that.

Theorem 3.2. *Let B be a nonnilpotent 2-block of a finite group with defect group*

$$D \cong \langle v, x, a \mid v^{2^n} = 1,\ a^{2^m} = x^2 = v^{2^{n-1}},\ {}^x v = {}^a v = v^{-1},\ {}^a x = vx \rangle$$
$$\cong Q_{2^{n+1}}.C_{2^m} \cong D_{2^{n+1}}.C_{2^m}$$

for some $n, m \geq 2$ and $m \neq n$. Then $k(B) = 2^{m+1}(2^{n-2} + 1)$, $k_0(B) = 2^{m+1}$, $k_1(B) = 2^{m-1}(2^n - 1)$, $k_n(B) = 2^{m-1}$ and $l(B) = 2$. In particular, Brauer's $k(B)$-conjecture and Alperin's weight conjecture are satisfied.

Proof. First observe that the proof of [Sambale 2012b, Theorem 4.20] shows that in fact

$$D \cong \langle v, x, a \mid v^{2^n} = x^2 = 1,\ a^{2^m} = v^{2^{n-1}},\ {}^x v = {}^a v = v^{-1},\ {}^a x = vx \rangle \cong D_{2^{n+1}}.C_{2^m}.$$

Let $\mathcal{F}$ be the fusion system of B, and let $y := v^{2^{n-2}}$ and $z := x^2$. Then by [Sambale 2012b, Theorem 4.19], $Q := \langle x, y, a^2 \rangle \cong Q_8 * C_{2^m}$ is the only $\mathcal{F}$-essential subgroup up to conjugation (since $n \neq m$, D is not a wreath product). Again we use Brauer's formula [Nagao and Tsushima 1989, Theorem 5.9.4] to get a lower bound for $k(B)$. The same argumentation as in Theorem 3.1 shows that D has $2^{m-1}(2^n + 3)$ conjugacy classes, and we need to know which of them are fused in Q. It is easy to

see that xa^{2^j} is conjugate to ya^{2^j} under $\mathrm{Aut}_{\mathcal{F}}(Q)$ for $j \in \mathbb{Z}$. Observe that xa^{2^j} is already conjugate to xya^{2^j} and $x^{-1}a^{2^j} = xa^{2^j+2^m}$ in D. Since $\mathrm{Z}(\mathcal{F}) = \langle a^2 \rangle$, this is the only fusion that occurs. Hence, the number of $\mathcal{F}$-conjugacy classes is again $2^m(2^{n-1}+1)$.

Again $D/\langle a^2 \rangle$ has maximal class and $l(b_{a^2}) = 2$ by [Brauer 1974; Olsson 1975]. The same is true for the odd powers of a^2. Now let $u := a^{2^j}$ for some $2 \le j \le m$. Then it turns out that $D/\langle u \rangle$ is isomorphic to the group $D_{2^n} \rtimes C_{2^j}$ as in Theorem 3.1. So we obtain $l(b_u) = 2$ as well. For the other nontrivial subsections (u, b_u), we have at least $l(b_u) \ge 1$. Finally, $l(B) \ge 2$ since B is centrally controlled (see [Külshammer and Okuyama ~ 2000, Theorem 1.1]). Therefore,

$$k(B) \ge 2^{m+1} + 2^m(2^{n-1}+1) - 2^m = 2^{m+1}(2^{n-2}+1). \tag{1}$$

Also, $k_0(B) \le 2^{m+1}$ by [Sambale 2012b, Theorem 5.3]. However, in this situation we cannot apply [Robinson 1991]. So we use [Héthelyi et al. 2013, Theorem 2.4] for the major subsection (a^2, b_{a^2}). Let us determine the isomorphism type of $\overline{D} := D/\langle a^2 \rangle$ precisely. Since $(ax)^2 = axax = vx^2a^2 \equiv v \pmod{\langle a^2 \rangle}$, ax generates a cyclic maximal subgroup $\overline{D}$. Since ${}^a(ax) = avx = axv^{-1} \equiv (ax)^{-1} \pmod{\langle a^2 \rangle}$, $\overline{D} \cong D_{2^{n+1}}$. Hence, the Cartan matrix of b_{a^2} is given by

$$2^m \begin{pmatrix} 2^{n-1}+1 & 2 \\ 2 & 4 \end{pmatrix}$$

up to basic sets (see [Erdmann 1990]). This gives $k(B) \le 2^m(2^{n-1}+3)$, which is not quite what we wanted. However, the restriction on $k_0(B)$ will show that this maximal value for $k(B)$ cannot be reached. For this, we use the same method as in [Sambale 2013a]; i.e., we analyze the generalized decomposition numbers $d^u_{\chi\varphi_i}$ for $u := a^2$ and $\mathrm{IBr}(b_u) = \{\varphi_1, \varphi_2\}$. Since the argument is quite similar except that n has a slightly different meaning, we only present some key observations here. As in [Sambale 2013a], we write

$$d^u_{\chi\varphi_i} = \sum_{j=0}^{2^{m-1}-1} a^i_j(\chi)\zeta^j,$$

where $\zeta := e^{2\pi i/2^m}$. It follows that

$$(a^1_i, a^1_j) = (2^n+2)\delta_{ij}, \quad (a^1_i, a^2_j) = 4\delta_{ij} \quad \text{and} \quad (a^2_i, a^2_j) = 8\delta_{ij}.$$

Moreover, $h(\chi) = 0$ if and only if $\sum_{j=0}^{2^{m-1}-1} a^2_j(\chi) \equiv 1 \pmod 2$. This gives three essentially different possibilities for a^1_j and a^2_j as in [Sambale 2013a]. Let the

numbers α, β, γ and δ be defined as there. Then

$$\begin{aligned}\gamma &= 2^{m-1} - \alpha - \beta,\\ k(B) &\le (2^n+6)\alpha + (2^n+4)\beta + (2^n+2)\gamma - \delta/2\\ &= 2^{m+n-1} + 6\alpha + 4\beta + 2\gamma - \delta/2\\ &= 2^{m+n-1} + 2^m + 4\alpha + 2\beta - \delta/2,\\ 8\alpha + 4\beta - \delta \le k_0(B) &\le 2^{m+1}.\end{aligned}$$

This shows $k(B) \le 2^{m+n-1} + 2^{m+1} = 2^{m+1}(2^{n-2}+1)$. Together with (1), we have $k(B) = 2^{m+1}(2^{n-2}+1)$ and $l(B) = 2$. The inequalities above also show $k_0(B) = 2^{m+1}$. Now we can carry over the further discussion in [Sambale 2013a] word by word. In particular, we get $\delta = 0$,

$$\begin{aligned}k_1(B) &= (2^n-2)\alpha + (2^n-1)\beta + 2^n\gamma = 2^{n+m-1} - 2\alpha - \beta\\ &= 2^{n+m-1} - 2^{m-1} = 2^{m-1}(2^n-1)\end{aligned}$$

and finally $k_n(B) = 2^{m-1}$. The conjectures follow as usual. □

Now we can also handle defect groups of type $Q_{2^{n+1}} \rtimes C_{2^m}$. It is interesting to see that we get the same number of characters although the groups are nonisomorphic as shown in [Sambale 2012b].

Theorem 3.3. *Let B be a nonnilpotent 2-block of a finite group with defect group*

$$D \cong \langle v, x, a \mid v^{2^n} = a^{2^m} = 1,\ x^2 = v^{2^{n-1}},\ {}^x v = {}^a v = v^{-1},\ {}^a x = vx\rangle \cong Q_{2^{n+1}} \rtimes C_{2^m}$$

for some $n, m \ge 2$. Then $k(B) = 2^{m+1}(2^{n-2}+1)$, $k_0(B) = 2^{m+1}$, $k_1(B) = 2^{m-1}(2^n-1)$, $k_n(B) = 2^{m-1}$ and $l(B) = 2$. In particular, Brauer's $k(B)$-conjecture and Alperin's weight conjecture are satisfied.

Proof. Let $\mathcal{F}$ be the fusion system of B, and let $y := v^{2^{n-2}}$ and $z := x^2$. Then by [Sambale 2012b, Theorem 4.19], $Q := \langle x, y, a^2\rangle \cong Q_{2^{n+1}} \times C_{2^{m-1}}$ is the only $\mathcal{F}$-essential subgroup up to conjugation. Again we use Brauer's formula [Nagao and Tsushima 1989, Theorem 5.9.4] to get a lower bound for $k(B)$.

The same argument as in Theorem 3.1 shows that D has $2^{m-1}(2^n+3)$ conjugacy classes and we need to know which of them are fused in Q. It is easy to see that xa^{2j} is conjugate to ya^{2j} under $\mathrm{Aut}_{\mathcal{F}}(Q)$ for $j \in \mathbb{Z}$. Since $\mathrm{Z}(\mathcal{F}) = \langle z, a^2\rangle$, this is the only fusion that occurs. Hence, the number of $\mathcal{F}$-conjugacy classes is again $2^m(2^{n-1}+1)$. In case $n = 2$, the group $D/\langle z\rangle \cong C_2^2 \rtimes C_{2^m}$ is minimal nonabelian, and we get $l(b_z) = 2$ from [Sambale 2011a]. Otherwise, $D/\langle z\rangle$ is isomorphic to one of the groups in Theorem 3.1. Hence, again $l(b_z) = 2$. As usual, the groups $D/\langle a^2\rangle$ and $D/\langle a^2 z\rangle$ have maximal class and it follows that $l(b_{a^2}) = l(b_{a^2 z}) = 2$. The same holds for all odd powers of a^2 and $a^2 z$. For $2 \le j \le m-1$, the group $D/\langle u\rangle$

with $u := a^{2^j}$ has the same isomorphism type as D where m has to be replaced by j. So induction on m shows $l(b_u) = 2$. It remains to deal with $u := a^{2^j} z$. Here $D/\langle u\rangle \cong Q_{2^{n+1}}.C_{2^j}$ is exactly the group from Theorem 3.2. Thus, for $j \neq n$ we have again $l(b_u) = 2$. In case $j = n$, $D/\langle u\rangle \cong C_{2^n} \wr C_2$. Then [Külshammer 1980, (7.G)] gives $l(b_u) = 2$ as well. Now Brauer's formula reveals

$$k(B) \geq 2^{m+1} + 2^m(2^{n-1} + 1) - 2^m = 2^{m+1}(2^{n-2} + 1).$$

For the opposite inequality, we apply [Héthelyi et al. 2013, Theorem 2.4] to the major subsection (u, b_u) where $u := a^2 z$. A similar calculation as in Theorem 3.2 shows that $D/\langle u\rangle \cong Q_{2^{n+2}}$. Hence, the Cartan matrix of b_u is given by

$$2^m \begin{pmatrix} 2^{n-1}+1 & 2 \\ 2 & 4 \end{pmatrix}$$

up to basic sets (see [Erdmann 1990]). This is the same matrix as in Theorem 3.2, but the following discussion is slightly different because a^2 has only order 2^{m-1} here. So we copy the proof of the main theorem in [Sambale 2013b]. In fact, we just have to replace m with $m + 1$ and n with $n - 2$ in order to use this proof word for word. The claim follows. □

We describe the structure of these group extensions in a more generic way.

Proposition 3.4. *Let D be an extension of the cyclic group $\langle a\rangle \cong C_{2^n}$ by a group M that has maximal class or is the four-group. Suppose that the corresponding coupling $\omega : \langle a\rangle \to \mathrm{Out}(M)$ satisfies the following: if $\omega \neq 0$, then the coset $\omega(a)$ of $\mathrm{Inn}(M)$ contains an involution that acts nontrivially on $M/\Phi(M)$. Moreover, assume that $D \ncong C_{2^m} \wr C_2$ for all $m \geq 3$. Then the invariants for every block of a finite group with defect group D are known.*

Proof. Assume first that $M \cong C_2^2$. Then in case $\omega = 0$, we get the groups $C_{2^n} \times C_2^2$ and $C_{2^{n+1}} \times C_2$ for which the block invariants can be calculated by [Usami 1988; Kessar et al. 2012]. So let $\omega \neq 0$. If D is nonsplit, it must contain a cyclic maximal subgroup. In particular, D is metacyclic and the block invariants are known. If the extension splits, we obtain the minimal nonabelian group $C_2^2 \rtimes C_{2^n}$. Here the block invariants are known by [Sambale 2011a].

Hence, let M be a 2-group of maximal class. Then $|\mathrm{Z}(M)| = 2$. Thus, for $\omega = 0$ we obtain precisely two extensions for every group M. All these cases were handled in [Sambale 2012a; 2013b; 2013a]. Let us now consider the case $\omega \neq 0$. Since the three maximal subgroups of a semidihedral group are pairwise nonisomorphic, M must be a dihedral or quaternion group. Write $M = \langle v, x \mid v^{2^m} = 1,\ x^2 \in \langle v^{2^{m-1}}\rangle,\ {}^x v = v^{-1}\rangle$. Let $\alpha \in \mathrm{Aut}(M)$ be an involution that acts nontrivially on $M/\Phi(M)$. Then there is an odd integer i such that ${}^\alpha x = v^i x$. Since $\alpha^2 = 1$, it follows that ${}^\alpha v = v^{-1}$. Hence, the coset $\alpha\,\mathrm{Inn}(M) \in \mathrm{Out}(M)$ is determined uniquely. Hence, ω is unique.

So we get four group extensions for every pair (n, m). Two of them are isomorphic, and all cases are covered in Theorems 3.1, 3.2 and 3.3 (and [Külshammer 1980] for $C_4 \wr C_2$). □

4. More examples

Since almost all block invariants for 2-blocks of defect 4 are known (see [Külshammer and Sambale 2013]), it is natural to look at 2-blocks of defect 5. Here for the abelian defect group $C_4 \times C_2^3$, the invariants are not known so far. We handle more general abelian defect groups in the next theorem. This result relies on the classification of the finite simple groups. We denote the inertial index of B by $e(B)$.

Theorem 4.1. *Let B be a block of a finite group G with defect group $C_{2^n} \times C_2^3$ for some $n \geq 2$. Then we have $k(B) = k_0(B) = |D| = 2^{n+3}$ and one of the following holds*:

(i) $e(B) = l(B) = 1$.

(ii) $e(B) = l(B) = 3$.

(iii) $e(B) = l(B) = 7$.

(iv) $e(B) = 21$ *and* $l(B) = 5$.

Proof. Let $D = C_{2^n} \times C_2^3$. Since $\mathrm{Aut}(D)$ acts faithfully on $\Omega(D)/\Phi(D) \cong C_2^3$, we have $e(B) \in \{1, 3, 7, 21\}$. In case $e(B) = 1$, the block is nilpotent and the result is clear. Now we consider the remaining cases.

Case 1: $e(B) = 3$. Then there are 2^{n+2} subsections (u, b_u) up to conjugation and 2^{n+1} of them satisfy $l(b_u) = 1$. For the other 2^{n+1} subsections, [Watanabe 1991, Theorem 1] implies $l(b_u) = 3$. This gives $k(B) = 2^{n+3} = |D|$. The height-0 conjecture follows from [Kessar and Malle 2013, Theorem 1.1].

Case 2: $e(B) = 7$. Here we have 2^{n+1} subsections (u, b_u) up to conjugation, where 2^n of them satisfy $l(b_u) = 1$. For the other 2^n subsections, we use [Watanabe 1991, Theorem 1] in connection with [Kessar et al. 2012, Theorem 1.1] (instead of [Kessar et al. 2012], we could also use [Kessar and Malle 2013], which we need anyway). This gives $l(b_u) = 7$ for these subsections. It follows that $k(B) = |D|$ and $k(B) = k_0(B)$ by [Kessar and Malle 2013, Theorem 1.1].

Case 3: $e(B) = 21$. Here we have again 2^{n+1} subsections (u, b_u) up to conjugation. But this time 2^n subsections satisfy $l(b_u) = 3$ and the other 2^n subsections satisfy $l(b_u) = 5$ by [Watanabe 1991; Kessar et al. 2012]. The result follows as before. □

Next we study another group of order 32 with an easy structure. For this, let $\mathrm{MNA}(r, s)$ be the minimal nonabelian group given by

$$\langle x, y \mid x^{2^r} = y^{2^s} = [x, y]^2 = [x, x, y] = [y, x, y] = 1 \rangle$$

for some $r \geq s \geq 1$ (see [Rédei 1947]). For the notion of a *constrained* fusion system, we refer to [Oliver and Ventura 2009, Definition 2.3].

Proposition 4.2. *Let B be a nonnilpotent block of a finite group with defect group $D \cong \mathrm{MNA}(2,1) \times C_2$. Then $k(B) = 20$, $k_0(B) = 16$, $k_1(B) = 4$ and $l(B) = 2$. In particular, Olsson's conjecture and Alperin's weight conjecture hold for B.*

Proof. Let $\mathcal{F}$ be the fusion system of B. Since $|D : \mathrm{Z}(D)| = 4$, every $\mathcal{F}$-essential subgroup is maximal, and there are three candidates for these groups. Let $\mathrm{Z}(D) < M < D$ such that $M \cong C_4 \times C_2^2$. Then $\mathrm{Aut}_{\mathcal{F}}(M)$ must act nontrivially on $\Omega(M)/\Phi(M)$. However, it can be seen that $\mathrm{N}_D(M)$ acts trivially on $\Omega(M)/\Phi(M)$. In particular, M is not $\mathcal{F}$-radical. Hence, there is only one $\mathcal{F}$-essential subgroup $Q \cong C_2^4$ (up to conjugation). Since $Q \trianglelefteq D$, $\mathcal{F}$ is constrained and thus uniquely determined by $\mathrm{Out}_{\mathcal{F}}(Q)$ (see [Linckelmann 2007, Theorem 4.6]). By [Sambale 2012b, Lemma 3.11], we have some possibilities for $\mathrm{Out}_{\mathcal{F}}(Q)$. However, a GAP calculation shows that only $\mathrm{Out}_{\mathcal{F}}(Q) \cong S_3$ is realizable. Then $\mathcal{F}$ is the fusion system on the group $\mathtt{SmallGroup}(96, 194) \cong (A_4 \rtimes C_4) \times C_2$. In particular, there are exactly 16 $\mathcal{F}$-conjugacy classes on D. Moreover, $\mathrm{Z}(\mathcal{F}) \cong C_2^2$, and for $1 \neq z \in \mathrm{Z}(\mathcal{F})$, we have $D/\langle z \rangle \in \{\mathrm{MNA}(2,1), D_8 \times C_2\}$. Hence, we get $l(b_z) = 2$ as usual. For all other nontrivial subsections (u, b_u), we have $l(b_u) \geq 1$. Since B is centrally controlled, [Külshammer and Okuyama ~ 2000, Theorem 1.1] implies $l(B) \geq 2$. Brauer's formula for $k(B)$ gives $k(B) \geq 20$. If $x \in D$ has order 4, then $\mathrm{C}_D(x)/\langle x \rangle$ has order 4. Hence, Olsson's conjecture follows from [Héthelyi et al. 2013, Theorem 2.5]; i.e., $k_0(B) \leq |D : D'| = 16$. For an element $z \in \mathrm{Z}(D) \setminus \mathrm{Z}(\mathcal{F})$, the block b_z is nilpotent. Thus, [Robinson 1991, Theorem 3.4] implies

$$|D| = 32 \leq k_0(B) + 4(k(B) - k_0(B)) \leq \sum_{i=0}^{\infty} 2^{2i} k_i(B) \leq |D|.$$

The claim follows as usual. □

In the classification of the simple groups of 2-rank 2, the sole exception $\mathrm{PSU}(3,4)$ shows up (see [Alperin et al. 1973]). This group has a Suzuki Sylow 2-subgroup P of order 64 (see [Craven and Glesser 2012, Definition 1.4]). The group P also occurs in the classification of the center-free fusion systems on 2-groups of 2-rank 2 (see [Craven and Glesser 2012]). It can also be described as the smallest 2-group with exactly three involutions and an automorphism of order 5. This answers a question raised in [Berkovich and Janko 2008, Exercise 82.3]. In fact, P admits an automorphism of order 15. Moreover, $\mathrm{Z}(P) = \Phi(P) = P' = \Omega(P) \cong C_2^2$, so P is *special* (see [Gorenstein 1968, p. 183]).

Using this as a motivation, it seems worthwhile to obtain the invariants of blocks with defect group P (this will be done in an upcoming diploma thesis). Doing so,

we need to handle the extraspecial group $P/\langle z\rangle \cong D_8 * Q_8$ for $1 \neq z \in \mathrm{Z}(P)$ for the induction step.

Proposition 4.3. *Let B be a block of a finite group G with defect group $D_8 * Q_8$ and inertial index 5. Then $l(B) = 5$, $k(B) = 13$, $k_0(B) = 8$ and $k_2(B) = 5$. Moreover, the Cartan matrix of B is given by*

$$2\begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 4 \end{pmatrix}$$

up to basic sets.

Proof. Let $D = D_8 * Q_8$, and let $\mathcal{F}$ be the fusion system of B. By [Stancu 2006, Theorem 5.3], $\mathcal{F}$ is controlled by $\mathrm{Aut}_{\mathcal{F}}(D)$. Let $\mathrm{Z}(D) = \langle z\rangle$. As usual, we denote the subsections by (u, b_u). Then b_z covers a block $\overline{b_z}$ with elementary abelian defect group of order 16. It follows from [Külshammer and Sambale 2013, Proposition 2.1] that $5 = e(B) = e(b_z) = e(\overline{b_z}) = l(\overline{b_z}) = l(b_z)$. Moreover, B is centrally controlled; in particular, [Külshammer and Okuyama ~ 2000, Theorem 1.1] implies $l(B) \geq 5$.

There are three nonmajor subsections (u_1, b_1), (u_2, b_2) and (u_3, b_3). Since $|D'| = 2$, every conjugacy class in D has at most two elements. In particular, $|\mathrm{C}_D(u_i)| = 16$ for $i = 1, 2, 3$. By [Héthelyi et al. 2013, Proposition 5.1], we have $l(b_i) = 1$ for $i = 1, 2, 3$. Now let us look at the major subsection (z, b_z). By the proof of [Sambale 2011c, Proposition 1], the Cartan matrix of b_z is given by

$$2\begin{pmatrix} 4 & 3 & 3 & 3 & 3 \\ 3 & 4 & 3 & 3 & 3 \\ 3 & 3 & 4 & 3 & 3 \\ 3 & 3 & 3 & 4 & 3 \\ 3 & 3 & 3 & 3 & 4 \end{pmatrix}$$

up to basic sets. If we change the basic set, we get the following matrix with smaller entries:

$$C_z := 2\begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 4 \end{pmatrix}.$$

Now we consider the matrix $D_z := (d_{ij}^z)$. Since z has order 2, D_z is an integral matrix such that $D_z^{\mathrm{T}} D_z = C$. Since all columns of D_z are orthogonal to the columns of ordinary decomposition numbers, we see that the first four columns consist of

exactly four entries ± 1 each. By way of contradiction, assume that the first two columns of D_z have the form

$$\begin{pmatrix} 1 & 1 & 1 & 1 & . & \cdots & . \\ 1 & 1 & 1 & -1 & . & \cdots & . \end{pmatrix}^{\mathrm{T}}.$$

Then there is at least one column of ordinary decomposition numbers that is not orthogonal to the difference of these two columns of D_z. This contradiction shows that D_z has the form

$$D_z = \begin{pmatrix} 1 & 1 & 1 & 1 & . & . & . & . & . & . & . & \cdots & . \\ 1 & 1 & . & . & 1 & 1 & . & . & . & . & . & \cdots & . \\ 1 & 1 & . & . & . & . & 1 & 1 & . & . & . & \cdots & . \\ 1 & 1 & . & . & . & . & . & . & 1 & 1 & . & \cdots & . \\ * & * & * & * & * & * & * & * & * & * & * & \cdots & * \end{pmatrix}^{\mathrm{T}}$$

up to signs and permutations. It holds that $k(B) = l(B) + l(b_z) + l(b_1) + l(b_2) + l(b_3) \geq 13$. Hence, for the last column of D_z we have essentially the following possibilities:

$$\begin{aligned} &\begin{pmatrix} 1 & 1 & . & . & . & . & . & . & . & . & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}^{\mathrm{T}}, \\ &\begin{pmatrix} 1 & . & 1 & . & 1 & . & 1 & . & 1 & . & 1 & 1 & 1 \end{pmatrix}^{\mathrm{T}}, \\ &\begin{pmatrix} 1 & 1 & 1 & -1 & . & . & . & . & . & . & 1 & 1 & 1 & 1 \end{pmatrix}^{\mathrm{T}}, \\ &\begin{pmatrix} 2 & . & . & . & . & . & . & . & . & . & 1 & 1 & 1 & 1 \end{pmatrix}^{\mathrm{T}}, \\ &\begin{pmatrix} 1 & 1 & . & . & . & . & . & . & . & . & 2 & 1 & 1 \end{pmatrix}^{\mathrm{T}}. \end{aligned}$$

This already implies $k(B) \in \{13, 14, 16\}$. In order to investigate the heights of the irreducible characters, we consider the matrix $M^z = (m^z_{ij}) = 32 D_z C_z^{-1} D_z^{\mathrm{T}}$ of contributions. We have

$$32 C_z^{-1} = \begin{pmatrix} 13 & -3 & -3 & -3 & -1 \\ -3 & 13 & -3 & -3 & -1 \\ -3 & -3 & 13 & -3 & -1 \\ -3 & -3 & -3 & 13 & -1 \\ -1 & -1 & -1 & -1 & 5 \end{pmatrix}.$$

By [Brauer 1968, (5G) and (5H)], we have

$$h(\chi) = 0 \iff m^z_{\chi\chi} \equiv 1 \pmod 2 \iff \sum_{\varphi \in \mathrm{IBr}(b_z)} d^z_{\chi\varphi} \equiv 1 \pmod 2.$$

This gives $k_0(B) \in \{8, 12, 16\}$ according to the last column of D^z. By [Broué 1980, Proposition 1], we also have $h(\chi) = 0 \iff d^{u_i}_{\chi\varphi_i} \equiv 1 \pmod 2$ for $i = 1, 2, 3$ where $\mathrm{IBr}(b_{u_i}) = \{\varphi_i\}$. Since the norm of these nonmajor columns is 16, we have the

following possibilities for the nonvanishing entries according to $k_0(B)$: sixteen ± 1; twelve ± 1 and one ± 2; eight ± 1 and two ± 2; or seven ± 1 and one ± 3.

Taking this together, we can enumerate all the possibilities for the decomposition numbers of nontrivial subsections with GAP. Then the ordinary decomposition matrix (up to multiplication with an invertible matrix) can be determined as the orthogonal space. Finally the square of the ordinary decomposition matrix is the Cartan matrix C of B. Now we determine the elementary divisors of C by considering the lower defect groups.

By [Brauer 1969, (7G)], the multiplicity $m(d)$ of the elementary divisor $d \in \mathbb{N}$ of C is given by

$$m(d) = \sum_{R \in \mathcal{R}} m_B^{(1)}(R)$$

where $\mathcal{R}$ is a set of representatives for the G-conjugacy classes of subgroups of G of order d. After combining this with [Broué and Olsson 1986, Formula (2S)], we get

$$m(d) = \sum_{(R, b_R) \in \mathcal{R}'} m_B^{(1)}(R, b_R)$$

where $\mathcal{R}'$ is a set of representatives for the G-conjugacy classes of B-subpairs (R, b_R) such that R has order d. We have to emphasize that in contrast to other papers we regard b_R as a block of $\mathrm{C}_G(R)$ instead of $R\,\mathrm{C}_G(R)$. Let b_D be a Brauer correspondent of B in $\mathrm{C}_G(D)$. Then after changing the representatives if necessary, we may assume $(R, b_R) \leq (D, b_D)$ for $(R, b_R) \in \mathcal{R}'$. Then it is well-known that b_R is uniquely determined by R. Since the fusion of these subpairs is controlled by $\mathrm{N}_G(D, b_D)$, we get

$$m(d) = \sum_{R \in \mathcal{R}''} m_B^{(1)}(R, b_R)$$

where $\mathcal{R}''$ is a set of representatives for the $\mathrm{Aut}_{\mathcal{F}}(D)$-conjugacy classes of subgroups of D of order d.

It is well-known that we have $m(32) = 1$. Now we discuss smaller values for d. We begin with the case $d = 2$. For this, let $m_B^{(1)}(Q, b_Q) > 0$ for some Q with $|Q| = 2$. Then (Q, b_Q) is in fact a subsection and 2 is also an elementary divisor of the Cartan matrix of b_Q. In particular, $l(b_Q) > 1$. This shows that $Q = \mathrm{Z}(D)$. One can show that 2 occurs as elementary divisor of C_z exactly four times. If we apply the same arguments to the block b_z instead of B, we see that $m(2) = m_B^{(1)}(Q, b_Q) = 4$.

Now let $2 < d < 32$ and $Q \leq D$ such that $|Q| = d$. Then by [Broué and Olsson 1986, (2Q)], we have $m_{B_Q}^{(1)}(Q) > 0$ where $B_Q := b_Q^{\mathrm{N}_G(Q, b_Q)}$. Since Q is fully $\mathcal{F}$-normalized, [Linckelmann 2006, Theorem 2.4] implies that $\mathrm{C}_D(Q)$ is a defect group of b_Q and $\mathrm{N}_D(Q)$ is a defect group of B_Q. By [An 2011, Proposition 2.1], also the block b_Q is controlled. If we follow the proof of this proposition more

closely, it turns out that $(\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)})$ is a Sylow b_Q-subpair. So the inertial quotient of b_Q is

$$\begin{aligned}\mathrm{N}_{\mathrm{C}_G(Q)}(\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)})/\,\mathrm{C}_D(Q)\,\mathrm{C}_{\mathrm{C}_G(Q)}(\mathrm{C}_D(Q))\\ \leq \mathrm{N}_G(Q\,\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)}) \cap \mathrm{C}_G(Q)/\,\mathrm{C}_D(Q)\,\mathrm{C}_G(Q\,\mathrm{C}_D(Q)).\end{aligned}$$

All odd-order automorphisms of

$$\mathrm{Aut}_{\mathcal{F}}(Q\,\mathrm{C}_D(Q)) = \mathrm{N}_G(Q\,\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)})/\,\mathrm{C}_G(Q\,\mathrm{C}_D(Q))$$

come from restrictions of $\mathrm{Aut}_{\mathcal{F}}(D)$. However, the automorphism of order 5 in $\mathrm{Aut}_{\mathcal{F}}(D)$ cannot centralize Q since $2 < d$. Hence, the inertial index of b_Q is 1 and $l(b_Q) = 1$. Finally, [Olsson 1980, Theorem 5.11] and the remark following it show

$$1 = l(b_Q) \geq m^{(1)}_{B_Q}(Q) + m^{(1)}_{B_Q}(\mathrm{N}_D(Q)) = m^{(1)}_{B_Q}(Q) + 1$$

and $m^{(1)}_{B_Q}(Q) = 0$. Taking these arguments together, we proved that the elementary divisors of C are $32, 2, 2, 2, 2, 1, \ldots, 1$ (including the possibility of no 1 at all).

Using this, our GAP program reveals that the only possibility for the generalized decomposition numbers is

$$\begin{pmatrix}
1 & 1 & 1 & 1 & . & . & . & . & . & . & . & . & . \\
1 & 1 & . & . & 1 & 1 & . & . & . & . & . & . & . \\
1 & 1 & . & . & . & . & 1 & 1 & . & . & . & . & . \\
1 & 1 & . & . & . & . & . & . & 1 & 1 & . & . & . \\
1 & . & 1 & . & 1 & . & 1 & . & 1 & . & 1 & 1 & 1 \\
-1 & . & . & 1 & . & 1 & . & 1 & . & 1 & 3 & -1 & -1 \\
-1 & . & . & 1 & . & 1 & . & 1 & . & 1 & -1 & 3 & -1 \\
-1 & . & . & 1 & . & 1 & . & 1 & . & 1 & -1 & -1 & 3
\end{pmatrix}^{\mathrm{T}}$$

(up to permutations and choosing signs as described earlier). In particular, $k(B) = 13$, $k_0(B) = 8$ and $l(B) = 5$. Moreover, C is uniquely determined up to basic sets. Hence, $C = C_z$ up to basic sets because in case $z \in \mathrm{Z}(G)$, B and b_z would coincide. It remains to determine $k_i(B)$ for $i > 0$. For this, let $\psi \in \mathrm{Irr}(B)$ be the fourth character in the numbering above. In particular, ψ has height 0. Then for a character $\chi \in \mathrm{Irr}(B)$ with $h(\chi) > 0$, we can see that $m^z_{\chi\psi}$ is divisible by 4 but not by 8. Thus, [Brauer 1968, (5H)] implies $k_2(B) = 5$. □

For the defect group in Proposition 4.3, the inertial index could also be 3. However, in this case the computational effort is too big.

In [Sambale 2011c], we verified Brauer's $k(B)$-conjecture for defect groups of order at most 32 but not isomorphic to the extraspecial group $D_8 * D_8$. We are finally able to handle this remaining group as well.

Theorem 4.4. *Brauer's $k(B)$-conjecture holds for defect groups with a central cyclic subgroup of index at most* 16. *In particular, the $k(B)$-conjecture holds for the* 2*-blocks of defect at most* 5.

Proof. Let B be a p-block with defect group D of the stated form. By [Sambale 2011c, Theorems 1 and 3], we may assume that there is a major B-subsection (z, b_z) such that $D/\langle z\rangle \cong C_2^4$ (in particular, $p = 2$) and B has inertial index 9. We apply [Feit 1982, Theorem V.9.17]. For this, it suffices to determine the Cartan matrix of b_z (only up to basic sets). Thus, we may consider a 2-block B with elementary abelian defect group D of order 16 and inertial index 9. As in [Külshammer and Sambale 2013, Lemma 2.2], we obtain a list of possible Cartan matrices of B. However, since we are considering 9×9 matrices it is very hard to see if two of these candidates only differ by basic sets. In order to reduce the set of possible Cartan matrices further, we apply various ad hoc matrix manipulations as permutations of rows and columns and elementary row/column operations. After this procedure, we end up with a list of only ten possible Cartan matrices of B that might be all equal up to basic sets. For the purpose of illustrating, we display one of these matrices:

$$\begin{pmatrix} 4 & -1 & 1 & . & 1 & 1 & 2 & . & . \\ 1 & 4 & . & 1 & -1 & 1 & . & 1 & 1 \\ 1 & . & 4 & 1 & -1 & 1 & 2 & -1 & -1 \\ . & 1 & 1 & 4 & . & . & . & 2 & . \\ 1 & -1 & -1 & . & 4 & . & 1 & 1 & 1 \\ 1 & 1 & 1 & . & . & 4 & 1 & 1 & 1 \\ 2 & . & 2 & . & 1 & 1 & 4 & . & -2 \\ . & 1 & -1 & 2 & 1 & 1 & . & 4 & . \\ . & 1 & -1 & . & 1 & 1 & -2 & . & 4 \end{pmatrix}.$$

It can be seen that all diagonal entries are 4 (for every one of these ten matrices). In order to apply [Feit 1982, Theorem V.9.17], let C be one of these ten matrices. Then we have a positive definite integral quadratic form q corresponding to the matrix $16C^{-1}$. We need to find the minimal nonzero value of q among all integral vectors. More precisely, we have to check if a value strictly smaller than 9 is assumed by q. By [Liebeck 1971, Theorem 1], it suffices to consider only vectors with entries in $\{0, \pm 1\}$ (observe that the notation of a quadratic form given by a matrix is the same in [Feit 1982] and [Liebeck 1971]). Hence, there are only 3^9 values to consider. An easy computer computation shows that in fact the minimum of q is at least 9. So Brauer's $k(B)$-conjecture follows from [Feit 1982, Theorem V.9.17]. □

We like to point out that we do not know a single Cartan matrix such that Brauer's $k(B)$-conjecture would not follow from [Héthelyi et al. 2013, Theorem 2.4] or from [Feit 1982, Theorem V.9.17]. Since these two results are somehow related, it seems

interesting to investigate the following problem: Let $C = (c_{ij}) \in \mathbb{Z}^{l\times l}$ be the Cartan matrix of a p-block with defect d. Assume that for all integral, positive definite quadratic forms $q(x_1, \dots, x_{l(b_u)}) = \sum_{1\le i\le j\le l} q_{ij}x_ix_j$ we have

$$\sum_{1\le i\le j\le l} q_{ij}c_{ij} > p^d.$$

Then prove that $xp^dC^{-1}x^{\mathrm{T}} \ge l$ for all $0 \ne x \in \mathbb{Z}^l$. If this can be done, the $k(B)$-conjecture would follow in full generality. A diagonal matrix shows that this argument fails for arbitrary positive definite, symmetric matrices C.

In the next proposition, we take a closer look at the defect group $D_8 * D_8$.

Proposition 4.5. *Let B be a block of a finite group G with defect group $D \cong D_8 * D_8$. Suppose that the inertial quotient* $\mathrm{Out}_{\mathcal{F}}(D)$ *has order* 3 *and acts freely on $D/\Phi(D)$. Then $k(B) = 11$, $k_0(B) = 8$ and $l(B) = 3$. Moreover, the Cartan matrix of B is given by*

$$2\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 6 \end{pmatrix}$$

up to basic sets. For the numbers $k_i(B)$ $(i \ge 1)$, we have the following cases: $(k_1(B), k_2(B)) \in \{(0, 3), (2, 1)\}$.

Proof. Let $\mathcal{F}$ be the fusion system of B. By [Stancu 2006, Theorem 5.3], $\mathcal{F}$ is controlled by $\mathrm{Aut}_{\mathcal{F}}(D)$. By hypothesis, $\mathrm{Out}_{\mathcal{F}}(D) \cong C_3$ acts freely on $D/\Phi(D)$. Hence, there are two major and five nonmajor subsections. The Cartan matrix of the nontrivial major subsection (z, b_z) is given by

$$2\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 6 \end{pmatrix}$$

up to basic sets. In particular, $k(B) \le 16$. The nonmajor subsections (u, b_u) all satisfy $l(b_u) = 1$. Since B is centrally controlled, we have $k(B) \ge 11$. The first two columns of the b_z decomposition numbers have the form

$$\begin{pmatrix} 1 & 1 & 1 & 1 & . & . & . & . & . & . & \cdots & . \\ 1 & 1 & . & . & 1 & 1 & . & . & . & . & \cdots & . \end{pmatrix}^{\mathrm{T}}$$

up to signs and permutations (compare with the proof of Proposition 4.3). For the third column, we have essentially 17 possibilities, which we do not list explicitly here. As in Proposition 4.3, we get $k_0(B) \in \{8, 12, 16\}$ and also the positions of the characters of height 0 depending on the third column of D_z. Since every element of order 4 in D is conjugate to its inverse, all generalized decomposition numbers are integers. For each of the 17 cases, we proceed by enumerating the five columns of

nonmajor subsections with the help of a computer. Entirely similar to the proof of Proposition 4.3, we see that the elementary divisors of the Cartan matrix of B are $32, 2, 2, 1, \ldots, 1$. Now the computations reveal $k(B) = 11$, $k_0(B) = 8$, $l(B) = 3$ and the Cartan matrix of B up to basic sets. However, the value of $k_1(B)$ does not follow immediately from these calculations. Instead we obtain the two cases $(k_1(B), k_2(B)) \in \{(0, 3), (2, 1)\}$. □

It is easy to construct examples for Proposition 4.5 such that $k_1(B) = 0$. In contrast, $k_1(B) = 2$ would contradict the ordinary weight conjecture (see [Robinson 2004]).

The next proposition concerns the Sylow 2-subgroup of PSU(3, 4) as mentioned above. This will result will be used in an upcoming diploma thesis.

Proposition 4.6. *Let B be a block of a finite group G with inertial index* 15 *and defect group $D \in \mathrm{Syl}_2(\mathrm{PSU}(3,4))$. Then the elementary divisors of the Cartan matrix of B lie in* $\{1, 4, 64\}$, *where* 4 *occurs with multiplicity at most* 4.

Proof. Since D is a Suzuki 2-group, [Craven and Glesser 2012, Theorem 4.4] tells us that the fusion system $\mathcal{F}$ of B is controlled. So as in the proof of Proposition 4.3, the multiplicity of d as an elementary divisor of the Cartan matrix C of B is given by

$$m(d) = \sum_{R \in \mathcal{R}''} m_B^{(1)}(R, b_R),$$

where $\mathcal{R}''$ is a set of representatives for the $\mathrm{Aut}_{\mathcal{F}}(D)$-conjugacy classes of subgroups of D of order d. Assume first $d = 2$ and $m_B^{(1)}(Q, b_Q) > 0$ for $|Q| = 2$. Then (Q, b_Q) is a subsection and $Q \subseteq \mathrm{Z}(D)$. One can show that b_Q has defect group D and inertial index 5. Moreover, b_Q covers a block $\overline{b_Q}$ of $\mathrm{C}_G(Q)/Q$ with defect group $D/Q \cong D_8 * Q_8$. Hence, Proposition 4.3 implies that all elementary divisors of the Cartan matrix of b_Q are divisible by 4. This contradiction shows that $m(2) = 0$. Now suppose that $2 < d < 64$. Again we assume $m_B^{(1)}(Q, b_Q) > 0$ such that $|Q| = d$. We argue as in the proof of Proposition 4.3. The inertial quotient of b_Q is given by

$$\begin{aligned}\mathrm{N}_{\mathrm{C}_G(Q)}(\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)})/\,\mathrm{C}_D(Q)\,\mathrm{C}_{\mathrm{C}_G(Q)}(\mathrm{C}_D(Q))\\ \leq \mathrm{N}_G(Q\,\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)}) \cap \mathrm{C}_G(Q)/\,\mathrm{C}_D(Q)\,\mathrm{C}_G(Q\,\mathrm{C}_D(Q)).\end{aligned}$$

Every odd order automorphism in

$$\mathrm{N}_G(Q\,\mathrm{C}_D(Q), b_{Q\,\mathrm{C}_D(Q)})/\,\mathrm{C}_G(Q\,\mathrm{C}_D(Q)) = \mathrm{Aut}_{\mathcal{F}}(Q\,\mathrm{C}_D(Q))$$

comes from a restriction of $\mathrm{Aut}_{\mathcal{F}}(D)$. Moreover, $\mathrm{Out}_{\mathcal{F}}(D)$ acts freely on $D/\Phi(D)$. So in case $d > 4$, we see that these odd order automorphisms cannot lie in $\mathrm{C}_G(Q)$. Hence, in this case $l(b_Q) = 1$ and $m(d) = 0$ (compare with Proposition 4.3). It remains to deal with the case $Q = \mathrm{Z}(D) = \Phi(D)$. Then we have $b_Q = b_D^{\mathrm{C}_G(\mathrm{Z}(D))}$.

Moreover, b_Q has defect group D and inertial index 5. Looking at the covered block of $\mathrm{C}_G(Q)/Q$, we see that $l(b_Q) = 5$. Hence,

$$5 = l(b_Q) \geq m^{(1)}_{B_Q}(Q) + m^{(1)}_{B_Q}(\mathrm{N}_D(Q)) = m^{(1)}_{B_Q}(Q) + 1$$

by [Olsson 1980, Theorem 5.11] and the remark following it. This gives $m(4) = m^{(1)}_{B_Q}(Q) \leq 4$, and the proof is complete. □

Our next result handles rather unknown groups of order 32. The key observation here is that the fusion system is constrained and thus quite easy to understand.

Proposition 4.7. *Let B be a nonnilpotent block of a finite group G with defect group $D \cong \mathtt{SmallGroup}(32, q)$ for $q \in \{28, 29\}$. Then $k(B) = 14$, $k_0(B) = 8$, $k_1(B) = 6$ and $l(B) = 2$.*

Proof. Let $\mathcal{F}$ be the fusion system of B. Using GAP, one can show that $\mathrm{Aut}(D)$ is a 2-group. In particular, $\mathrm{Out}_{\mathcal{F}}(D) = 1$. Moreover, one can show using general results in [Sambale 2012b] that D contains only one $\mathcal{F}$-essential subgroup Q. Here $C_2^2 \times C_4 \cong Q \trianglelefteq D$. In particular, $\mathcal{F}$ is constrained. Another GAP calculation shows that $\mathcal{F}$ is the fusion system of the groups $\mathtt{SmallGroup}(96, 187)$ or $\mathtt{SmallGroup}(96, 185)$ for $q \in \{28, 29\}$, respectively. We have ten B-subsections up to conjugation. The center of D is a four-group and $\Phi(Q) \subseteq \mathrm{Z}(D)$. Hence, an odd order automorphism of Q cannot act on $\mathrm{Z}(D)$. It follows that we have four major subsections $(1, B)$, (z, b_z), (v, b_v) and (w, b_w) up to conjugation. Here we may assume that $l(b_v) = l(b_w) = 1$. On the other hand, b_z dominates a nonnilpotent block with defect group $D/\langle z\rangle \cong D_8 \times C_2$. Thus, by [Sambale 2011c, Proposition 3] we have $l(b_z) = 2$. Also we find an element $u \in Q$ such that b_u is nonnilpotent with defect group Q. Here [Sambale 2011c, Proposition 2] implies $l(b_u) = 3$. The remaining nonmajor subsections split in one subsection (u_1, b_1) of defect 16 and four subsections (u_i, b_i) $(i = 2, 3, 4, 5)$ of defect 8. Here $l(b_i) = 1$ for $i = 1, \dots, 5$. In particular, Olsson's conjecture $k_0(B) \leq 8 = |D : D'|$ follows at once. Since B is centrally controlled, we also obtain $l(B) \geq 2$ and $k(B) \geq 14$. So the generalized decomposition numbers d^v_{ij} consist of eight entries ± 1 and six entries ± 2. Hence, $k(B) = 14$, $k_0(B) = 8$, $k_1(B) = 6$ and $l(B) = 2$. □

Also in the next proposition, the corresponding fusion system is easy to understand since it is controlled. Another advantage here is that $k(B)$ is relatively small so that the computational effort is small as well.

Proposition 4.8. *Let D be a central cyclic extension of $\mathtt{SmallGroup}(32, q)$ for $q \in \{33, 34\}$. Then Brauer's $k(B)$-conjecture holds for all blocks with defect group D.*

Proof. It suffices to consider a block B with defect group $D \cong \mathtt{SmallGroup}(32, q)$ for $q \in \{33, 34\}$ as usual. GAP shows that B is a controlled block with inertial index 3. Hence, the fusion system of B is the same as the fusion system of the

group $D \rtimes C_3$. It follows that there are only six B-subsections up to conjugation; two of them are major. For $1 \neq z \in \mathrm{Z}(D)$, we have $l(b_z) = 1$. Let us denote the four nonmajor subsections by (u_i, b_i) for $i = 1, \ldots, 4$. We may assume that b_1 has defect group C_2^3. It is easy to see that $\mathrm{Aut}_{\mathcal{F}}(D)$ restricts to the inertial group of b_1. In particular, $l(b_1) = e(b_1) = 3$. Moreover, the Cartan matrix of b_1 is given by

$$2\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

up to basic sets (see [Sambale 2011b, Theorem 3]). Moreover, b_2 has defect 3 and b_3 and b_4 have defect 4. Here, $l(b_2) = l(b_3) = l(b_4) = 1$. In particular, Olsson's conjecture $k_0(B) \leq 8 = |D : D'|$ follows. Looking at d_{ij}^z, we get $k(B) \leq 14$. The numbers $d_{ij}^{u_1}$ can certainly be arranged in the form

$$\begin{pmatrix} 1 & 1 & 1 & 1 & . & . & . & . & . & \cdots & . \\ 1 & 1 & . & . & 1 & 1 & . & . & . & \cdots & . \\ 1 & 1 & . & . & . & . & 1 & 1 & . & \cdots & . \end{pmatrix}^{\mathrm{T}}.$$

Using the contributions, it follows that $k_0(B) = 8$. We can easily add the column for (u_2, b_2) as

$$\begin{pmatrix} 1 & 1 & -1 & \cdots & -1 & 0 & \cdots & 0 \end{pmatrix}^{\mathrm{T}} \quad \text{or} \quad \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}^{\mathrm{T}}.$$

We investigate next the elementary divisors of the Cartan matrix of B. For this, we consider the multiplicity of $\langle u_1 \rangle$ as a lower defect group. The multiplicity of 2 as an elementary divisor of the Cartan matrix of b_1 is certainly 2. Since $\langle u_1 \rangle$ is the only lower defect group of order 2 of b_1, we have $m(2) = m_B^{(1)}(\langle u_1 \rangle, b_1) = m_{b_1}^{(1)}(\langle u_1 \rangle, b_1) = 2$. This shows $l(B) \geq 3$ and $k(B) \geq 10$. Every automorphism of order 3 of D fixes only two elements in D. Thus, it follows as in Proposition 4.3 that $m(d) = 0$ for $2 < d < 32$. We have essentially four possibilities for the numbers d_{ij}^z:

- eight entries ± 1 and six entries ± 2,
- eight entries ± 1, two entries ± 2 and one entry ± 4,
- seven entries ± 1, four entries ± 2 and one entry ± 3 or
- six entries ± 1, two entries ± 2 and two entries ± 3.

In particular, $k(B)$ determines $k_i(B)$ for $i \geq 1$ uniquely. It remains to add the generalized decomposition numbers corresponding to (u_3, b_3) and (u_4, b_4). Here the situation is distinguished by $q \in \{33, 34\}$. Assume first that $q = 34$. Then u_3^{-1} and u_4^{-1} are conjugate to u_3 and u_4, respectively. Hence, the numbers $d_{ij}^{u_3}$ and $d_{ij}^{u_4}$ are integers. It is easy to see that such a column must consist of the following (nonzero) entries:

- eight entries ± 1 and two entries ± 2 and
- seven entries ± 1 and one entry ± 3.

In contrast, for $q = 33$ the elements u_3^{-1} and u_4 are conjugate. So we may assume $u_4 := u_3^{-1}$, and it suffices to consider the column $d_{ij}^{u_3}$ whose entries are Gaussian integers. Let us write $d_{\chi\varphi_3}^{u_3} := a(\chi) + b(\chi)i$, where $\mathrm{IBr}(b_3) = \{\varphi_3\}$, $a, b \in \mathbb{Z}^{k(B)}$ and $i := \sqrt{-1}$. Then $(a, a) = (b, b) = 8$ and $(a, b) = 0$. Since we have only one pair of algebraically conjugate subsections, there is only one pair of 2-conjugate characters (see [Feit 1982, Lemma IV.6.10]). This shows that b consists of two entries ± 2. Now $k_0(B) = 8$ implies that a has eight entries ± 1.

As usual, we enumerate all these configurations of the generalized decomposition matrix and obtain the Cartan matrix of B as orthogonal space. However, we get two possibilities $l(B) \in \{3, 4\}$. We are not able to exclude the case $l(B) = 4$ despite its contradiction of Alperin's weight conjecture. Anyway in both cases $l(B) \in \{3, 4\}$, all candidates for the Cartan matrix satisfy [Héthelyi et al. 2013, Theorem 2.4]. The claim follows. □

We add a short discussion about the defect group

$$\begin{aligned} D &:= \texttt{SmallGroup}(32, 27) \\ &\cong \langle a, b, c \mid a^2 = b^2 = c^2 = [a, b] = [a, {}^c a] = [{}^c a, b] = [b, {}^c b] = 1 \rangle \cong C_2^4 \rtimes C_2. \end{aligned}$$

Let $\mathcal{F}$ be a nonnilpotent fusion system on D. One can show $Q := \langle a, b, {}^c a, {}^c b \rangle \cong C_2^4$ is the only possible $\mathcal{F}$-essential subgroup. In particular, $\mathcal{F}$ is controlled or constrained (note that controlled is a strong form of constrained). In the controlled case, we have $\mathcal{F} = \mathcal{F}_D(D \rtimes C_3) = \mathcal{F}_D(\texttt{SmallGroup}(96, 70))$. In the noncontrolled case, we have various possibilities for $\mathcal{F}$ according to $\mathrm{Out}_{\mathcal{F}}(Q) \in \{S_3, D_{10}, S_3 \times C_3, \texttt{SmallGroup}(18, 4), D_{10} \times C_3\}$ (see [Sambale 2012b, Lemma 3.11]). These possibilities are represented by the following groups:

$$\begin{array}{ll} \texttt{SmallGroup}(96, 195), & \texttt{SmallGroup}(288, 1025), \\ \texttt{SmallGroup}(96, 227), & \texttt{SmallGroup}(288, 1026), \\ \texttt{SmallGroup}(160, 234), & \texttt{SmallGroup}(480, 1188). \end{array}$$

Here observe that in case $\mathrm{Out}_{\mathcal{F}}(Q) = S_3$ there are essentially two different actions of $\mathrm{Out}_{\mathcal{F}}(Q)$ on Q. The cases $\mathrm{Out}_{\mathcal{F}}(Q) \in \{S_3 \times C_3, \texttt{SmallGroup}(18, 4)\}$ also differ by $\mathrm{Out}_{\mathcal{F}}(D) \in \{C_3, 1\}$, respectively. Additionally, for $\mathrm{Out}_{\mathcal{F}}(Q) = \texttt{SmallGroup}(18, 4)$ there exists a nontrivial 2-cocycle on $\mathrm{Out}_{\mathcal{F}}(Q)$ (on the other hand, the Künneth formula implies $\mathrm{H}^2(S_3 \times C_3, F^\times) = 0$ for an algebraically closed field F of characteristic 2). This gives even more examples for blocks with defect group D. For example, a nonprincipal 2-block of $\texttt{SmallGroup}(864, 3996)$ has defect group D and only one irreducible Brauer character. In all these examples, $l(B)$ assumes the

values 1, 2, 3, 5, 6, 9. We will not consider the block invariants in full generality although it might be possible. We also end the discussion about the remaining groups of order 32. In most cases (especially when 9×9 Cartan matrices show up), the computational effort to compute the corresponding block invariants is too big.

In the following table, we enumerate all groups of order 32 by using the small groups library and give information about blocks with corresponding defect groups. In many cases, it can be shown with GAP that there are no nontrivial fusion systems. These cases were also determined in [van der Waall 1991], but with the enumeration of [Hall and Senior 1964]. Using a conversion between both enumerations provided by Eamonn O'Brien, we confirm the results in [van der Waall 1991]. We denote the modular group of order $2^n \geq 16$ by M_{2^n}, i.e., the unique group of class 2 with a cyclic maximal subgroup.

small group #	description	invariants	comments	reference
1	C_{32}	known	nilpotent	
2	MNA(2, 2)	known	controlled	[Sambale 2011a; Eaton et al. 2012]
3	$C_8 \times C_4$	known	nilpotent	
4	$C_8 \rtimes C_4$	known	nilpotent	[Sambale 2012c]
5	MNA(3, 1)	known		[Sambale 2011a]
6	$\mathrm{MNA}(2, 1) \rtimes C_2$	known	nilpotent	GAP
7	$M_{16} \rtimes C_2$	known	nilpotent	GAP
8	$C_2.\mathrm{MNA}(2, 1)$	known	nilpotent	GAP
9	$D_8 \rtimes C_4$	known	bicyclic	Theorem 3.1
10	$Q_8 \rtimes C_4$	known	bicyclic	Theorem 3.3
11	$C_4 \wr C_2$	known		[Külshammer 1980]
12	$C_4 \rtimes C_8$	known	nilpotent	[Sambale 2012c]
13	$C_8 \rtimes C_4$	known	nilpotent	[Sambale 2012c]
14	$C_8 \rtimes C_4$	known	nilpotent	[Sambale 2012c]
15	$C_8.C_4$	known	nilpotent	[Sambale 2012c]
16	$C_{16} \times C_2$	known	nilpotent	
17	M_{32}	known	nilpotent	[Sambale 2012c]
18	D_{32}	known	maximal class	[Brauer 1974]
19	SD_{32}	known	maximal class	[Olsson 1975]
20	Q_{32}	known	maximal class	[Olsson 1975]
21	$C_4^2 \times C_2$	known	controlled	[Usami 1988]
22	$\mathrm{MNA}(2, 1) \times C_2$	known	constrained	Proposition 4.2
23	$(C_4 \rtimes C_4) \times C_2$	known	nilpotent	GAP
24	$C_4^2 \rtimes C_2$	known	nilpotent	GAP
25	$D_8 \times C_4$	known		[Sambale 2012a]
26	$Q_8 \times C_4$	known		[Sambale 2013b]
27	$C_2^4 \rtimes C_2$			

small group #	description	invariants	comments	reference
28	$(C_4 \times C_2^2) \rtimes C_2$	known	constrained	Proposition 4.7
29	$(Q_8 \times C_2) \rtimes C_2$	known	constrained	Proposition 4.7
30	$(C_4 \times C_2^2) \rtimes C_2$	known	nilpotent	GAP
31	$(C_4 \times C_4) \rtimes C_2$	known	nilpotent	GAP
32	$C_2^2.C_2^3$	known	nilpotent	GAP
33	$(C_4 \times C_4) \rtimes C_2$		controlled	
34	$(C_4 \times C_4) \rtimes C_2$		controlled	
35	$C_4 \rtimes Q_8$	known	nilpotent	GAP
36	$C_8 \times C_2^2$	known	controlled	[Usami 1988]
37	$M_{16} \times C_2$	known	nilpotent	GAP
38	$D_8 * C_8$	known		[Sambale 2013a]
39	$D_{16} \times C_2$	known		[Sambale 2012a]
40	$SD_{16} \times C_2$	known		[Sambale 2013b]
41	$Q_{16} \times C_2$	known		[Sambale 2013b]
42	$D_{16} * C_4$	known		[Sambale 2013a]
43	$(D_8 \times C_2) \rtimes C_2$			
44	$(Q_8 \times C_2) \rtimes C_2$			
45	$C_4 \times C_2^3$	known	controlled	Theorem 4.1
46	$D_8 \times C_2^2$			
47	$Q_8 \times C_2^2$		controlled	
48	$(D_8 * C_4) \times C_2$		controlled	
49	$D_8 * D_8$		controlled	
50	$D_8 * Q_8$		controlled	
51	C_2^5		controlled	

We apply these results to Theorem 2.2.

Theorem 4.9. *Let D be a cyclic central extension of one of the following groups*:

(i) *a metacyclic group*,

(ii) *a minimal nonabelian group*,

(iii) *a group of order at most* 16,

(iv) $\prod_{i=1}^{n} C_{2^{m_i}}$ *where* $|\{m_i : i = 1, \dots, n\}| \geq n - 1$,

(v) $M \times C$ *where M has maximal class and C is cyclic*,

(vi) $M * C$ *where M has maximal class and C is cyclic*,

(vii) $D_{2^n} \rtimes C_{2^m}$, $Q_{2^n} \rtimes C_{2^m}$ *and* $D_{2^n}.C_{2^m}$ *as in Theorems 3.1, 3.3 and 3.2*,

(viii) `SmallGroup`$(32, q)$ *for* $q \in \{11, 22, 28, 29, 33, 34\}$, *or*

(ix) *a group that admits only the nilpotent fusion system.*

Then Brauer's $k(B)$-conjecture holds for every 2-block with defect group D.

Proof. The case (iii) follows from Theorem 4.4. In case (viii), the result follows from Propositions 4.2, 4.7 and 4.8 and [Külshammer 1980]. In all other cases, it suffices to show $l(B) \leq 3$ for every block B with defect group given in the remaining list of the statement. For the abelian defect group $\prod_{i=1}^{n} C_{2^{m_i}}$ where $|\{m_i : i = 1, \ldots, n\}| \geq n-1$, it is easy to see that the inertial index $e(B)$ is at most 3. Thus, results of Puig and Usami [Usami 1988] imply Alperin's weight conjecture in this case. Now $l(B) \leq 3$ follows easily. For the remaining cases, the claim was shown in [Sambale 2012c; Eaton et al. 2012; Sambale 2011c; 2012a; 2013b; 2013a] and the present paper. □

One can show with GAP that Theorem 4.9 suffices to verify Brauer's $k(B)$-conjecture for 244 of the 267 defect groups of order 64. Here we also use the following elementary observation: Let $z \in \mathrm{Z}(D)$ such that every fusion system on $D/\langle z\rangle$ is controlled. If $\mathrm{C}_{\mathrm{Aut}(D)}(z)$ is a 2-group, then Brauer's $k(B)$-conjecture holds for every block with defect group D.

For the group $D \cong \mathtt{SmallGroup}(64, 265)$, we can argue even more subtly. Every block B with defect group D fulfills $e(B) \in \{1, 3, 5\}$. In case $e(B) = 3$, we find an element $z \in \mathrm{Z}(D)$ such that $D/\langle z\rangle$ is elementary abelian. Then [Usami 1988] implies $k(B) \leq 64$. On the other hand if $e(B) = 5$, we choose $z \in \mathrm{Z}(D)$ such that $D/\langle z\rangle \cong D_8 * Q_8$. Here the $k(B)$-conjecture follows from Proposition 4.3.

For the purpose of further research, we state all indices q such that Brauer's $k(B)$-conjecture for the defect group $\mathtt{SmallGroup}(64, q)$ is *not* known so far:

$$134, 135, 136, 137, 138, 139, 202, 224, 229, 230, 231,$$
$$238, 239, 242, 254, 255, 257, 258, 259, 261, 262, 264, 267.$$

This implies the following corollary:

Corollary 4.10. *Let B be a 2-block with defect group D of order at most 64. If D is generated by two elements, then Brauer's $k(B)$-conjecture holds for B.*

One can also formulate a version of Theorem 4.9 for $k_0(B)$ using Theorem 2.4. Compare also with [Héthelyi et al. 2013, Theorem 2.5].

Corollary 4.11. *Let D be a 2-group containing a cyclic subgroup of index at most 4. Then Brauer's $k(B)$-conjecture holds for every block with defect group D.*

Proof. We may assume that D is not metacyclic. In particular, $|D|/\exp D = 4$. If D is abelian, the result follows from [Sambale 2011c, Corollary 2]. Hence, let us assume that D is nonabelian. Then D is one of the groups given in [Ninomiya 1994, Theorem 2]. We will consider this list of groups case by case and apply Theorem 4.9. We remark that the terms "quasidihedral" and "semidihedral" have different meanings in [Ninomiya 1994].

The group G_1 is metacyclic.

For the groups G_2 and G_3, we even know the block invariants precisely.

Now consider G_4. Here the element a lies in the center. In particular, the group is a cyclic central extension of a group of order 4. The $k(B)$-conjecture follows.

For the group G_5, the element b lies in the center. Moreover, $G_5/\langle b\rangle$ is abelian and has a cyclic subgroup of index 2. Again the claim holds.

The groups G_6, G_7, G_8 and G_9 are metacyclic.

The groups G_{10} and G_{11} are cyclic central extensions of metacyclic groups.

In G_{12}, the subgroup $\langle a\rangle$ is normal; in particular, $a^{2^{m-3}} \in \mathrm{Z}(G_{12})$. Moreover, b is central in $G_{12}/\langle a^{2^{m-3}}\rangle$ and $G_{12}/\langle a^{2^{m-3}}\rangle \cong D_{2^{m-2}} \times C_2$. The claim follows.

In G_{13} and G_{14}, we see that b is central and the corresponding quotient is certainly metacyclic.

Next, $a^{2^{m-3}} \in \mathrm{Z}(G_{15})$ and $G_{15}/\langle a^{2^{m-3}}\rangle \cong D_{2^{m-2}} \times C_2$. Exactly the same argument applies to G_{16}.

For G_{17}, we have $c^{-1}a^2c = abab = a^{2+2^{m-3}}$ and $a^4 \in \mathrm{Z}(G_{17})$. Since $G_{17}/\langle a^4\rangle$ has order 16, the claim follows.

The group G_{18} is slightly more complicated. In general, the core of $\langle a\rangle$ has index at most 8. Thus, $a^{2^{m-3}}$ is always central (in all of these groups). Adjusting notation slightly gives

$$G_{18}/\langle a^{2^{m-3}}\rangle \cong \langle a, b, c \mid a^{2^{m-3}} = b^2 = c^2 = [a, b] = 1,\ cac = a^{-1}b\rangle.$$

We define new elements in this quotient by $\tilde{v} := a^2b$, $\tilde{x} := bc$ and $\tilde{a} := ac$. Then $\tilde{v}^{2^{m-4}} = 1$, $\tilde{a}^2 = b$ and $\tilde{a}^4 = 1$. Moreover, $cbc = c(acac)c = b$. It follows that $\tilde{x}^2 = 1$ and $\tilde{x}\tilde{v}\tilde{x} = \tilde{v}^{-1}$. Hence, $\langle \tilde{v}, \tilde{x}\rangle \cong D_{2^{m-3}}$. Now $\tilde{a}\tilde{v}\tilde{a}^{-1} = ca^2bc = a^{-2}b = \tilde{v}^{-1}$ and finally $\tilde{a}\tilde{x}\tilde{a}^{-1} = a^2c = \tilde{v}\tilde{x}$. Since $G_{18}/\langle a^{2^{m-3}}\rangle = \langle \tilde{v}, \tilde{x}, \tilde{a}\rangle$, we see that this is precisely the group from Theorem 3.1. The claim follows.

The groups G_{19}, G_{20} and G_{21} are metacyclic.

In G_{22}, the element a^4 is central and $G_{22}/\langle a^4\rangle$ has order 16.

Let us consider G_{23}. As above, we have

$$G_{23}/\langle a^{2^{m-3}}\rangle \cong \langle a, b, c \mid a^{2^{m-3}} = b^2 = c^2 = [a, b] = 1,\ cac = a^{-1+2^{m-4}}b\rangle$$

(observe that the relation $[b, c] \equiv 1 \pmod{\langle a^{2^{m-3}}\rangle}$ follows from $b \equiv a^{1+2^{m-4}}cac$). Here we define $\tilde{v} := a^{2+2^{m-4}}b$, $\tilde{x} := bc$ and $\tilde{a} := ac$. Then again $\langle \tilde{v}, \tilde{x}\rangle \cong D_{2^{m-3}}$. Moreover, $\tilde{a}^2 = a^{2^{m-4}}b$, $\tilde{a}^4 = 1$ and $\tilde{a}\tilde{x}\tilde{a}^{-1} = bca^{-1}cac = a^{2+2^{m-4}}c = \tilde{v}\tilde{x}$. So $G_{23}/\langle a^{2^{m-3}}\rangle$ is the group from Theorem 3.1.

Now it is easy to see that $G_{24}/\langle a^{2^{m-3}}\rangle \cong G_{25}/\langle a^{2^{m-3}}\rangle \cong G_{23}/\langle a^{2^{m-3}}\rangle$.

Finally the group G_{26} has order 32, so also here the $k(B)$-conjecture holds. This completes the proof. □

For every integer $n \geq 6$, there are exactly 33 groups of order 2^n satisfying the hypothesis of Corollary 4.11.

5. Olsson's conjecture

We have seen in [Héthelyi et al. 2013] that Olsson's conjecture holds for all controlled 2-blocks of defect at most 5. Using the table above, we remove the controlled condition.

Theorem 5.1. *Olsson's conjecture holds for all* 2*-blocks of defect at most* 5.

Proof. By the remark above it, suffices to consider only the defect groups $D := \mathtt{SmallGroup}(32, m)$ where $m \in \{27, 43, 44, 46\}$. Let B be a block with defect group D and fusion system $\mathcal{F}$. Then we can find (with GAP) an element $u \in D$ such that $|\mathrm{C}_D(u)| = |D : D'|$. Moreover, we can choose u such that every element $v \in D$ of the same order also satisfies $|\mathrm{C}_D(u)| = |D : D'|$. Hence, the subgroup $\langle u \rangle$ is fully $\mathcal{F}$-centralized. In particular, $\mathrm{C}_D(u)$ is a defect group of the block b_u. Now the claim follows from [Héthelyi et al. 2013, Proposition 2.5(ii)]. □

In [Héthelyi et al. 2013], we also verified Olsson's conjecture for defect groups of p-rank 2 provided $p > 3$. We use the opportunity to explore the case $p = 3$ in more detail.

Theorem 5.2. *Let B be a* 3*-block of a finite group G with defect group D. Assume that D has* 3*-rank* 2 *but not maximal class. Then Olsson's conjecture holds for B.*

Proof. By [Héthelyi et al. 2013, Theorem 5.6], we may assume that the fusion system $\mathcal{F}$ of B is not controlled. Then $|D| \geq 3^4$ since D does not have maximal class. By [Díaz et al. 2007, Theorems 4.1 and 4.2], it remains to handle the groups $D = G(3, r; \epsilon)$ of order 3^r where $r \geq 5$ and $\epsilon \in \{\pm 1\}$ as in [Díaz et al. 2007, Theorem 4.7] (by [Díaz et al. 2007, Remark A.3], $G(3, 4; \epsilon)$ has maximal class). Assume that D is given by generators and relations as in Theorem A.1 of the same paper. Consider the element $x := ac$. By [Díaz et al. 2007, Lemma A.8], x is not contained in the unique $\mathcal{F}$-essential ($\mathcal{F}$-Alperin) subgroup $C(3, r-1) = \langle a, b, c^3 \rangle$. In particular, $\langle x \rangle$ is fully $\mathcal{F}$-centralized, and the block b_x of the subsection (x, b_x) has defect group $\mathrm{C}_D(x)$. It is easy to see that $D' = \langle b, c^{3^{r-3}} \rangle \cong C_p \times C_p$. It follows that $x^{3^{r-4}} \equiv c^{3^{r-4}} \not\equiv 1 \pmod{D'}$ and $|\langle x \rangle| \geq 3^{r-3}$. As usual, we have $|\mathrm{C}_D(x)| \geq |D : D'| = 3^{r-2}$. In case $|\mathrm{C}_D(x)| \geq 3^{r-1}$, we get the contradiction $b \in D' \subseteq \mathrm{C}_D(x)$. Hence, $|\mathrm{C}_D(x)| = |D : D'|$ and $\mathrm{C}_D(x)/\langle x \rangle$ is cyclic. Now Olsson's conjecture for B follows from [Héthelyi et al. 2013, Proposition 2.5]. □

Theorem 5.3. *Let B be a* 3*-block of a finite group with defect group D of order at least* 3^4*. Assume that D has maximal class but is* not *isomorphic to the group*

$$\begin{aligned} B(3, r; 0, 0, 0) = \big\langle s, s_1, \ldots, s_{r-1} \;\big|\; & s^3 = s_{r-2}^3 = s_{r-1}^3 = [s_1, s_2] = \cdots = [s_1, s_{r-1}] \\ & = s_1^3 s_2^3 s_3 = \cdots = s_{r-3}^3 s_{r-2}^3 s_{r-1} = 1,\ s_i = [s_{i-1}, s] \quad \textit{for } i = 2, \ldots, r-1 \big\rangle \end{aligned}$$

of order 3^r. Then Olsson's conjecture holds for B.

Proof. By [Héthelyi et al. 2013, Theorem 5.6], we may assume that the fusion system $\mathcal{F}$ of B is not controlled. Then $\mathcal{F}$ is given as in [Díaz et al. 2007, Theorem 5.10]. In particular, $D = B(3, r; 0, \gamma, 0)$ is given by generators and relations as in [Díaz et al. 2007, Theorem A.2] where $\gamma \in \{1, 2\}$. Let D_1 be as in [Huppert 1967, Definition III.14.3]. Observe that in the notation of [Díaz et al. 2007; Blackburn 1958] we have $D_1 = \gamma_1(D)$. From [Díaz et al. 2007, Proposition A.9] we see that $x := ss_1 \notin D_1$. Moreover, we have $x^3 \neq 1$ also by the same proposition. Then by [Díaz et al. 2007, Lemma A.15], x does not lie in one of the centric subgroups D_1, E_i or V_i for $i \in \{-1, 0, 1\}$. This shows that x is not $\mathcal{F}$-conjugate to an element in D_1. By [Huppert 1967, Satz III.14.17], D is not an exceptional group. In particular, [Huppert 1967, Hilfssatz III.14.13] implies $|\mathrm{C}_D(y)| = 9 = |D : D'|$ for all $y \in D \setminus D_1$. Hence, $\langle x \rangle$ is fully $\mathcal{F}$-centralized. Thus, the block b_x of the subsection (x, b_x) has defect group $\mathrm{C}_D(x)$. Now Olsson's conjecture follows from [Héthelyi et al. 2013, Proposition 2.5]. □

We remark that the method in Theorem 5.3 does not work for the groups $B(3, r; 0, 0, 0)$. For example, every block of a subsection of the principal 3-block of ${}^3D_4(2)$ has defect at least 3 (here $r = 4$). However, $|D : D'| = 3^2$ for every 3-group of maximal class.

Acknowledgments

This work is supported by the German Academic Exchange Service (DAAD) and the German Research Foundation (DFG) and was written partly in Santa Cruz, USA. I thank the University of California for its hospitality.

References

[Alperin et al. 1973] J. L. Alperin, R. Brauer, and D. Gorenstein, "Finite simple groups of 2-rank two", *Scripta Math.* **29**:3-4 (1973), 191–214. MR 53 #5728 Zbl 0274.20021

[An 2011] J. An, "Controlled blocks of the finite quasisimple groups for odd primes", *Adv. Math.* **227**:3 (2011), 1165–1194. MR 2012f:20023 Zbl 1247.20008

[Berkovich and Janko 2008] Y. Berkovich and Z. Janko, *Groups of prime power order, II*, de Gruyter Expositions in Mathematics **47**, de Gruyter, 2008. MR 2009m:20026b Zbl 1168.20002

[Blackburn 1958] N. Blackburn, "On a special class of p-groups", *Acta Math.* **100** (1958), 45–92. MR 21 #1349 Zbl 0083.24802

[Brauer 1968] R. Brauer, "On blocks and sections in finite groups, II", *Amer. J. Math.* **90** (1968), 895–925. MR 39 #5713 Zbl 0174.05401

[Brauer 1969] R. Brauer, "Defect groups in the theory of representations of finite groups", *Illinois J. Math.* **13** (1969), 53–73. MR 40 #248 Zbl 0167.29801

[Brauer 1974] R. Brauer, "On 2-blocks with dihedral defect groups", pp. 367–393 in *Symposia Mathematica, XIII* (Rome, 1972), Academic Press, London, 1974. MR 50 #7315 Zbl 0288.20010

[Broué 1980] M. Broué, "On characters of height zero", pp. 393–396 in *The Santa Cruz Conference on Finite Groups* (Santa Cruz, CA, 1979), edited by B. Cooperstein and G. Mason, Proc. Sympos. Pure Math. **37**, Amer. Math. Soc., Providence, R.I., 1980. MR 82g:20021 Zbl 0475.20009

[Broué and Olsson 1986] M. Broué and J. B. Olsson, "Subpair multiplicities in finite groups", *J. Reine Angew. Math.* **371** (1986), 125–143. MR 87j:20026 Zbl 0596.20009

[Craven and Glesser 2012] D. A. Craven and A. Glesser, "Fusion systems on small p-groups", *Trans. Amer. Math. Soc.* **364**:11 (2012), 5945–5967. MR 2946938 Zbl 06199346

[Díaz et al. 2007] A. Díaz, A. Ruiz, and A. Viruel, "All p-local finite groups of rank two for odd prime p", *Trans. Amer. Math. Soc.* **359**:4 (2007), 1725–1764. MR 2008a:20037 Zbl 1113.55010

[Eaton 2003] C. W. Eaton, "Generalisations of conjectures of Brauer and Olsson", *Arch. Math. (Basel)* **81**:6 (2003), 621–626. MR 2004i:20012 Zbl 1046.20007

[Eaton and Moretó 2013] C. W. Eaton and A. Moretó, "Extending Brauer's height zero conjecture to blocks with nonabelian defect groups", *Int. Math. Res. Not.* (2013), Art. ID rnt131.

[Eaton et al. 2012] C. W. Eaton, B. Külshammer, and B. Sambale, "2-blocks with minimal nonabelian defect groups, II", *J. Group Theory* **15**:3 (2012), 311–321. MR 2920888 Zbl 1253.20008

[Erdmann 1990] K. Erdmann, *Blocks of tame representation type and related algebras*, Lecture Notes in Mathematics **1428**, Springer, Berlin, 1990. MR 91c:20016 Zbl 0696.20001

[Feit 1982] W. Feit, *The representation theory of finite groups*, North-Holland Mathematical Library **25**, North-Holland, Amsterdam, 1982. MR 83g:20001 Zbl 0493.20007

[GAP 2012] *GAP – Groups, Algorithms, and Programming, Version 4.5.6*, The GAP Group, 2012, Available at http://www.gap-system.org.

[Gorenstein 1968] D. Gorenstein, *Finite groups*, Harper & Row, New York, 1968. MR 38 #229 Zbl 0185.05701

[Hall and Senior 1964] M. Hall, Jr. and J. K. Senior, *The groups of order* 2^n ($n \leq 6$), Macmillan, New York, 1964. MR 29 #5889 Zbl 0192.11701

[Héthelyi et al. 2013] L. Héthelyi, B. Külshammer, and B. Sambale, "A note on Olsson's Conjecture", preprint, 2013, Available at https://www.minet.uni-jena.de/algebra/personen/sambale/olssoncon.pdf. To appear in *J. Algebra*.

[Huppert 1967] B. Huppert, *Endliche Gruppen, I*, Grundlehren der Math. Wissenschaften **134**, Springer, Berlin, 1967. MR 37 #302 Zbl 0217.07201

[Kessar 2007] R. Kessar, "Introducton to block theory", pp. 47–77 in *Group representation theory*, edited by M. Geck et al., EPFL Press, Lausanne, 2007. MR 2008f:20020 Zbl 1165.20007

[Kessar and Malle 2013] R. Kessar and G. Malle, "Quasi-isolated blocks and Brauer's height zero conjecture", *Ann. of Math. (2)* **178**:1 (2013), 321–384. MR 3043583 Zbl 06190562

[Kessar et al. 2012] R. Kessar, S. Koshitani, and M. Linckelmann, "Conjectures of Alperin and Broué for 2-blocks with elementary abelian defect groups of order 8", *J. Reine Angew. Math.* **671** (2012), 85–130. MR 2983198 Zbl 06100515

[Külshammer 1980] B. Külshammer, "On 2-blocks with wreathed defect groups", *J. Algebra* **64**:2 (1980), 529–555. MR 81i:20008 Zbl 0442.20012

[Külshammer and Okuyama ∼ 2000] B. Külshammer and T. Okuyama, "On centrally controlled blocks of finite groups", ∼ 2000. unpublished.

[Külshammer and Sambale 2013] B. Külshammer and B. Sambale, "The 2-blocks of defect 4", *Represent. Theory* **17** (2013), 226–236. MR 3048571 Zbl 06183358

[Liebeck 1971] H. Liebeck, "The location of the minimum of a positive definite integral quadratic form", *J. London Math. Soc.* (2) **3** (1971), 477–484. MR 44 #1193 Zbl 0211.07801

[Linckelmann 2006] M. Linckelmann, "Simple fusion systems and the Solomon 2-local groups", *J. Algebra* **296**:2 (2006), 385–401. MR 2006i:20024 Zbl 1097.20014

[Linckelmann 2007] M. Linckelmann, "Introduction to fusion systems", pp. 79–113 in *Group representation theory*, edited by M. Geck et al., EPFL Press, Lausanne, 2007. MR 2008f:20021 Zbl 1161.20007

[Malle and Navarro 2006] G. Malle and G. Navarro, "Inequalities for some blocks of finite groups", *Arch. Math.* (*Basel*) **87**:5 (2006), 390–399. MR 2007i:20073 Zbl 1109.20011

[Murai 2000] M. Murai, "On subsections of blocks and Brauer pairs", *Osaka J. Math.* **37**:3 (2000), 719–733. MR 2001k:20018 Zbl 0969.20008

[Nagao and Tsushima 1989] H. Nagao and Y. Tsushima, *Representations of finite groups*, Academic Press, Boston, 1989. MR 90h:20008 Zbl 0673.20002

[Ninomiya 1994] Y. Ninomiya, "Finite p-groups with cyclic subgroups of index p^2", *Math. J. Okayama Univ.* **36** (1994), 1–21. MR 96h:20044 Zbl 0838.20017

[Oliver and Ventura 2009] B. Oliver and J. Ventura, "Saturated fusion systems over 2-groups", *Trans. Amer. Math. Soc.* **361**:12 (2009), 6661–6728. MR 2010i:20018 Zbl 1244.20015

[Olsson 1975] J. B. Olsson, "On 2-blocks with quaternion and quasidihedral defect groups", *J. Algebra* **36**:2 (1975), 212–241. MR 51 #13016 Zbl 0314.20009

[Olsson 1980] J. B. Olsson, "Lower defect groups", *Comm. Algebra* **8**:3 (1980), 261–288. MR 81g:20024 Zbl 0426.20008

[Olsson 1981] J. B. Olsson, "Inequalities for block-theoretic invariants", pp. 270–284 in *Representations of algebras* (Puebla, 1980), edited by M. Auslander and E. Lluis, Lecture Notes in Math. **903**, Springer, Berlin, 1981. MR 83i:20013 Zbl 0478.20008

[Rédei 1947] L. Rédei, "Das „schiefe Produkt" in der Gruppentheorie", *Comment. Math. Helv.* **20** (1947), 225–264. MR 9,131a Zbl 0035.01503

[Robinson 1991] G. R. Robinson, "On the number of characters in a block", *J. Algebra* **138**:2 (1991), 515–521. MR 92h:20022a Zbl 0727.20010

[Robinson 1992] G. R. Robinson, "On Brauer's $k(B)$ problem", *J. Algebra* **147**:2 (1992), 450–455. MR 93b:20020 Zbl 0796.20006

[Robinson 2004] G. R. Robinson, "Weight conjectures for ordinary characters", *J. Algebra* **276**:2 (2004), 761–775. MR 2005c:20021 Zbl 1121.20007

[Robinson 2008] G. R. Robinson, "On the number of characters in a block and the Brauer–Feit matrix", 2008. unpublished.

[Sambale 2011a] B. Sambale, "2-blocks with minimal nonabelian defect groups", *J. Algebra* **337** (2011), 261–284. MR 2012d:20019 Zbl 1247.20010

[Sambale 2011b] B. Sambale, "Cartan matrices and Brauer's $k(B)$-conjecture", *J. Algebra* **331** (2011), 416–427. MR 2012c:20023 Zbl 1230.20010

[Sambale 2011c] B. Sambale, "Cartan matrices and Brauer's $k(B)$-conjecture, II", *J. Algebra* **337** (2011), 345–362. MR 2012e:20018 Zbl 1243.20013

[Sambale 2012a] B. Sambale, "Blocks with defect group $D_{2^n} \times C_{2^m}$", *J. Pure Appl. Algebra* **216**:1 (2012), 119–125. MR 2012f:20031 Zbl 1241.20013

[Sambale 2012b] B. Sambale, "Fusion systems on bicyclic 2-groups", preprint, 2012. arXiv 1206.2962v1

[Sambale 2012c] B. Sambale, "Fusion systems on metacyclic 2-groups", *Osaka J. Math.* **49**:2 (2012), 325–329. MR 2945751 Zbl 1247.20025

[Sambale 2013a] B. Sambale, "Blocks with central product defect group $D_{2^n} * C_{2^m}$", *Proc. Amer. Math. Soc.* **141**:12 (2013), 4057–4069. MR 3105851 Zbl 06218137

[Sambale 2013b] B. Sambale, "Blocks with defect group $Q_{2^n} \times C_{2^m}$ and $SD_{2^n} \times C_{2^m}$", *Algebr. Represent. Theory* **16**:6 (2013), 1717–1732. MR 3127355

[Stancu 2006] R. Stancu, "Control of fusion in fusion systems", *J. Algebra Appl.* **5**:6 (2006), 817–837. MR 2007j:20025 Zbl 1118.20020

[Usami 1988] Y. Usami, "On p-blocks with abelian defect groups and inertial index 2 or 3, I", *J. Algebra* **119**:1 (1988), 123–146. MR 89i:20024 Zbl 0659.20008

[van der Waall 1991] R. W. van der Waall, "On p-nilpotent forcing groups", *Indag. Math.* (*N.S.*) **2**:3 (1991), 367–384. MR 93b:20040 Zbl 0749.20004

[Watanabe 1991] A. Watanabe, "Notes on p-blocks of characters of finite groups", *J. Algebra* **136**:1 (1991), 109–116. MR 92i:20016 Zbl 0725.20008

Communicated by David Benson
Received 2012-09-30 Revised 2012-10-16 Accepted 2013-03-23

benjamin.sambale@uni-jena.de *Mathematisches Institut, Friedrich-Schiller-Universität, D-07737 Jena, Germany*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2275

Network parametrizations for the Grassmannian

Kelli Talaska and Lauren Williams

Deodhar introduced his decomposition of partial flag varieties as a tool for understanding Kazhdan–Lusztig polynomials. The Deodhar decomposition of the Grassmannian is also useful in the context of soliton solutions to the KP equation, as shown by Kodama and the second author. Deodhar components $\mathcal{R}_D$ of the Grassmannian are in bijection with certain tableaux D called *Go-diagrams*, and each component is isomorphic to $(\mathbb{K}^*)^a \times (\mathbb{K})^b$ for some nonnegative integers a and b.

Our main result is an explicit parametrization of each Deodhar component in the Grassmannian in terms of networks. More specifically, from a Go-diagram D we construct a weighted network N_D and its *weight matrix* W_D, whose entries enumerate directed paths in N_D. By letting the weights in the network vary over $\mathbb{K}$ or $\mathbb{K}^*$ as appropriate, one gets a parametrization of the Deodhar component $\mathcal{R}_D$. One application of such a parametrization is that one may immediately determine which Plücker coordinates are vanishing and nonvanishing, by using the Lindström–Gessel–Viennot lemma. We also give a (minimal) characterization of each Deodhar component in terms of Plücker coordinates. A main tool for us is the work of Marsh and Rietsch [*Represent. Theory* **8** (2004), 212–242] on Deodhar components in the flag variety.

The first author was partially supported by NSF grant DMS-1004532. The second author was partially supported by an NSF CAREER award and an Alfred Sloan fellowship.
MSC2010: primary 14M15; secondary 05E10.
Keywords: Grassmannian, network, total positivity, Deodhar decomposition.

1. Introduction

There is a remarkable subset of the real Grassmannian $\mathrm{Gr}_{k,n}(\mathbb{R})$ called its *totally nonnegative part* $(\mathrm{Gr}_{k,n})_{\geq 0}$ [Lusztig 1998; Postnikov 2006], which may be defined as the subset of the real Grassmannian where all Plücker coordinates have the same sign. Postnikov showed that $(\mathrm{Gr}_{k,n})_{\geq 0}$ has a decomposition into *positroid cells*, which are indexed by certain tableaux called ɹ-diagrams. He also gave explicit parametrizations of each cell. In particular, he showed that from each ɹ-diagram one can produce a planar *network*, and that one can write down a parametrization of the corresponding cell using the *weight matrix* of that network. This parametrization shows that the cell is isomorphic to $\mathbb{R}^d_{>0}$ for some d. Such a parametrization is convenient because, for example, one may read off formulas for Plücker coordinates from nonintersecting paths in the network, using the Lindström–Gessel–Viennot lemma.

A natural question is whether these network parametrizations for positroid cells can be extended from $(\mathrm{Gr}_{k,n})_{\geq 0}$ to the entire real Grassmannian $\mathrm{Gr}_{k,n}(\mathbb{R})$. In this paper we give an affirmative answer to this question, by replacing the positroid cell decomposition with the Deodhar decomposition of the Grassmannian $\mathrm{Gr}_{k,n}(\mathbb{K})$ (here $\mathbb{K}$ is an arbitrary field).

The components of the Deodhar decomposition are not in general cells, but nevertheless have a simple topology: by [Deodhar 1985; 1987], each one is isomorphic to $(\mathbb{K}^*)^a \times (\mathbb{K})^b$. The relation of the Deodhar decomposition of $\mathrm{Gr}_{k,n}(\mathbb{R})$ to Postnikov's cell decomposition of $(\mathrm{Gr}_{k,n})_{\geq 0}$ is as follows: the intersection of a Deodhar component $\mathcal{R}_D \cong (\mathbb{R}^*)^a \times (\mathbb{R})^b$ with $(\mathrm{Gr}_{k,n})_{\geq 0}$ is precisely one positroid cell isomorphic to $(\mathbb{R}_{>0})^a$ if $b = 0$, and is empty otherwise. In particular, when one intersects the Deodhar decomposition with $(\mathrm{Gr}_{k,n})_{\geq 0}$, one obtains the positroid cell decomposition of $(\mathrm{Gr}_{k,n})_{\geq 0}$. There is a related *positroid stratification* of the real Grassmannian, and each positroid stratum is a union of Deodhar components.

As for the combinatorics, components of the Deodhar decomposition are indexed by *distinguished subexpressions* [Deodhar 1985; 1987], or equivalently, by certain tableaux called *Go-diagrams* [Kodama and Williams 2013], which generalize ɹ-diagrams. In this paper we associate a network to each Go-diagram, and write down a parametrization of the corresponding Deodhar component using the weight matrix of that network. Our construction generalizes Postnikov's, but our networks are no longer planar in general.

Our main results can be summed up as follows. See Theorems 3.16 and 7.8 and the constructions preceding them for complete details.

Theorem. *Let $\mathbb{K}$ be an arbitrary field.*

- *Every point in $\mathrm{Gr}_{k,n}(\mathbb{K})$ can be realized as the weight matrix of a unique network associated to a Go-diagram, and we can explicitly construct the*

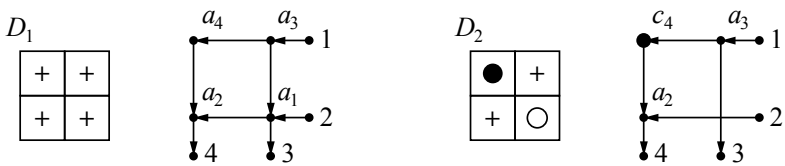


Figure 1. The diagrams and networks associated to $\mathcal{R}_{D_1}$ and $\mathcal{R}_{D_2}$ in Example 1.1.

corresponding network. The networks corresponding to points in the same Deodhar component have the same underlying graph, but different weights.

- *Every Deodhar component may be characterized by the vanishing and nonvanishing of certain Plücker coordinates. Using this characterization, we can also explicitly construct the network associated to a point given either by a matrix representative or by a list of Plücker coordinates.*

To illustrate the main results, we provide a small example here. More complicated examples may be seen throughout the rest of the paper.

Example 1.1. Consider the Grassmannian $\mathrm{Gr}_{2,4}$. The large Schubert cell in this Grassmannian can be characterized as

$$\Omega_\lambda = \{A \in \mathrm{Gr}_{2,4} \mid \Delta_{1,2}(A) \neq 0\},$$

where Δ_J denotes the Plücker coordinate corresponding to the column set J in a matrix representative of a point in $\mathrm{Gr}_{2,4}$. This Schubert cell contains multiple positroid strata, including $S_{\mathcal{I}}$, where $\mathcal{I}$ is the Grassmann necklace $\mathcal{I} = (12, 23, 34, 14)$. This positroid stratum can also be characterized by the nonvanishing of certain Plücker coordinates:

$$S_{\mathcal{I}} = \{A \in \mathrm{Gr}_{2,4} \mid \Delta_{1,2}(A) \neq 0,\ \Delta_{2,3}(A) \neq 0,\ \Delta_{3,4}(A) \neq 0,\ \Delta_{1,4}(A) \neq 0\}.$$

Figure 1 shows two Go-diagrams D_1 and D_2 and their associated networks. Note that the network on the right is not planar. The weight matrices associated to these diagrams are

$$\begin{pmatrix} 1 & 0 & -a_3 & -(a_3a_4 + a_3a_2) \\ 0 & 1 & a_1 & a_1a_2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & -a_3 & -a_3c_4 \\ 0 & 1 & 0 & a_2 \end{pmatrix}.$$

The positroid stratum $S_{\mathcal{I}}$ is the disjoint union of the two corresponding Deodhar components $\mathcal{R}_{D_1}$ and $\mathcal{R}_{D_2}$, which can be characterized in terms of vanishing and nonvanishing of minors as

$$\mathcal{R}_{D_1} = \{A \in S_{\mathcal{I}} \mid \Delta_{1,3} \neq 0\} \quad \text{and} \quad \mathcal{R}_{D_2} = \{A \in S_{\mathcal{I}} \mid \Delta_{1,3} = 0\}.$$

Note that if one lets the a_i's range over $\mathbb{K}^*$ and lets c_4 range over $\mathbb{K}$, then we see

that $\mathcal{R}_{D_1} \cong (\mathbb{K}^*)^4$ and $\mathcal{R}_{D_2} \cong (\mathbb{K}^*)^2 \times \mathbb{K}$.

There are several applications of our construction. First, as a special case of our theorem, one may parametrize all $k \times n$ matrices using networks. Second, by applying the Lindström–Gessel–Viennot lemma to a given network, one may write down explicit formulas for Plücker coordinates in terms of collections of nonintersecting paths in the network. Third, building upon [Kodama and Williams 2013], we obtain (minimal) descriptions of Deodhar components in the Grassmannian in terms of vanishing and nonvanishing of Plücker coordinates. It follows that each Deodhar component is a union of matroid strata.

Although less well known than the Schubert decomposition and matroid stratification, the Deodhar decomposition is very interesting in its own right. Deodhar's original motivation for introducing his decomposition was the desire to understand Kazhdan–Lusztig polynomials. In the flag variety, one may intersect two opposite Schubert cells, obtaining a Richardson variety, which Deodhar showed is a union of Deodhar components. Each Richardson variety $\mathcal{R}_{v,w}(q)$ may be defined over a finite field $\mathbb{K} = \mathbb{F}_q$, and in this case the number of points determines the R-polynomials $R_{v,w}(q) = \#(\mathcal{R}_{v,w}(\mathbb{F}_q))$, introduced by Kazhdan and Lusztig [1979] to give a recursive formula for the Kazhdan–Lusztig polynomials. Since each Deodhar component is isomorphic to $(\mathbb{F}_q^*)^a \times (\mathbb{F}_q)^b$ for some a and b, if one understands the decomposition of a Richardson variety into Deodhar components, then in principle one may compute the R-polynomials and hence Kazhdan–Lusztig polynomials.

Another reason for our interest in the Deodhar decomposition is its relation to soliton solutions of the KP equation. It is well known that from each point A in the real Grassmannian one may construct a soliton solution $u_A(x, y, t)$ of the KP equation. It was shown in [Kodama and Williams 2013] that when the time variable t tends to $-\infty$, the combinatorics of the solution $u_A(x, y, t)$ depends precisely on which Deodhar component A lies in.

One final result of this paper is the verification that two notions of total positivity for the Grassmannian coincide. Lusztig [1998] defined the totally nonnegative part of any partial flag variety in a Lie-theoretic way. He also conjectured a cell decomposition for it, proved by Rietsch [1998]. Independently, Postnikov defined the totally nonnegative part of the real Grassmannian in terms of Plücker coordinates, and gave a cell decomposition of it. It is not obvious that Lusztig's definitions (for $\mathrm{Gr}_{k,n}(\mathbb{R})$) coincide with Postnikov's; however, this has been verified by Rietsch (personal communication, 2009). In this paper we give a new proof that the two notions of total positivity coincide.

Corollary 1.2. *Lusztig's definition of the totally nonnegative part of* $\mathrm{Gr}_{k,n}(\mathbb{R})$ *and its cell decomposition coincides with Postnikov's definition of the totally nonnegative part of* $\mathrm{Gr}_{k,n}(\mathbb{R})$ *and its cell decomposition.*

The outline of this paper is as follows. In Section 2 we give some background on the Grassmannian and its decompositions, including the Schubert decomposition, the positroid stratification, and the matroid stratification. In Section 3 we present our main construction: we explain how to construct a network from each diagram, then use that network to write down a parametrization of a subset of the Grassmannian that we call a network component. In Section 4 we define Deodhar's decomposition of the flag variety, and its projection to the Grassmannian. We also describe parametrizations of Deodhar components in the flag variety which are due to Marsh and Rietsch [2004]. In Sections 5 and 6 we prove that after a rational transformation of variables, our network parametrizations coincide with the projections of the Marsh–Rietsch parametrizations. Finally, in Section 7 we give a characterization of Deodhar components in terms of the vanishing and nonvanishing of certain Plücker coordinates.

2. Background on the Grassmannian

The *Grassmannian* $\mathrm{Gr}_{k,n}$ is the space of all k-dimensional subspaces of an n-dimensional vector space $\mathbb{K}^n$. In this paper we will usually let $\mathbb{K}$ be an arbitrary field, though we will often think of it as $\mathbb{R}$ or $\mathbb{C}$. An element of $\mathrm{Gr}_{k,n}$ can be viewed as a full-rank $k \times n$ matrix modulo left multiplication by nonsingular $k \times k$ matrices. In other words, two $k \times n$ matrices represent the same point in $\mathrm{Gr}_{k,n}$ if and only if they can be obtained from each other by row operations. Let $\binom{[n]}{k}$ be the set of all k-element subsets of $[n] := \{1, \dots, n\}$. For $I \in \binom{[n]}{k}$, let $\Delta_I(A)$ be the *Plücker coordinate*, that is, the maximal minor of the $k \times n$ matrix A located in the column set I. The map $A \mapsto (\Delta_I(A))$, where I ranges over $\binom{[n]}{k}$, induces the *Plücker embedding* $\mathrm{Gr}_{k,n} \hookrightarrow \mathbb{KP}^{\binom{n}{k}-1}$ into projective space.

We now describe several useful decompositions of the Grassmannian: the Schubert decomposition, the positroid stratification, and the matroid stratification. Note that the matroid stratification refines the positroid stratification, which refines the Schubert decomposition. The main subject of this paper is the *Deodhar decomposition* of the Grassmannian, which refines the positroid stratification, and is refined by the matroid stratification (as we prove in Corollary 7.9).

2A. *The Schubert decomposition of* $\mathrm{Gr}_{k,n}$. Throughout this paper, we identify partitions with their Young diagrams. Recall that the partitions λ contained in a $k \times (n-k)$ rectangle are in bijection with k-element subset $I \subset [n]$. The boundary of the Young diagram of such a partition λ forms a lattice path from the upper-right corner to the lower-left corner of the rectangle. Let us label the n steps in this path by the numbers $1, \dots, n$, and define $I = I(\lambda)$ as the set of labels on the k vertical steps in the path. Conversely, we let $\lambda(I)$ denote the partition corresponding to the subset I.

Definition 2.1. For each partition λ contained in a $k \times (n-k)$ rectangle, we define the *Schubert cell*

$$\Omega_\lambda = \{A \in \mathrm{Gr}_{k,n} \mid I(\lambda) \text{ is the lexicographically minimal subset such that } \Delta_{I(\lambda)}(A) \neq 0\}.$$

As λ ranges over the partitions contained in a $k \times (n-k)$ rectangle, this gives the *Schubert decomposition* of the Grassmannian $\mathrm{Gr}_{k,n}$, that is,

$$\mathrm{Gr}_{k,n} = \bigsqcup_{\lambda \subset (n-k)^k} \Omega_\lambda.$$

We now define the *shifted linear order* $<_i$ (for $i \in [n]$) to be the total order on $[n]$ defined by

$$i <_i i+1 <_i i+2 <_i \cdots <_i n <_i 1 <_i \cdots <_i i-1.$$

One can then define *cyclically shifted Schubert cells*:

Definition 2.2. For each partition λ contained in a $k \times (n-k)$ rectangle, and each $i \in [n]$, we define the *cyclically shifted Schubert cell*

$$\Omega^i_\lambda = \{A \in \mathrm{Gr}_{k,n} \mid I(\lambda) \text{ is the lexicographically minimal subset with respect to } <_i \text{ such that } \Delta_{I(\lambda)} \neq 0\}.$$

2B. *The positroid stratification of* $\mathbf{Gr}_{k,n}$. The *positroid stratification* of the Grassmannian $\mathrm{Gr}_{k,n}$ is obtained by taking the simultaneous refinement of the n Schubert decompositions with respect to the n shifted linear orders $<_i$. This stratification was first considered by Postnikov [2006], who showed that the strata are conveniently described in terms of *Grassmann necklaces*, as well as *decorated permutations* and *Ⅎ-diagrams*. Postnikov coined the terminology *positroid* because the intersection of the positroid stratification of the real Grassmannian with the *totally nonnegative part of the Grassmannian* $(\mathrm{Gr}_{k,n})_{\geq 0}$ gives a cell decomposition of $(\mathrm{Gr}_{k,n})_{\geq 0}$ (whose cells are called *positroid cells*).

Definition 2.3 [Postnikov 2006, Definition 16.1]. A *Grassmann necklace* is a sequence $\mathcal{I} = (I_1, \ldots, I_n)$ of subsets $I_r \subset [n]$ such that, for $i \in [n]$, if $i \in I_i$ then $I_{i+1} = (I_i \setminus \{i\}) \cup \{j\}$, for some $j \in [n]$ (j may coincide with i); and if $i \notin I_i$ then $I_{i+1} = I_i$. (Here indices i are taken modulo n.) In particular, we have $|I_1| = \cdots = |I_n|$, which is equal to some $k \in [n]$. We then say that $\mathcal{I}$ is a Grassmann necklace of *type* (k, n).

Example 2.4. $\mathcal{I} = (1345, 3456, 3456, 4567, 4567, 1467, 1478, 1348)$ is an example of a Grassmann necklace of type $(4, 8)$.

+	0	+	+
+	0	+	
0	0	+	
+	0		

Figure 2. A ⅃-diagram $L = (\lambda, D)_{k,n}$.

Lemma 2.5 [Postnikov 2006, Lemma 16.3]. *For $A \in \mathrm{Gr}_{k,n}$, let $\mathcal{I}(A) = (I_1, \dots, I_n)$ be the sequence of subsets in $[n]$ such that, for $i \in [n]$, I_i is the lexicographically minimal subset of $\binom{[n]}{k}$ with respect to the shifted linear order $<_i$ such that $\Delta_{I_i}(A) \neq 0$. Then $\mathcal{I}(A)$ is a Grassmann necklace of type (k, n).*

The *positroid stratification* of $\mathrm{Gr}_{k,n}$ is defined as follows.

Definition 2.6. Let $\mathcal{I} = (I_1, \dots, I_n)$ be a Grassmann necklace of type (k, n). The *positroid stratum* $S_{\mathcal{I}}$ is defined to be

$$S_{\mathcal{I}} = \{A \in \mathrm{Gr}_{k,n} \mid \mathcal{I}(A) = \mathcal{I}\}.$$

Equivalently, each positroid stratum is an intersection of n cyclically shifted Schubert cells, that is,

$$S_{\mathcal{I}} = \bigcap_{i=1}^{n} \Omega^{i}_{\lambda(I_i)}.$$

Grassmann necklaces are in bijection with tableaux called ⅃-*diagrams*.

Definition 2.7 [Postnikov 2006, Definition 6.1]. Fix k, n. A ⅃-*diagram* $(\lambda, D)_{k,n}$ of type (k, n) is a partition λ contained in a $k \times (n-k)$ rectangle together with a filling $D\colon \lambda \to \{0, +\}$ of its boxes which has the ⅃-*property*: there is no 0 which has a + above it and a + to its left.[1] (Here, "above" means above and in the same column, and "to its left" means to the left and in the same row.)

In Figure 2 we give an example of a ⅃-diagram.

2C. *The matroid stratification of* $\mathrm{Gr}_{k,n}$.

Definition 2.8. A *matroid* of *rank* k on the set $[n]$ is a nonempty collection $\mathcal{M} \subset \binom{[n]}{k}$ of k-element subsets in $[n]$, called *bases* of $\mathcal{M}$, that satisfies the *exchange axiom*: For any $I, J \in \mathcal{M}$ and $i \in I$ there exists $j \in J$ such that $(I \setminus \{i\}) \cup \{j\} \in \mathcal{M}$.

Given an element $A \in \mathrm{Gr}_{k,n}$, there is an associated matroid $\mathcal{M}_A$ whose bases are the k-subsets $I \subset [n]$ such that $\Delta_I(A) \neq 0$.

[1] This forbidden pattern is in the shape of a backwards L, and hence is denoted ⅃, pronounced "Le."

Definition 2.9. Let $\mathcal{M} \subset \binom{[n]}{k}$ be a matroid. Define the *matroid stratum* $S_{\mathcal{M}}$ as

$$S_{\mathcal{M}} = \{A \in \mathrm{Gr}_{k,n} \mid \Delta_I(A) \neq 0 \text{ if and only if } I \in \mathcal{M}\}.$$

This gives a stratification of $\mathrm{Gr}_{k,n}$ called the *matroid stratification*, or *Gelfand–Serganova stratification*.

Remark 2.10. Clearly the matroid stratification refines the positroid stratification, which in turn refines the Schubert decomposition.

3. The main result: network parametrizations from Go-diagrams

In this section we define certain tableaux called *Go-diagrams*, then explain how to parametrize the Grassmannian using networks associated to Go-diagrams. First we will define more general tableaux called *diagrams*.

3A. *Diagrams and networks.*

Definition 3.1. Let λ be a partition contained in a $k \times (n-k)$ rectangle. A *diagram* in λ is an arbitrary filling of the boxes of λ with pluses $+$, black stones $\bullet$, and white stones $\circ$.

To each diagram D we associate a network N_D as follows.

Definition 3.2. Let λ be a partition with ℓ boxes contained in a $k \times (n-k)$ rectangle, and let D be a diagram in λ. Label the boxes of λ from 1 to ℓ, starting from the rightmost box in the bottom row, then reading right to left across the bottom row, then right to left across the row above that, etc. The *(weighted) network* N_D associated to D is a directed graph obtained as follows:

- Associate an *internal vertex* to each $+$ and each $\bullet$.
- After labeling the southeast border of the Young diagram with the numbers $1, 2, \ldots, n$ (from northeast to southwest), associate a *boundary vertex* to each number.
- From each internal vertex, draw an edge right to the nearest $+$-vertex or boundary vertex.
- From each internal vertex, draw an edge down to the nearest $+$-vertex or boundary vertex.
- Direct all edges left and down. After doing so, k of the boundary vertices become *sources* and the remaining $n-k$ boundary vertices become *sinks*.
- If e is a horizontal edge whose left vertex is a $+$-vertex (respectively $\bullet$-vertex) in box b, assign e the weight a_b (respectively c_b). We think of a_b and c_b as indeterminates, but later they will be elements of $\mathbb{K}^*$ and $\mathbb{K}$, respectively.
- If e is a vertical edge, assign e the weight 1.

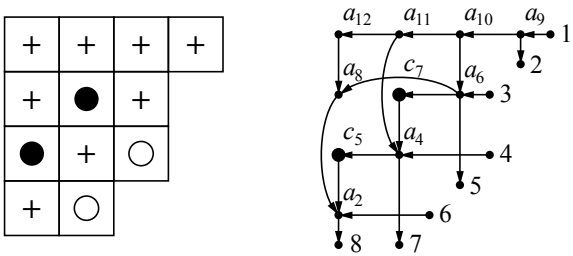


Figure 3. An example of a diagram and its associated network.

Note that in general such a directed graph is not planar, as two edges may cross over each other without meeting at a vertex. See Figure 3 for an example of a diagram and its associated network.

We now explain how to associate a *weight matrix* to such a network.

Definition 3.3. Let N_D be a network as in Definition 3.2. Let

$$I = \{i_1 < i_2 < \cdots < i_k\} \subset [n]$$

denote the sources. If P is a directed path in the network, let $w(P)$ denote the product of all weights along P. If P is the empty path which starts and ends at the same boundary vertex, we let $w(P) = 1$. If s is a source and t is any boundary vertex, define

$$W_{st} = \pm \sum_P w(P),$$

where the sum is over all paths P from s to t. The sign is chosen (uniquely) so that

$$\Delta_{I \setminus \{s\} \cup \{t\}}(W_D) = \sum_P w(P),$$

where

$$W_D = (W_{st})$$

is the $k \times n$ *weight matrix*. We make the convention that the rows of W_D are indexed by the sources $i_1, \ldots, i_k$ from top to bottom, and its columns are indexed by $1, 2, \ldots, n$ from left to right. An equivalent way to define the sign of W_{st} is to let $q = |\{s+1, s+2, \ldots, t-1\} \cap I|$, that is, the number of sources which are strictly between s and t. Then the sign of W_{st} is $(-1)^q$.

Example 3.4. The weight matrix associated to the network in Figure 3 is

$$\begin{pmatrix} 1 & a_9 & 0 & 0 & a_9a_{10} & 0 & -a_9a_{10}(a_{11}+c_7) & -a_9a_{10}(a_{11}a_{12}+a_{11}c_5+a_8+c_7c_5) \\ 0 & 0 & 1 & 0 & -a_6 & 0 & a_6c_7 & a_6a_8+a_6c_7c_5 \\ 0 & 0 & 0 & 1 & 0 & 0 & -a_4 & -a_4c_5 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & a_2 \end{pmatrix}.$$

3B. *Distinguished expressions.* We now review the notion of distinguished subexpressions, as in [Deodhar 1985] and [Marsh and Rietsch 2004]. This definition will be essential for defining Go-diagrams. We assume the reader is familiar with the (strong) Bruhat order $<$ on $W = \mathfrak{S}_n$, and the basics of reduced expressions, as in [Björner and Brenti 2005].

Let $\boldsymbol{w} := s_{i_1} \cdots s_{i_m}$ be a reduced expression for $w \in W$. A *subexpression* $\boldsymbol{v}$ of $\boldsymbol{w}$ is a word obtained from the reduced expression $\boldsymbol{w}$ by replacing some of the factors with 1. For example, consider a reduced expression in $\mathfrak{S}_4$, say $s_3 s_2 s_1 s_3 s_2 s_3$. Then $s_3 s_2\, 1\, s_3 s_2\, 1$ is a subexpression of $s_3 s_2 s_1 s_3 s_2 s_3$. Given a subexpression $\boldsymbol{v}$, we set $v_{(k)}$ to be the product of the leftmost k factors of $\boldsymbol{v}$, if $k \geq 1$, and $v_{(0)} = 1$.

Definition 3.5 [Marsh and Rietsch 2004; Deodhar 1985]. Given a subexpression $\boldsymbol{v}$ of a reduced expression $\boldsymbol{w} = s_{i_1} s_{i_2} \cdots s_{i_m}$, we define

$$J_{\boldsymbol{v}}^{\circ} := \{k \in \{1, \dots, m\} \mid v_{(k-1)} < v_{(k)}\},$$
$$J_{\boldsymbol{v}}^{+} := \{k \in \{1, \dots, m\} \mid v_{(k-1)} = v_{(k)}\},$$
$$J_{\boldsymbol{v}}^{\bullet} := \{k \in \{1, \dots, m\} \mid v_{(k-1)} > v_{(k)}\}.$$

The expression $\boldsymbol{v}$ is called *nondecreasing* if $v_{(j-1)} \leq v_{(j)}$ for all $j = 1, \dots, m$, for example, if $J_{\boldsymbol{v}}^{\bullet} = \varnothing$.

Definition 3.6 (distinguished subexpressions [Deodhar 1985, Definition 2.3]). A subexpression $\boldsymbol{v}$ of $\boldsymbol{w}$ is called *distinguished* if we have

$$v_{(j)} \leq v_{(j-1)}\, s_{i_j} \quad \text{for all } j \in \{1, \dots, m\}. \tag{3-1}$$

In other words, if right multiplication by s_{i_j} decreases the length of $v_{(j-1)}$, then in a distinguished subexpression we must have $v_{(j)} = v_{(j-1)} s_{i_j}$.

We write $\boldsymbol{v} \prec \boldsymbol{w}$ if $\boldsymbol{v}$ is a distinguished subexpression of $\boldsymbol{w}$.

Definition 3.7 (positive distinguished subexpressions). We call a subexpression $\boldsymbol{v}$ of $\boldsymbol{w}$ a *positive distinguished subexpression* (or a PDS for short) if

$$v_{(j-1)} < v_{(j-1)} s_{i_j} \quad \text{for all } j \in \{1, \dots, m\}. \tag{3-2}$$

In other words, it is distinguished and nondecreasing.

Lemma 3.8 [Marsh and Rietsch 2004]. *Given $v \leq w$ and a reduced expression $\boldsymbol{w}$ for w, there is a unique PDS $\boldsymbol{v}_+$ for v in $\boldsymbol{w}$.*

3C. *Go-diagrams.* In this section we explain how to index distinguished subexpressions by certain tableaux called *Go-diagrams*, which were introduced in [Kodama and Williams 2013]. Go-diagrams are certain fillings of Young diagrams by pluses $+$, *black stones* $\bullet$, and *white stones* $\circ$.[2]

[2] In [Kodama and Williams 2013], we used a slightly different convention and used blank boxes in place of $+$'s.

S_5	S_4	S_3	S_2	S_1
S_6	S_5	S_4	S_3	S_2
S_7	S_6	S_5	S_4	S_3

15	14	13	12	11
10	9	8	7	6
5	4	3	2	1

15	12	9	6	3
14	11	8	5	2
13	10	7	4	1

Figure 4. The labeling of a the boxes of a partition by simple generators s_i, and two reading orders.

Fix k and n. Let $W_k = \langle s_1, s_2, \ldots, \hat{s}_{n-k}, \ldots, s_{n-1}\rangle$ be a parabolic subgroup of $W = \mathfrak{S}_n$. Let W^k denote the set of minimal-length coset representatives of W/W_k. Recall that a *descent* of a permutation π is a position j such that $\pi(j) > \pi(j+1)$. Then W^k is the subset of permutations of $\mathfrak{S}_n$ which have at most one descent; and that descent must be in position $n-k$.

It is well known that elements w of W^k can be identified with partitions λ_w contained in a $k \times (n-k)$ rectangle: if $w = (w_1, \ldots, w_n) \in W^k$ then

$$\{w_{n-k+1}, w_{n-k+2}, \ldots, w_n\}$$

is a subset of size k, which gives rise to a partition, as described at the beginning of Section 2A. We refer to this partition as λ_w.

Moreover, it follows from [Stembridge 1996] and [Proctor 1984] that the reduced expressions of $w \in W^k$ correspond to certain *reading orders* of the boxes of the partition λ_w. Specifically, let Q^k be the poset whose elements are the boxes of a $k \times (n-k)$ rectangle; if b_1 and b_2 are two adjacent boxes such that b_2 is immediately to the left or immediately above b_1, we have a cover relation $b_1 \lessdot b_2$ in Q^k. The partial order on Q^k is the transitive closure of $\lessdot$. Now label the boxes of the rectangle with simple generators s_i as in Figure 4. If b is a box of the rectangle, then let s_b denote its label by a simple generator. Let $w_0^k \in W^k$ denote the longest element in W^k. Then the set of reduced expressions of w_0^k can be obtained by choosing a linear extension of Q^k and writing down the corresponding word in the s_i's. We call such a linear extension a *reading order*; two linear extensions are shown in Figure 4. Additionally, given a partition λ_w contained in the $k \times (n-k)$ rectangle (chosen so that the upper-left corner of its Young diagram is aligned with the upper-left corner of the rectangle), and a linear extension of the subposet of Q^k comprised of the boxes of λ, the corresponding word in s_i's is a reduced expression of the permutation $w \in W^k$. Moreover, all reduced expressions of w can be obtained by varying the linear extension.

Definition 3.9 [Kodama and Williams 2013, Section 4]. Fix k and n. Let $w \in W^k$, let $\boldsymbol{w}$ be a reduced expression for w, and let $\boldsymbol{v}$ be a distinguished subexpression of $\boldsymbol{w}$. Then w and $\boldsymbol{w}$ determine a partition λ_w contained in a $k \times (n-k)$ rectangle together with a reading order of its boxes. The *Go-diagram* associated to $\boldsymbol{v}$ and $\boldsymbol{w}$

is a filling of λ_w with pluses and black and white stones, such that: for each $k \in J_v^\circ$ we place a white stone in the corresponding box; for each $k \in J_v^\bullet$ we place a black stone in the corresponding box of λ_w; and for each $k \in J_v^+$ we place a plus in the corresponding box of λ_w.

Remark 3.10. By [Kodama and Williams 2013, Section 4], whether or not a filling of a partition λ_w is a Go-diagram does not depend on the choice of reading order of the boxes of λ_w.

Definition 3.11. We define the *standard reading order* of the boxes of a partition to be the reading order which starts at the rightmost box in the bottom row, then reads right to left across the bottom row, then right to left across the row above that, then right to left across the row above that, etc. This reading order is illustrated at the right of the figure below.

By default, we will use the standard reading order in this paper.

Example 3.12. Let $k = 3$ and $n = 7$, and let $\lambda = (4, 3, 1)$. The standard reading order is shown at the right of the figure below.

s_4	s_3	s_2	s_1
s_5	s_4	s_3	
s_6			

8	7	6	5
4	3	2	
1			

Then the following diagrams are Go-diagrams of shape λ.

○	○	○	○
○	○	○	
○			

+	○	○	+
+	○	+	
○			

●	+	+	○
+	○	○	
+			

They correspond to the expressions

$$s_6 s_3 s_4 s_5 s_1 s_2 s_3 s_4, \quad s_6 1 s_4 1 1 s_2 s_3 1, \quad 1 s_3 s_4 1 s_1 1 1 s_4.$$

The first and second are positive distinguished subexpressions (PDS's), and the third one is a distinguished subexpression (but not a PDS).

Note that the following diagram of shape λ is not a Go-diagram. It corresponds to the word $11 s_4 1 s_1 s_2 11$, which is not distinguished.

+	+	○	○
+	○	+	
+			

Remark 3.13. The Go-diagrams associated to PDS's are in bijection with ɹ-diagrams; see [Kodama and Williams 2013, Section 4]. Note that the Go-diagram associated to a PDS contains only pluses and white stones. This is precisely a ɹ-diagram.

If we choose a reading order of λ_w, then we will also associate to a Go-diagram of shape λ_w a *labeled Go-diagram*, as defined below. Equivalently, a labeled Go-diagram is associated to a pair $(\boldsymbol{v}, \boldsymbol{w})$.

Definition 3.14 [Kodama and Williams 2013, Definition 4.15]. Given a reading order of λ_w and a Go-diagram of shape λ_w, we obtain a *labeled Go-diagram* by replacing each $\circ$ with a 1, replacing each box b containing a $\bullet$ with a -1 *and* an m_i, and replacing each box b containing a $+$ by a p_i, where the subscript i corresponds to the label of b inherited from the reading order.

The labeled Go-diagrams corresponding to the examples above using the standard reading order are:

1	1	1	1
1	1	1	
1			

p_8	1	1	p_5
p_4	1	p_2	
1			

$-1 / m_8$	p_7	p_6	1
p_4	1	1	
p_1			

3D. ***The main result.*** To state the main result, we now consider Go-diagrams (not arbitrary diagrams), the corresponding networks (*Go-networks*), and the corresponding weight matrices.

Definition 3.15. Let D be a Go-diagram contained in a $k \times (n-k)$ rectangle. We define a subset $\mathcal{R}_D$ of the Grassmannian $\mathrm{Gr}_{k,n}$ by letting each variable a_i of the weight matrix (Definition 3.3) range over all nonzero elements $\mathbb{K}^*$, and letting each variable c_i of the weight matrix range over all elements $\mathbb{K}$. We call $\mathcal{R}_D$ the *network component associated to* D.

Theorem 3.16. *Let D be a Go-diagram contained in a $k \times (n-k)$ rectangle. Suppose that D has t pluses and u black stones. Then $\mathcal{R}_D$ is isomorphic to $(\mathbb{K}^*)^t \times \mathbb{K}^u$. Furthermore, $\mathrm{Gr}_{k,n}$ is the disjoint union of the network components $\mathcal{R}_D$, as D ranges over all Go-diagrams contained in a $k \times (n-k)$ rectangle. In other words, each point in the Grassmannian $\mathrm{Gr}_{k,n}$ can be represented uniquely by a weighted network associated to a Go-diagram.*

A more refined version of Theorem 3.16 is given in Corollary 6.9.

Corollary 3.17. *Every $k \times \ell$ matrix M can be represented by a unique weighted network associated to a Go-diagram contained in a $k \times \ell$ rectangle.*

Proof. Let $n = k + \ell$. Suppose $M = (m_{i,j})$ is a $k \times (n-k)$ matrix. Let $A(M) = (a_{i,j})$ be the full rank $k \times n$ matrix with an identity submatrix in the first k columns and the remaining columns given by

$$a_{i,j+k} = (-1)^{i+1} m_{n+1-i,j}.$$

Then $A(M)$ represents an element in the Grassmannian $\mathrm{Gr}_{k,n}$, so Theorem 3.16 applies. The minors of M are in bijection with the $k \times k$ minors of $A(M)$, so if $A(M)$ is represented by the network N, we see that $m_{i,j}$ enumerates paths from the boundary source i to the boundary vertex j in N. □

We will prove Theorem 3.16 by showing that each network component $\mathcal{R}_D$ from a Go-diagram coincides with a (*projected*) *Deodhar component* $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$ in the Grassmannian. (Therefore we may refer to each $\mathcal{R}_D$ as a *Deodhar component*.) More specifically, such Deodhar components have parametrizations due to Marsh and Rietsch [2004], and we will show that after an invertible transformation of variables, our network parametrizations coincide with theirs.

4. The Deodhar decomposition of the Grassmannian

In this section we review Deodhar's decomposition of the flag variety G/B [Deodhar 1985], and the parametrizations of the components due to [Marsh and Rietsch 2004]. The Deodhar decomposition of the Grassmannian is obtained by projecting the Deodhar decomposition of G/B to the Grassmannian [Deodhar 1987].

4A. ***The flag variety.*** Let $\mathbb{K}$ be a field, and let G denote the special linear group $\mathrm{SL}_n = \mathrm{SL}_n(\mathbb{K})$. Fix a maximal torus T and opposite Borel subgroups B^+ and B^-; thus T, B^+, B^- consist respectively of the diagonal, upper-triangular, and lower-triangular matrices in SL_n. Let U^+ and U^- be the unipotent radicals of B^+ and B^-; these are the subgroups of upper-triangular and lower-triangular matrices with 1's on the diagonals. For each $1 \le i \le n-1$ we have a homomorphism $\phi_i : \mathrm{SL}_2 \to \mathrm{SL}_n$ such that

$$\phi_i \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & a & b & \\ & & c & d & \\ & & & & \ddots \\ & & & & & 1 \end{pmatrix} \in \mathrm{SL}_n;$$

that is, ϕ_i replaces a 2×2 block of the identity matrix with $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Here a is at the $(i+1)$-st diagonal entry counting from the southeast corner. (Correspondingly, we will label the rows of such a matrix from bottom to top, and the columns of such a

matrix from right to left.) We have 1-parameter subgroups of G defined by

$$x_i(m) = \phi_i \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad y_i(m) = \phi_i \begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix}, \quad \text{where } m \in \mathbb{K}.$$

Let W denote the Weyl group $N_G(T)/T$, where $N_G(T)$ is the normalizer of T. The simple reflections $s_i \in W$ are given by $s_i := \dot{s}_i T$, where $\dot{s}_i := \phi_i \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and any $w \in W$ can be expressed as a product $w = s_{i_1} s_{i_2} \cdots s_{i_\ell}$ with $\ell = \ell(w)$ factors. We set $\dot{w} = \dot{s}_{i_1} \dot{s}_{i_2} \cdots \dot{s}_{i_\ell}$. In our setting W is isomorphic to $\mathfrak{S}_n$, the symmetric group on n letters, and s_i corresponds to the transposition exchanging i and $i+1$.

We can identify the flag variety G/B with the variety $\mathcal{B}$ of Borel subgroups via

$$gB \longleftrightarrow g \cdot B^+ := gB^+g^{-1}.$$

We have two opposite Bruhat decompositions of $\mathcal{B}$:

$$\mathcal{B} = \bigsqcup_{w \in W} B^+ \dot{w} \cdot B^+ = \bigsqcup_{v \in W} B^- \dot{v} \cdot B^+.$$

We define the intersection of opposite Bruhat cells

$$\mathcal{R}_{v,w} := B^+ \dot{w} \cdot B^+ \cap B^- \dot{v} \cdot B^+,$$

which is nonempty precisely when $v \leq w$. The strata $\mathcal{R}_{v,w}$ are often called *Richardson varieties*.

4B. ***Deodhar components in the flag variety.*** We now describe the Deodhar decomposition of the flag variety. Marsh and Rietsch [2004] gave explicit parametrizations for each Deodhar component, identifying each one with a subset in the group.

Definition 4.1 [Marsh and Rietsch 2004, Definition 5.1]. Let $\boldsymbol{w} = s_{i_1} \cdots s_{i_m}$ be a reduced expression for w, and let $\boldsymbol{v}$ be a distinguished subexpression. Define a subset $G_{\boldsymbol{v},\boldsymbol{w}}$ in G by

$$G_{\boldsymbol{v},\boldsymbol{w}} := \left\{ g_1 g_2 \cdots g_m \;\middle|\; \begin{array}{lll} g_\ell = x_{i_\ell}(m_\ell)\dot{s}_{i_\ell}^{-1} & \text{for some } m_l \in \mathbb{K} & \text{if } \ell \in J_{\boldsymbol{v}}^{\bullet}, \\ g_\ell = y_{i_\ell}(p_\ell) & \text{for some } p_l \in \mathbb{K}^* & \text{if } \ell \in J_{\boldsymbol{v}}^{+}, \\ g_\ell = \dot{s}_{i_\ell} & & \text{if } \ell \in J_{\boldsymbol{v}}^{\circ}. \end{array} \right\}.$$

There is an obvious map $(\mathbb{K}^*)^{|J_{\boldsymbol{v}}^+|} \times \mathbb{K}^{|J_{\boldsymbol{v}}^\bullet|} \to G_{\boldsymbol{v},\boldsymbol{w}}$ defined by the parameters p_ℓ and m_ℓ. For $v = w = 1$ we define $G_{\boldsymbol{v},\boldsymbol{w}} = \{1\}$.

Example 4.2. Let $W = \mathfrak{S}_8$, $\boldsymbol{w} = s_6 s_7 s_4 s_5 s_6 s_3 s_4 s_5 s_1 s_2 s_3 s_4$ and $\boldsymbol{v} = s_6 1 s_4 1 s_6 1 s_4 11111$. This is the distinguished expression $\boldsymbol{v}$ encoded by the diagram from Figure 3 (which is a Go-diagram). Then the corresponding element $g \in G_{\boldsymbol{v},\boldsymbol{w}}$ (the MR-matrix) is

given by

$$g = \dot{s}_6 y_7(p_2)\dot{s}_4 y_5(p_4) x_6(m_5)\dot{s}_6^{-1} y_3(p_6) x_4(m_7)\dot{s}_4^{-1} \cdot y_5(p_8) y_1(p_9) y_2(p_{10}) y_3(p_{11}) y_4(p_{12}), \tag{4-1}$$

which is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & -m_5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_8 & 1 & 0 & 0 & 0 & 0 \\ 0 & -p_4 & -m_7 p_8 & -m_7+p_{12} & 1 & 0 & 0 & 0 \\ 0 & 0 & -p_6 p_8 & -p_6+p_{11}p_{12} & p_{11} & 1 & 0 & 0 \\ 0 & 0 & 0 & p_{10}p_{11}p_{12} & p_{10}p_{11} & p_{10} & 1 & 0 \\ 0 & 0 & 0 & p_9 p_{10} p_{11} p_{12} & p_9 p_{10} p_{11} & p_9 p_{10} & p_9 & 1 \end{pmatrix}. \tag{4-2}$$

The following result gives an explicit parametrization for the Deodhar component $\mathcal{R}_{\boldsymbol{v},\boldsymbol{w}}$. We use Proposition 4.3 as the *definition* of $\mathcal{R}_{\boldsymbol{v},\boldsymbol{w}}$.

Proposition 4.3 [Marsh and Rietsch 2004, Proposition 5.2]. *The map*

$$(\mathbb{K}^*)^{|J_{\boldsymbol{v}}^+|} \times \mathbb{K}^{|J_{\boldsymbol{v}}^\bullet|} \to G_{\boldsymbol{v},\boldsymbol{w}}$$

from Definition 4.1 is an isomorphism. The map $g \mapsto g \cdot B^+$ defines an isomorphism

$$G_{\boldsymbol{v},\boldsymbol{w}} \xrightarrow{\sim} \mathcal{R}_{\boldsymbol{v},\boldsymbol{w}} \tag{4-3}$$

between the subset $G_{\boldsymbol{v},\boldsymbol{w}}$ of the group and the Deodhar component $\mathcal{R}_{\boldsymbol{v},\boldsymbol{w}}$ in G/B.

Suppose that for each $w \in W$ we choose a reduced expression $\boldsymbol{w}$ for w. Then it follows from [Deodhar 1985] and [Marsh and Rietsch 2004, Section 4.4] that

$$\mathcal{R}_{v,w} = \bigsqcup_{\boldsymbol{v} \prec \boldsymbol{w}} \mathcal{R}_{\boldsymbol{v},\boldsymbol{w}} \quad \text{and} \quad G/B = \bigsqcup_{w \in W} \Big(\bigsqcup_{\boldsymbol{v} \prec \boldsymbol{w}} \mathcal{R}_{\boldsymbol{v},\boldsymbol{w}} \Big), \tag{4-4}$$

where in the first sum $\boldsymbol{v}$ ranges over all distinguished subexpressions for v in w, and in the second sum $\boldsymbol{v}$ ranges over all distinguished subexpressions of $\boldsymbol{w}$. These two decompositions are called the *Deodhar decompositions* of $\mathcal{R}_{v,w}$ and G/B.

Remark 4.4. Although the Deodhar decomposition of $\mathcal{R}_{v,w}$ depends on the choice $\boldsymbol{w}$ of reduced expression for w, its projection to the Grassmannian does not depend on $\boldsymbol{w}$ [Kodama and Williams 2013, Proposition 4.16].

4C. ***Projections of Deodhar components to the Grassmannian.*** Following [Kodama and Williams 2013], we now consider the projection of the Deodhar decomposition to the Grassmannian $\mathrm{Gr}_{k,n}$ for $k < n$. Given the permutation

$$w = \big(w(1), w(2), \dots, w(n)\big) \in W^k,$$

we let

$$I(w) := \{w(n-k+1), w(n-k+2), \dots, w(n)\},$$

a k-element subset of $[n]$. The map I gives a bijection between W^k and k-element subsets of $[n]$.

Let $\pi_k \colon G/B \to \mathrm{Gr}_{k,n}$ be the projection from the flag variety to the Grassmannian; this is an isomorphism on each $\mathcal{R}_{v,w}$. For each $w \in W^k$ and $v \le w$, define $\mathcal{P}_{v,w} = \pi_k(\mathcal{R}_{v,w})$. Then by [Lusztig 1998] we have a decomposition

$$\mathrm{Gr}_{k,n} = \bigsqcup_{w \in W^k} \Big(\bigsqcup_{v \le w} \mathcal{P}_{v,w} \Big). \tag{4-5}$$

For each reduced decomposition $\boldsymbol{w}$ for $w \in W^k$, and each $\boldsymbol{v} \prec \boldsymbol{w}$, we define $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}} = \pi_k(\mathcal{R}_{\boldsymbol{v},\boldsymbol{w}})$. Now if for each $w \in W^k$ we choose a reduced decomposition $\boldsymbol{w}$, then we have

$$\mathcal{P}_{v,w} = \bigsqcup_{\boldsymbol{v} \prec \boldsymbol{w}} \mathcal{P}_{\boldsymbol{v},\boldsymbol{w}} \quad \text{and} \quad \mathrm{Gr}_{k,n} = \bigsqcup_{w \in W^k} \Big(\bigsqcup_{\boldsymbol{v} \prec \boldsymbol{w}} \mathcal{P}_{\boldsymbol{v},\boldsymbol{w}} \Big), \tag{4-6}$$

where in the first sum $\boldsymbol{v}$ ranges over all distinguished subexpressions for v in w, and in the second sum $\boldsymbol{v}$ ranges over all distinguished subexpressions of $\boldsymbol{w}$.

Proposition 4.3 gives a concrete way to think about the projected Deodhar components $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$. The projection $\pi_k \colon G/B \to \mathrm{Gr}_{k,n}$ maps $g \cdot B^+ \in R_{\boldsymbol{v},\boldsymbol{w}}$ (where $g \in G_{\boldsymbol{v},\boldsymbol{w}}$) to the span of the leftmost k columns of g. More specifically, it maps

$$g = \begin{pmatrix} g_{n,n} & \cdots & g_{n,n-k+1} & \cdots & g_{n,1} \\ \vdots & & \vdots & & \vdots \\ g_{1,n} & \cdots & g_{1,n-k+1} & \cdots & g_{1,1} \end{pmatrix} \to M = \begin{pmatrix} g_{1,n-k+1} & \cdots & g_{n,n-k+1} \\ \vdots & & \vdots \\ g_{1,n} & \cdots & g_{n,n} \end{pmatrix}$$

We call the resulting $k \times n$ matrix $M = (M_{st})$ the *MR-matrix*. To simplify the notation later, we will label its rows from top to bottom by $i_1, i_2, \dots, i_k$, where $\{i_1 < \cdots < i_k\} = I(w)$.

Remark 4.5. Recall from Section 3C that in the Grassmannian setting (that is, $W_k = \langle s_1, s_2, \dots, \hat{s}_{n-k}, \dots, s_{n-1} \rangle$ is a parabolic subgroup of $W = \mathfrak{S}_n$), the distinguished subexpressions of W^k are in bijection with Go-diagrams. Therefore each Go-diagram gives rise to an MR-matrix.

Example 4.6. We continue Example 4.2. Note that $w \in W^k$, where $k = 2$. Then the map $\pi_2 \colon G_{\boldsymbol{v},\boldsymbol{w}} \to \mathrm{Gr}_{2,5}$ is given by

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & -m_5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_8 & 1 & 0 & 0 & 0 & 0 \\ 0 & -p_4 & -m_7 p_8 & -m_7+p_{12} & 1 & 0 & 0 & 0 \\ 0 & 0 & -p_6 p_8 & -p_6+p_{11}p_{12} & p_{11} & 1 & 0 & 0 \\ 0 & 0 & 0 & p_{10}p_{11}p_{12} & p_{10}p_{11} & p_{10} & 1 & 0 \\ 0 & 0 & 0 & p_9 p_{10} p_{11} p_{12} & p_9 p_{10} p_{11} & p_9 p_{10} & p_9 & 1 \end{pmatrix} \longrightarrow$$

$$M = \begin{pmatrix} p_9 p_{10} p_{11} p_{12} & p_{10} p_{11} p_{12} & -p_6+p_{11}p_{12} & -m_7+p_{12} & 1 & 0 & 0 & 0 \\ 0 & 0 & -p_6 p_8 & -m_7 p_8 & p_8 & 1 & 0 & 0 \\ 0 & 0 & 0 & -p_4 & 0 & -m_5 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_2 & 0 & 1 \end{pmatrix}$$

We label the rows of M from top to bottom by the index set $\{1, 3, 4, 6\}$, and the columns from left to right by the index set $\{1, 2, \dots, 8\}$, so for example $M_{34} = -m_7 p_8$.

The following lemma is a consequence of Section 5.1 (and Corollary 5.8 in particular) in [Kodama and Williams 2013].

Lemma 4.7. *Let $M = M_D$ be the MR-matrix associated to the diagram D. The leftmost nonzero entry in row i_ℓ of M is in column i_ℓ. Furthermore, that entry is equal to $(-1)^b \prod p_i$, where b is the number of black stones in the row i_ℓ of D, and the product is over all boxes in the row i_ℓ of the labeled Go-diagram of D containing a p_i.*

5. Formulas for entries of the MR-matrices

In this section we consider arbitrary diagrams (not necessarily Go-diagrams) contained in a $k \times (n-k)$ rectangle and the corresponding MR-matrices, obtained by multiplying factors $\dot{s}_i$, $y_i(p_j)$, $x_i(m_j)\dot{s}_i^{-1}$ as specified by the filling of the diagram, and then projecting the resulting $n \times n$ matrix to a $k \times n$ matrix. We will give formulas for the entries of the MR-matrices in terms of *pseudopaths* in the corresponding network. For the purpose of giving this formula, we will replace weights a_i and c_j on the edges of the network by weights p_i and m_j.

Recall that if D is a diagram, its network N_D has three types of vertices: $+$-vertices, $\bullet$-vertices, and boundary vertices. A *step* on a network is an edge between two vertices. Let W denote a single step west, S denote a single step south, and E denote either a single step east, or an *east-west combination step* consisting of a step east, followed by a step west ending at a $\bullet$. Let A^* indicate 0 or more instances of a step of type A.

Definition 5.1. A *pseudopath* $\widetilde{P}$ on a network is a path on the (undirected version of the) network such that:

- it starts and ends at two different boundary vertices, or else is the empty path from a boundary vertex to itself;
- it does not cross the same edge twice;
- its sequence of steps (for a nonempty path) has the form

$$WW^*S(EE^*S)^*E^*.$$

In particular, a pseudopath may not take two consecutive steps south.

Definition 5.2. The *weight* $w(\widetilde{P})$ of a pseudopath $\widetilde{P}$ in a network is a Laurent monomial in p_i's and m_j's, which is obtained by multiplying the following terms:

- $1/p_i$ for every step west along an edge weighted p_i;
- p_i for every step east along an edge weighted p_i which is preceded by a step east;
- m_j for every step west along an edge weighted m_j;
- $(-1)^{b+w}$, where b (respectively w) is the number of black (resp. white) stones that the pseudopath skips over in the horizontal (resp. vertical) direction when we superimpose the Go-diagram onto the network.

Example 5.3. In Figure 5, there are two pseudopaths from 1 to 4, with weights $1/(p_9p_{10}p_{11})$ and $-m_7/(p_9p_{10}p_{11}p_{12})$, and there is one pseudopath from 1 to 5, with weight $1/(p_9p_{10}p_{11}p_{12})$.

Definition 5.4. If M is an MR-matrix, we will let $\widetilde{M}$ denote the matrix obtained from M by rescaling rows so that the leftmost nonzero entry in each row is 1.

Definition 5.5. If D is a diagram contained in a $k \times (n-k)$ rectangle, then we let $i_1 < i_2 < \cdots < i_k$ denote the labels of the sources in the corresponding network. If

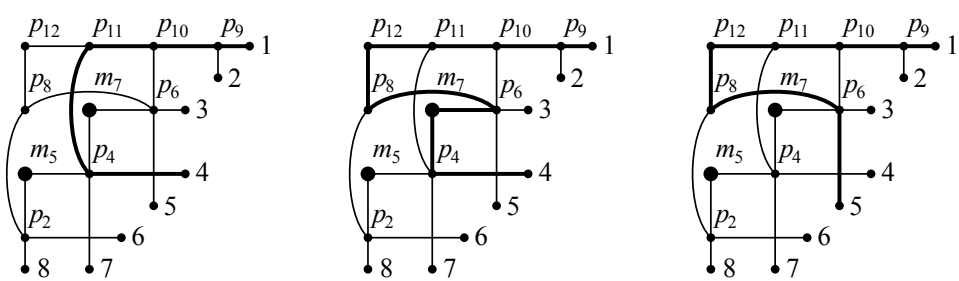


Figure 5. The two pseudopaths from 1 to 4 and the unique pseudopath from 1 to 5, indicated in bold. Note that the pseudopath in the middle figure contains an east-west combination step.

M and $\widetilde{M}$ are the corresponding $k \times n$ MR and rescaled MR-matrices associated to D, then we will index their rows by $i_1, \ldots, i_k$ from top to bottom and their columns by $1, 2, \ldots, n$ from left to right.

Theorem 5.6. *Let D be a diagram contained in a $k \times (n-k)$ rectangle, and let $\widetilde{M} = (\widetilde{M}_{st})$ be the corresponding $k \times n$ rescaled MR-matrix. Then*

$$\widetilde{M}_{st} = \sum_{\widetilde{P}} w(\widetilde{P}),$$

where the sum is over all pseudopaths from the source s to the boundary vertex t in the network.

Theorem 5.6 will follow from Theorem 5.10 and Lemma 5.11.

Example 5.7. The MR-matrix M from Example 4.6 corresponds to the network from Figure 5. The rows of M are indexed by $1, 3, 4, 6$ from top to bottom. Note that after we rescale the rows of M, obtaining $\widetilde{M}$, we have

$$\widetilde{M}_{14} = \frac{1}{p_9 p_{10} p_{11}} - \frac{m_7}{p_9 p_{10} p_{11} p_{12}} \quad \text{and} \quad \widetilde{M}_{15} = \frac{1}{p_9 p_{10} p_{11} p_{12}}.$$

This agrees with our pseudopath computation from Example 5.3.

Next we will give a formula for entries of MR-matrices, in terms of pseudopaths in *modified networks*.

Definition 5.8. Given a network N_D with k sources labeled $i_1, \ldots, i_k$ and n boundary vertices, we obtain from it a corresponding *modified network* N'_D by:

- adding k new boundary vertices to the left of N_D, labeled $i'_1, \ldots, i'_k$ from top to bottom;
- adding a new horizontal edge which connects i'_j to the nearest vertex of the network to its right.

See Figure 6 for the modified network associated to the network from Figure 5.

Definition 5.9. A *pseudopath* P on a modified network is a path on the modified network which:

- starts at one of the boundary vertices labeled $i'_1, \ldots, i'_k$, and ends at one of the boundary vertices labeled $1, 2, \ldots, n$;
- takes a sequence of steps which has the form

$$(EE^*S)^*E^*.$$

The arrows in Figure 6 indicate the allowed directions in which a path may travel.

The *weight* of a pseudopath in a modified network is defined the same way as the weight of a pseudopath in a network (see Definition 5.2). Note that since

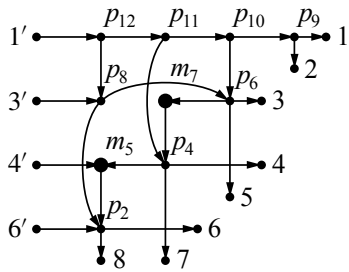


Figure 6. Example of a modified network.

a pseudopath in a modified network does not contain steps west along edges weighted p_i, its weight is a monomial (not a Laurent monomial).

Theorem 5.10. *Let D be a diagram contained in a $k \times (n-k)$ rectangle, and let $M = (M_{st})$ be the corresponding $k \times n$ MR-matrix. Then*

$$M_{st} = \sum_P w(P),$$

where the sum is over all pseudopaths in the modified network from the boundary vertex s' to the boundary vertex t.

Lemma 5.11. *Theorems 5.6 and 5.10 are equivalent.*

Proof. There is an obvious bijection between pseudopaths in a network starting at boundary vertex s, and pseudopaths in the corresponding modified network starting at boundary vertex s'. The weights of the corresponding pseudopaths are the same except for a factor of $(-1)^b \prod p_i$, where b is the number of $\bullet$-vertices in row s of the network, and the p_i's range over all edge weights in row s.

On the other hand, Lemma 4.7 implies that the leftmost nonzero entry of row s of the MR-matrix M is precisely the quantity $(-1)^b \prod p_i$ above. Therefore Theorems 5.6 and 5.10 are equivalent. □

By Lemma 5.11, in order to prove Theorem 5.6 it suffices to prove Theorem 5.10. Our strategy for proving Theorem 5.10 will be to interpret entries of the MR-matrix in terms of paths in a chip network, and then construct a weight-preserving bijection between these paths and between pseudopaths in the modified network.

Definition 5.12. A *chip* is one of the three configurations shown in Figure 7. We call the three configurations $y_i(p)$- or y_i-*chips*, s_i-*chips*, and $x_i(m)$- or x_i-*chips*, respectively.

Definition 5.13. A *chip network* is a concatenation of chips. Note that it has n boundary vertices at the left and n boundary vertices at the right. Let g be any product of factors of the form $y_i(p)$, $\dot{s}_i$, and $x_i(m)\dot{s}_i^{-1}$. We associate a chip network

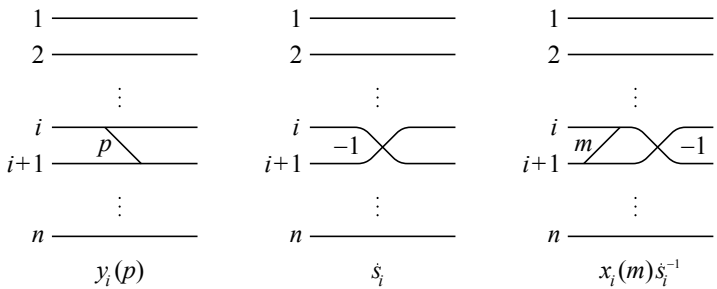


Figure 7. The three types of chips: y_i-chip (left), s_i-chip (center), and x_i-chip (right).

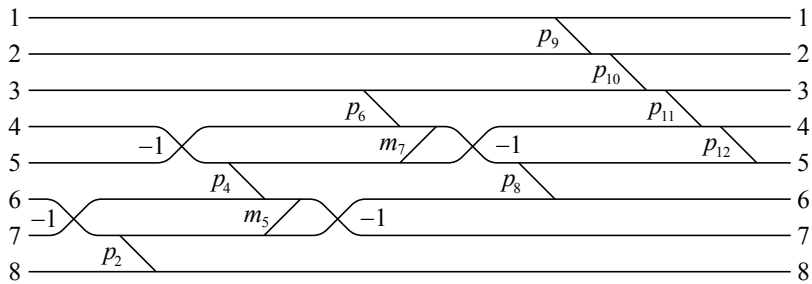
C_g to g by concatenating the chips corresponding to the factors of g in the order given by the factorization.

Example 5.14. The chip network C_g associated to the product g from (4-1) is shown in Figure 8.

Definition 5.15. A *route* Q in a chip network is a path in the network whose steps all travel east (or southeast or northeast for slanted edges). The *weight* $w(Q)$ of such a route is the product of all weights on its edges. To each chip network C we associate a *weight matrix* $x(C) = x_{ij}$, where $x_{ij} = \sum_Q w(Q)$, and the sum is over all routes from the boundary vertex i at the west to the boundary vertex j at the east.

It is simple to verify the following result. Recall our convention from Section 4A that the rows of g are labeled from bottom to top, and the columns of g are labeled from right to left.

Lemma 5.16. *Let g be a product of factors of the form $y_i(p)$, $\dot{s}_i$, and $x_i(m)\dot{s}_i^{-1}$. Then the weight matrix $x(C_g)$ of the chip network C_g associated to g coincides with the matrix g.*



$$g = \dot{s}_6 y_7(p_2)\dot{s}_4 y_5(p_4) x_6(m_5)\dot{s}_6^{-1} y_3(p_6) x_4(m_7)\dot{s}_4^{-1} y_5(p_8) y_1(p_9) y_2(p_{10}) y_3(p_{11}) y_4(p_{12})$$

Figure 8. The chip network corresponding to the product g of (4-1).

Proof of Theorem 5.10. Let D be a diagram contained in a $k \times (n-k)$ rectangle, and N'_D the corresponding modified network. Let $i'_1 < \cdots < i'_k$ be the labels of the sources of N'_D. Let g be the product of factors of the form $y_i(p)$, $\dot{s}_i$, and $x_i(m)\dot{s}_i^{-1}$ which is encoded by D, and let M be the corresponding MR-matrix, whose rows are indexed from top to bottom by $i_1, \ldots, i_k$. Recall that the projection from g to M switches rows and columns, and the columns labeled $n-k+1, n-k+2, \ldots, n$ in g become rows labeled $i_1, i_2, \ldots, i_k$ in M. Therefore, to prove Theorem 5.10 it suffices to prove that for all $1 \le t \le n$ and $1 \le s \le k$ we have

$$g_{t,s+(n-k)} = \sum_P w(P), \tag{5-1}$$

where the sum is over all pseudopaths P from i'_s to t in the modified network.

By Lemma 5.16, the matrix g coincides with the weight matrix $x(C_g) = (x_{st})$ of the chip network associated to g. Therefore by (5-1) it suffices to prove that for all $1 \le t \le n$ and $1 \le s \le k$ we have

$$x_{t,s+(n-k)} = \sum_P w(P), \tag{5-2}$$

where the sum is over all pseudopaths P from i'_s to t in the modified network.

Recall from Definition 5.15 that $x_{ij} = \sum_Q w(Q)$, where the sum is over all routes Q in the chip network from the boundary vertex i at the west to the boundary vertex j at the east. To prove (5-2), we will give a weight-preserving bijection between pseudopaths P in the modified network from i'_s to t, and routes Q in the chip network from the boundary vertex t at the west to the boundary vertex $s+(n-k)$ at the east. More specifically, given a pseudopath P, we will examine its sequence of steps from source to sink, and explain how to build the corresponding route Q in the chip network. As illustrated in Figure 9, each step in a pseudopath corresponds to a portion of a route in a chip network. (Note that our bijection will build the route in the chip network from east to west, rather than west to east.)

Recall from Figure 4 that each modified network comes from a diagram, and that every box of a diagram is naturally associated with a simple generator s_i. Therefore every internal vertex in a modified network is naturally associated with a simple generator s_i for some i. We will call this the *position* of the vertex.

Let us consider the various kinds of steps in a pseudopath. Such steps naturally fall into one of the following categories (illustrated in Figure 9):

0. A single step east, which starts at a source and ends at position s_i.

I. A single step east, which is preceded by a south step and followed by an east or south step. Such a step starts and ends at positions s_i and s_j (for $i > j$), and is labeled by some weight p. It may skip over some (positions corresponding to) white and black stones in the Go-diagram.

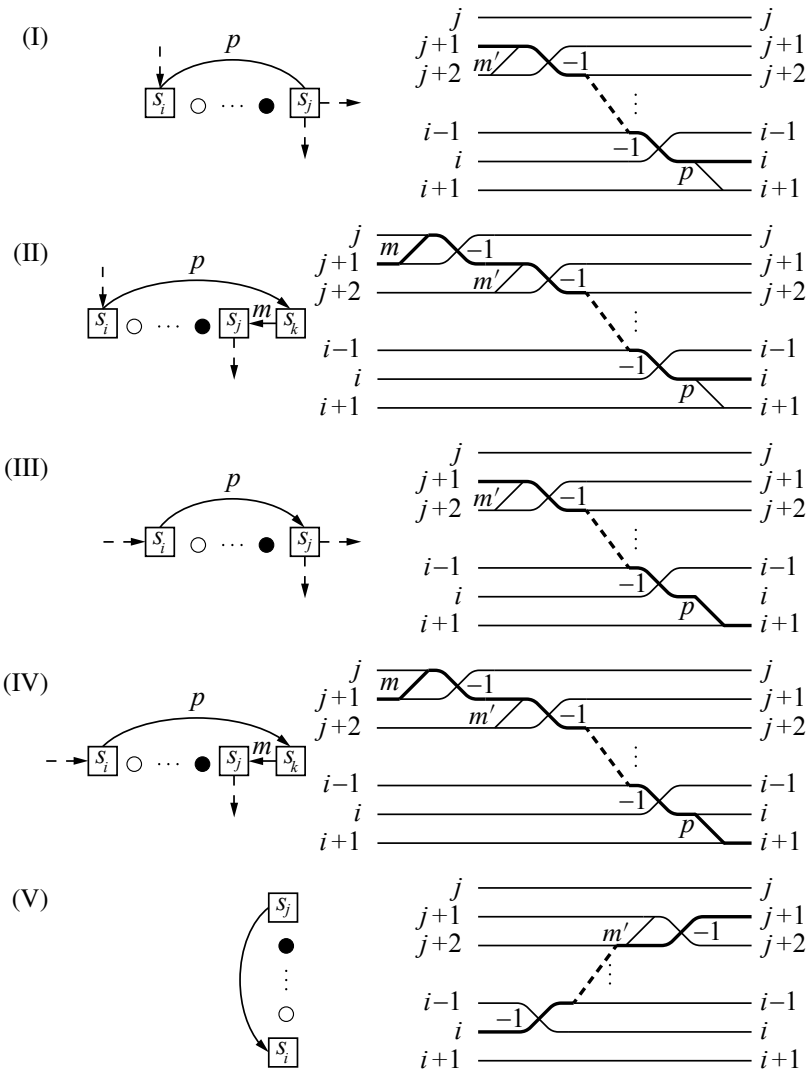


Figure 9. Steps in pseudopaths and their corresponding fragments of the chip network.

II. An east-west combination step, which is preceded by a south step and travels from position s_i to s_k to s_j (where $i > j > k$). The two components of such a step are labeled by some weights p and m, and may skip over some white and black stones.

III. A single step east, which is preceded by an east step and followed by an east or south step. Such a step starts and ends at positions s_i and s_j (for $i > j$), and is labeled by some weight p. It may skip over some white and black stones.

IV. An east-west combination step, which is preceded by an east step and travels from position s_i to s_k to s_j (where $i > j > k$). The two components of such a step are labeled by some weights p and m, and may skip over some white and black stones.

V. A south step, which starts and ends at positions s_i and s_j (for $i < j$). Such a step may skip over some white and black stones.

The above steps in a pseudopath correspond to the following portion of a route in a chip network:

0. Start at the boundary vertex $i+1$ at the east of the chip network.

I. Start at level i, then travel west straight across the $y_i(p)$ chip. For each white or black stone (say in position s_ℓ) which lies in between positions s_i and s_j, travel northwest through the corresponding s_ℓ or x_ℓ-chip, ending at level $j+1$.

II. Start at level i, then travel west straight across the $y_i(p)$ chip. For each white or black stone (say in position s_ℓ) which lies in between positions s_i and s_j, travel northwest through the corresponding s_ℓ or x_ℓ-chip. Finally, travel along the -1-edge and then the m-edge of the $x_j(m)$ chip, ending at level $j+1$.

III. Start at level $i+1$, then travel northwest along the p-edge in the $y_i(p)$ chip. For each white or black stone (say in position s_ℓ) which lies in between positions s_i and s_j, travel northwest through the corresponding s_ℓ or x_ℓ-chip, ending at level $j+1$.

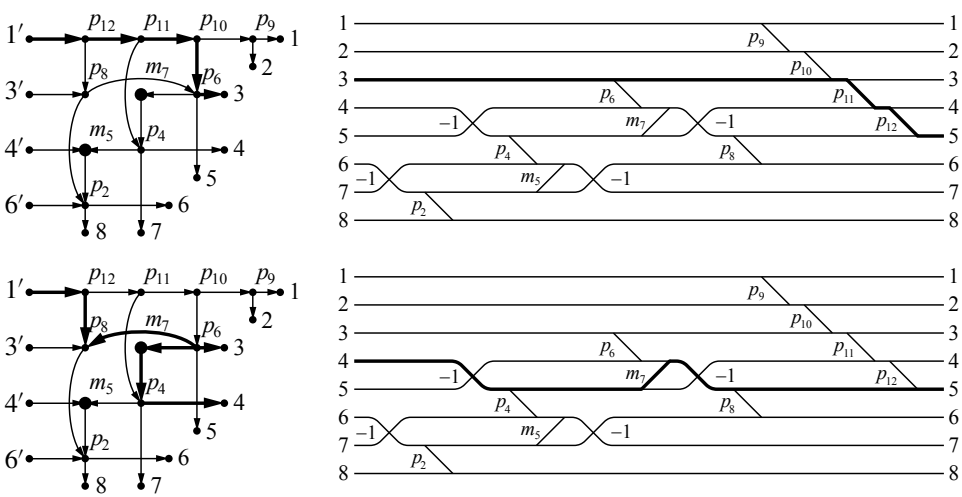


Figure 10. Pseudopaths in the modified network and their corresponding routes in the chip network.

IV. Start at level $i+1$, then travel northwest along the p-edge in the $y_i(p)$ chip. For each white or black stone (say in position s_ℓ) which lies in between positions s_i and s_j, travel northwest through the corresponding s_ℓ or x_ℓ-chip. Finally, travel along the -1-edge and the m-edge of the $x_j(m)$ chip, ending at level $j+1$.

V. Start at level $i+1$. For each white or black stone (say in position s_ℓ) which lies in between positions s_i and s_j, travel southwest through the corresponding s_ℓ or x_ℓ-chip, ending at level j.

It is a straightforward exercise to verify that this map is a bijection between pseudopaths P from i'_s to t in the modified network, and routes Q between the t vertex at the west and the $s+(n-k)$ vertex at the east in the chip network. Moreover, the weights of P and Q are equal. See Figure 10 for examples of entire pseudopaths and routes. □

6. Proof of the main result

Let D be a diagram which contains t pluses and u black stones. In Section 6A we will define an isomorphism

$$\Psi = \Psi_D \colon (\mathbb{K}^*)^t \times \mathbb{K}^u \to (\mathbb{K}^*)^t \times \mathbb{K}^u$$

which maps each parameter from the weight matrix of the network N_D to a Laurent monomial in the parameters used in the MR-matrix $M = M_D$. Then in Section 6B we will show that after applying Ψ, our network parametrization of the network component $\mathcal{R}_D$ coincides with the corresponding MR-parametrization of the projected Deodhar component $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$. Combining this fact with Proposition 4.3 yields Theorem 3.16.

6A. *A rational transformation of parameters.*

Definition 6.1. Let D be a diagram, and let b_0 be a box of D containing a $+$ or $\bullet$. Let b_1 be the nearest box to the right of b_0 which contains a $+$ (if it exists). Let R_t be the set of boxes in the same row as b_0 which are to the right of b_0 and left of b_1. Let R_ℓ be the set of boxes in the same column as b_0 and below b_0. If b_1 exists, let R_r be the set of boxes in the same column as b_1 and below b_1 (otherwise $R_r = \varnothing$). (See Figure 11.) Let R_r^+ (resp. R_ℓ^+) be the set of boxes in R_r (resp. R_ℓ) containing a $+$. Let $R = R_r \cup R_\ell \cup R_t$ and let $R^\bullet$ be the set of boxes in R containing a $\bullet$.

If b_0 contains a $+$, then define

$$\Psi(a_{b_0}) = \frac{(-1)^{|R^\bullet|} \prod_{b \in R_r^+} p_b}{p_{b_0} \prod_{b \in R_\ell^+} p_b}.$$

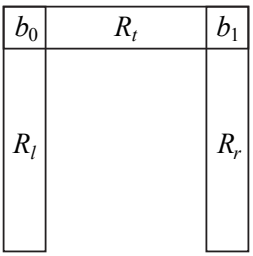


Figure 11. The definition of b_0, b_1, R_t, R_ℓ, and R_r.

And if b_0 contains a $\bullet$, then define

$$\Psi(c_{b_0}) = \frac{m_{b_0}(-1)^{|R^\bullet|}\prod_{b\in R_r^+} p_b}{\prod_{b\in R_\ell^+} p_b}.$$

We also extend the definition of Ψ to all polynomials in the a_b's and c_b's by requiring it to be a ring homomorphism.

Remark 6.2. Clearly $\Psi = \Psi_D$ is an isomorphism from $(\mathbb{K}^*)^t \times \mathbb{K}^u$ to itself.

Example 6.3. Consider the network from Figure 3 (shown again in Figure 12). Then we have $\Psi(a_2) = 1/p_2$, $\Psi(a_4) = 1/p_4$, $\Psi(c_5) = m_5/p_2$, $\Psi(a_6) = 1/p_6$, $\Psi(c_7) = m_7/p_4$, $\Psi(a_8) = 1/(p_2p_8)$, $\Psi(a_9) = 1/p_9$, $\Psi(a_{10}) = 1/(p_6p_{10})$, $\Psi(a_{11}) = -p_6/(p_4p_{11})$, and $\Psi(a_{12}) = p_4/(p_2p_8p_{12})$.

From Definition 6.1, one may easily deduce a formula for $\Psi(w(P))$, where P is a path in N_D. We will state this formula in terms of the Go-diagram. Note that one may identify a path P in N_D with a connected sequence $\mathcal{P}$ of boxes in the Go-diagram, where any two adjacent boxes must share a side. We call a box b in $\mathcal{P}$ a *corner box* if the path P turns from west to south, or from south to west, at the vertex associated to b. (Such a box b in D must contain a $+$ or $\bullet$.)

The next result, whose proof is left to the reader, is a simple consequence of Definition 6.1.

Proposition 6.4. *Let P be a path in the network N_D, which we identify with a sequence $\mathcal{P}$ of boxes in the Go-diagram D. Among the boxes in $\mathcal{P}$, let B_1 denote the subset containing a $+$; let B_2 denote the subset containing a $\bullet$ which are corner boxes of $\mathcal{P}$; and let B_3 denote the subset containing a $\bullet$ which are not corner boxes of $\mathcal{P}$. Then*

$$\Psi(w(P)) = \frac{(-1)^{|B_3|}\prod_{b\in B_2} m_b}{\prod_{b\in B_1} p_b}. \tag{6-1}$$

Example 6.5. Let P_1 and P_2 be the paths shown in Figure 12. Then

$$\Psi(w(P_1)) = \Psi(a_9a_{10}a_{11}c_5) = \frac{-m_5}{p_2p_4p_9p_{10}p_{11}}$$

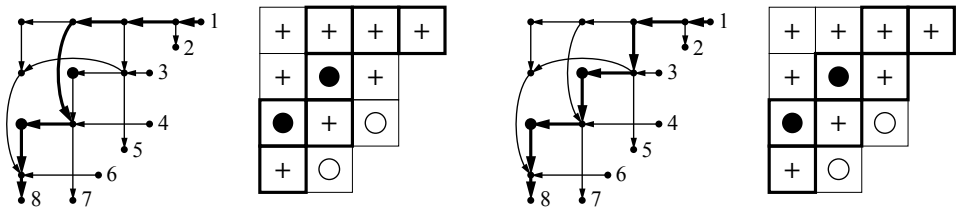


Figure 12. Two paths in N_D and their corresponding sequences of boxes in D.

and

$$\Psi(w(P_2)) = \Psi(a_9 a_{10} c_5 c_7) = \frac{m_5 m_7}{p_2 p_4 p_6 p_9 p_{10}}.$$

6B. ***Applying row operations to the rescaled MR-matrix.***

Theorem 6.6. *Let D be a diagram, let $M = M_D$ be the corresponding MR-matrix, and let L be the matrix we obtain by putting M into reduced row-echelon form. Let $W_D = (W_{ij})$ be the weight matrix associated to D, and let $\Psi(W_D)$ be the matrix obtained from W_D by applying the rational map Ψ to each entry. Then $\Psi(W_D) = L$.*

Proof. To prove Theorem 6.6, we start by considering the rescaled MR-matrix $\widetilde{M}$. Its rows are indexed by the set $i_1, \dots, i_k$, the set of sources of the network N_D, and the leftmost nonzero entry in every row is a 1. Moreover, by Lemma 4.7, the 1 in row i_ℓ is located in column i_ℓ. The entries of the reduced row-echelon matrix L obtained from $\widetilde{M}$ are given by the formula

$$L_{st} = \widetilde{M}_{st} + \sum_{s<j_1<\cdots<j_r<t} (-1)^r \widetilde{M}_{sj_1} \widetilde{M}_{j_1 j_2} \cdots \widetilde{M}_{j_r t}, \tag{6-2}$$

where the sum ranges over all nonempty subsets $\{j_1, \dots, j_r\} \subset \{i_1, \dots, i_k\}$ of sources of the network between s and t.

By Theorem 5.6, the entry $\widetilde{M}_{st}$ equals $\sum_{\widetilde{P}} w(\widetilde{P})$, where the sum is over all pseudopaths in the network N_D from the source s to the boundary vertex t. Therefore, the right-hand side of (6-2) can be interpreted as a generating function for all concatenations of pseudopaths, where the first pseudopath starts at s and the last pseudopath ends at t.

Let us identify a pseudopath with its collection of directed edges. Given a set of pseudopaths on N_D, we define its *signed union* to be the union of directed edges that one obtains by taking the multiset of all directed edges in the pseudopaths, and then cancelling pairs which traverse the same edge but in opposite directions. We define the *weight* of a set of pseudopaths to be the product of the weights of each pseudopath in the set.

Our goal is to show that after cancellation, the only terms which survive on the right-hand side of (6-2) correspond to concatenations of pseudopaths whose signed

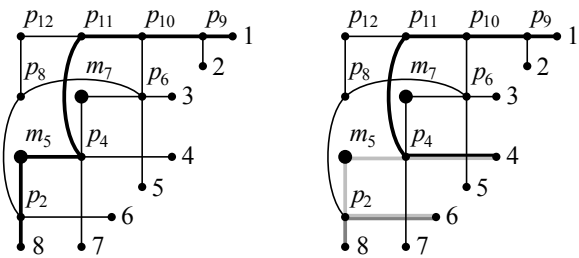


Figure 13. A path in N_D and its decomposition into U-turn pseudopaths.

union forms a directed path (and not merely a pseudopath) from s to t in N_D. This will allow us to relate (6-2) to W_{st}, which is defined as a sum over all paths from s to t in N_D.

Definition 6.7. A *U-turn* pseudopath in a network is a pseudopath whose sequence of steps has the form $(WW^*S)E^*$.

First note that any path P in a network N_D has a unique decomposition as a signed union of U-turn pseudopaths. Moreover, the products of the weights of the pseudopaths is precisely the quantity on the right-hand side of (6-1). See Figure 13 for an example of the decomposition into U-turn pseudopaths.

This observation on the decomposition of paths may be generalized to pseudopaths. Consider a pseudopath $\widetilde{P}$ which is not a path, and turns from south to east precisely q times (for $q \geq 1$). Then for each $0 \leq r \leq q$, there are $\binom{q}{r}$ decompositions of $\widetilde{P}$ as a signed union of r pseudopaths. Moreover, each set of pseudopaths forming a decomposition of $\widetilde{P}$ has the same weight. See Figure 14 for an example of all decompositions of a pseudopath as a signed union of pseudopaths. It is easy to check that each decomposition has the same weight.

Note that since all decompositions of a pseudopath have the same weight, and because

$$\binom{q}{0} - \binom{q}{1} + \binom{q}{2} - \cdots \pm \binom{q}{q} = 0 \quad \text{for } q \geq 1,$$

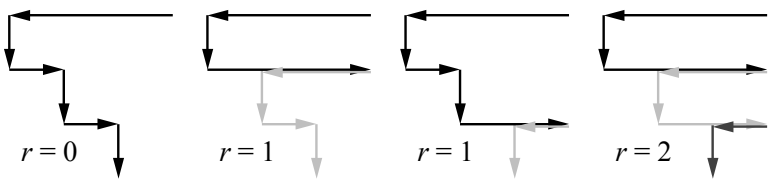


Figure 14. An example of the 2^q decompositions of a path in N_D into pseudopaths.

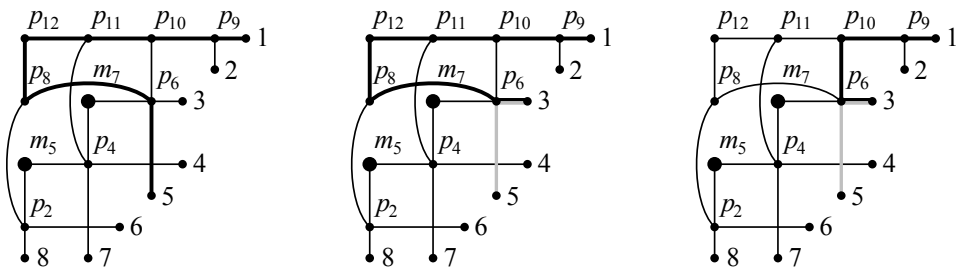


Figure 15. The three possible concatenations of pseudopaths from 1 to 5.

the net contribution of the corresponding concatenations of pseudopaths in (6-2) is 0.

More generally, a term on the right-hand side of (6-2) corresponds to a concatenation of pseudopaths (whose signed union may not be a pseudopath). However, just as before, one may decompose the signed union $\widetilde{P}$ in 2^q ways, where q is the number of times that $\widetilde{P}$ turns from south to east. And again, for $q \geq 1$, the net contribution of the corresponding concatenations of pseudopaths in (6-2) is 0. Therefore when one interprets the right-hand side of (6-2) as a sum over concatenations of pseudopaths, the only terms that are not cancelled are the terms corresponding to concatenations of pseudopaths whose signed union is a directed path.

Example 6.8. In Figure 15, the left and middle diagrams show a pseudopath from 1 to 5, and a concatenation of two pseudopaths from 1 to 5, whose weights cancel each other. The right diagram shows a path from 1 to 5, written as a signed union of two pseudopaths, which will not be cancelled by any other term.

It is a simple exercise to verify that the absolute value of the weight of such a concatenation is precisely the absolute value of the right-hand side of (6-1). We also need to verify that the signs agree. Once we do this, then since W_{st} is the sum of all weights of paths in N_D from s to t, $\Psi(W_{st})$ equals the corresponding term in the expression for L_{st} in (6-2), so the proof is done.

We now check that the signs agree. More specifically, consider a concatenation of $(r+1)$ pseudopaths whose signed union is a directed path P. Using Definition 5.2, the total sign associated to the concatenation of pseudopaths from (6-2) is $(-1)^{r+b+w}$, where b (resp. w) is the number of black (resp. white) stones that P skips over in the horizontal (resp. vertical) direction. (Note that for the purpose of computing b and w, we can count stones skipped by P here, as opposed to the set of pseudopaths whose signed union is P.) Meanwhile, using Definition 3.3 and Proposition 6.4, the total sign associated to the directed path P in the expression $\Psi(W_{st})$ is $(-1)^{q+|B_3|}$, where q is the number of sources in the network which are strictly between s and t, and $|B_3|$ is the number of noncorner

black stones in the boxes of the Go-diagram which P traces out. We need to show that $(-1)^{r+b+w} = (-1)^{q+|B_3|}$.

Note that $|B_3|$ is the number of black stones skipped either vertically or horizontally by P. Let b_v (resp. w_v) denote the number of black stones (resp. white stones) skipped vertically by P. Similarly, let b_h (resp. w_h) denote the number of black stones (resp. white stones) skipped horizontally by P. With this notation, we need to show that $(-1)^{b_h+w_v+r} = (-1)^{b_h+b_v+q}$, i.e., that $(-1)^{w_v+r} = (-1)^{b_v+q}$.

To prove this, we will show that $q - r = w_v + b_v$. Note that $q - r$ is the number of sources strictly between s and t which are *not* sources of any pseudopath in the pseudopath decomposition of P, that is, which are not in the set $\{j_1, \dots, j_r\}$ from (6-2). Recall that the pseudopath decomposition of P is a U-turn pseudopath decomposition; therefore, a source between s and t lies in $\{j_1, \dots, j_r\}$ if and only if P has a vertical edge ending at a $+$ in this row. Otherwise P skips a black or white stone in this row. This proves that $q - r = w_v + b_v$, and hence completes the proof of Theorem 6.6. □

We have now shown that after an invertible transformation of the parameters, our network parametrization of $\mathcal{R}_D$ coincides with the corresponding MR-parametrization of the projected Deodhar component $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$. Combining this result with Proposition 4.3 yields Theorem 3.16.

Our proof yields the following statement.

Corollary 6.9. *Let D be the Go-diagram associated to the distinguished subexpression $\boldsymbol{v}$ of $\boldsymbol{w}$. Then $\mathcal{R}_D = \mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$ as subsets of* $\mathrm{Gr}_{k,n}$. *Furthermore,*

$$\mathrm{Gr}_{k,n} = \bigsqcup_D \mathcal{R}_D,$$

where the union is over all Go-diagrams D contained in a $k \times (n-k)$ rectangle.

Finally, we explain how our proof also yields Corollary 1.2.

Proof. We first note that the Marsh–Rietch parametrizations of Deodhar components restrict to parametrizations of cells in the totally nonnegative part of the complete flag variety (using Lusztig's definition of total positivity), if $\boldsymbol{v}$ is a positive distinguished subexpression of $\boldsymbol{w}$, and the parameters p_i range over $\mathbb{R}_{>0}$ [Marsh and Rietsch 2004].

Our proof of Theorem 3.16 shows that if one takes a (particular) Marsh–Rietsch parametrization of a cell in the nonnegative part of the complete flag variety, then projects it to $(\mathrm{Gr}_{k,n})_{\geq}$ (using Lusztig's definition of total positivity), and then uses an invertible transformation of variables, one gets a network parametrization of a cell of $(\mathrm{Gr}_{k,n})_{\geq 0}$ (using Postnikov's definition of total positivity). It follows that Lusztig's definition of $(\mathrm{Gr}_{k,n})_{\geq 0}$ coincides with Postnikov's definition of $(\mathrm{Gr}_{k,n})_{\geq 0}$, and moreover that the cell decompositions coincide. □

7. A characterization of Deodhar components in terms of Plücker coordinates

In this section we characterize Deodhar components in the Grassmannian by a list of vanishing and nonvanishing Plücker coordinates. Our main result in this section is Theorem 7.8. The proof uses results from [Kodama and Williams 2013], which gave formulas for Plücker coordinates of Deodhar components.

7A. ***Plücker coordinates of Deodhar components in terms of the MR parameters.*** Consider the Deodhar component $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}} \subset \mathrm{Gr}_{k,n}$, where $\boldsymbol{w}$ is a reduced expression for $w \in W^k$ and $\boldsymbol{v} \prec \boldsymbol{w}$. In this section we will review some formulas from [Kodama and Williams 2013] for the Plücker coordinates of the elements of $\mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$ in terms of the parameters which Marsh and Rietsch [2004] used to define $G_{\boldsymbol{v},\boldsymbol{w}}$.

Theorem 7.1 [Kodama and Williams 2013, Lemma 5.1 and Theorem 5.2]. *Let $\boldsymbol{w} = s_{i_1} \cdots s_{i_m}$ be a reduced expression for $w \in W^k$ and $\boldsymbol{v} \prec \boldsymbol{w}$ be a distinguished subexpression for v. Let $A = \pi_k(g) \in \mathcal{P}_{\boldsymbol{v},\boldsymbol{w}}$ for any $g \in G_{\boldsymbol{v},\boldsymbol{w}}$. Then the lexicographically minimal and maximal nonzero Plücker coordinates of A are Δ_I and $\Delta_{I'}$, where $I = w\{n, n-1, \dots, n-k+1\}$ and $I' = v\{n, n-1, \dots, n-k+1\}$. If we write $g = g_1 \cdots g_m$ as in Definition 4.1, then*

$$\Delta_I(A) = (-1)^{|J_{\boldsymbol{v}}^{\bullet}|} \prod_{i \in J_{\boldsymbol{v}}^{+}} p_i \quad \text{and} \quad \Delta_{I'}(A) = 1. \tag{7-1}$$

Remark 7.2. If we write $I = \{i_1, \dots, i_k\}$, then $I' = vw^{-1}\{i_1, \dots, i_k\}$.

Definition 7.3 [Kodama and Williams 2013, Definition 5.4]. Let $W = \mathfrak{S}_n$, let $\boldsymbol{w} = s_{i_1} \cdots s_{i_m}$ be a reduced expression for $w \in W^k$ and choose $\boldsymbol{v} \prec \boldsymbol{w}$. This determines a Go-diagram D of shape $\lambda = \lambda_w$. Let b be any box of D. Note that the set of all boxes of D which are weakly southeast of b forms a Young diagram λ_b^{in}; also the complement of λ_b^{in} in λ is a Young diagram which we call λ_b^{out} (see Example 7.4 below). By looking at the restriction of $\boldsymbol{w}$ to the positions corresponding to boxes of λ_b^{in}, we obtain a reduced expression $\boldsymbol{w}_b^{\mathrm{in}}$ for some permutation w_b^{in}, together with a distinguished subexpression $\boldsymbol{v}_b^{\mathrm{in}}$ for some permutation v_b^{in}. Similarly, by using the positions corresponding to boxes of λ_b^{out}, we obtain $\boldsymbol{w}_b^{\mathrm{out}}$, w_b^{out}, $\boldsymbol{v}_b^{\mathrm{out}}$, and v_b^{out}. When the box b is understood, we will often omit the subscript b.

For any box b, note that it is always possible to choose a reading order of $\lambda = \lambda_w$ which orders all the boxes of λ^{out} after those of λ^{in}. We can then adjust $\boldsymbol{w}$ accordingly; this does not affect whether the corresponding expression $\boldsymbol{v}$ is distinguished. Having chosen such a reading order, we can then write $\boldsymbol{w} = \boldsymbol{w}^{\mathrm{in}}\boldsymbol{w}^{\mathrm{out}}$ and $\boldsymbol{v} = \boldsymbol{v}^{\mathrm{in}}\boldsymbol{v}^{\mathrm{out}}$. We then use g^{in} and g^{out} to denote the corresponding factors of $g \in G_{\boldsymbol{v},\boldsymbol{w}}$. We define $J_{\boldsymbol{v}^{\mathrm{out}}}^{+}$ to be the subset of $J_{\boldsymbol{v}}^{+}$ coming from the factors of $\boldsymbol{v}$ contained in $\boldsymbol{v}^{\mathrm{out}}$. Similarly for $J_{\boldsymbol{v}^{\mathrm{out}}}^{\circ}$ and $J_{\boldsymbol{v}^{\mathrm{out}}}^{\bullet}$.

Example 7.4. Let $W = \mathfrak{S}_7$ and $\boldsymbol{w} = s_4 s_5 s_2 s_3 s_4 s_6 s_5 s_1 s_2 s_3 s_4$ be a reduced expression for $w \in W^3$. Let $\boldsymbol{v} = s_4 s_5 1 1 s_4 1 s_5 s_1 1 1 s_4$ be a distinguished subexpression. So $w = (3, 5, 6, 7, 1, 2, 4)$ and $v = (2, 1, 3, 4, 6, 5, 7)$. We can represent this data by the poset λ_w and the corresponding Go-diagram:

s_4	s_3	s_2	s_1
s_5	s_4	s_3	s_2
s_6	s_5	s_4	

●	+	+	○
●	○	+	+
+	○	○	

Let b be the box of the Young diagram which is in the second row and the second column (counting from left to right). Then the diagram below shows: the boxes of λ^{in} and λ^{out}; a reading order which puts the boxes of λ^{out} after those of λ^{in}; and the corresponding labeled Go-diagram. Using this reading order, $\boldsymbol{w}^{\text{in}} = s_4 s_5 s_2 s_3 s_4$, $\boldsymbol{w}^{\text{out}} = s_6 s_5 s_1 s_2 s_3 s_4$, $\boldsymbol{v}^{\text{in}} = s_4 s_5 1 1 s_4$, and $\boldsymbol{v}^{\text{out}} = 1 s_5 s_1 1 1 s_4$.

out	out	out	out
out	in	in	in
out	in	in	

11	10	9	8
7	5	4	3
6	2	1	

-1 / m_{11}	p_{10}	p_9	1
-1 / m_7	1	p_4	p_3
p_6	1	1	

Note that $J^{\bullet}_{\boldsymbol{v}^{\text{out}}} = \{7, 11\}$ and $J^{+}_{\boldsymbol{v}^{\text{out}}} = \{6, 9, 10\}$. Then $g \in G_{\boldsymbol{v},\boldsymbol{w}}$ has the form

$$g = g^{\text{in}} g^{\text{out}} = \big(\dot{s}_4 \dot{s}_5 y_2(p_3) y_3(p_4) \dot{s}_4\big)\big(y_6(p_6) x_5(m_7) \dot{s}_5^{-1} \dot{s}_1 y_2(p_9) y_3(p_{10}) x_4(m_{11}) s_4^{-1}\big).$$

When we project the resulting 7×7 matrix to its first three columns, we get the matrix

$$A = \begin{pmatrix} -p_9 p_{10} & -p_3 p_{10} & -p_{10} & -m_{11} & 0 & -1 & 0 \\ 0 & -p_3 p_4 & -p_4 & -m_7 & 1 & 0 & 0 \\ 0 & 0 & 0 & p_6 & 0 & 0 & 1 \end{pmatrix}$$

Theorem 7.5 [Kodama and Williams 2013, Theorem 5.6]. *Let $\boldsymbol{w} = s_{i_1} \cdots s_{i_m}$ be a reduced expression for $w \in W^k$ and $\boldsymbol{v} \prec \boldsymbol{w}$, and let D be the corresponding Go-diagram. Choose any box b of D, and let $v^{\text{in}} = v^{\text{in}}_b$ and $w^{\text{in}} = w^{\text{in}}_b$, and $v^{\text{out}} = v^{\text{out}}_b$ and $w^{\text{out}} = w^{\text{out}}_b$. Let $A = \pi_k(g)$ for any $g \in G_{\boldsymbol{v},\boldsymbol{w}}$, and let $I = w\{n, n-1, \dots, n-k+1\}$. If b contains a $+$, define $I_b = v^{\text{in}}(w^{\text{in}})^{-1} I \in \binom{[n]}{k}$. If b contains a white or black stone, define $I_b = v^{\text{in}} s_b (w^{\text{in}})^{-1} I \in \binom{[n]}{k}$. If we write $g = g_1 \cdots g_m$ as in Definition 4.1, then:*

(1) *If b contains a $+$, then $\Delta_{I_b}(A) = (-1)^{|J^{\bullet}_{\boldsymbol{v}^{\text{out}}}|} \prod_{i \in J^{+}_{\boldsymbol{v}^{\text{out}}}} p_i$.*

(2) *If b contains a white stone, then $\Delta_{I_b}(A) = 0$.*

(3) *If b contains a black stone, then*

$$\Delta_{I_b}(A) = (-1)^{|J^{\bullet}_{\boldsymbol{v}^{\text{out}}}|+1} m_b \prod_{i \in J^{+}_{\boldsymbol{v}^{\text{out}}}} p_i + \Delta_{I_b}(A_b),$$

where m_b is the parameter corresponding to b, and A_b is the matrix A with $m_b = 0$.

Example 7.6. We continue Example 7.4. By Theorem 7.1, $I = w\{5, 6, 7\} = \{1, 2, 4\}$ and $I' = v\{5, 6, 7\} = \{5, 6, 7\}$, and the lexicographically minimal and maximal nonzero Plücker coordinates for A are $\Delta_I(A) = p_3 p_4 p_6 p_9 p_{10}$ and $\Delta_{I'}(A) = 1$; this can be verified for the matrix A above.

We now verify Theorem 7.5 for the box b labeled 5 in the reading order. Then $I_b = v^{\text{in}}(w^{\text{in}})^{-1} I = \{1, 4, 6\}$. Theorem 7.5 says that $\Delta_{I_b}(A) = 0$, since this box contains a white stone. The analogous computations for the boxes labeled 7, 6, 4, 3, 2, 1, respectively, yield $\Delta_{1,5,7} = -p_9 p_{10}$, $\Delta_{1,2,7} = p_3 p_4 p_9 p_{10}$, $\Delta_{1,4,5} = p_6 p_9 p_{10}$, $\Delta_{1,3,4} = p_4 p_6 p_9 p_{10}$, $\Delta_{1,2,4} = p_3 p_4 p_6 p_9 p_{10}$, and $\Delta_{1,2,4} = p_3 p_4 p_6 p_9 p_{10}$. These can be checked for the matrix A above.

Proposition 7.7 [Kodama and Williams 2013, Corollary 5.11]. *For any box b, the rescaled Plücker coordinate*

$$\frac{\Delta_{I_b}(A)}{\prod_{i \in J_{\boldsymbol{v}}^+} p_i}$$

depends only on the parameters $p_{b'}$ and $m_{b'}$ which correspond to boxes b' weakly southeast of b in the Go-diagram.

7B. *The characterization of Deodhar components in terms of minors.* Given a Go-diagram D of shape λ, contained in a $k \times (n-k)$ rectangle, let $I = I(\lambda)$. It is not hard to check that if D corresponds to the distinguished subexpression $\boldsymbol{v}$ of the reduced expression $\boldsymbol{w}$, then $I = w\{n, n-1, \dots, n-k+1\}$.

Theorem 7.8. *Let D be a Go-diagram of shape λ contained in a $k \times (n-k)$ rectangle. Let $A \in \mathrm{Gr}_{k,n}$. Then A lies in the Deodhar component $\mathcal{R}_D$ if and only if the following conditions are satisfied*:

(1) $\Delta_{I_b}(A) = 0$ *for all boxes in D containing a white stone.*

(2) $\Delta_{I_b}(A) \neq 0$ *for all boxes in D containing a $+$.*

(3) $\Delta_{I(\lambda)}(A) \neq 0$.

(4) $\Delta_J(A) = 0$ *for all k-subsets J which are lexicographically smaller than $I(\lambda)$.*

Proof. Suppose that A lies in the Deodhar component $\mathcal{R}_D$. Then by Theorem 7.1, conditions (3) and (4) hold. And by Theorem 7.5, conditions (1) and (2) hold.

Now suppose that $A \in \mathrm{Gr}_{k,n}$, and conditions (1)–(4) hold. We want to show that $A \in \mathcal{R}_D$. Since the Deodhar components partition $\mathrm{Gr}_{k,n}$, it suffices to show that A cannot lie in $\mathcal{R}_{D'}$ for any other Go-diagram D'. For the sake of contradiction, assume that $A \in \mathcal{R}_{D'}$. Then by conditions (3) and (4), and Theorem 7.1, it follows that D and D' must be Go-diagrams of the same shape. Therefore D and D' correspond to distinguished subexpressions $\boldsymbol{v}$ and $\boldsymbol{v}'$ of the same reduced expression $\boldsymbol{w}$.

Choose a reading order for the boxes of the Go-diagrams D and D', and let b be the first box in that order where D and D' differ. Then without loss of generality, in D the box b must contain a $+$, and in D' the box b must contain a stone. (Because $\boldsymbol{v}$ and $\boldsymbol{v}'$ are distinguished subexpressions of the same reduced word $\boldsymbol{w}$, which agree in the first ℓ factors and differ in the $(\ell+1)$-st factor, one of $\boldsymbol{v}$ and $\boldsymbol{v}'$ must use the $(\ell+1)$-st simple generator s_b and one must omit it.) In fact, it follows from the definition of distinguished subexpression and the fact that D corresponds to a distinguished subexpression that the box b in D' must contain a white stone, not a black one. (When building a distinguished subexpression from left to right, if choosing the next simple generator s_b would decrease the length of the word so far, then one must choose s_b.)

Now note that the minor which Theorem 7.5 associates to box b in D is Δ_{I_b}, where $I_b = v^{\text{in}}(w^{\text{in}})^{-1}(I)$, and the minor which the theorem associates to box b in D' is $\Delta_{I'_b}$, where $I'_b = v'^{\text{in}} s_b (w^{\text{in}})^{-1}(I)$. But now note that $v'^{\text{in}} s_b = v^{\text{in}}$. Therefore $I_b = I'_b$. So then by Theorem 7.5, if $A \in \mathcal{R}_{D'}$, then $\Delta_{I_b}(A) = 0$, while by condition (2), $\Delta_{I_b}(A) \neq 0$. This is a contradiction. □

Corollary 7.9. *The Deodhar decomposition of the Grassmannian is a coarsening of the matroid stratification*: *in other words*, *each Deodhar component is a union of matroid strata.*

Proof. Each matroid stratum is defined by specifying that certain Plücker coordinates are nonzero while the rest are zero. Therefore the corollary is an immediate consequence of Theorem 7.8. □

There is also an oriented version of Corollary 7.9. To state this, we need a little preparation. First we define the *oriented matroid stratification* of the real Grassmannian to be the decomposition into strata based on which Plücker coordinates are 0, positive, and negative. Next note that from Definition 4.1 and Proposition 4.3, it is immediate that if we are working over $\mathbb{K} = \mathbb{R}$, then the Deodhar component coming from a Go-diagram D has 2^r connected components, where r is the number of boxes in D which contain a $+$. We have the following result.

Corollary 7.10. *Consider the decomposition of the real Grassmannian into connected components of Deodhar components. This is a coarsening of the oriented matroid stratification*: *in other words*, *each connected component of a Deodhar component is a union of oriented matroid strata.*

Proof. From Definition 4.1 and Proposition 4.3, we see that the connected components of a Deodhar component coming from D are in bijection with the 2^r ways of choosing a sign (either positive or negative) for each of the r parameters p_ℓ corresponding to the $+$ boxes of D. By Theorems 7.1 and 7.5, such a choice of signs determines the sign pattern for the Plücker coordinates $\Delta_{I(\lambda)}$ and Δ_{I_b},

where b contains a $+$. Conversely, suppose we know the signs for those Plücker coordinates. Then we may algorithmically determine the signs of the p_ℓ's: first we use Theorem 7.1 to determine the sign of the product of all of the p_ℓ's; then we apply Theorem 7.5 to each box b containing a $+$, reading the boxes in an order that proceeds from southeast to northwest. (For example, by reading the rows from bottom to top, and the boxes within each row from right to left.) □

Remark 7.11. Theorem 7.8 implicitly gives an algorithm for determining the Deodhar component and corresponding network of a point of the Grassmannian, given by a matrix representative or by a list of its Plücker coordinates. The steps are as follows.

(1) Find the lexicographically minimal nonzero Plücker coordinate Δ_I. Then the Go-diagram has shape $\lambda(I)$. Fix a reading order for this shape.

(2) We determine how to fill each box, working in the reading order, as follows. First check whether the box b is forced to contain a black stone. If so, proceed to the next box. If not, look at Δ_{I_b}. If this Plücker coordinate is zero, b must contain a white stone, and if it is nonzero, b must contain a $+$. Proceed to the next box. This process will completely determine the Go-diagram.

(3) Given the Go-diagram, we know what the underlying graph of the network must be. To determine the weights on horizontal edges, work through them in the reading order again. The Plücker coordinate Δ_{I_b} will only use edge weights a_b (when b contains a $+$) or c_b (when b contains a black stone) and weights $a_{b'}$ and $c_{b'}$ corresponding to boxes b' which are earlier than b in the reading order. Thus, we may use the Lindström–Gessel–Viennot lemma recursively to determine each weight a_b or c_b.

Acknowledgements

We are grateful to Sara Billey for numerous helpful comments on the first version of this paper, as well as to an anonymous referee for useful comments. L. Williams is also grateful to Yuji Kodama for their joint work on soliton solutions of the KP equation, which provided motivation for this project.

References

[Björner and Brenti 2005] A. Björner and F. Brenti, *Combinatorics of Coxeter groups*, Graduate Texts in Mathematics **231**, Springer, New York, 2005. MR 2006d:05001 Zbl 1110.05001

[Deodhar 1985] V. V. Deodhar, "On some geometric aspects of Bruhat orderings, I: A finer decomposition of Bruhat cells", *Invent. Math.* **79**:3 (1985), 499–511. MR 86f:20045 Zbl 0563.14023

[Deodhar 1987] V. V. Deodhar, "On some geometric aspects of Bruhat orderings, II: The parabolic analogue of Kazhdan–Lusztig polynomials", *J. Algebra* **111**:2 (1987), 483–506. MR 89a:20054 Zbl 0656.22007

[Kazhdan and Lusztig 1979] D. Kazhdan and G. Lusztig, "Representations of Coxeter groups and Hecke algebras", *Invent. Math.* **53**:2 (1979), 165–184. MR 81j:20066 Zbl 0499.20035

[Kodama and Williams 2013] Y. Kodama and L. Williams, "The Deodhar decomposition of the Grassmannian and the regularity of KP solitons", *Adv. Math.* **244** (2013), 979–1032. MR 3077895

[Lusztig 1998] G. Lusztig, "Total positivity in partial flag manifolds", *Represent. Theory* **2** (1998), 70–78. MR 2000b:20060 Zbl 0895.14014

[Marsh and Rietsch 2004] R. J. Marsh and K. Rietsch, "Parametrizations of flag varieties", *Represent. Theory* **8** (2004), 212–242. MR 2005c:14061 Zbl 1053.14057

[Postnikov 2006] A. Postnikov, "Total positivity, Grassmannians, and networks", preprint, 2006. arXiv math/0609764

[Proctor 1984] R. A. Proctor, "Bruhat lattices, plane partition generating functions, and minuscule representations", *European J. Combin.* **5**:4 (1984), 331–350. MR 86h:17007 Zbl 0562.05003

[Rietsch 1998] K. C. Rietsch, *Total positivity and real flag varieties*, thesis, Massachusetts Institute of Technology, 1998, Available at http://hdl.handle.net/1721.1/46139. MR 2716793

[Stembridge 1996] J. R. Stembridge, "On the fully commutative elements of Coxeter groups", *J. Algebraic Combin.* **5**:4 (1996), 353–385. MR 97g:20046 Zbl 0864.20025

Communicated by Andrei Zelevinsky
Received 2012-10-19 Revised 2013-03-15 Accepted 2013-03-24

kellicar@umich.edu *Department of Mathematics, University of California, Berkeley, 889 Evans Hall, Berkeley, CA 94720-3840, United States*

williams@math.berkeley.edu *Department of Mathematics, University of California, Berkeley, 913 Evans Hall, Berkeley, CA 94720-3840, United States*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2313

Chow quotients of toric varieties as moduli of stable log maps

Qile Chen and Matthew Satriano

Let X be a projective normal toric variety and T_0 a rank-1 subtorus of the defining torus T of X. We show that the normalization of the Chow quotient $X /\!\!/ T_0$, in the sense of Kapranov, Sturmfels, and Zelevinsky, coarsely represents the moduli space of stable log maps to X with discrete data given by $T_0 \subset X$. We also obtain similar results when $T_0 \to T$ is a homomorphism that is not necessarily an embedding.

1. Introduction

Throughout, we work over an algebraically closed field k of characteristic 0.

Chow quotients of toric varieties were introduced by Kapranov, Sturmfels, and Zelevinsky in [Kapranov et al. 1991]. Given a projective normal toric variety X and a subtorus T_0 of the defining torus T, the *Chow quotient* $X /\!\!/ T_0$ has the property that its normalization is the smallest toric variety that maps onto all GIT quotients of X by T_0. We show in this paper that when T_0 has rank 1, the normalization of $X /\!\!/ T_0$ can be reinterpreted as the coarse moduli space of the stack of stable log maps introduced in [Chen 2011; Abramovich and Chen 2011] and independently in [Gross and Siebert 2013]. We also obtain similar results by replacing $T_0 \subset T$ with a homomorphism $T_0 \to T$ and the Chow quotient in the sense of [Kapranov et al. 1991] with that of [Kollár 1996].

MSC2010: primary 14H10; secondary 14N35.
Keywords: toric, Kontsevich, stable log map, Chow quotient.

Let X be a normal toric variety of dimension n with defining torus T. Denoting by $N \cong \mathbb{Z}^n$ the cocharacter lattice of T, we see that every point $v \in N$ corresponds to a morphism of multiplicative groups

$$\iota_v : T_0 := \mathbb{G}_m \to T. \tag{1-1}$$

It is convenient to view this map as the action of T_0 on the identity element $1 \in T$. Let $v = k\omega$ for some positive integer k and primitive lattice point $\omega \in N$. Note that ι is an embedding if and only if $k = 1$.

We begin by introducing the Chow quotient $X /\!\!/ T_0$. For every point $x \in T$, the closure $Z_x := \overline{T_0 x}$ of the orbit of x under T_0 with the reduced scheme structure is a subvariety of X. Thus, we obtain a Chow cycle $k \cdot [Z_x]$. For $x \in T$, the orbit closures Z_x have the same dimension and homology class. Denoting by $T' := [T/T_0]$ the stack quotient, we therefore obtain a morphism from T' to the Chow variety $C(X)$ of algebraic cycles of the given dimension and homology class. For the definition and construction of the Chow variety $C(X)$, we refer to [Kollár 1996, Chapter I]. Since the Chow variety is not actually a moduli space for cycles as above, one may initially be worried that we only obtain a map on the level of closed points. However, we will later see that there is a family of stable maps over T' whose image is precisely the Chow cycle we obtained here; it then follows from [Kollár 1996, Chapter I, 3.17 and 3.21] that there is a natural map $T' \to C(X)$. We define the *Chow quotient* $X /\!\!/ T_0$ to be the closure of the image T' in $C(X)$ with the reduced scheme structure.

Note that when $k = 1$, T' is a variety and $X /\!\!/ T_0$ is the Chow quotient introduced by Kapranov et al. [1991]. In this case, it is a toric variety and the fan of its normalization is given explicitly in [Kapranov et al. 1991, §1].

As mentioned above, the goal of this paper is to relate $X /\!\!/ T_0$ to moduli spaces of stable log maps. Notice that by compactifying ι, we obtain a stable map $f_1 : \mathbb{P}^1 \to X$, where $\mathbb{P}^1$ is marked at the points $\{0, \infty\} = \mathbb{P}^1 \setminus T_0$. By viewing X as a log scheme with its canonical log structure $\mathcal{M}_X$ given by the boundary $X \setminus T$ and $\mathbb{P}^1$ as a log curve with log structure $\mathcal{M}_{\mathbb{P}^1}$ given by the two markings $\{0, \infty\}$, we obtain a stable log map

$$f_1 : (\mathbb{P}^1, \mathcal{M}_{\mathbb{P}^1}) \to (X, \mathcal{M}_X).$$

Let β_0 be the curve class of the stable map f_1, and let c_0 and c_∞ be the contact orders of 0 and ∞ with respect to the toric boundary $X \setminus T$. Roughly speaking, c_0 and c_∞ are functions that assign to the marked points their orders of tangency with the components of $X \setminus T$ (see [Abramovich et al. 2011] for more details). In the toric case, the contact orders can be explained as the slopes and weights of the unbounded edges of tropical curves associated to stable log maps; see Section 3.3. Let $\mathcal{H}_{\Gamma_0}(X)$ be the stack parametrizing stable log maps from rational curves with

two marked points to X such that the curve class is β_0 and the marked points have contact orders given by c_0 and c_∞; here the notation

$$\Gamma_0 := (0, \beta_0, 2, \{c_0, c_\infty\}) \tag{1-2}$$

keeps track of the discrete data consisting of genus, curve class, number of marked points, and their tangency conditions. Our main result is:

Theorem 1.1. *The normalization of $X /\!\!/ T_0$ is the coarse moduli space of $\mathcal{K}_{\Gamma_0}(X)$.*

Remark 1.2. In particular, we see that $\mathcal{K}_{\Gamma_0}(X)$ is irreducible.

Remark 1.3. In Proposition 2.3, we prove that for any $\Gamma = (0, \beta, 2, \{c_0, c_\infty\})$, either the stack $\mathcal{K}_\Gamma(X)$ is empty or $\Gamma = \Gamma_0$ for some Γ_0 as in (1-2). Thus, our discussion covers all two-pointed stable log maps to toric varieties.

In the process of proving Theorem 1.1, we obtain an alternative description of $\mathcal{K}_{\Gamma_0}(X)$ that is more akin to the construction of the Chow quotient. As we saw above, $X /\!\!/ T_0$ is defined as the closure of T' in the Chow variety $C(X)$. Replacing $C(X)$ by other moduli spaces, we obtain alternate spaces analogous to $X /\!\!/ T_0$. For each point $x \in T$, letting T_0 act on x via the group morphism ι and taking the closure, we obtain a stable log map

$$f_x : (\mathbb{P}^1, \mathcal{M}_{\mathbb{P}^1}) \to (X, \mathcal{M}_X)$$

again with curve class β_0 and contact orders c_0 and c_∞. Note that for any point $x' \in \overline{T_0 x}$, the two stable log maps f_x and $f_{x'}$ are canonically isomorphic. We thus obtain a family of stable log maps over the stack quotient T'. It is important to notice that the log structure on T' is trivial (and is denoted by $\mathcal{O}^*$). The stack $\mathcal{K}_{\Gamma_0}(X)$ comes equipped with a log structure, and the above discussion defines a morphism of log stacks

$$(T', \mathcal{O}^*_{T'}) \to (\mathcal{K}_{\Gamma_0}(X), \mathcal{M}_{\mathcal{K}_{\Gamma_0}(X)}).$$

Forgetting the log structures, we obtain an immersion

$$T' \to \mathfrak{M}_{0,2}(X, \beta_0),$$

where $\mathfrak{M}_{0,2}(X, \beta_0)$ denotes the Kontsevich space of stable maps to X with genus 0, curve class β_0, and two marked points. In analogy with the construction of the Chow variety, we let $\mathfrak{M}$ denote the closure of T' in $\mathfrak{M}_{0,2}(X, \beta_0)$. Then we have:

Theorem 1.4. *$\mathcal{K}_{\Gamma_0}(X)$ is the normalization of $\mathfrak{M}$.*

Remark 1.5. There is an analogous picture if one assumes that X is an affine normal toric variety and replaces $\mathfrak{M}_{0,2}(X, \beta_0)$ above by the toric Hilbert scheme, as defined in [Peeva and Stillman 2002]. That is, for all $x \in T$, the Z_x are T'-invariant closed subschemes of X that have the same discrete invariants. We therefore

obtain an immersion from T' to an appropriate toric Hilbert scheme. The closure of T' in this toric Hilbert scheme is called the main component. Olsson [2008, Theorem 1.7] shows that the normalization of the main component has a natural moduli interpretation in terms of log geometry. Theorem 1.4 above can therefore be viewed as an analogue of Olsson's theorem, replacing his use of the toric Hilbert scheme by the Kontsevich space. That is, we show that the normalization of $\mathfrak{M}$ carries a moduli interpretation in terms of stable log maps.

Recall that given any collection of discrete data $\Gamma = (g, \beta, n, \{c_i\}_{i=1}^n)$, it is shown in [Chen 2011; Abramovich and Chen 2011; Gross and Siebert 2013] that there is a proper Deligne–Mumford stack $\mathcal{K}_\Gamma(X)$ that parametrizes stable log maps to X from genus-g curves with n marked points having curve class β and contact orders given by the c_i.[1] We show in Proposition 2.1 that if $g = 0$, then $\mathcal{K}_\Gamma(X)$ is log smooth and in particular normal. This is a key ingredient in the proof of Theorem 1.4, which we give in Section 2. In Section 3, following [Nishinou and Siebert 2006; Gross and Siebert 2013], we explain the relationship between tropical curves and stable log maps to toric varieties. While the use of tropical curves is not strictly necessary for this paper, they serve as a convenient tool to study the boundary of $\mathcal{K}_\Gamma(X)$. Theorem 1.1 is then proved in Section 4.

Remark 1.6. One of the purposes of the theory of stable log maps is to define and compute Gromov–Witten invariants with tangency conditions. The authors plan to calculate the Gromov–Witten invariants in the case of this paper once the forthcoming paper [Abramovich et al. $\geq$ 2013] is ready to use; this latter paper will carefully treat the virtual cycle of the space of stable log maps as well as a version of the degeneration formula of Gromov–Witten invariants.

Prerequisites. We assume the reader is familiar with logarithmic geometry in the sense of Fontaine, Illusie, and Kato (see for example [Kato 1989] or [Ogus 2006]).

2. Log smoothness and irreducibility

Throughout this section, X is a projective normal toric variety of dimension d and Γ is an arbitrary choice of discrete data $(0, \beta, n, \{c_i\})$. Let T be the defining torus of X and M be the character lattice of T.

Proposition 2.1. $(\mathcal{K}_\Gamma(X), \mathcal{M}_{\mathcal{K}_\Gamma(X)})$ *is log smooth over* $(k, \mathbb{G}_k^*)$. *Also*, $\dim \mathcal{K}_\Gamma(X) = \dim X + n - 3$.

[1] Strictly speaking, [Chen 2011; Abramovich and Chen 2011] only consider log schemes that are generalized Deligne–Faltings (see Definition A.1), so to apply their theory, one must first show that the natural log structure on X satisfies this hypothesis. This is done in Proposition A.4, which we relegate to the Appendix since the theory developed in [Gross and Siebert 2013] is already known to apply to toric varieties.

Proof. The universal curve on $\mathcal{K}_\Gamma(X)$ induces a morphism of log stacks

$$\pi : (\mathcal{K}_\Gamma(X), \mathcal{M}_{\mathcal{K}_\Gamma(X)}) \to (\mathfrak{M}_{0,n}, \mathcal{M}_{\mathfrak{M}_{0,n}}),$$

where $(\mathfrak{M}_{g,n}, \mathcal{M}_{\mathfrak{M}_{g,n}})$ denotes the log stack of (g,n)-prestable curves; see [Kato 2000] and [Olsson 2007, Theorem 1.10] for the definition and construction of this log stack. Since $(\mathfrak{M}_{g,n}, \mathcal{M}_{\mathfrak{M}_{g,n}})$ is log smooth over $(k, \mathbb{O}_k^*)$, it suffices to show that π is log smooth. By [Olsson 2003, Theorem 4.6], this is equivalent to showing that the induced morphism

$$\pi' : \mathcal{K}_\Gamma(X) \to \mathcal{L}og_{(\mathfrak{M}_{0,n}, \mathcal{M}_{\mathfrak{M}_{0,n}})}$$

of stacks is smooth, where $\mathcal{L}og_{(S,\mathcal{M}_S)}$ is the stack of log morphisms to a log scheme $(S, \mathcal{M}_S)$ as defined in the introduction of [loc. cit.].

Let $i : \operatorname{Spec} A \to \operatorname{Spec} A'$ be a square zero thickening of Artin local rings, and let

$$\begin{tikzcd} \operatorname{Spec} A \arrow[r] \arrow[d, "i"'] & \mathcal{K}_\Gamma(X) \arrow[d, "\pi'"] \\ \operatorname{Spec} A' \arrow[r] & \mathcal{L}og_{(\mathfrak{M}_{0,n}, \mathcal{M}_{\mathfrak{M}_{0,n}})} \end{tikzcd}$$

be a commutative diagram. We may view this as a commutative diagram of log stacks by endowing the Artin local rings with the log structure pulled back from $\mathcal{L}og_{(\mathfrak{M}_{0,n}, \mathcal{M}_{\mathfrak{M}_{0,n}})}$. Hence, the two vertical arrows are strict. Denote the induced log structures on Spec A and Spec A' by $\mathcal{M}_A$ and $\mathcal{M}_{A'}$, respectively. We therefore have a log smooth curve h', a cartesian diagram

$$\begin{tikzcd} (C, \mathcal{M}_C) \arrow[r] \arrow[d, "h"'] & (C', \mathcal{M}_{C'}) \arrow[d, "h'"] \\ (\operatorname{Spec} A, \mathcal{M}_A) \arrow[r] & (\operatorname{Spec} A', \mathcal{M}_{A'}) \end{tikzcd}$$

and a minimal stable log map $f : (C, \mathcal{M}_C) \to (X, \mathcal{M}_X)$, which we must show deforms to a minimal stable log map $f' : (C', \mathcal{M}_{C'}) \to (X, \mathcal{M}_X)$. Since the minimality condition is open by [Chen 2011, Proposition 3.5.2], it suffices to show that f deforms as a morphism of log schemes.

By standard arguments in deformation theory, it is enough to consider the case where the kernel $\mathcal{I}$ of $A' \to A$ is principal and killed by the maximal ideal $\mathfrak{m}$ of A'. Then the obstruction to deforming f to a morphism of log schemes lies in

$$\operatorname{Ext}^1(f_0^* \Omega^1_{(X,\mathcal{M}_X)/k}, \mathbb{O}_{C_0}) \otimes_k \mathcal{I},$$

where f_0 denotes the reduction of f mod $\mathfrak{m}$ and C_0 denotes the fiber of C over $A'/\mathfrak{m} = k$. By [Kato 1996, Example 5.6], $\Omega^1_{(X,\mathcal{M}_X)/k} \simeq \mathbb{O}_X \otimes_{\mathbb{Z}} M$. Therefore,

$$\mathrm{Ext}^1(f_0^*\Omega^1_{(X,\mathcal{M}_X)/k}, \mathcal{O}_{C_0}) = H^1(\mathcal{O}_{C_0}^d) = 0,$$

where the last equality holds because C_0 is a curve of arithmetic genus 0. This shows that $(\mathcal{K}_\Gamma(X), \mathcal{M}_{\mathcal{K}_\Gamma(X)})$ is log smooth.

To prove the claim about the dimension of $\mathcal{K}_\Gamma(X)$, note that

$$\dim \mathrm{Ext}^0(f_0^*\Omega^1_{(X,\mathcal{M}_X)/k}, \mathcal{O}_{C_0}) = \dim H^0(\mathcal{O}_{C_0}^d) = d,$$

and so π has relative dimension d. Since $\dim \mathfrak{M}_{0,n} = n-3$, we see $\dim \mathcal{K}_\Gamma(X) = d+n-3$. □

Let $\mathcal{K}^\circ_\Gamma(X)$ denote the nondegeneracy locus, that is, the locus of $\mathcal{K}_\Gamma(X)$ where the log structure $\mathcal{M}_{\mathcal{K}_\Gamma(X)}$ is trivial. By Proposition 2.1 and [Nizioł 2006, Proposition 2.6], $\mathcal{K}^\circ_\Gamma(X)$ is an open dense subset of $\mathcal{K}_\Gamma(X)$. Consider the Kontsevich moduli space of stable maps $\mathfrak{M}_{0,n}(X,\beta)$. The forgetful map

$$\Phi : \mathcal{K}_\Gamma(X) \to \mathfrak{M}_{0,n}(X,\beta)$$

sending a stable log map to its underlying stable map induces a locally closed immersion

$$\mathcal{K}^\circ_\Gamma(X) \to \mathfrak{M}_{0,n}(X,\beta).$$

Since the forgetful map does not change the underlying markings or the underlying maps, no stabilization of the underlying curve is needed here. Let $\mathfrak{M}_\Gamma(X)$ be the closure of $\mathcal{K}^\circ_\Gamma(X)$ in $\mathfrak{M}_{0,n}(X,\beta)$. Then Φ factors through a morphism

$$\phi : \mathcal{K}_\Gamma(X) \to \mathfrak{M}_\Gamma(X).$$

Lemma 2.2. *ϕ is the normalization map.*

Proof. By [Abramovich and Chen 2011, Corollary 3.10] and Proposition A.4, the morphism Φ is representable and finite and so is ϕ. Since $(\mathcal{K}_\Gamma(X), \mathcal{M}_{\mathcal{K}_\Gamma(X)})$ is fs and log smooth over $(k, \mathcal{O}_k^*)$ by Proposition 2.1, it follows that $\mathcal{K}_\Gamma(X)$ is normal. Since ϕ is an isomorphism over $\mathcal{K}^\circ_\Gamma(X)$, it is birational, and so by Zariski's main theorem, ϕ is the normalization map. □

Now we consider the case $\Gamma = (0, \beta, 2, \{c_0, c_\infty\})$, where β is an arbitrary curve class, and c_0 and c_∞ are two arbitrary contact orders along the two different markings. Note that both c_0 and c_∞ are nontrivial. Otherwise, there is a curve in toric variety intersect the boundary at only one point, which is impossible. Then we have the following result:

Proposition 2.3. (1) *If $\mathcal{K}_\Gamma(X) \neq \varnothing$, then $\Gamma = \Gamma_0$ for some Γ_0 as in (1-2), obtained from a group morphism (1-1).*

(2) *$\mathcal{K}_{\Gamma_0}(X)$ is irreducible.*

Proof. Assume that $\mathcal{K}_\Gamma(X) \neq \varnothing$; hence, $\mathcal{K}^\circ_\Gamma(X) \neq \varnothing$ by Proposition 2.1. Let $s \in \mathcal{K}^\circ_\Gamma(X)(k)$ and $f : \mathbb{P}^1 \to X$ be the stable log map corresponding to s. Note that the log structure of the boundary of X is everywhere nontrivial. Since the log structure is trivial at s, the image of f necessarily meets T. After acting on f by some element of T, we may assume that f sends $1 \in \mathbb{P}^1$ to $1 \in T \subset X$. Choose a maximal cone σ in the fan of X such that the associated affine open toric variety $U \subset X$ contains $f(0)$. Restricting f to U, we obtain a map $f' : V = \operatorname{Spec} k[t] \to U$.

Let P be the monoid $\sigma^\vee \cap M$, and let $e_1, \dots, e_l$ be the irreducible elements of P. We see that for each i,

$$f^*(e_i) = t^{c_i} a_i,$$

where c_i is the contact order prescribed by Γ and a_i is some element of $k[t]$. Note that if $\alpha \in k$ is a zero of a_i, then the point $t = \alpha$ is mapped to the toric boundary; however, the contact order given by Γ implies that $t = 0$ is the only point in V that maps to the boundary. Hence, a_i must be a power of t. But if a_i is divisible by t, then the contact order of $t = 0$ along $e_i = 0$ is greater than c_i. Therefore, a_i must be a nonzero constant.

Now observe that the point $1 \in T \subset U$ is given by $e_i = 1$ for all i. Since $f(1) = 1$, the equation $f^*(e_i) = t^{c_i} a_i$ shows that $a_i = 1$. Note that such f defines a group morphism ι_v as in (1-1). This implies that first statement.

To prove the second statement, it is enough to show that $\mathcal{K}^\circ_\Gamma(X)$ is irreducible, which again follows from the above statement. □

Now we set $\Gamma = \Gamma_0$ as in (1-2) and use the setting and notation of the introduction. As discussed in the introduction, we have an immersion $T' \to \mathcal{K}_\Gamma(X)$. Let $\mathfrak{X}_\Gamma$ be the closure of T' in $\mathcal{K}_\Gamma(X)$. The forgetful morphism Φ then induces a map

$$\phi' : \mathfrak{X}_\Gamma \to \mathfrak{M}.$$

Since $\mathcal{K}_\Gamma(X)$ is irreducible, Theorem 1.4 follows from the next lemma.

Lemma 2.4. *$\mathfrak{X}_\Gamma$ is an open substack of $\mathcal{K}_\Gamma(X)$, and so ϕ' is the normalization map.*

Proof. As in the proof of Lemma 2.2, ϕ' is representable and finite. If $\mathfrak{X}_\Gamma$ is an open substack of $\mathcal{K}_\Gamma(X)$, it is then normal. Since ϕ' is an isomorphism over T', Zariski's main theorem shows that it is the normalization map.

To show that $\mathfrak{X}_\Gamma$ is open in $\mathcal{K}_\Gamma(X)$, it suffices to prove that $\mathfrak{X}^\circ_\Gamma := \mathfrak{X}_\Gamma \cap \mathcal{K}^\circ_\Gamma(X)$ has the same dimension as $\mathcal{K}^\circ_\Gamma(X)$. Since T' is dense in $\mathfrak{X}_\Gamma$, we see that $\mathfrak{X}_\Gamma$ has dimension $d - 1$. On the other hand, the map

$$\pi : (\mathcal{K}_\Gamma(X), \mathcal{M}_{\mathcal{K}_\Gamma(X)}) \to (\mathfrak{M}_{0,2}, \mathcal{M}_{\mathfrak{M}_{0,2}})$$

in the proof of Proposition 2.1 induces a map

$$\mathcal{K}^\circ_\Gamma(X) \to \mathfrak{M}^\circ_{0,2},$$

where $\mathfrak{M}^\circ_{0,2}$ denotes the open substack of $\mathfrak{M}_{0,2}$ with smooth fiber curves. By Proposition 2.1, we see that $\mathcal{K}^\circ_\Gamma(X)$ has dimension $d-1$. $\square$

3. Tropical curves associated to stable log maps

The goal of this section is to prove Proposition 3.8. Following [Nishinou and Siebert 2006; Gross and Siebert 2013], we explain the connection between tropical curves and stable log maps to toric varieties.

3.1. ***Review of tropical curves.*** Let $\overline{G}$ be the geometric realization of a weighted, connected finite graph with weight function ω. That is, $\overline{G}$ is the CW complex associated to a finite connected graph with vertex set $\overline{G}^{[0]}$ and edge set $\overline{G}^{[1]}$, and

$$\omega : \overline{G}^{[1]} \to \mathbb{N}$$

is a function. Here we allow $\overline{G}$ to have divalent vertices. Given an edge $l \in \overline{G}^{[1]}$, we denote its set of adjacent vertices by ∂l. If l is a loop, then we require $\omega(l) = 0$.

Let $G^{[0]}_\infty \subset \overline{G}^{[0]}$ be the set of one-valent vertices, and let

$$G := \overline{G} \setminus \overline{G}^{[0]}_\infty.$$

Let $G^{[1]}_\infty$ be the set of noncompact edges in G, which we refer to as *unbounded edges*. A *flag* of G is a pair (v, l) where l is an edge and $v \in \partial l$. We let FG be the set of flags of G, and for each vertex v, we let

$$FG(v) := \{(v, l) \in FG\}.$$

Let N be a lattice and $M = N^\vee$. We let $N_\mathbb{Q} := N \otimes_\mathbb{Z} \mathbb{Q}$ and $N_\mathbb{R} := N \otimes_\mathbb{Z} \mathbb{R}$.

Definition 3.2. A *parametrized tropical curve* in $N_\mathbb{Q}$ is a proper map $\varphi : G \to N_\mathbb{R}$ of topological spaces satisfying the following conditions:

(1) For every edge l of G, the restriction $\varphi|_l$ acts as dilation by a factor $\omega(l)$ with image $\varphi(l)$ contained in an affine line with rational slope. If $\omega(l) = 0$, then $\varphi(l)$ is a point.

(2) For every vertex v of G, we have $\varphi(v) \in N_\mathbb{Q}$.

(3) For each $(v, l) \in FG(v)$, let $u_{v,l}$ be a primitive integral vector emanating from $\varphi(v)$ along the direction of $h(l)$. Then

$$\epsilon_v := \sum_{(v,l)\in FG(v)} \omega(l) u_{v,l} = 0,$$

which we refer to as the *balancing condition*.

An *isomorphism* of tropical curves $\varphi : G \to N_\mathbb{R}$ and $\varphi' : G' \to N_\mathbb{R}$ is a homeomorphism $\Phi : G \to G'$ compatible with the weights of the edges such that $\varphi = \varphi' \circ \Phi$.

A *tropical curve* is an isomorphism class of parametrized tropical curves.

3.3. ***Tropical curves from nondegenerate stable log maps.*** Let $(X, \mathcal{M}_X)$ be a toric variety with its standard log structure, and let $T \subset X$ be its defining torus. We denote by N the lattice of one-parameter subgroups of T. Let $f : (C, \mathcal{M}_C) \to (X, \mathcal{M}_X)$ be a stable log map over $(S, \mathcal{M}_S)$ with S a geometric point. Further assume that f is nondegenerate; that is, the log structure $\mathcal{M}_S$ is trivial.

In this subsection, we show how to assign a tropical curve $\mathrm{Trop}(f) : G \to N_{\mathbb{R}}$ to any such nondegenerate stable log map f. Note that in this case, the points on the source curve with nontrivial log structures are marked points or nodal points. To begin, let G be the graph with a single vertex v, which we think of as being associated to the unique component of C, and with one unbounded edge for each marked point of C. We let $\mathrm{Trop}(f)(v) = 0$.

Let l be an edge corresponding to a marked point p of C. If p has trivial contact orders, then we set $\omega(l) = 0$ and let $\mathrm{Trop}(f)$ contract l to 0. Otherwise, the contact order is equivalent to giving a nontrivial map

$$c_l : \overline{\mathcal{M}}_{X, f(p)} \to \overline{\mathcal{M}}_{C,p} = \mathbb{N}.$$

Note that we have a surjective cospecialization map of groups

$$M := N^\vee \to \overline{\mathcal{M}}^{gp}_{X, f(p)}$$

corresponding to the specialization of the generic point of T to $f(p)$. Composing with c_l^{gp}, we obtain a map

$$\mu_l : M \to \mathbb{Z},$$

which defines an element $\mu_l \in N$. Let u_l be the primitive vector with slope given by $\mu_l \in N$. We define $\omega(l)$ to be the positive integer such that $\mu_l = \omega(l) u_l$ and define the image $\mathrm{Trop}(f)(l)$ to be the unbounded ray emanating from 0 along the direction of u_l. This defines our desired map $\mathrm{Trop}(f) : G \to N_{\mathbb{R}}$ up to reparametrization.

Proposition 3.4. $\mathrm{Trop}(f) : G \to N_{\mathbb{R}}$ *defines a tropical curve.*

Proof. It remains to check that the balancing condition holds. That is, we must show $\epsilon_v = 0$. Note that every $m \in M$ defines a rational function on C and that the degree of the associated Cartier divisor is $0 = \epsilon_v(m)$. Therefore, $\epsilon_v \in N = M^\vee$ is 0. □

3.5. ***Tropical curves from stable log maps over the standard log point.*** Suppose $(X, \mathcal{M}_X)$ is a toric variety with its standard log structure, and let $T \subset X$ be its defining torus. Fix discrete data $\Gamma = (g, \beta, n, \{c_i\})$, and let $f : (C, \mathcal{M}_C) \to (X, \mathcal{M}_X)$ be a stable log map with discrete data Γ over the standard log point $(S, \mathcal{M}_S)$; that is, S is a geometric point and $\mathcal{M}_S$ is the log structure associated to the map $\mathbb{N} \to \mathbb{C}_S$ sending 1 to 0. This is equivalent to giving a (not necessarily strict) log map

$$(S, \mathcal{M}_S) \to (\mathcal{H}_\Gamma(X), \mathcal{M}_{\mathcal{H}_\Gamma(X)}),$$

and the stable log map f is obtained by pulling back the universal stable log map over $(\mathcal{K}_\Gamma(X), \mathcal{M}_{\mathcal{K}_\Gamma(X)})$. In this subsection, we associate a tropical curve

$$\mathrm{Trop}(f) : G \to N_{\mathbb{R}}$$

to f by modifying the construction given in [Gross and Siebert 2013, §1.3].

We define G to be the dual graph of C where we attach an unbounded edge for each marked point. Given a vertex v, let t be the generic point of the corresponding component of C. We therefore have a morphism

$$\overline{\mathcal{M}}_{X,f(t)} \to \overline{\mathcal{M}}_{C,t} = \mathbb{N}$$

of monoids. Taking the associated groups and composing with the cospecialization map $M \to \overline{\mathcal{M}}^{gp}_{X,f(t)}$ yields a map

$$\tau_v : M \to \mathbb{Z}$$

and hence a point in N. We define $\mathrm{Trop}(f)(v) = \tau_v$.

Let l be an edge of G. If $\partial l = \{v, v'\}$ and $v \neq v'$, then we define the image of l under $\mathrm{Trop}(f)$ to be the line segment joining τ_v and $\tau_{v'}$. In this case, $\tau_{v'} - \tau_v = e_l \mu_l$, where $e_l \in \overline{\mathcal{M}}_S = \mathbb{N}$ is the section that smooths the node corresponding to l, and μ_l is an element of N. We define $\omega(l)$ to be the positive integer such that $\mu_l = \omega(l) u_l$, where u_l is a primitive integral vector.

Suppose now that l is an unbounded edge corresponding to a marked point p. If p has trivial contact orders, then we set $\omega(l) = 0$ and let $\mathrm{Trop}(f)$ contract l to τ_v, where $\partial l = \{v\}$. Otherwise, the contact orders of p define a nontrivial map

$$c_l : \overline{\mathcal{M}}_{X,f(p)} \to \overline{\mathcal{M}}_{C,p} = \mathbb{N} \oplus \overline{\mathcal{M}}_S \to \mathbb{N},$$

where the last map is the projection. Again taking the associated groups and composing with the cospecialization map $M \to \overline{\mathcal{M}}^{gp}_{X,f(p)}$, we obtain

$$\mu_l : M \to \mathbb{Z}.$$

We define $\omega(l)$ to be the positive integer such that $\mu_l = \omega(l) u_l$, where $u_l \in N$ is a primitive integral vector, and we let $\mathrm{Trop}(f)(l)$ be the unbounded ray emanating from τ_v in the direction of u_l.

Proposition 3.6. $\mathrm{Trop}(f) : G \to N_{\mathbb{R}}$ *defines a tropical curve.*

Proof. We must check that the balancing condition holds for each vertex v of G. As in the proof of Proposition 3.4, every $m \in M$ defines a rational function on the irreducible component of C corresponding to v. The associated Cartier divisor has degree $0 = \epsilon_v(m)$, and so $\epsilon_v = 0$; see [Gross and Siebert 2013, Proposition 1.14]. □

Remark 3.7. Let R be the complete local ring of $\mathbb{A}^1$ at the origin, and let $\mathcal{M}_R$ be the log structure on R induced by the standard log structure on $\mathbb{A}^1$. Denote the closed and generic points of Spec R by 0 and η, respectively. Suppose $h : (\mathscr{C}, \mathcal{M}_{\mathscr{C}}) \to (X, \mathcal{M}_X)$ is a stable log map over R with discrete data Γ such that $h_0 = f$. Note that h_η is a nondegenerate stable log map. For each marked section $p : \operatorname{Spec} R \to \mathscr{C}$, let l_0 and l_η be the edges of the dual graphs of $\mathscr{C}_0$ and $\mathscr{C}_\eta$ corresponding to the marked points p_0 and p_η, respectively. Consider the morphism

$$\overline{\mathcal{M}}_X|_{h(p)} \to \overline{\mathcal{M}}_{\mathscr{C}}|_p = \mathbb{N} \oplus \overline{\mathcal{M}}_R \to \mathbb{N},$$

where the last map is the projection. Taking associated groups and precomposing with the map $M \to \overline{\mathcal{M}}_X^{gp}|_{h(p)}$, we obtain a map $M \to \mathbb{Z}$ of constant sheaves on Spec R whose special and generic fibers are μ_{l_0} and μ_{l_η}. Hence, we see $\mu_{l_0} = \mu_{l_\eta}$.

Denote by $\operatorname{Trop}(f_\eta)$ and $\operatorname{Trop}(f_0)$ the tropical curves associated to the generic and closed fiber of f. Then the above argument implies that $\operatorname{Trop}(f_\eta)$ is the asymptotic fan of $\operatorname{Trop}(f_0)$ as defined in [Nishinou and Siebert 2006, Definition 3.1].

The following result plays an important role in the proof of Theorem 1.1:

Proposition 3.8. *If the discrete data Γ is given by $g = 0$, $n = 2$, and $\beta \neq 0$, then* $\operatorname{Trop}(f)$ *is an embedding whose image is a line. Moreover, C is a chain of $\mathbb{P}^1$s and f does not contract any components of C.*

Proof. Since $\mathcal{K}_\Gamma(X)$ is log smooth by Proposition 2.1, there exists a stable log map $h : (\mathscr{C}, \mathcal{M}_{\mathscr{C}}) \to (X, \mathcal{M}_X)$ over $(R, \mathcal{M}_R)$ as in Remark 3.7. Let $p, p' : \operatorname{Spec} R \to \mathscr{C}$ be the two marked sections, and let l_0, l_0', l_η, and l_η' be the corresponding edges of the dual graphs of C and $\mathscr{C}_\eta$. Since $\beta \neq 0$, the two marked points p_η and p_η' of $\mathscr{C}_\eta$ have nontrivial contact orders. The balancing condition for $\operatorname{Trop}(h_\eta)$ then shows $\mu_{l_\eta'} = -\mu_{l_\eta} \neq 0$. By Remark 3.7, we therefore have $\mu_{l_0'} = -\mu_{l_0} \neq 0$. In particular, $\operatorname{Trop}(f)$ maps l_0 and l_0' to unbounded rays.

We next show that if l is an edge of G, then $\operatorname{Trop}(f)(l)$ is a point or it is a line segment or ray parallel to μ_{l_0}. Suppose $\operatorname{Trop}(f)(l)$ is not a point. If $\operatorname{Trop}(f)(l)$ is unbounded, then l is l_0 or l_0', and so $\operatorname{Trop}(f)(l)$ is parallel to μ_{l_0}. Otherwise, $\operatorname{Trop}(f)(l)$ is a line segment and $\partial l = \{v, v_1\}$ with $v \neq v_1$. If $\operatorname{Trop}(f)(l)$ is not parallel to μ_{l_0}, then the balancing condition shows that there is an edge $l_1 \neq l$ such that $v_1 \in \partial l_1$ and $\operatorname{Trop}(f)(l_1)$ is not parallel to μ_{l_0}. Hence, l_1 is a line segment with endpoints v_1 and v_2. Again, the balancing condition shows that there is an edge l_2 containing v_2 such that $\operatorname{Trop}(f)(l_2)$ is a line segment which is not parallel to μ_{l_0}. Since C has genus 0, we see l, l_1, and l_2 are distinct. Continuing in this manner, we produce an infinite sequence of distinct edges l_i of the dual graph of C. This is a contradiction.

Lastly, we show that every irreducible component A of C has exactly two special points. Hence, C is a chain of $\mathbb{P}^1$s, f does not contract any component of C,

and $\mathrm{Trop}(f)(G)$ is a line parallel to μ_{l_0}. Suppose A is a component with at least three special points, and let v be the vertex of G corresponding to A. Then $G \setminus v$ is a disjoint union of nonempty trees $T_1, T_2, \dots, T_m$ with $m \geq 3$. Without loss of generality, T_1 only contains bounded edges. The argument in the preceding paragraph then shows that $\mathrm{Trop}(f)$ maps every edge of T_1 to a single point. If C_1 denotes the subcurve of C corresponding to T_1, then we see that every special point of C_1 has a trivial contact order, and so f contracts C_1. Since T_1 is a tree, C_1 contains components with only two special points. This contradicts the stability of f. $\square$

4. The Chow quotient as the coarse moduli space

Throughout this section, we let $\Gamma = \Gamma_0$ and $C(X)$ denote the Chow variety as in the introduction. Let K be the normalization of $X /\!\!/ T_0$. Since the stack $\mathcal{H}_\Gamma(X)$ is normal, it follows from [Kollár 1996, Chapter I, 3.17 and 3.21] that there is a map

$$F : \mathcal{H}_\Gamma(X) \to C(X)$$

sending a stable log map $f : (C, \mathcal{M}_C) \to (X, \mathcal{M}_X)$ to the image cycle $f_*[C]$. Since $\mathcal{H}_\Gamma(X)$ is irreducible by Theorem 1.4, F factors as

$$\mathcal{H}_\Gamma(X) \xrightarrow{F'} X /\!\!/ T_0 \xrightarrow{i} C(X),$$

where i is the natural inclusion. Since F is an isomorphism over T' and $\mathcal{H}_\Gamma(X)$ is normal, by Proposition 2.1, we obtain an induced morphism

$$G : \mathcal{H}_\Gamma(X) \to K.$$

To prove Theorem 1.1, we show:

Proposition 4.1. *G is a coarse space morphism.*

Proof. Note that both $\mathcal{H}_\Gamma(X)$ and K are normal and proper, and G is bijective on the level of closed points over T'. To show that K is the coarse moduli space of $\mathcal{H}_\Gamma(X)$, by Zariski's main theorem, it suffices to show G is quasifinite. To do so, it is enough to show F' is quasifinite at the level of closed points. That is, we show that if $x \in X /\!\!/ T_0$ is a closed point and E_x denotes the corresponding cycle of X, then there are finitely many stable log maps whose image cycles are given by E_x. Let

$$E_x = \sum a_i Z_i,$$

where the a_i are positive integers and the Z_i are reduced irreducible closed subschemes of X. Let $\widetilde{Z}_i$ be the normalization of Z_i. Since E_x is of dimension 1, we have $\widetilde{Z}_i \simeq \mathbb{P}^1$.

We claim that if $f : (C, \mathcal{M}_C) \to (X, \mathcal{M}_X)$ is a stable log map that defines a closed point of $\mathcal{H}_\Gamma(X)$ such that the image cycle of f is E_x, then f can only be

ramified at the special points of C. Given this claim, F' is quasifinite. Indeed, since Proposition 3.8 shows that no component of C is contracted under f, the number of irreducible components of C is bounded by $\sum a_i$. For each irreducible component A of C, the restriction $f|_A$ factors as

$$A \to \widetilde{Z}_i \to X$$

for some i. Since the first map $A \to \widetilde{Z}_i$ can only be ramified at the two fixed special points, it is determined by the degree of $f|_A$. This implies that there are only finitely many choices for the underlying map $C \to X$. Since the forgetful morphism $\Phi : \mathcal{K}_\Gamma(X) \to \mathfrak{M}_{0,2}(X, \beta)$ is finite, there are finitely many choices for the stable log map f.

It remains to prove the claim. By Proposition 2.3, $\mathcal{K}_\Gamma(X)$ is irreducible and T' is dense, so there exists a toric morphism $\mathbb{A}^1 \to \mathcal{K}_\Gamma(X)$ such that the fiber over $0 \in \mathbb{A}^1$ is our given stable log map $f : (C, \mathcal{M}_C) \to (X, \mathcal{M}_X)$ whose image cycle is E_x. Let R denote the complete local ring $\widehat{\mathcal{O}}_{\mathbb{A}^1,0}$ and let

$$\begin{tikzcd} \mathscr{C} \arrow[r, "h"] \arrow[d] & X \\ \operatorname{Spec} R & \end{tikzcd}$$

be the associated underlying stable map. Let $\eta \in \operatorname{Spec} R$ be the generic point.

We first handle the case when X is smooth. Let $\mathscr{C}^\circ$ be the open subset of $\mathscr{C}$ obtained by removing the special points. Note that $\mathscr{C}^\circ$ is normal, and $h|_{\mathscr{C}^\circ}$ is quasifinite by Proposition 3.8. By the purity of the branch locus theorem [Altman and Kleiman 1971, p. 461], if $h|_{\mathscr{C}^\circ}$ is ramified, then the ramification locus D is pure of codimension 1. Since $h|_{\mathscr{C}^\circ}$ is not everywhere ramified over the central fiber, D must intersect the generic fiber. However, $h|_{\mathscr{C}^\circ}$ is unramified over the generic fiber, so we conclude that D is empty.

We now consider the case when X is singular. Let $p : \widetilde{X} \to X$ be a toric resolution. We may replace R by a ramified extension as this does not affect the set of closed points. By the properness of $\mathcal{K}_\Gamma(X)$, we can assume we have a stable log map $\tilde{h} : (\widetilde{\mathscr{C}}, \mathcal{M}_{\widetilde{\mathscr{C}}}) \to (\widetilde{X}, \mathcal{M}_{\widetilde{X}})$ and a commutative diagram of the underlying maps

$$\begin{tikzcd} \widetilde{\mathscr{C}} \arrow[r, "\tilde{h}"] \arrow[d, "q"'] & \widetilde{X} \arrow[d, "p"] \\ \mathscr{C} \arrow[r, "h"] & X \end{tikzcd}$$

over R. Here h is the underlying map of the stable log map to X, which can be also obtained by taking the stabilization of the prestable map $p \circ \tilde{h}$. The previous

paragraph shows that $\tilde{h}$ only ramifies at the special points. Since Proposition 3.8 shows that $\widetilde{\mathscr{C}}$ and $\mathscr{C}$ are both chains of $\mathbb{P}^1$s, we see that h only ramifies at the special points as well. □

Appendix: Toric varieties have generalized Deligne–Faltings log structures

The theory of moduli spaces of stable log maps $\mathcal{K}_\Gamma(Y, \mathcal{M}_Y)$ is developed in [Chen 2011; Abramovich and Chen 2011] and [Gross and Siebert 2013] for different classes of log schemes $(Y, \mathcal{M}_Y)$. In [Chen 2011; Abramovich and Chen 2011], Abramovich and the first author consider log schemes that are generalized Deligne–Faltings (see Definition A.1); Gross and Siebert [2013] consider log schemes that are quasigenerated Zariski. It is shown in [Abramovich and Chen 2011, Proposition 4.8] that when $(Y, \mathcal{M}_Y)$ is both generalized Deligne–Faltings and quasigenerated Zariski, the Abramovich–Chen and Gross–Siebert constructions are identical. Gross and Siebert show that the standard log structure $\mathcal{M}_X$ on a normal toric variety X is always quasigenerated Zariski. Here we show that if X is also projective, then $\mathcal{M}_X$ is generalized Deligne–Faltings. Therefore, the two theories agree for projective normal toric varieties.

Definition A.1. A log structure $\mathcal{M}_Y$ on a scheme Y is called *generalized Deligne–Faltings* if there exists a fine saturated sharp monoid P and a morphism $P \to \overline{\mathcal{M}}_Y$ that locally lifts to a chart $P \to \mathcal{M}_Y$.

Remark A.2. Given a fine saturated sharp monoid P, let $A_P = \operatorname{Spec} k[P]$ with its standard log structure $\mathcal{M}_{A_P}$. Then there is a natural action of $T_P := \operatorname{Spec} k[P^{gp}]$ on $(A_P, \mathcal{M}_{A_P})$ induced by the morphism $P \to P \oplus P^{gp}$ sending p to (p, p). The log structure $\mathcal{M}_{A_P}$ descends to yield a log structure $\mathcal{M}_{[A_P/T_P]}$ on the quotient stack $[A_P/T_P]$. By [Olsson 2003, Remark 5.15], a log scheme $(Y, \mathcal{M}_Y)$ is generalized Deligne–Faltings if and only if there exists a strict morphism

$$(Y, \mathcal{M}_Y) \to ([A_P/T_P], \mathcal{M}_{[A_P/T_P]})$$

for some fine saturated sharp monoid P.

Let X be a projective normal toric variety, and let $\mathcal{M}_X$ be its standard log structure. Let $Q \subset \mathbb{R}^n$ be a polytope associated to a sufficiently positive projective embedding of X. Placing Q at height 1 in $\mathbb{R}^n \times \mathbb{R}$ and letting P be the monoid of lattice points in the cone over Q, we have $X = \operatorname{Proj} k[P]$. Note that P is fine, saturated, and sharp. Let $(A_P, \mathcal{M}_{A_P})$ be as in Remark A.2, let U be the compliment of the closed subscheme of A_P defined by the irrelevant ideal of $k[P]$, and let $\mathcal{M}_U = \mathcal{M}_{A_P}|_U$. The function $\deg : P \to \mathbb{Z}$ sending an element to its height induces a $\mathbb{G}_m$-action on $(A_P, \mathcal{M}_{A_P})$. Hence, $\mathcal{M}_U$ descends to yield a log structure $\mathcal{M}_P$ on X.

Lemma A.3. *$\mathcal{M}_P$ is generalized Deligne–Faltings.*

Proof. We have a cartesian diagram

$$\begin{array}{ccc} (U, \mathcal{M}_U) & \longrightarrow & (A_P, \mathcal{M}_P) \\ \downarrow & & \downarrow \\ (X, \mathcal{M}_P) & \longrightarrow & ([A_P/\mathbb{G}_m], \mathcal{M}_{[A_P/\mathbb{G}_m]}) \end{array}$$

where all morphisms are strict and the vertical morphisms are smooth covers. Note that the $\mathbb{G}_m$-action on $(A_P, \mathcal{M}_{A_P})$ is induced from the morphism $\sigma : P \to P \oplus \mathbb{Z}$ defined by $p \mapsto (p, \deg p)$. Since σ factors as

$$P \to P \oplus P^{gp} \to P \oplus \mathbb{Z}$$

where the first map is $p \mapsto (p, p)$ and the second is $(p, \xi) \mapsto (p, \deg \xi)$, we see that there is a strict smooth cover

$$([A_P/\mathbb{G}_m], \mathcal{M}_{[A_P/\mathbb{G}_m]}) \to ([A_P/T_P], \mathcal{M}_{[A_P/T_P]}).$$

Hence, Remark A.2 shows that $\mathcal{M}_P$ is generalized Deligne–Faltings. □

Note that $\mathcal{M}_P|_T = \mathbb{O}_T^*$, where T is the torus of X. We therefore obtain a map

$$\psi : \mathcal{M}_P \to j_*^{\log}\mathbb{O}_T^* =: \mathcal{M}_X.$$

Proposition A.4. *ψ is an isomorphism, and so $(X, \mathcal{M}_X)$ is generalized Deligne–Faltings.*

Proof. To show ψ is an isomorphism, it is enough to look Zariski locally on X. Note that X has an open cover by the $X_v := \operatorname{Spec} k[Q_v]$, where v is a vertex of the polytope Q and Q_v is the monoid of lattice points in the cone over $Q - v := \{q - v \mid q \in Q \subset \mathbb{R}^n\}$. Let P_v be the submonoid of P^{gp} generated by P and $-v$. Then we have a cartesian diagram

$$\begin{array}{ccc} A_{P_v} & \xrightarrow{i} & U \\ \pi \downarrow & & \downarrow \\ X_v & \longrightarrow & X \end{array}$$

where π is induced from the map $Q_v \to P_v$ embedding Q_v at height 0 in P_v and where the composite of i and $U \to A_P$ is induced from the inclusion $P \to P_v$. Hence,

$$\mathcal{M}_{Q_v} = (\mathcal{M}_{P_v})^{\mathbb{G}_m},$$

and so ψ is an isomorphism over X_v. □

Acknowledgments

We would like to thank Dan Abramovich, Dustin Cartwright, Anton Geraschenko, Noah Giansiracusa, and Martin Olsson. We also thank the anonymous referee for helpful comments. Chen was partially supported by the Simons Foundation. Satriano was partially supported by NSF grant DMS-0943832 and an NSF postdoctoral fellowship (DMS-1103788).

References

[Abramovich and Chen 2011] D. Abramovich and Q. Chen, "Stable logarithmic maps to Deligne–Faltings pairs, II", preprint, 2011. arXiv 1102.4531v2

[Abramovich et al. 2011] D. Abramovich, Q. Chen, W. D. Gillam, and S. Marcus, "The evaluation space of logarithmic stable maps", preprint, 2011. arXiv 1012.5416v1

[Abramovich et al. $\geq$ 2013] D. Abramovich, Q. Chen, M. Gross, and B. Siebert, in preparation.

[Altman and Kleiman 1971] A. Altman and S. L. Kleiman, "On the purity of the branch locus", *Compos. Math.* **23** (1971), 461–465. MR 46 #7233 Zbl 0242.14001

[Chen 2011] Q. Chen, "Stable logarithmic maps to Deligne–Faltings pairs, I", preprint, 2011. arXiv 1008.3090v4

[Gross and Siebert 2013] M. Gross and B. Siebert, "Logarithmic Gromov–Witten invariants", *J. Amer. Math. Soc.* **26**:2 (2013), 451–510. MR 3011419 Zbl 06168513

[Kapranov et al. 1991] M. M. Kapranov, B. Sturmfels, and A. V. Zelevinsky, "Quotients of toric varieties", *Math. Ann.* **290**:4 (1991), 643–655. MR 92g:14050 Zbl 0762.14023

[Kato 1989] K. Kato, "Logarithmic structures of Fontaine–Illusie", pp. 191–224 in *Algebraic analysis, geometry, and number theory* (Baltimore, 1988), edited by J.-I. Igusa, Johns Hopkins Univ. Press, Baltimore, MD, 1989. MR 99b:14020 Zbl 0776.14004

[Kato 1996] F. Kato, "Log smooth deformation theory", *Tohoku Math. J.* (2) **48**:3 (1996), 317–354. MR 99a:14012 Zbl 0876.14007

[Kato 2000] F. Kato, "Log smooth deformation and moduli of log smooth curves", *Internat. J. Math.* **11**:2 (2000), 215–232. MR 2001d:14016 Zbl 1100.14502

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **32**, Springer, Berlin, 1996. MR 98c:14001 Zbl 00833161

[Nishinou and Siebert 2006] T. Nishinou and B. Siebert, "Toric degenerations of toric varieties and tropical curves", *Duke Math. J.* **135**:1 (2006), 1–51. MR 2007h:14083 Zbl 1105.14073

[Nizioł 2006] W. Nizioł, "Toric singularities: log-blow-ups and global resolutions", *J. Algebraic Geom.* **15**:1 (2006), 1–29. MR 2006i:14015 Zbl 1100.14011

[Ogus 2006] A. Ogus, "Lectures on logarithmic algebraic geometry", preprint, 2006, available at http://math.berkeley.edu/∼ogus/preprints/log_book/logbook.pdf.

[Olsson 2003] M. C. Olsson, "Logarithmic geometry and algebraic stacks", *Ann. Sci. École Norm. Sup.* (4) **36**:5 (2003), 747–791. MR 2004k:14018 Zbl 1069.14022

[Olsson 2007] M. C. Olsson, "(Log) twisted curves", *Compos. Math.* **143**:2 (2007), 476–494. MR 2008d:14021 Zbl 1138.14017

[Olsson 2008] M. Olsson, "Logarithmic interpretation of the main component in toric Hilbert schemes", pp. 231–252 in *Curves and abelian varieties*, edited by V. Alexeev et al., Contemp. Math. **465**, Amer. Math. Soc., Providence, RI, 2008. MR 2009j:14007 Zbl 1153.14007

[Peeva and Stillman 2002] I. Peeva and M. Stillman, "Toric Hilbert schemes", *Duke Math. J.* **111**:3 (2002), 419–449. MR 2003m:14008 Zbl 1067.14005

Communicated by Ravi Vakil
Received 2012-10-22 Revised 2013-02-04 Accepted 2013-03-12

q_chen@math.columbia.edu *Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027, United States*

satriano@umich.edu *Department of Mathematics, University of Michigan, 2074 East Hall, Ann Arbor, MI 48109, United States*

ALGEBRA AND NUMBER THEORY 7:9 (2013)
dx.doi.org/10.2140/ant.2013.7.2331

Vinberg's representations and arithmetic invariant theory

Jack A. Thorne

Recently, Bhargava and others have proved very striking results about the average size of Selmer groups of Jacobians of algebraic curves over $\mathbb{Q}$ as these curves are varied through certain natural families. Their methods center around the idea of counting integral points in coregular representations, whose rational orbits can be shown to be related to Galois cohomology classes for the Jacobians of these algebraic curves.

In this paper we construct for each simply laced Dynkin diagram a coregular representation (G, V) and a family of algebraic curves over the geometric quotient $V//G$. We show that the arithmetic of the Jacobians of these curves is related to the arithmetic of the rational orbits of G. In the case of type A_2, we recover the correspondence between orbits and Galois cohomology classes used by Birch and Swinnerton-Dyer and later by Bhargava and Shankar in their works concerning the 2-Selmer groups of elliptic curves over $\mathbb{Q}$.

1. Introduction

This paper is a contribution to *arithmetic invariant theory*. Let G be a reductive group over a field k, and let V be a linear representation of G. Then the ring $k[V]^G$ is a k-algebra of finite type, and we can define the quotient $V/\!\!/G = \operatorname{Spec} k[V]^G$ and a quotient map $\pi : V \to V/\!\!/G$. The determination of the structure of $k[V]^G$ and

This research was partially conducted during the period the author served as a Clay Research Fellow.
MSC2010: primary 20G30; secondary 11E72.
Keywords: Arithmetic invariant theory, Galois cohomology, Arithmetic of algebraic curves.

the fibers of π falls under the rubric of geometric invariant theory, and is important in algebraic geometry.

In the case where k is not algebraically closed, a further layer of difficulty is obtained by considering the $G(k)$-orbits in the fibers of π over k-points of $V/\!\!/G$. This problem can be translated into the language of Galois cohomology, and as such often has close ties to arithmetic.

Bhargava has singled out those representations which are coregular, in the sense that $k[V]^G$ is isomorphic to a polynomial ring, as promising candidates for representations which may have interesting connections to arithmetic. For example, he has studied together with Shankar the case $G = \mathrm{SL}_2$ and $V = \mathrm{Sym}^4 2^\vee$, the space of binary quartic forms. In this case there are two independent polynomial invariants I and J, and k-rational orbits with given values of I and J are related to classes in the Galois cohomology group $H^1(k, E[2])$ for the elliptic curve

$$E\colon y^2 = x^3 + Ix + J.$$

These considerations have had very striking applications; see [Bhargava and Shankar 2010], or [Poonen 2013] for a beautiful summary. See also [Ho 2009] for a variety of similar orbit parametrizations associated to other representations, and [Bhargava and Ho 2013] for an exhaustive study of coregular representations related to genus-one curves. For each choice of pair (G, V), one makes a construction in algebraic geometry which relates orbits in the given representation to algebraic curves, possibly with marked points, given line bundles, or other types of extra data.

By contrast, this paper represents a first effort to describe some of the phenomena appearing in arithmetic invariant theory through the lens of representation theory. We take as our starting point certain representations arising from Vinberg theory, whose role in arithmetic invariant theory has been emphasized by Gross. If G is a reductive group over k endowed with an automorphism θ of finite order m, then the fixed group G^θ acts on the $\theta = \zeta$ eigenspace $\mathfrak{g}_1 \subset \mathfrak{g} = \mathrm{Lie}\, G$ for any choice $\zeta \in k^\times$ of primitive m-th root of unity. Vinberg theory describes the geometric invariant theory of these representations. In the case when θ is regular and elliptic, in the sense of [Reeder et al. 2012], the generic element of $\mathfrak{g}_1$ will have a finite abelian stabilizer, and orbits in the representation are thus related to interesting Galois cohomology.

If G is a split reductive group over k, then it has a unique $G^{\mathrm{ad}}(k)$-conjugacy class of regular elliptic involutions θ, characterized by the requirement that $\mathfrak{g}_1$ contain a regular nilpotent element. It is the representations associated to these canonical involutions for simple G of type A, D or E that we study in this paper. We associate to each of these groups a family of algebraic curves, namely the smooth nearby fibers of a semiuniversal deformation of the corresponding simple plane curve

singularity. The arithmetic of the Jacobians of these curves turns out to be related to the arithmetic of the rational orbits in the Vinberg representations. In each case the families of curves are universal families with marked points of fixed type. In types A and D we obtain families of hyperelliptic curves, while in types E_6, E_7 and E_8 we obtain families of nonhyperelliptic curves of genus respectively 3, 3 and 4.

Remark 1.1. In the forthcoming work [Bhargava and Ho 2013], the authors construct families of related coregular representations through the operations of symmetrization and skew-symmetrization. These representations are all related to the arithmetic of curves of genus one. For example, they consider the natural representation of $\mathrm{SL}_2 \times \mathrm{SL}_2 \times \mathrm{SL}_2 \times \mathrm{SL}_2$ on $2 \otimes 2 \otimes 2 \otimes 2$, and its quadruple symmetrization yields the representation of SL_2 on the space of binary quartic forms described above.

The quadruple skew-symmetrization, however, is not directly related to curves of genus one. In fact, the Vinberg representation we associate to E_7 is the representation of SL_8/μ_4 on $\wedge^4 8$. This is the quadruple skew-symmetrization of the above representation of SL_2^4, and our work shows that its orbits are related to the arithmetic of the universal family of nonhyperelliptic curves of genus 3 with a rational flex in the canonical embedding.

Results. Let us now turn to a precise statement of our main results. For any unfamiliar notation relating to algebraic groups and their Lie algebras, we refer to the section on notation (page 2335). Let k be a field of characteristic zero, and let G be a split adjoint group over k of type A, D or E. We choose a regular elliptic involution θ of G as described above, and set $G_0 = (G^\theta)^\circ$, $\mathfrak{g}_1 = \mathfrak{g}^{\theta=-1}$. Then G_0 acts on $\mathfrak{g}_1$ and a Chevalley-type restriction theorem holds for the pair $(G_0, \mathfrak{g}_1)$. In particular, the space $B = \mathfrak{g}_1 /\!\!/ G_0$ is isomorphic to affine r-space: $\mathfrak{g}_1$ is coregular. We write $\Delta \subset B$ for the discriminant divisor. Thus Δ is the image under $\pi : \mathfrak{g}_1 \to B$ of the set of elements which are not regular semisimple.

Proposition 2.27 below implies that $\mathfrak{g}_1$ contains subregular nilpotent elements. We choose a subregular normal $\mathfrak{sl}_2$-triple (e, h, f). (See Definition 2.16 for the definition of a normal $\mathfrak{sl}_2$-triple. To say that it is subregular simply means that e and f are subregular nilpotent elements of $\mathfrak{g}$.) Define $X = e + \mathfrak{z}_{\mathfrak{g}}(f)_1 = e + \mathfrak{z}_{\mathfrak{g}}(f) \cap \mathfrak{g}_1$. Our first theorem concerns the natural map $X \hookrightarrow \mathfrak{g}_1 \to B$.

Theorem 1.2. *The morphism $X \to B$ is a flat family of reduced connected curves, smooth away from Δ. The equations of these curves are given in the statement of Theorem 3.8.*

In fact, X is a transverse slice to the G_0-orbit of e inside $\mathfrak{g}_1$, and $X \to B$ realizes a semiuniversal deformation of the central fiber X_0, which is an affine plane curve with a unique simple singularity of type equal to that of G. (For the definition of a simple curve singularity, we refer, for example, to [Cook 1998].)

The following theorem incorporates results of Section 2 and Theorem 4.10.

Theorem 1.3. *Let $x \in \mathfrak{g}_1$ be a regular semisimple element; equivalently, suppose that $b = \pi(x) \in B(k)$ does not lie inside Δ.*

- *The stabilizer $Z_{G_0}(x) = Z_b$ is a finite abelian k-group and depends only on b up to canonical isomorphism. This group is endowed with a nondegenerate alternating pairing $Z_b \times Z_b \to \mu_2$.*
- *Let Y_b denote the smooth projective curve containing X_b as a dense open subset. Let J_{Y_b} denote its Jacobian variety. Then there is a canonical isomorphism of finite k-groups $J_{Y_b}[2] \cong Z_b$. Under this isomorphism the above pairing corresponds to the Weil pairing of J_{Y_b}.*

Given $b \in (B \setminus \Delta)(k)$, we write $\mathfrak{g}_{1,b} = \pi^{-1}(b)$. If K is a separable closure of k, then $\mathfrak{g}_{1,b}(K)$ consists of a single $G_0(K)$-orbit; the rational orbits in $\mathfrak{g}_{1,b}(k)$ are therefore classified by a suitable Galois cohomology set, with coefficients in $Z_b \cong J_{Y_b}[2]$. The inclusion $X_b \subset \mathfrak{g}_{1,b}$ induces a map on rational points $X_b(k) \to \mathfrak{g}_{1,b}(k)/G_0(k)$. Our main theorem asserts that this map can in fact be interpreted in terms of 2-descent on the Jacobian J_{Y_b}:

Theorem 1.4. *There is a commutative diagram, functorial in k:*

$$\begin{tikzcd} X_b(k) \arrow[r] \arrow[d] & \mathfrak{g}_{1,b}(k)/G_0(k) \arrow[d] \\ J_{Y_b}(k) \arrow[r] & H^1(k, J_{Y_b}[2]). \end{tikzcd}$$

For the definitions of the arrows in this diagram, we refer to the statement of Theorem 4.15. If $G = \mathrm{PGL}_3$, then the family of curves $X \to B$ is the family $y^2 = x^3 + Ix + J$ of genus-one curves described above, and we then recover the correspondence between orbits and Galois cohomology classes used by Bhargava and Shankar in their work on the average size of the 2-Selmer group of an elliptic curve over $\mathbb{Q}$. We are hopeful that the ideas discussed in this paper will have applications to the study of the average size of 2-Selmer groups beyond this case; compare the discussion following Conjecture 4.16.

Methods. Our methods are inspired primarily by work of Slodowy. Rational double point singularities of surfaces can be classified in terms of the Dynkin diagrams of simply laced simple algebraic groups. Grothendieck conjectured that one could give a representation-theoretic construction of this correspondence by looking at the generic singularity of the nilpotent cone of the corresponding group G. A proof of this conjecture was announced in a famous ICM lecture of Brieskorn [1971], but the first detailed proofs were given by Esnault [1980] and Slodowy [1980b]. Our work is what one obtains on combining the respective ideas of Slodowy and Vinberg.

(After this work was completed, J. Sekiguchi informed us of some related earlier work [Sekiguchi and Shimizu 1981] where our families of curves also appear.)

Let us say a few words about the limits of our methods. Essential to our work is the use of $\mathfrak{sl}_2$-triples, whose existence relies in turn on the Jacobson–Morozov lemma. We must therefore work over a field of sufficiently large characteristic, relative to the Coxeter number of G. In this paper we choose for simplicity to work over a field of characteristic zero.

More serious is the lack of information we obtain about the image of the map $\mathfrak{g}_{1,b}(k) \to H^1(k, J_{Y_b}[2])$ constructed above. It follows from the above considerations that it contains the elements in the image under the 2-descent map $\delta\colon J_{Y_b}(k) \to H^1(k, J_{Y_b}[2])$ of $X_b(k)$; we conjecture (page 2363) that it moreover contains the image under δ of the whole group $J_{Y_b}(k)$ of rational points of the Jacobian. In other words, we currently lack a way to construct sufficiently many orbits in the representations we study. We hope to return to this question in a future work.

Outline. Let us now outline the contents of this paper. In Section 2, we prove some basic properties of the so-called stable involutions θ, and define the Vinberg representations to which they correspond. An important point here is the calculation of the stabilizers of the regular elements in $\mathfrak{g}_1$ in terms of the root datum of the ambient reductive group G. We also introduce the subregular nilpotent elements, and address the question of when $\mathfrak{g}_1$ contains subregular nilpotent elements which are defined over the base field k.

In Section 3, we construct the families of curves mentioned above inside a suitable transverse slice to the subregular nilpotent orbit.

Finally, in Section 4, we show how to relate the 2-torsion in the Jacobians of our curves and the stabilizers of regular elements, and prove our main theorem relating the 2-descent map to the classifying map for orbits in non-abelian Galois cohomology.

Other groups. In this paper we restrict to simple groups G arising from simply laced Dynkin diagrams, and the corresponding Vinberg representations. One can try to apply our constructions to groups of nonsimply laced type. The families of curves thus obtained are versal deformations of planar curve singularities "with fixed symmetries"; this is the direct analogue for our context of the results in [Slodowy 1980b, §6.2]. The Jacobians of these curves admit a family of isogenies ϕ, and it seems likely that some version of our main result continues to hold, with the groups $J_{Y_b}[\phi]$ now playing the role of the groups $J_{Y_b}[2]$.

Notation. As mentioned above, we work throughout over a field k of characteristic zero. We assume basic familiarity with the theory of reductive groups over k, as

studied, for example, in [Humphreys 1975] or [Springer 2009]. We assume that reductive groups are connected.

If G is a reductive group acting linearly on a k-vector space V, then the ring of invariants $k[V]^G$ is a k-algebra of finite type (see, for example, [Springer 1977, Theorem 2.4.9]). We define $V/\!\!/G = \operatorname{Spec} k[V]^G$ and call it the categorical quotient. It in fact satisfies a universal property, but we will not need this here. We will write $\mathcal{N}(V)$ for the closed subscheme of V cut out by the augmentation ideal of $k[V]^G$.

If $G, H, \dots$ are algebraic groups then we will use gothic letters $\mathfrak{g}, \mathfrak{h}, \dots$ to denote their Lie algebras. Let G be a reductive group, and $T \subset G$ a split maximal torus. Then we shall write $\Phi_{\mathfrak{t}} \subset X^*(T)$ for the set of roots of T in $\mathfrak{g}$, and $\Phi_{\mathfrak{t}}^\vee \subset X_*(T)$ for the set of coroots. The assignment $\alpha \in \Phi_{\mathfrak{t}} \mapsto d\alpha \in \mathfrak{t}^*$ identifies $\Phi_{\mathfrak{t}}$ with the set of roots of $\mathfrak{t}$ in $\mathfrak{g}$, and we will use this identification without comment. We write $W(\mathfrak{t}) = N_G(T)/T$ for the Weyl group of G with respect to $\mathfrak{t}$. We have the Cartan decomposition

$$\mathfrak{g} = \mathfrak{t} \oplus \bigoplus_{\alpha \in \Phi_{\mathfrak{t}}} \mathfrak{g}^\alpha,$$

where $\dim \mathfrak{g}^\alpha = 1$ for each $\alpha \in \Phi_{\mathfrak{t}}$. We write $U_\alpha \subset G$ for the unique T-invariant closed subgroup with Lie algebra $\mathfrak{g}_\alpha$ (see [Humphreys 1975, §26.3]). The tuple

$$\big(X^*(T), \Phi_{\mathfrak{t}}, X_*(T), \Phi_{\mathfrak{t}}^\vee\big)$$

is a root datum in the sense of [Springer 2009, §7.4]. We write A_G for the center of G, and $\mathfrak{a}_{\mathfrak{g}}$ for its Lie algebra.

We will write $L_G = \mathbb{Z}\Phi_{\mathfrak{t}}$ for the root lattice of G and $\Lambda_G \subset L_G \otimes_{\mathbb{Z}} \mathbb{Q}$ for the weight lattice of L_G. (These are the groups Q and P, respectively, of [Bourbaki 1968, Chapter VI, §1.9].) If the group G is clear from the context, we will omit the subscript G. We understand these to depend only on G and not on T, so that L_G and Λ_G are defined up to (nonunique) isomorphism. We write $W_G \subset \operatorname{Aut}(L_G)$ for the corresponding Weyl group.

If $x \in \mathfrak{g}$, we write $Z_G(x)$ for its centralizer in G under the adjoint representation, and $\mathfrak{z}_{\mathfrak{g}}(x)$ for its centralizer in $\mathfrak{g}$. If x is semisimple, then $Z_G(x)$ is reductive. Let $T \subset G$ be a maximal torus, and suppose that $x \in \mathfrak{t}$. Then $T \subset Z_G(x)$ is a maximal torus. Let

$$\Phi_{\mathfrak{t}}(x) = \{\alpha \in \Phi_{\mathfrak{t}} \mid \alpha(x) = 0\} \quad \text{and} \quad \Phi_{\mathfrak{t}}^\vee(x) = \{\alpha^\vee \in \Phi_{\mathfrak{t}}^\vee \mid \alpha \in \Phi_{\mathfrak{t}}(x)\}.$$

Let $W(x) = Z_{W(\mathfrak{t})}(x)$. Then the root datum of $Z_G(x)$ is

$$\big(X^*(T), \Phi_{\mathfrak{t}}(x), X_*(T), \Phi_{\mathfrak{t}}^\vee(x)\big),$$

and the Weyl group of $Z_G(x)$ with respect to T can be identified in a natural way with $W(x)$.

2. Preliminaries: Vinberg theory, stable involutions, subregular elements

Throughout this section, G is a split reductive group over a field k of characteristic zero.

Elements of Vinberg theory. Let $\theta \in \mathrm{Aut}(G)$ be an automorphism of exact order $m > 1$, and let $\zeta \in k$ be a primitive m-th root of unity. We will also write θ for the induced automorphism of $\mathfrak{g}$. We associate to θ the grading $\mathfrak{g} = \oplus_{i \in \mathbb{Z}/m\mathbb{Z}} \mathfrak{g}_i$, where by definition we have

$$\mathfrak{g}_i = \{x \in \mathfrak{g} \mid \theta(x) = \zeta^i x\}.$$

We write G^θ for the fixed subgroup of θ, and G_0 for its connected component. Then $\mathrm{Lie}\, G_0 = \mathfrak{g}_0$, so the notation is consistent. The action of G^θ on $\mathfrak{g}$ leaves each $\mathfrak{g}_i$ invariant.

In what follows, we shall consider the representation of G_0 on the subspace $\mathfrak{g}_1 \subset \mathfrak{g}$. The study of such representations is what we call Vinberg theory. For the basic facts about Vinberg theory, and in particular for proofs of the unproved assertions in this section, we refer to the papers [Vinberg 1976] or [Levy 2009].

Lemma 2.1. *Let $x \in \mathfrak{g}_1$. Then x can be written uniquely as $x = x_s + x_n$, where x_s, x_n both lie in $\mathfrak{g}_1$ and are respectively semisimple and nilpotent.*

Definition 2.2. A Cartan subspace $\mathfrak{c} \subset \mathfrak{g}_1$ is a maximal subalgebra consisting of semisimple elements. Note that $\mathfrak{c}$ is automatically abelian.

Proposition 2.3. *Suppose that k is algebraically closed. Then an element $x \in \mathfrak{g}_1$ is semisimple if and only if it is contained in a Cartan subspace, and all Cartan subspaces are $G_0(k)$-conjugate.*

Let $\mathfrak{c} \subset \mathfrak{g}_1$ be a Cartan subspace, and define $W(\mathfrak{c}, \theta) = N_{G_0}(\mathfrak{c})/Z_{G_0}(\mathfrak{c})$. This is the "little Weyl group" of the pair (G, θ). We define $\mathrm{rank}\, \theta = \dim \mathfrak{c}$. This is well-defined by Proposition 2.3.

The following result is contained in [Panyushev 2005, Theorem 1.1]. It is Vinberg's main result concerning the invariant theory of the representations considered here.

Theorem 2.4. 1. *Restriction of functions induces an isomorphism*

$$k[\mathfrak{g}_1]^{G_0} \to k[\mathfrak{c}]^{W(\mathfrak{c},\theta)}.$$

Moreover, $W(\mathfrak{c}, \theta)$ is a (pseudo-)reflection group and $k[\mathfrak{c}]^{W(\mathfrak{c},\theta)}$ is a polynomial ring in $\mathrm{rank}\, \theta$ *indeterminates.*

2. *Let $\pi : \mathfrak{g}_1 \to \mathfrak{g}_1 /\!\!/ G_0$ denote the quotient map. Then π is flat. If k is algebraically closed, then for all $x \in \mathfrak{g}_1$, $\pi^{-1}\pi(x)$ consists of only finitely many $G_0(k)$-orbits.*

We say that $v \in \mathfrak{g}_1$ is stable if $G_0 \cdot v$ is closed in $\mathfrak{g}_1$, and $Z_{G_0}(v)$ is finite. We say that θ is stable if $\mathfrak{g}_1$ contains stable elements. The property of being stable is hereditary, in the following sense.

Lemma 2.5. *Suppose that θ is a stable automorphism. Let $x \in \mathfrak{g}_1$ be semisimple. Let $H = Z_G(x)$ and $\mathfrak{h} = \mathrm{Lie}\, H$. Then $\theta(H) = H$, and $\theta|_H$ is a stable automorphism.*

Proof. Given x as in the lemma, choose a Cartan subspace $\mathfrak{c}$ containing it. Then $\mathfrak{c}$ contains a stable vector, which is also stable when considered as an element of $\mathfrak{h}$; the result follows. □

Stable involutions. In this paper we shall be particularly interested in the stable involutions.

Lemma 2.6. *Suppose that k is algebraically closed. There is a unique $G(k)$-conjugacy class of stable involutions θ.*

Proof. To show uniqueness, we reduce immediately to the case that G is adjoint. By [Reeder et al. 2012, Lemma 5.6], any stable vector $v \in \mathfrak{g}_1$ is regular semisimple, and θ acts as -1 on its centralizer $\mathfrak{c} = \mathfrak{z}_{\mathfrak{g}}(v)$. In particular, we have $\mathfrak{c} \subset \mathfrak{g}_1$. It follows that the trace of θ on $\mathfrak{g}$ is equal to $-\dim \mathfrak{c} = -\operatorname{rank} G$, and a well-known theorem of E. Cartan asserts that this determines θ up to $G(k)$-conjugacy. We can also reduce existence to the case of G adjoint. We will prove existence (even when k is not algebraically closed) in this case below. □

Lemma 2.7. *Let θ be a stable involution of G. Then θ satisfies the following.*

1. $\operatorname{rank} \theta = \operatorname{rank} G$.
2. *There exists a maximal torus C in G on which θ acts by $x \mapsto x^{-1}$.*
3. *For all $x \in A_G$, we have $\theta(x) = x^{-1}$.*
4. *Let $\mathfrak{c}$ be a Cartan subspace (and hence, a Cartan subalgebra). Then the natural map $W(\mathfrak{c}, \theta) \to W(\mathfrak{c})$ is an isomorphism.*

Proof. The first and second properties follow from the proof of Lemma 2.6. For the third property, we recall that A_G is contained in any maximal torus of G. The final property is [Reeder et al. 2012, Corollary 7.4]. □

Suppose for the rest of this section that θ is a stable involution.

Proposition 2.8. *Let $x = x_s + x_n \in \mathfrak{g}_1$ be a regular element. Then $Z_{G^\theta}(x) = A_{Z_G(x_s)}[2]$. In particular, this group is always finite and abelian.*

Proof. We have $Z_G(x) = Z_G(x_s) \cap Z_G(x_n)$, so after replacing G by $Z_G(x_s)$, we may assume that $x = x_n$ is a regular nilpotent element.

Then $Z_G(x) = A_G \cdot Z_U(x)$, a direct product, where U is the unipotent radical of the unique Borel subgroup containing x. Quotienting by A_G, we may suppose that

G is adjoint and must show that $Z_U(x)^\theta$ is trivial. But since x is regular, this is a finite unipotent group, so the result follows. □

Corollary 2.9. *Let $x = x_s + x_n$ be a regular element, and let $\mathfrak{c}$ be a Cartan subspace containing x_s. Let $C \subset G$ denote the maximal torus with Lie algebra $\mathfrak{c}$. Then*

$$Z_{G^\theta}(x) \cong \mathrm{Hom}\big(X^*(C)/2X^*(C) + \mathbb{Z}\Phi_{\mathfrak{c}}(x), \mathbb{G}_m\big).$$

Proof. For any reductive group G with root datum $\big(X^*(T), \Phi_{\mathfrak{t}}, X_*(T), \Phi_{\mathfrak{t}}^\vee\big)$, there is a canonical isomorphism $X^*(A_G) \cong X^*(T)/\mathbb{Z}\Phi_{\mathfrak{t}}$. Now apply Proposition 2.8. □

Corollary 2.10. *Suppose that G is adjoint and that k is algebraically closed. Let $x \in \mathfrak{g}_1$ be a regular semisimple element. Let L denote the root lattice of G, and $\Lambda \subset L \otimes_{\mathbb{Z}} \mathbb{Q}$ the weight lattice. Then there is an isomorphism*

$$Z_{G_0}(x) \cong \mathrm{Hom}(N, \mathbb{G}_m),$$

well-defined up to conjugacy by the Weyl group W of L, where N denotes the image of L in $\Lambda/2\Lambda$.

Proof. Let G^{sc} denote the simply connected cover of G. Then θ acts on G^{sc}. A theorem of Steinberg — [Onishchik and Vinberg 1988, Chapter 4.4.8, Theorem 9] — states that $(G^{\mathrm{sc}})^\theta$ is connected, and hence G_0 is the image of the map $(G^{\mathrm{sc}})^\theta \to G$. The present corollary now follows from the previous one. □

Now suppose that the simple components of G are simply laced (that is, their root systems are all of type A, D, or E), and let L, Λ and W be as in the statement of the corollary. Then there is a W-invariant quadratic form $\langle\cdot,\cdot\rangle\colon L \times L \to \mathbb{Z}$ uniquely determined by the requirement that $\langle\alpha,\alpha\rangle = 2$ for every root α. The pairing $\langle\cdot,\cdot\rangle$ on L induces a pairing $(\cdot,\cdot)\colon L/2L \times L/2L \to \mathbb{F}_2$. An easy calculation shows this pairing is alternating. In fact, we have the following:

Lemma 2.11. *The pairing $(\cdot,\cdot)$ descends to a nondegenerate alternating pairing on N.*

Proof. Suppose $x \in L$. Then the image of x in $L/2L$ lies in the radical of $(\cdot,\cdot)$ if and only if $\langle x, L\rangle \subset 2\mathbb{Z}$, if and only if $x \in 2\Lambda$, since Λ is the $\mathbb{Z}$-dual of L with respect to the pairing $\langle\cdot,\cdot\rangle$. □

Pairings of this type, associated to regular elliptic elements of Weyl groups, were first considered in [Reeder 2011].

Corollary 2.12. *Suppose that G is an adjoint group, and that the simple components of G are simply laced. Then for any regular semisimple element $x \in \mathfrak{g}_1$, there is a canonical nondegenerate alternating form $(\cdot,\cdot)\colon Z_{G_0}(x) \times Z_{G_0}(x) \to \mu_2$.*

We now show how to construct a stable involution over an arbitrary field k of characteristic 0. We let G be a simple split adjoint group, and fix a split maximal torus T and a Borel subgroup B containing it. This determines a set $\Phi^+ \subset \Phi = \Phi_{\mathfrak{t}}$ of positive roots, and a root basis $R \subset \Phi^+$. We fix moreover for each $\alpha \in R$ a basis X_α of the one-dimensional vector space $\mathfrak{g}^\alpha \subset \mathfrak{g}$. The tuple $(T, B, \{X_\alpha\}_{\alpha \in R})$ is called a pinning of G.

This choice of data determines a splitting $\mathrm{Aut}(G) = G \rtimes \Sigma$, where Σ is the group of *pinned automorphisms* induced by automorphisms of the Dynkin diagram of G. On the other hand, writing $L = X^*(T) = \mathbb{Z}\Phi$ for the root lattice of $\mathfrak{g}$, the choice of root basis determines a splitting $\mathrm{Aut}(L) = W \rtimes \Sigma$ in a similar manner; see [Bourbaki 1975, Chapter VIII, §5.2]. We write $\sigma \in \Sigma$ for the image of $-1 \in \mathrm{Aut}(L)$, and define $\theta = \rho^\vee(-1) \rtimes \sigma \in \mathrm{Aut}(G)(k)$, where $\rho^\vee \in X_*(T)$ is the sum of the fundamental coweights.

Lemma 2.13. *The automorphism θ is a stable involution.*

Proof. This follows immediately from Corollary 5.7 of [Reeder et al. 2012]. □

This stable involution has good rationality properties. This is based on the following fact.

Lemma 2.14. *Let θ be as above. Then $\mathfrak{g}_1$ contains a regular nilpotent element. Any two regular nilpotent elements of $\mathfrak{g}_1$ are conjugate by a unique element of $G^\theta(k)$.*

Proof. The element $\sum_{\alpha \in R} X_\alpha$ is regular nilpotent and, by construction, lies in $\mathfrak{g}_1$. Fix a separable closure K of k. If $E, E' \in \mathfrak{g}_1$ are two regular nilpotent elements then they are conjugate by an element of $G^\theta(K)$. (This follows from [Levy 2007, Theorem 5.16].)

For any such E, the group $Z_{G^\theta}(E)$ is a finite unipotent group, and therefore trivial. It follows that E, E' are conjugate by a unique element of $G^\theta(K)$, which must therefore lie in $G^\theta(k)$. □

Corollary 2.15. *There is a unique $G(k)$-conjugacy class of stable involutions θ_1 of G such that there exists a regular nilpotent element $E_1 \in \mathfrak{g}$ with $\theta_1(E_1) = -E_1$.*

Proof. We have already proved the existence of such an element. For the uniqueness, fix again a separable closure K of k. We have seen that $G(K)$ acts transitively on pairs (θ_1, E_1). On the other hand, the stabilizer of such a pair in $G(K)$ is trivial. It follows that any two such pairs are conjugate by a unique element of $G(k)$. □

Definition 2.16. We call a tuple (E, H, F) of elements of $\mathfrak{g}$ a normal $\mathfrak{sl}_2$-triple if it is an $\mathfrak{sl}_2$-triple, and moreover we have $E \in \mathfrak{g}_1$, $H \in \mathfrak{g}_0$, and $F \in \mathfrak{g}_1$.

Note that if (E, H, F) is a normal $\mathfrak{sl}_2$-triple, then the restriction of θ to the subalgebra spanned by these elements is a stable involution.

Lemma 2.17. 1. *Any nilpotent element $E \in \mathfrak{g}_1$ is contained in a normal $\mathfrak{sl}_2$-triple.*

2. *Any two normal $\mathfrak{sl}_2$-triples (E, H, F) and (E, H', F') are $Z_{G_0}(E)(k)$-conjugate.*

Proof. Fix a separable closure K of k. For the first part, choose an arbitrary $\mathfrak{sl}_2$-triple (E, h, f) containing E, and decompose $h = h_0 + h_1$ into θ-eigenvectors. The argument of [Kostant and Rallis 1971, Proposition 4] implies that there is a unique $F \in \mathfrak{g}_1 \otimes_k K$ such that (E, h_0, F) is an $\mathfrak{sl}_2$-triple. But an $\mathfrak{sl}_2$-triple is determined uniquely by any 2 of its 3 elements, so descent implies that $F \in \mathfrak{g}_1$, and (E, h_0, F) is the desired triple.

For the second part, we argue as in the proof of [Kostant and Rallis 1971, Proposition 4] and apply [Bourbaki 1975, Chapter VIII, §11.1, Lemma 4] to obtain the desired rationality property. □

Corollary 2.18. *The group $G(k)$ acts simply transitively on the set of pairs*

$$\big((\theta_1), (E, H, F)\big),$$

where θ_1 is a stable involution of G and (E, H, F) is a normal $\mathfrak{sl}_2$-triple with respect to θ_1 in which E is a regular nilpotent element.

Example 2.19. We illustrate some of the concepts introduced so far in the case where G is a split adjoint group of type A_{2r}. Let V be a vector space of dimension $2r+1$, with basis $\{e_1, e_2, \dots, e_r, v, f_r, \dots, f_2, f_1\}$. We define an inner product $\langle \cdot, \cdot \rangle$ on V by the formulae

$$\langle e_i, e_j \rangle = 0 = \langle f_i, f_j \rangle = \langle e_i, v \rangle = \langle f_i, v \rangle$$

for all i, j and

$$\langle v, v \rangle = 1, \quad \langle e_i, f_j \rangle = \delta_{ij}.$$

If $T \in \mathrm{End}(V)$, write T^* for the adjoint of T with respect to this inner product. Then we take $G = \mathrm{PGL}_{2r+1} = \mathrm{PGL(V)}$, and $\theta : \mathfrak{sl}_{2r+1} \to \mathfrak{sl}_{2r+1}$ to be the involution $X \mapsto -X^*$. It is easy to check that $-\theta$ is just reflection in the antidiagonal. In particular, fixing the standard pinning $(T, B, \{X_\alpha\}_{\alpha \in R})$ of $\mathfrak{sl}_{2r+1}$, this θ is exactly the stable involution constructed of Lemma 2.13.

Then we see that $G^\theta = G_0 = \mathrm{SO(V)}$ is connected, and we have

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1, \quad \mathfrak{g}_0 = \{X \in \mathrm{End}(V) \mid \mathrm{tr}\, X = 0, X = -X^*\} = \mathfrak{so}(V).$$

In particular, $\mathfrak{g}_1 = \{X \in \mathrm{End}(V) \mid \mathrm{tr}\, X = 0, X = X^*\}$ consists of the space of trace zero operators self-adjoint with respect to $\langle \cdot, \cdot \rangle$.

The regular nilpotent element determined by the pinning is

$$E = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

These representations of odd orthogonal groups are exactly the ones used in [Bhargava and Gross 2013] to deal with the Selmer groups of hyperelliptic Jacobians.

Subregular elements. We recall that $x \in \mathfrak{g}$ is called subregular if $\dim \mathfrak{z}_{\mathfrak{g}}(x) = \operatorname{rank} G + 2$.

Proposition 2.20. *The Lie algebra $\mathfrak{g}$ contains subregular nilpotent elements. Suppose that G is simple and that k is algebraically closed. Then there is a unique $G(k)$-orbit of subregular nilpotent elements in $\mathfrak{g}$, and these are dense in the complement of the regular nilpotent orbit in the nilpotent variety of $\mathfrak{g}$.*

Proof. This follows from [Steinberg 1974, §3.10, Theorem 1]. □

Thus if $\mathfrak{g}$ is simple, then its nilpotent variety has a unique open orbit, consisting of regular nilpotent elements; its complement again has a unique open orbit, consisting of the subregular nilpotents. If $\mathfrak{g} = \mathfrak{g}_1 \times \cdots \times \mathfrak{g}_s$ is a product of simple Lie algebras, then any nilpotent element n can be written uniquely as a sum $n = n_1 + \cdots + n_s$, where $n_i \in \mathfrak{g}_i$. It is then easy to see that n is regular if and only if each n_i is regular in $\mathfrak{g}$; and n is subregular if and only if some n_i is subregular in $\mathfrak{g}_i$, and all other n_j are regular nilpotent elements. In particular, when k is algebraically closed there are exactly s $G(k)$-orbits of subregular nilpotent elements, and there is a canonical bijection between these and the set of connected components of the Dynkin diagram of $\mathfrak{g}$.

Now suppose that θ is a stable involution of G. Before we continue, it is helpful to note the following.

Lemma 2.21. *Let $x \in \mathfrak{g}_1$. Then*

$$\dim \mathfrak{z}_{\mathfrak{g}_0}(x) = (\dim \mathfrak{z}_{\mathfrak{g}}(x) - \operatorname{rank} G)/2 \quad \textit{and} \quad \dim G_0 \cdot x = (\dim G \cdot x)/2.$$

Proof. This follows from [Kostant and Rallis 1971, Proposition 5]. □

Our next goal is to show that $\mathfrak{g}_1$ contains subregular nilpotent elements. We use a trick based on the Kostant–Sekiguchi correspondence, which we now recall:

Theorem 2.22. *Suppose that $k = \mathbb{R}$ and that G is semisimple. Let τ be a Cartan involution of G. Then each of the following three sets is in canonical bijection with the others*:

1. *The set of nilpotent $G(\mathbb{R})^\circ$-orbits in $\mathfrak{g}$.*
2. *The set of nilpotent $G^\tau(\mathbb{C})^\circ$-orbits in $\mathfrak{g}^{\tau=-1} \otimes_{\mathbb{R}} \mathbb{C}$.*
3. *The set of nilpotent $G_0(\mathbb{C})$-orbits in $\mathfrak{g}_1 \otimes_{\mathbb{R}} \mathbb{C}$.*

(*Here we write $G(\mathbb{R})^\circ$ and $G^\tau(\mathbb{C})^\circ$ for the connected components of these groups in the analytic topology.*) *The map $G(\mathbb{R})^\circ \cdot X \mapsto G^\tau(\mathbb{C})^\circ \cdot X'$ satisfies $G(\mathbb{C}) \cdot X = G(\mathbb{C}) \cdot X'$.*

Proof. The bijection between the first two sets is constructed in [Collingwood and McGovern 1993, §9.5]. The existence of the bijection between the latter two follows since τ is a stable involution, and all such are conjugate over $\mathbb{C}$. □

Corollary 2.23. *Suppose that k is algebraically closed. Then $\mathfrak{g}_1$ contains subregular nilpotent elements.*

Proof. This is implied by Theorem 2.22 since, if $k = \mathbb{R}$ and $\mathfrak{g}$ is split, all conjugacy classes of nilpotent elements have an element defined over k. □

To obtain more information, we must argue on a case-by-case basis. For the rest of this section, we assume that G is adjoint, and that $\mathfrak{g}_1$ contains a regular nilpotent element. We first recall the following (see [Slodowy 1980b, §7.5, Lemma 4]).

Proposition 2.24. *Suppose that G is simple and simply laced, and let $x \in \mathfrak{g}$ be a subregular nilpotent element. Then $Z_G(x)$ is the semidirect product of a unipotent group with either $\mathbb{G}_m$ (if G is type A_r) or the trivial group (if G is of type D_r or E_r). In particular, this centralizer is connected.*

Corollary 2.25. *Suppose that k is algebraically closed, and that G is of type D_r or E_r. Then $(G^\theta/G_0)(k)$ acts simply transitively on the set of $G_0(k)$-orbits of subregular nilpotent elements of $\mathfrak{g}_1$.*

Proof. Let x be a subregular nilpotent element. Then $Z_{G^\theta}(x) = Z_{G_0}(x)$, by Proposition 2.24. It therefore suffices to show that $\#(G^\theta/G_0)(k)$ is equal to the number of real subregular nilpotent orbits. This can be accomplished, for example, by inspection of the tables in [Collingwood and McGovern 1993]. □

Proposition 2.26. *Suppose that k is algebraically closed, and that G is of type A_r. Then there is a unique $G_0(k)$-conjugacy class of subregular nilpotent elements in $\mathfrak{g}_1$.*

Proof. We note that when $k = \mathbb{R}$, there is a unique real orbit of subregular nilpotents in $\mathfrak{g}$. □

We now treat the case where k is not necessarily algebraically closed.

Proposition 2.27. *The space $\mathfrak{g}_1$ contains a subregular nilpotent element. In particular, we can find normal $\mathfrak{sl}_2$-triples (e, h, f) in $\mathfrak{g}$ with e a subregular nilpotent element.*

Proof. Let K denote a separable closure of k. It suffices to find a normal $\mathfrak{sl}_2$-triple (e, h, f) in $\mathfrak{g} \otimes_k K$ such that e is subregular nilpotent and $h \in \mathfrak{g}$. For then the set of subregular elements is Zariski dense in $\mathfrak{g}_1^{\mathrm{ad}\, h=2}$ (see [de Graaf 2011, Proposition 7]) and our chosen field k is infinite.

Since $\mathfrak{g}_1$ contains a regular nilpotent element, we may assume that G is equipped with a pinning $(T, B, \{X_\alpha\}_{\alpha \in R})$ and that θ is the involution of Lemma 2.13, constructed in terms of this pinning. In particular, $\mathfrak{t}_0 = \mathfrak{t}^\theta \subset \mathfrak{g}_0$ is a split Cartan subalgebra of G_0.

Let (e, h, f) be a subregular normal $\mathfrak{sl}_2$-triple in $\mathfrak{g} \otimes_k K$. After conjugating by an element of $G_0(K)$, we can assume that h lies in $\mathfrak{t}_0 \otimes_k K \subset \mathfrak{t} \otimes_k K$. Now we have $\alpha(h) \in \mathbb{Z}$ for every root α, since h embeds in an $\mathfrak{sl}_2$-triple, and hence h lies in $\mathfrak{t}_0$. The result follows. □

Definition 2.28. We refer to a normal $\mathfrak{sl}_2$-triple (e, h, f) with e subregular as a subregular normal $\mathfrak{sl}_2$-triple.

Proposition 2.29. 1. *Suppose that G is of type D_r or E_r. Then all subregular nilpotent elements in $\mathfrak{g}_1$ are $G^\theta(k)$-conjugate.*

2. *Suppose that G is of type A_{2r}. Then there is a bijection between $k^\times/(k^\times)^2$ and the set of $G_0(k)$-orbits of subregular nilpotent elements in $\mathfrak{g}_1$, given by sending $d \cdot (k^\times)^2$ to the orbit of the element* (*in the notation of Example 2.19 above*):

$$(f_1 \mapsto f_2 \mapsto f_3 \mapsto \cdots \mapsto f_n \mapsto de_n, e_n \mapsto e_{n-1} \mapsto \cdots \mapsto e_1, v \mapsto 0).$$

3. *Suppose that G is of type A_{2r+1}. Then all subregular nilpotent elements in $\mathfrak{g}_1$ are $G_0(k)$-conjugate.*

Proof. Let $x \in \mathfrak{g}_1$ be a subregular nilpotent element. The first part follows since $Z_{G^\theta}(x)$ is a unipotent group, and hence has vanishing first Galois cohomology. To prove the second and third parts, we make an explicit calculation using the results of Kawanaka [1987]. Briefly, if (e, h, f) is a normal $\mathfrak{sl}_2$-triple, let $\overline{G}_0$ denote the connected subgroup of G with Lie algebra $\mathfrak{g}_0 \cap \mathfrak{g}^{\mathrm{ad}\, h=0}$. Then Kawanaka shows that $Z_{G_0}(e)$ has the form $C \ltimes R$, where R is connected unipotent and $C = Z_{\overline{G}_0}(e)$ has reductive connected component. We summarize the results of this calculation here.

If $\mathfrak{g}$ is of type A_{2r}, a choice of subregular nilpotent x is the transformation given by the formula (in the notation of Example 2.19)

$$f_1 \mapsto f_2 \mapsto f_3 \mapsto \cdots \mapsto f_n \mapsto e_n \mapsto e_{n-1} \mapsto \cdots \mapsto e_1, v \mapsto 0.$$

If $d \in k^\times$, we define another element x_d by the formula

$$f_1 \mapsto f_2 \mapsto f_3 \mapsto \cdots \mapsto f_n \mapsto de_n, e_n \mapsto e_{n-1} \mapsto \cdots \mapsto e_1, v \mapsto 0.$$

One calculates that $Z_{G_0}(e)$ is a semidirect product of μ_2 by a connected unipotent group, with Galois cohomology isomorphic (via the Kummer isomorphism) to

$k^\times/(k^\times)^2$. With appropriate identifications the element $d \in k^\times/(k^\times)^2$ corresponds to the $G_0(k)$-orbit of the element x_d.

If $\mathfrak{g}$ is of type A_{2r+1}, then one calculates that $Z_{G_0}(e)$ is connected unipotent, so has vanishing first Galois cohomology. □

Proposition 2.30. *Suppose that k is algebraically closed. If G is of type A_r, D_{2r+1} or E_r then the closure of every regular nilpotent $G_0(k)$-orbit in $\mathfrak{g}_1$ contains every subregular nilpotent orbit.*

If G is of type D_{2r}, then the closure of each regular nilpotent $G_0(k)$-orbit contains exactly 3 subregular nilpotent orbits, and each subregular nilpotent orbit is contained in the closure of exactly 3 regular nilpotent orbits.

Proof. The only cases needing proof are A_{2r+1}, D_r, and E_7. The case of A_{2r+1} follows immediately, since $(G^\theta/G_0)(k)$ permutes the regular nilpotent orbits. The cases of D_r and E_7 follow from the descriptions given in [Đoković and Litvinov 2003] and [Đoković 2001], respectively. □

3. Subregular curves

For the rest of this paper, we fix the following notation. We suppose that G is a split simple group over k, of type A_r, D_r, or E_r. We fix also a stable involution θ of G and a regular nilpotent element $E \in \mathfrak{g}_1$. We recall that the pair (θ, E) is determined uniquely up to $G^{\mathrm{ad}}(k)$-conjugacy. In this section we construct a family of curves over the categorical quotient $\mathfrak{g}_1/\!\!/G_0$. The construction is based on the notion of transverse slice to the action of an algebraic group, which we now briefly review.

Transverse slices. For the moment, let H be an algebraic group acting on a variety X, both defined over k. Let $x \in X(k)$. By a transverse slice in X to the orbit of x (or more simply, a transverse slice at x), we mean a locally closed subvariety $S \subset X$ satisfying the following:

1. $x \in S(k)$.
2. The orbit map $H \times S \to X$, $(h, s) \mapsto h \cdot s$, is smooth.
3. S has minimal dimension with respect to the above properties.

It is easy to show that if X is smooth, then transverse slices of the above kind always exist and have dimension equal to the codimension of the orbit $H \cdot x$ in X. (Here we use that k is of characteristic zero; in general, one should assume also that the orbit maps are separable.) An important property of transverse slices is the following slight extension of [Slodowy 1980b, §5.2, Lemma 3]:

Proposition 3.1. *Let H be an algebraic group acting on a smooth variety X. Let S_1, S_2 be transverse slices at points $x_1, x_2 \in X(k)$, respectively, where x_1, x_2 lie in*

the same $H(k)$-orbit of X. Let $f\colon X \to Y$ be an H-equivariant morphism, where H acts trivially on Y.

- *S_1, S_2 are étale locally isomorphic over Y in the sense that there exists a variety S over Y with a geometric point $\bar{s}$ and étale Y-morphisms $\phi_1\colon S \to S_1$, $\phi_2\colon S \to S_2$ with $\phi_1(\bar{s}) = x_1$, $\phi_2(\bar{s}) = x_2$.*
- *Suppose further that $k = \mathbb{C}$. Then $S_1(\mathbb{C}), S_2(\mathbb{C})$ are locally isomorphic over $Y(\mathbb{C})$ in the analytic topology. Furthermore, there exist arbitrarily small neighborhoods $U_1 \subset S_1(\mathbb{C}), U_2 \subset S_2(\mathbb{C})$ of x and analytic isomorphisms $\psi\colon U_1 \to U_2$ over $Y(\mathbb{C})$ such that the induced maps*

$$U_1 \hookrightarrow X(\mathbb{C}), \quad U_1 \cong U_2 \hookrightarrow X(\mathbb{C})$$

are homotopic over $Y(\mathbb{C})$.

An important special case where we can construct transverse slices explicitly is the case of a reductive group H acting via the adjoint representation on its Lie algebra $\mathfrak{h}$. Let (e, h, f) be an $\mathfrak{sl}_2$-triple in $\mathfrak{h}$.

Proposition 3.2. *$S = e + \mathfrak{z}_{\mathfrak{h}}(f)$ is a transverse slice to the action of H at every point of S. In other words, the multiplication map $\mu\colon H \times S \to \mathfrak{h}$ is everywhere smooth.*

The proof is based on the following construction of Slodowy. First, we may decompose $\mathfrak{h} = \oplus_i V_i$ into a direct sum of irreducible submodules under the adjoint action of the $\mathfrak{sl}_2$ spanned by e, h, and f. We let $\lambda\colon \mathbb{G}_m \to H$ be the cocharacter with $d\lambda(1) = h$. Let $p_1, \dots, p_r$ be algebraically independent homogeneous polynomials generating the ring of invariants $k[\mathfrak{h}]^H$. (We remind the reader that the adjoint representation of H on $\mathfrak{h}$ is coregular, so such elements certainly exist.) We suppose that they have degrees $d_1, \dots, d_r$. We suppose that V_i has dimension m_i, and choose for each i a basis vector v_i of the lowest weight space of V_i.

A general element $v \in S$ can be written in the form $v = e + \sum_i x_i v_i$, and we have

$$\lambda(t)(v) = t^2 e + \sum_i t^{1-m_i} x_i v_i, \quad tv = te + \sum_i t x_i v_i$$

and

$$p_i\big(\lambda(t)(v)\big) = p_i(v), \quad p_i(tv) = t^{d_i} p_i(v).$$

Defining an action ρ of $\mathbb{G}_m$ on $\mathfrak{h}$ by $\rho(t)(v) = t^2\lambda(t^{-1})(v)$, we see that S is ρ-invariant, and the ρ-action contracts S to e. If we let $\mathbb{G}_m$ act on $\mathfrak{h}/\!\!/H$ by the square of its usual action, then the composite $S \hookrightarrow \mathfrak{h} \to \mathfrak{h}/\!\!/H$ becomes $\mathbb{G}_m$-equivariant. In other words, writing $w_1, \dots, w_n$ for the weights of the ρ-action on S, the morphism $S \to \mathfrak{h}/\!\!/H$ is quasihomogeneous of type $(d_1, \dots, d_r; w_1, \dots, w_n)$. The weights w_i are given by the formula $w_i = m_i + 1$.

Proof of Proposition 3.2. Define an action of $\mathbb{G}_m \times H$ on $H \times S$ by

$$(t, g) \cdot (k, s) = \big(gk\lambda(t), \rho(t)(s)\big),$$

and let $\mathbb{G}_m \times H$ act on $\mathfrak{h}$ by $(t, g) \cdot X = t^2 g(X)$. Then the map $\mu\colon H \times S \to \mathfrak{h}$ is equivariant for these actions, and smooth in a neighborhood of $H \times \{e\} \subset H \times S$; since the $\mathbb{G}_m$-actions are contracting, it follows that μ is smooth everywhere. □

Corollary 3.3. *The composite $S \hookrightarrow \mathfrak{h} \to \mathfrak{h}/\!\!/H$ is faithfully flat.*

Proof. The composite $H \times S \to S \to \mathfrak{h}/\!\!/H$ is equal to the composite

$$H \times S \to \mathfrak{h} \to \mathfrak{h}/\!\!/H,$$

which is a composition of flat morphisms, hence flat ($H \times S \to \mathfrak{h}$ is flat since we have just proved it to be smooth). Since the second projection $H \times S \to S$ is flat, $S \to \mathfrak{h}/\!\!/H$ must also be flat.

The image is a $\mathbb{G}_m$-stable open subset of $\mathfrak{h}/\!\!/H$ containing the origin, hence the whole of $\mathfrak{h}/\!\!/H$. The faithful flatness follows. □

Let us now return to our group G with stable involution θ, and let (e, h, f) now denote a normal $\mathfrak{sl}_2$-triple. From the above, we see that there is a direct sum decomposition $\mathfrak{g} = [e, \mathfrak{g}] \oplus \mathfrak{z}_{\mathfrak{g}}(f)$. Both summands are θ-stable so we deduce that $\mathfrak{g}_1 = [e, \mathfrak{g}_0] \oplus \mathfrak{z}_{\mathfrak{g}}(f)_1$, where by definition $\mathfrak{z}_{\mathfrak{g}}(f)_1 = \mathfrak{z}_{\mathfrak{g}}(f) \cap \mathfrak{g}_1$. It follows that $X = e + \mathfrak{z}_{\mathfrak{g}}(f)_1$ is a transverse slice at $e \in \mathfrak{g}_1$, and the contracting $\mathbb{G}_m$-action on $e + \mathfrak{z}_{\mathfrak{g}}(f)$ leaves X invariant. Identical arguments to those above now prove the following.

Proposition 3.4. *The map $\mu\colon G_0 \times X \to \mathfrak{g}_1$ is smooth and the composite $X \hookrightarrow \mathfrak{g}_1 \to \mathfrak{g}_1/\!\!/G_0$ is faithfully flat.*

We now examine two special cases of this construction in more detail.

The regular $\mathfrak{sl}_2$ and the Kostant section. Let $d_1, \dots, d_r$ denote the degrees of algebraically independent homogeneous generators of the polynomial ring $k[\mathfrak{g}_1]^{G_0}$. We let (E, H, F) be the unique normal $\mathfrak{sl}_2$-triple containing the element E, and set $\kappa = E + \mathfrak{z}_{\mathfrak{g}}(F)_1$. We call κ the Kostant section. It has the following remarkable properties.

Lemma 3.5. *The composite $\kappa \hookrightarrow \mathfrak{g}_1 \to \mathfrak{g}_1/\!\!/G_0$ is an isomorphism. Every element of κ is regular. In particular, the map $\mathfrak{g}_1(k) \to (\mathfrak{g}_1/\!\!/G_0)(k)$ is surjective, and if k is algebraically closed then κ meets every $G_0(k)$-conjugacy class of regular semisimple elements.*

Proof. It is well-known that in this case the map $\kappa \to \mathfrak{g}_1/\!\!/G_0$ is quasihomogeneous of type $(2d_1, \dots, 2d_r; 2d_1, \dots, 2d_r)$; compare [Panyushev 2005, proof of Theorem 3.3]. Lemma 3.14 now implies that it must be an isomorphism. The remaining claims are immediate. □

A subregular $\mathfrak{sl}_2$. Fix a normal subregular $\mathfrak{sl}_2$-triple (e, h, f), and set $X = e + \mathfrak{z}_{\mathfrak{g}}(f)_1$. (Note that if G is of type A_1, then there is no nonzero subregular nilpotent element, and therefore no subregular $\mathfrak{sl}_2$-triple, since by definition an $\mathfrak{sl}_2$-triple consists of 3 linearly independent elements. In this case, we just take $X = \mathfrak{g}_1$.) Recall that we have defined an action of $\mathbb{G}_m$ on X.

Proposition 3.6. *We have* $\dim X = r + 1$. *We write* $w_1, \dots, w_{r+1}$ *for the weights of the* $\mathbb{G}_m$*-action. After reordering, we have* $w_i = 2d_i$ *for* $i = 1, \dots, r-1$. *The* $2d_i$, $i = 1, \dots, r-1$ *and* w_r *and* w_{r+1} *are given in the following table*:

	$2d_1$	$2d_2$	$2d_3$	$\cdots$	$\cdots$	$2d_{r-2}$	$2d_{r-1}$	$2d_r$	w_r	w_{r+1}
A_r	4	6	8	$\cdots$	$\cdots$	$2r-2$	$2r$	$2r+2$	2	$r+1$
D_r	4	8	12	$\cdots$	$\cdots$	$4r-8$	$2r$	$4r-4$	4	$2r-4$
E_6	4	10	12			16	18	24	6	8
E_7	4	12	16		20	24	28	36	8	12
E_8	4	16	24	28	36	40	48	60	12	20

Proof. The proof is by explicit calculation, along similar lines to the proof of [Slodowy 1980b, §7.4, Proposition 2]. We describe the method. If $V \subset \mathfrak{g}$ is a θ-stable simple $\mathfrak{sl}_2$-submodule, then its highest weight space is θ-invariant. Moreover, the eigenvalue of θ on this highest weight space determines the action of θ on every weight space. We can calculate a decomposition of $\mathfrak{g}$ into a direct sum of θ-stable simple $\mathfrak{sl}_2$-modules by calculating the dimension of each weight space of h, and the trace of θ on each weight space. This can be accomplished by using the explicit θ constructed in Lemma 2.13 and a list of the roots of $\mathfrak{g}$. We can then fill in the table by reading off the lowest weight spaces which have θ-eigenvalue equal to -1. □

Example 3.7. We illustrate the method of proof in the case that G is of type A_2. Then a choice of h is

$$h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

in the notation of Example 2.19. We can write the weights of h on $\mathfrak{g}$ with multiplicity as follows:

$$\begin{array}{ccccc} -2 & & 0 & & 2 \\ & -1 & & 1 & \\ & -1 & & 1 & \\ & & 0 & & \end{array}$$

Thus $\mathfrak{g}$ decomposes as a direct sum $V(3) \oplus V(2) \oplus V(2) \oplus V(1)$, where $V(i)$ denotes the unique isomorphism class of $\mathfrak{sl}_2$-modules of dimension i. In this case -1 is an eigenvalue of θ of multiplicity 1 on each weight space. (Recall that $-\theta$ is

reflection in the antidiagonal.) We can now decorate each weight space with a $+$ or $-$, according to its θ-eigenvalue:

$$\begin{array}{ccccc} -2^- & & 0^+ & & 2^- \\ & -1^+ & & 1^- & \\ & -1^- & & 1^+ & \\ & & 0^- & & \end{array}$$

It follows that $\dim \mathfrak{z}_{\mathfrak{g}}(f)_1 = 3$, as expected, and the eigenvalues of h on $\mathfrak{z}_{\mathfrak{g}}(f)_1$ are -2, -1 and 0, hence the weights on $e + \mathfrak{z}_{\mathfrak{g}}(f)_1$ are 2, 3 and 4.

Henceforth we write $\mathfrak{g}_1 /\!\!/ G_0 = B$, and $\varphi\colon X \to B$ for the restriction of the quotient map $\pi : \mathfrak{g}_1 \to \mathfrak{g}_1 /\!\!/ G_0$ to X. The main result of this section is the following.

Theorem 3.8. *The fibers of φ are reduced curves. The central fiber $X_0 = \varphi^{-1}(0)$ has a unique singular point which is a simple singularity of type A_r, D_r, E_r, corresponding to that of G. We can choose homogeneous coordinates $(p_{d_1}, \dots, p_{d_r})$ on B and $(p_{d_1}, \dots, p_{d_{r-1}}, x, y)$ on X such that the family $X \to B$ of curves is as given by the following table*:

G	X
A_r	$y^2 = x^{r+1} + p_2 x^{r-1} + \cdots + p_{r+1}$
D_r	$y(xy + p_r) = x^{r-1} + p_2 x^{r-2} + \cdots + p_{2r-2}$
E_6	$y^3 = x^4 + y(p_2 x^2 + p_5 x + p_8) + p_6 x^2 + p_9 x + p_{12}$
E_7	$y^3 = x^3 y + p_{10} x^2 + x(p_2 y^2 + p_8 y^2 + p_{14}) + p_6 y^2 + p_{12} y + p_{18}$
E_8	$y^3 = x^5 + y(p_2 x^3 + p_8 x^2 + p_{14} x + p_{20}) + p_{12} x^3 + p_{18} x^2 + p_{24} x + p_{30}$

(This means, for example, that when G is of type A_r, the relation $p_{r+1} = y^2 - (x^{r+1} + p_2 x^{r-1} + \cdots + p_r x)$ holds on X.) The proof of Theorem 3.8 follows closely the one in [Slodowy 1980b], with some simplifications due to the fact that we work with curves, rather than surfaces. We begin with some general considerations, and reduce to a case-by-case calculation using the invariant degrees of G.

The possibility of choosing coordinates as above is a consequence of the next result.

Lemma 3.9 [Slodowy 1980b, §8.1, Lemma 2]. *Let V and U be k-vector spaces of dimensions m and n, respectively, on which $\mathbb{G}_m$ acts linearly. Let $\phi\colon V \to U$ be a morphism equivariant for these actions. Suppose that $d\phi_0$ has rank s and that $\mathbb{G}_m$ acts with strictly positive weights on U and V.*

Then there exist $\mathbb{G}_m$-invariant decompositions $V = V_0 \oplus W$, $U = U_0 \oplus W$, $\dim W = s$, and a regular automorphism α of V such that $\phi \circ \alpha$ has the form $(v_0, w) \mapsto (\psi(v_0, w), w)$ for some morphism $\psi\colon V_0 \oplus W \to U_0$.

To apply this to the map $\varphi\colon X \to B$, we need the following result.

Proposition 3.10. *Let $x \in X$. Then $d\varphi_x$ has maximal rank $r = \operatorname{rank} G$ if and only if x is a regular element. The map $d\varphi_0 \colon T_e X \to T_0 B$ has rank $r-1$.*

Proof. Let $p \colon \mathfrak{g} \to \mathfrak{g}/\!\!/G$ denote the adjoint quotient map. For any $y \in \mathfrak{g}_1$, we have $dp_y(\mathfrak{g}_0) = 0$. This is true if y is regular, since then $\mathfrak{g}_0 = [y, \mathfrak{g}_1] \subset [y, \mathfrak{g}]$ is contained in the tangent space to the orbit $G \cdot y$. It then follows for any $y \in \mathfrak{g}_1$, since the regular elements are dense. In particular, we have $\operatorname{rank} dp_y = \operatorname{rank} d\pi_y = \operatorname{rank} d\varphi_y$ for any $y \in X$. The first part of the proposition now follows, since $y \in \mathfrak{g}_1$ is regular if and only if dp_y has maximal rank.

For the second part, we remark that $\operatorname{rank} dp_e = r-1$ if e is subregular nilpotent, by [Slodowy 1980b, §8.3, Proposition 1]. □

We thus obtain a decomposition of affine spaces $X = V_0 \oplus W$, $B = U_0 \oplus W$, where $\dim W = r-1$, $\dim V_0 = 2$, and $\dim U_0 = 1$. With respect to these decompositions we write $\varphi \colon V_0 \oplus W \to U_0 \oplus W$ in the form $\varphi(v_0, w) = (\psi(v_0, w), w)$.

Recall that φ is $\mathbb{G}_m$-equivariant of type $(2d_1, \dots, 2d_r; w_1, \dots, w_{r+1})$. By inspection of the tables above, we have $2d_r > w_i$, each $i = 1, \dots, r+1$, and hence the weights occurring in W are $2d_1, \dots, 2d_{r-1}$. Moreover, the unique weight of U_0 is given by $2d_r$ and the weights of V_0 are w_r, w_{r+1}. Let x, y be homogeneous coordinates on V_0 of weight w_r and w_{r+1}, respectively. It follows that $X_0 \subset V_0$ is cut out by a quasihomogeneous polynomial $f(x, y)$ of type $(2d_r; w_r, w_{r+1})$.

Proposition 3.11. *After possibly making a linear change of variables, the polynomial $f(x, y)$ is as given by the following table.*

G	$f(x, y)$
$A_r, r \geq 1$	$y^2 - x^{r+1}$
$D_r, r \geq 4$	$xy^2 - x^{r-1}$
E_6	$y^3 - x^4$
E_7	$y^3 - x^3 y$
E_8	$y^3 - x^5$

Proof. We suppose first that k is algebraically closed. Then the induced map $G_0 \times X_0 \to \pi^{-1}(0)$ is smooth, since X is a transverse slice and this property is preserved under passage to fibers (see [Slodowy 1980b, §5, Lemma 2]). Since $\pi^{-1}(0)$ is smooth along the regular locus, X_0 is generically smooth, hence reduced. We now proceed by direct computation. Let us treat for example the case of A_r. Then $f(x, y)$ is quasihomogeneous of type $(2r+2; 2, r+1)$, where we suppose that the weights of x and y are 2 and $r+1$, respectively.

Since f defines a reduced curve, it must have the form $ay^2 - bx^{r+1}$, with a, b nonzero constants. After rescaling we may assume that f has the form given in the statement of the proposition. The same argument works for the other cases as well.

Now suppose that k is not algebraically closed. The same argument suffices, except in the cases A_{2r+1} and D_{2r}. For example, in the case A_{2r+1} one must rule out the possibility $f(x, y) = y^2 - ax^{2r+2}$, where $a \in k^\times$ is a nonsquare. But the natural action map $G_0 \times X_0 \to \pi^{-1}(0)$ induces an injection on geometric irreducible components — see Lemma 4.14. The irreducible components of $\pi^{-1}(0)$ are geometrically irreducible, so it follows that the same must be true for X_0, hence a must be a square. The same argument works for the case of type D_{2r}. □

At this point we have identified the central fiber of φ with the desired curve. We will obtain the identification over the whole of B via a deformation argument. Before doing this, we must determine the singularities appearing in the other fibers of φ.

Proposition 3.12. *Let $t \in \mathfrak{g}_1$ be a semisimple element, and let b denote its image in B. Let D denote the Dynkin diagram of $Z_G(t)$, and write it as a disjoint union $D = D_1 \cup \cdots \cup D_k$ of its connected components.*

Let $y \in \varphi^{-1}(b)(k) = X_b(k)$ be a singular point. Then y is a simple singularity of type D_i for some $i = 1, \dots, s$.

Proof. We have an isomorphism

$$G_0 \times^{Z_{G_0}(t)} \big(t + \mathcal{N}(\mathfrak{z}_{\mathfrak{g}}(t)_1)\big) \cong \pi^{-1}(b),$$

induced by the map $(g, t+n) \mapsto g \cdot (t+n)$. Let y have Jordan decomposition $y = y_s + y_n$. Without loss of generality, we may suppose that k is algebraically closed and that $y_s = t$. Then $y_n \in \mathfrak{z}_{\mathfrak{g}}(t)$ is a subregular nilpotent element. If we decompose $[\mathfrak{z}_{\mathfrak{g}}(t), \mathfrak{z}_{\mathfrak{g}}(t)] = \mathfrak{l}^1 \times \cdots \times \mathfrak{l}^k$ into a product of simple, θ-stable subalgebras then y_n has a decomposition $y_n = y_1 + \cdots + y_k$, where $y_i \in \mathfrak{l}^i$ is a nilpotent element. After renumbering, we can assume that $y_1 \in \mathfrak{l}^1$ is a subregular nilpotent element, and all of the other $y_i \in \mathfrak{l}^i$ are regular nilpotent. Moreover, the restriction of θ to each $\mathfrak{l}^i$ is a stable involution.

Now fix a transverse slice S_1 to the $Z_{G_0}(t)$-orbit of y_1 in $\mathfrak{l}^1_1$. It then follows that $S_1 + \sum_{j \geq 2} y_j$ is a transverse slice to the $Z_{G_0}(t)$-orbit of y_n in $\mathcal{N}(\mathfrak{l}^1_1)$ and hence $X_1 = t + S_1 + \sum_{j \geq 2} y_j$ is a transverse slice at y to the G_0 action in $\pi^{-1}(b)$, as the above isomorphism makes $\pi^{-1}(b)$ into a fiber bundle over $G_0/Z_{G_0}(t)$ with fiber $\mathcal{N}(\mathfrak{z}_{\mathfrak{g}}(t)_1)$.

On the other hand, we know that X_b is also a transverse slice at y to the G_0 action in $\pi^{-1}(b)$. The result now follows from Proposition 3.1 and Proposition 3.11. □

Semiuniversal deformations and the proof of Theorem 3.8. We can now complete the proof of Theorem 3.8. There exists a semiuniversal deformation $\widehat{Z} \to \widehat{D}$ of the central fiber X_0 as a $\mathbb{G}_m$-scheme, where $\widehat{Z} \to \widehat{D}$ is a morphism of formal schemes with underlying reduced schemes given by $X_0 \to \operatorname{Spec} k$ [Slodowy 1980b, §2.7].

The proof of the theorem is based on the fact that, since X_0 is given as the zero set of an explicit polynomial $f(x, y)$, $\widehat{Z} \to \widehat{D}$ admits a canonical algebraization $Z \to D$ which we can calculate explicitly and then compare with $X \to B$.

Proposition 3.13. *Let $f(x, y)$ be a polynomial in two variables, quasihomogeneous of type $(d; w_1, w_2)$. Let $X_0 \subset \mathbb{A}^2$ denote the closed subscheme defined by f, and suppose that X_0 has an isolated singularity at the origin. Then we can construct a semiuniversal $\mathbb{G}_m$-deformation of X_0: Let $J = (\partial f/\partial x, \partial f/\partial y) \subset k[x, y]$ denote the Jacobian ideal of f. Then $k[x, y]/J$ is a finite-dimensional k-vector space, and receives an action of $\mathbb{G}_m$. Choose $\mathbb{G}_m$-invariant polynomials $g_1(x, y), \dots, g_n(x, y)$ projecting to a k-basis of $\mathbb{G}_m$-eigenvectors of $k[x, y]/J$. Now define*

$$Z = \{f + t_1 g_1 + \cdots + t_n g_n = 0\} \subset \mathbb{A}^2 \times \mathbb{A}^n,$$

and let $\Phi\colon Z \to D$ denote the natural projection to the $\mathbb{A}^n$ factor.

Suppose that g_i has weight r_i, and let $\mathbb{G}_m$ act on t_i by the character $t \mapsto t^{d-r_i}$. Then Φ is a $\mathbb{G}_m$-equivariant morphism, and the formal completion $\widehat{\Phi}\colon \widehat{Z} \to \widehat{D}$ of this morphism is a semiuniversal $\mathbb{G}_m$-deformation of X_0.

Proof. See [Slodowy 1980b, §2.4]. □

Applying this to our fixed polynomial f, we obtain a family of curves $Z \to D$, where D is an affine space of dimension n, and a Cartesian diagram of formal schemes:

$$\begin{tikzcd} \widehat{X} \arrow[r] \arrow[d] & \widehat{Z} \arrow[d] \\ \widehat{B} \arrow[r] & \widehat{D} \end{tikzcd}$$

An elementary calculation shows that in each case A_r, D_r, or E_r, we have $n = r$ and $Z \to D$ is the family of curves appearing in the statement of Theorem 3.8. The morphism $\widehat{B} \to \widehat{D}$ is given by power series and respects the $\mathbb{G}_m$-actions on either side, which both have strictly positive weights; it follows that these power series are in fact polynomials, so this morphism has a canonical algebraization. We obtain a second Cartesian diagram:

$$\begin{tikzcd} X \arrow[r] \arrow[d] & Z \arrow[d] \\ B \arrow[r] & D \end{tikzcd}$$

The bottom horizontal arrow is a $\mathbb{G}_m$-equivariant polynomial map between affine spaces of the same dimension and the weights on the domain and codomain are the same. We now apply the following lemma:

Lemma 3.14 [Slodowy 1980b, §8.1, Lemma 3]. *Let $\mathbb{G}_m$ act on affine spaces V, U of dimension n, and let $\phi\colon V \to U$ be an equivariant morphism. Suppose that*:

- *$\mathbb{G}_m$ acts on V and U with the same strictly positive weights.*
- *The central fiber $\phi^{-1}(0)$ is zero dimensional.*

Then ϕ is an isomorphism.

We must verify that the second condition holds. If $b \in B$ is mapped to $0 \in D$, then $X_b \cong X_0$. Proposition 3.12 implies that all singularities in the noncentral fibers of φ are simple singularities belonging to simply laced root systems of rank strictly less than r, and so we must have $b = 0$. This completes the proof of Theorem 3.8.

A lemma. The results of this section will be used later. Let $S = e + \mathfrak{z}_{\mathfrak{g}}(f)$, and let τ denote the involution of S induced by $-\theta$. Thus S is an affine space of dimension $r+2$, and we have $S^\tau = X$.

Lemma 3.15. *We can choose global coordinates $z_1, \dots, z_{r+2}$ on S, $u_1, \dots, u_r$ on B such that $z_1, \dots, z_{r+1}$ are fixed by τ, $\tau(z_{r+2}) = -z_{r+2}$, and such that the following holds*: *the morphism $X \to B$ is given by the formula*

$$(z_1, \dots, z_{r+1}) \mapsto \big(z_1, \dots, z_{r-1}, f(z_1, \dots, z_{r+1})\big)$$

for some polynomial function f, and the morphism $S \to B$ is given by the formula

$$(z_1, \dots, z_{r+2}) \mapsto \big(z_1, \dots, z_{r-1}, f(z_1, \dots, z_{r+1}) + z_{r+2}^2\big).$$

Proof. We recall that there is a contracting action of $\mathbb{G}_m$ on S, and that this action sends X to itself. Applying Lemma 3.9, we see that we can find $\mathbb{G}_m$ and τ-invariant decompositions $S = V_0 \oplus V_1 \oplus U$, $B = U_0 \oplus U$ such that the map $S \to B$ is given by $(v_0, v_1, u) \mapsto \big(\psi(v_0, v_1, u), u\big)$ for some $\mathbb{G}_m$-equivariant morphism ψ. Moreover, τ acts trivially on $V_0 \oplus U$ and as -1 on V_1. We have $\dim V_0 = 2$, $\dim V_1 = \dim U_0 = 1$, $\dim U = r-1$. Moreover, ψ is quasihomogeneous of some degree.

We choose coordinates as follows: let $z_1, \dots, z_{r-1}$ be arbitrary linear coordinates on U, z_r, z_{r+1} coordinates which are eigenfunctions for the $\mathbb{G}_m$-action, and z_{r+2} an arbitrary linear coordinate on V_1. Then [Slodowy 1980b, §7.4, Proposition 2] implies that z_{r+2} has degree equal to half the degree of ψ. It follows that we must have $\psi(v_0, v_1, u) = \psi(v_0, 0, u) + z_{r+2}^2$, after possibly rescaling coordinates. (The coefficient of z_{r+2}^2 must be nonzero since S_0 has a unique isolated singularity.) $\square$

Corollary 3.16. *Let $b \in B(k)$, and let $t \in \pi^{-1}(b)(k)$ be a semisimple element. Then there is a bijection between the connected components of the Dynkin diagram of $Z_G(t)$ and the singularities of the fiber X_b, which takes each* (*connected*, *simply laced*) *Dynkin diagram to a singularity of corresponding type.*

Proof. Lemma 3.15 implies that the singular locus of S_b is equal to the singular locus of X_b. We have seen that the singular points of X_b are precisely the subregular elements of X_b. It therefore suffices to show that X_b meets each G-orbit of subregular elements in $p^{-1}(b)$ exactly once, or equivalently that S_b meets each G-orbit of subregular elements in $p^{-1}(b)$ exactly once. This follows immediately from [Slodowy 1980b, §6.6, Proposition 2] and the remark following. □

4. Jacobians and stabilizers of regular elements

We continue with the notation of the previous section. Thus G is a split simple group of type A_r, D_r, or E_r, θ is a stable involution of G, and $E \in \mathfrak{g}_1$ is a regular nilpotent element. The pair (θ, E) is uniquely determined up to $G^{\mathrm{ad}}(k)$-conjugacy. This data determines a regular normal $\mathfrak{sl}_2$-triple (E, H, F). We choose further a subregular normal $\mathfrak{sl}_2$-triple (e, h, f). Our chosen $\mathfrak{sl}_2$-triples give two special transverse slices: first, the Kostant section $\kappa = E + \mathfrak{z}_{\mathfrak{g}}(F)_1$, which is a section of the categorical quotient $\pi : \mathfrak{g}_1 \to B$ by regular elements; second, a transverse slice to the G_0-orbit of e, $X = e + \mathfrak{z}_{\mathfrak{g}}(f)_1$. The fibers of the induced map $\varphi\colon X \to B$ are reduced connected curves.

In this section we shall write $\mathfrak{g}_1^{\mathrm{rs}}$ for the open subvariety of regular semisimple elements, and B^{rs} for its image in B. For any variety $Z \to B$ we will write $Z^{\mathrm{rs}} = Z \times_B B^{\mathrm{rs}}$. Thus the morphism $X^{\mathrm{rs}} \to B^{\mathrm{rs}}$ is a family of smooth curves.

Homology. Fix a separable closure K of k. In the following if X is a k-scheme of finite type, we will write $H_1(X, \mathbb{F}_2)$ for $H^1_{\text{ét}}(X \otimes_k K, \mathbb{F}_2)^*$, the dual of the first étale cohomology of $X \otimes_k K$. This is a finite group, and receives an action of the Galois group $\mathrm{Gal}(K/k)$.

Suppose that A is a finite group scheme over k, killed by 2, and that $Y \to X$ is an A-torsor. This defines a class in $H^1_{\text{ét}}(X \otimes_k K, A \otimes_k K) \cong \mathrm{Hom}\big(H_1(X, \mathbb{F}_2), A(K)\big)$. Viewing $H_1(X, \mathbb{F}_2)$ as a finite group scheme over k, this class defines a homomorphism $H_1(X, \mathbb{F}_2) \to A$.

Now suppose given an embedding $K \hookrightarrow \mathbb{C}$. Then there is a canonical isomorphism $H_1(X, \mathbb{F}_2) \cong H_1(X(\mathbb{C}), \mathbb{F}_2)$ with the topological homology. If $X(\mathbb{C})$ is connected and $x \in X(\mathbb{C})$, then the homomorphism $\pi_1(X(\mathbb{C}), x) \to A(\mathbb{C})$ induced by the torsor $Y \to X$ factors through the Hurewicz map $\pi_1(X(\mathbb{C}), x) \to H_1(X(\mathbb{C}), \mathbb{F}_2)$ and the induced map $H_1(X(\mathbb{C}), \mathbb{F}_2) \to A(\mathbb{C})$ agrees with the previous one, up to applying the comparison isomorphism. In particular, this map does not depend on the choice of basepoint.

If X is a geometrically connected smooth projective curve over k, then there is a canonical isomorphism $H_1(X, \mathbb{F}_2) \cong J_X[2]$, where J_X denotes the Jacobian of the curve X.

Stabilizers of regular elements. Let $\mathfrak{g}_1^{\mathrm{reg}} \subset \mathfrak{g}_1$ denote the open subset of regular elements. We write $Z \to \mathfrak{g}_1^{\mathrm{reg}}$ for the stabilizer scheme, defined as the equalizer of the diagram

$$G_0 \times \mathfrak{g}_1^{\mathrm{reg}} \underset{(g,x)\mapsto x}{\overset{(g,x)\mapsto g\cdot x}{\rightrightarrows}} \mathfrak{g}_1^{\mathrm{reg}} .$$

Proposition 4.1. 1. *Z is a commutative group scheme, quasifinite over $\mathfrak{g}_1^{\mathrm{reg}}$.*

2. *Z admits a canonical descent to B. In particular, for any two $x, y \in \mathfrak{g}_1^{\mathrm{reg}}$ with the same image in B, there is a canonical isomorphism $Z_{G_0}(x) \cong Z_{G_0}(y)$.*

Proof. The first part can be checked on geometric fibers.

For the second part, we show that $\kappa^* Z$ is the sought-after descent. The map $(G^{\mathrm{ad}})^\theta \times \kappa \to \mathfrak{g}_1^{\mathrm{reg}}$ is faithfully flat. In fact, it is étale, and [Kostant and Rallis 1971, Theorem 7] shows it to be surjective. It is now easy to construct an isomorphism between $\pi^*\kappa^* Z$ and Z over this faithfully flat cover. This defines a morphism of descent data since Z is commutative. □

We henceforth write Z for the descent to a commutative group scheme over B. Consider the orbit map $\mu^{\mathrm{rs}} \colon G_0 \times \kappa^{\mathrm{rs}} \to \mathfrak{g}_1^{\mathrm{rs}}$. This map is finite and étale, and we can form the pullback square:

$$\begin{tikzcd} \Gamma \arrow[r] \arrow[d] & G_0 \times \kappa^{\mathrm{rs}} \arrow[d] \\ X^{\mathrm{rs}} \arrow[r] & \mathfrak{g}_1^{\mathrm{rs}} \end{tikzcd}$$

Concretely, for $b \in B^{\mathrm{rs}}(k)$, $\Gamma_b \to X_b$ is the Z_b-torsor given by

$$\Gamma_b = \{g \in G_0 \mid g \cdot \kappa(b) \in X_b\}. \tag{4-1}$$

We thus obtain a Galois-equivariant map $H_1(X_b, \mathbb{F}_2) \to Z_b$.

Theorem 4.2. *If G is simply connected, the map just defined is an isomorphism.*

Example 4.3. Let us first illustrate the theorem in the case $G = \mathrm{SL}_2$. We can take θ to be conjugation by the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Then we have

$$\mathfrak{g}_0 = \left\{ \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} \right\} \quad \text{and} \quad \mathfrak{g}_1 = \left\{ \begin{pmatrix} cc0 & x \\ y & 0 \end{pmatrix} \right\}.$$

The regular nilpotents in $\mathfrak{g}_1$ are those with x or y zero but not both, and the only subregular nilpotent element in $\mathfrak{g}_1$ is zero. The quotient map $\mathfrak{g}_1 \to \mathfrak{g}_1 /\!\!/ G_0 \cong \mathbb{A}^1$ sends the above matrix to $xy \in \mathbb{A}^1$. In particular $X = \mathfrak{g}_1$ in this case, with the smooth fibers of the map $\varphi \colon X \to \mathfrak{g}_1 /\!\!/ G_0$ isomorphic to the punctured affine line.

The group G_0 is isomorphic to $\mathbb{G}_m$, and $t \in \mathbb{G}_m$ acts by

$$t \cdot \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix} = \begin{pmatrix} 0 & t^2 x \\ t^{-2} y & 0 \end{pmatrix}.$$

The stabilizer of any regular semisimple element is $\mu_2 \subset \mathbb{G}_m$, and it is clear that for any $b \in \mathbb{A}^1 - \{0\}$, the induced map $H_1(X_b, \mathbb{F}_2) \to \mu_2$ is an isomorphism.

We now consider the proof of the theorem in the general case. It suffices to prove the theorem when $k = \mathbb{C}$, which we now assume. In what follows, we simplify notation by identifying all varieties with their complex points. Fix a choice $\mathfrak{c}$ of Cartan subspace, and let $C \subset G$ denote the corresponding maximal torus.

Now choose $x \in \mathfrak{c}$, and let $b = \pi(x) \in B$. Let $L = Z_G(x)$ and $\mathfrak{l} = \operatorname{Lie} L$. We write L^1 for the derived group of $Z_G(x)$, which is simply connected, since G is. In the following, given $y \in \mathfrak{c}$, we shall write $\mathfrak{g}_{1,y}$ for the fiber of the map $\mathfrak{g}_1 \times_{\mathfrak{c}/W} \mathfrak{c} \to \mathfrak{c}$ above y, and $\mathfrak{l}_{1,y}$ for the fiber of the map $\mathfrak{l}_1 \times_{\mathfrak{c}/W(x)} \mathfrak{c} \to \mathfrak{c}$ above y.

Lemma 4.4. *Let $y \in \mathfrak{c}^{\mathrm{rs}}$. Then there is a commutative diagram*

$$\begin{tikzcd} H_1(\mathfrak{l}_{1,y}, \mathbb{F}_2) \arrow[r] \arrow[d] & Z_{L_0^1}(y) \arrow[d] \\ H_1(\mathfrak{g}_{1,y}, \mathbb{F}_2) \arrow[r] & Z_{G_0}(y) \end{tikzcd}$$

Proof. This follows from the existence of a commutative diagram

$$\begin{tikzcd} L_0^1 \arrow[r] \arrow[d] & \mathfrak{l}_{1,y} \arrow[d] \\ G_0 \arrow[r] & \mathfrak{g}_{1,y}, \end{tikzcd}$$

where the top row is a $Z_{L_0^1}(y)$-torsor and the bottom row is a $Z_{G_0}(y)$-torsor. The vertical arrows are compatible with the homomorphism $Z_{L_0^1}(y) \to Z_{G_0}(y)$. $\square$

Suppose that X_b has a singular point $u = u_s + u_n$. Choose $g \in G_0$ such that $g \cdot u_s = x \in \mathfrak{c}$, and set $v = g \cdot u$. The Jordan decomposition of v is $v = v_s + v_n = x + v_n$. Then $v_n \in \mathfrak{l}_1$ is a subregular nilpotent, corresponding to a connected component $D(v_n)$ of the Dynkin diagram of L. We choose a normal subregular $\mathfrak{sl}_2$-triple (v_n, t, w) in $\mathfrak{l}$ containing v_n, and define $X^1 = v_n + \mathfrak{z}_{\mathfrak{l}}(w)_1$. X^1 is a transverse slice to the L_0-orbit of v in $\mathfrak{l}_1$, by Proposition 3.4.

Proposition 4.5. *The dimension of X^1 is* $\operatorname{rank} G + 1$. *$X^1 \subset \mathfrak{g}_1$ is a transverse slice to the G_0-orbit of v in $\mathfrak{g}_1$.*

Proof. X^1 has the correct dimension to be a transverse slice to the orbit of a subregular element, so it suffices to check the infinitesimal condition $[v, \mathfrak{g}_0] \cap \mathfrak{z}_{\mathfrak{l}}(w)_1 = 0$. In fact, we show that $[v, \mathfrak{g}] \cap \mathfrak{z}_{\mathfrak{l}}(w) = 0$. Define

$$V = \bigoplus_{\substack{\alpha \in \Phi_{\mathfrak{c}} \\ \alpha(x) \neq 0}} \mathfrak{g}^{\alpha}.$$

Then V is the orthogonal complement of $\mathfrak{l}$ with respect to the Killing form of $\mathfrak{g}$, and so is $\mathfrak{l}$-invariant. It follows that $[v, \mathfrak{g}] = [v, V] \oplus [v_n, \mathfrak{l}] \subset V \oplus [v_n, \mathfrak{l}]$. We thus have $[v, \mathfrak{g}] \cap \mathfrak{z}_{\mathfrak{l}}(w) = [v_n, \mathfrak{l}] \cap \mathfrak{z}_{\mathfrak{l}}(w) = 0$. □

Proposition 4.6. *For all sufficiently small open neighborhoods U of u in X, there exists an open neighborhood U_0 of $b \in \mathfrak{c}/W$ such that for all $y \in \pi^{-1}(U_0) \cap \mathfrak{c}$ there is a commutative diagram*

$$\begin{array}{ccc} H_1(X^1_y, \mathbb{F}_2) & \longrightarrow & H_1(\mathfrak{l}_{1,y}, \mathbb{F}_2) \\ \downarrow & & \downarrow \\ H_1(U_y, \mathbb{F}_2) & \longrightarrow & H_1(\mathfrak{g}_{1,y}, \mathbb{F}_2). \end{array}$$

Proof. If U is a sufficiently small open set around u in X, then by Proposition 3.1 we can find an isomorphism ψ between U and an open neighborhood V of v in X^1 over $\mathfrak{c}/W$, such that $\psi(u) = v$ and the two induced maps $V \hookrightarrow \mathfrak{l}_1 \hookrightarrow \mathfrak{g}_1$ and $V \cong U \hookrightarrow \mathfrak{g}_1$ are homotopic over $\mathfrak{c}/W$. After possibly shrinking U, we can assume that the image of V in $\mathfrak{c}/W(x)$ maps injectively to $\mathfrak{c}/W$.

In particular, for c sufficiently close to b we have a commutative diagram

$$\begin{array}{ccc} H_1(V_c, \mathbb{F}_2) & \longrightarrow & H_1(\mathfrak{l}_{1,c}, \mathbb{F}_2) \\ \downarrow & & \downarrow \\ H_1(U_c, \mathbb{F}_2) & \longrightarrow & H_1(\mathfrak{g}_{1,c}, \mathbb{F}_2). \end{array}$$

To obtain the statement in the proposition, we note that for c sufficiently close to b and $y \in \pi^{-1}(c) \cap \mathfrak{c}$, we can find an open subset $V'_c \subset V_c$ such that the inclusion $V'_c \subset X^1_y$ induces an isomorphism on H_1. (Use the contracting $\mathbb{G}_m$-action.) This completes the proof. □

Corollary 4.7. *With hypotheses as in Proposition 4.6, suppose in addition that $y \in \mathfrak{c}^{\mathrm{rs}}$. Let $C(x) \subset L^1$ be the maximal torus with Lie algebra $\mathfrak{c} \cap \mathfrak{l}^1$. Then there is a commutative diagram*

$$\begin{array}{ccc} H_1(X^1_y, \mathbb{F}_2) & \longrightarrow & X_*(C(x))/2X_*(C(x)) \\ \downarrow & & \downarrow \\ H_1(X_y, \mathbb{F}_2) & \longrightarrow & X_*(C)/2X_*(C). \end{array}$$

Proof. There is an isomorphism

$$Z_{G_0}(y) \cong X_*(C)/2X_*(C),$$

and similarly for $Z_{L_0^1}(y)$. The corollary now follows from Proposition 4.6, on noting that the map $U_y \to \mathfrak{g}_{1,y}$ factors through the inclusion $X_y \subset \mathfrak{g}_{1,y}$. □

To go further, it is helpful to compare this with another description of the homology of the curves X_y.

Theorem 4.8. 1. *The map $X^{\mathrm{rs}} \to \mathfrak{c}^{\mathrm{rs}}/W$ is a locally trivial fibration (in the analytic topology), and so the homology groups $H_1(X_c, \mathbb{F}_2)$ for $c \in B^{\mathrm{rs}}$ fit into a local system $\mathcal{H}_1(X)$ over $\mathfrak{c}^{\mathrm{rs}}/W$. The pullback of this local system to $\mathfrak{c}^{\mathrm{rs}}$ is constant.*

2. *Suppose $x \in \mathfrak{c}$ has been chosen so that $\alpha(x) = 0$ for some $\alpha \in \Phi_{\mathfrak{c}}$, and the only roots vanishing on x are $\pm\alpha$. Then for each $y \in \mathfrak{c}^{\mathrm{rs}}$ there is a vanishing cycle $\gamma_\alpha \in H_1(X_y, \mathbb{F}_2)$, associated to the specialization $X_y \to X_x$. This element defines a global section of the pullback of $\mathcal{H}_1(X)$ to $\mathfrak{c}^{\mathrm{rs}}$.*

3. *Let $R_{\mathfrak{c}} \subset \Phi_{\mathfrak{c}}$ denote a choice of root basis. Then for each $y \in \mathfrak{c}^{\mathrm{rs}}$ the set $\{\gamma_\alpha \mid \alpha \in R_{\mathfrak{c}}\}$ is a basis of $H_1(X_y, \mathbb{F}_2)$.*

It seems likely that this description of the local system $\mathcal{H}_1(X)$ is well-known to experts, but we have not been able to find an adequate reference in the literature. The proof of this theorem is given in Section 4 below. See in particular Lemma 4.20 for the definition of the vanishing cycle γ_α.

Now suppose $x \in \mathfrak{c}$ has been chosen so that $\alpha(x) = 0$ for some $\alpha \in \Phi_{\mathfrak{c}}$, and the only roots vanishing on x are $\pm\alpha$. Then the derived group of L is isomorphic to SL_2. By Corollary 3.16, the fiber X_x has a unique singularity of type A_1. For $y \in \mathfrak{c}^{\mathrm{rs}}$ sufficiently close to x, we have by Corollary 4.7 a diagram

$$\begin{tikzcd} H_1(X_y^1, \mathbb{F}_2) \arrow[r] \arrow[d] & X_*(C(x))/2X_*(C(x)) \arrow[d] \\ H_1(X_y, \mathbb{F}_2) \arrow[r] & X_*(C)/2X_*(C). \end{tikzcd}$$

It follows from the calculations in Example 4.3 for $G = \mathrm{SL}_2$ that the top arrow is an isomorphism, while the right vertical arrow has image equal to the image of the set $\{0, \alpha^\vee\}$ in $X_*(C)/2X_*(C)$. Moreover, it is clear from the proof of Proposition 4.6 and the definition of the vanishing cycle (Lemma 4.20) that the image of the nontrivial element of $H_1(X_y^1, \mathbb{F}_2)$ in $H_1(X_y, \mathbb{F}_2)$ is exactly the vanishing cycle γ_α. Applying the commutativity of the above diagram, we deduce that the image of γ_α in $X_*(C)/2X_*(C)$ is just $\alpha^\vee \bmod 2X_*(C)$. Since γ_α comes from a global section of the local system $\mathcal{H}_1(X)$, we deduce the result for any $y \in \mathfrak{c}^{\mathrm{rs}}$, not just y sufficiently close to x.

It follows that for any $y \in \mathfrak{c}^{\mathrm{rs}}$, the map

$$H_1(X_y, \mathbb{F}_2) \to Z_{G_0}(y) \cong X_*(C)/2X_*(C)$$

takes a basis of $H_1(X_y, \mathbb{F}_2)$, namely the set of γ_α as α ranges over a set of simple roots, to a basis of $X_*(C)/2X_*(C)$, namely the corresponding set of simple coroots. This completes the proof of the theorem.

The case of G adjoint. We now introduce a compactification of the family $X \to B$ of affine curves.

Lemma 4.9. *The family $\varphi\colon X \to B$ admits a compactification to a family $Y \to B$ of projective curves. Endow $Y \setminus X$ with its reduced closed subscheme structure. Then $Y \setminus X$ is a disjoint union of smooth nonintersecting open subschemes $P_1, \dots, P_s$, each of which maps isomorphically onto B. Moreover, $Y \to B$ is smooth in a Zariski neighborhood of each P_i. For each $b \in B^{\mathrm{rs}}(k)$, Y_b is the unique smooth projective curve containing X_b as a dense open subset. Each irreducible component of Y_0 meets exactly one of the sections P_i.*

Proof. We take the projective closure of the equations given in Theorem 3.8, and blow up any singularities at infinity. An easy calculation shows in each case that the induced family $Y \to B$ satisfies the required properties. □

Let us now suppose that G is adjoint, and let $G^{\mathrm{sc}} \to G$ denote its simply connected cover. We write Z^{sc} for the stabilizer scheme of G^{sc} over B. The natural map $Z^{\mathrm{sc}} \to Z$ is fiberwise surjective. Fix $b \in B^{\mathrm{rs}}(k)$. In Theorem 4.2, we saw that the inclusion $X_b \hookrightarrow \mathfrak{g}_{1,b}$ induces an isomorphism $H_1(X_b, \mathbb{F}_2) \to Z_b^{\mathrm{sc}}$ of finite k-groups. On the other hand, we have a surjection $H_1(X_b, \mathbb{F}_2) \to H_1(Y_b, \mathbb{F}_2)$.

Theorem 4.10. *The composite*

$$H_1(X_b, \mathbb{F}_2) \to Z_b^{\mathrm{sc}} \to Z_b$$

factors through this surjection, and induces an isomorphism $H_1(Y_b, \mathbb{F}_2) \cong Z_b$.

By Corollary 2.12, there is a canonical alternating pairing on Z_b^{sc}, with radical equal to the kernel of the map $Z_b^{\mathrm{sc}} \to Z_b$. On the other hand, there is a pairing $(\cdot\,,\cdot)$ on $H_1(X_b, \mathbb{F}_2)$, namely the intersection product, whose radical is exactly the kernel of the map $H_1(X_b, \mathbb{F}_2) \to H_1(Y_b, \mathbb{F}_2)$. The theorem is therefore a consequence of the following result.

Theorem 4.11. *The isomorphism $H_1(X_b, \mathbb{F}_2) \cong Z_b^{\mathrm{sc}}$ preserves these alternating pairings.*

Corollary 4.12. *There is an isomorphism $J_{Y_b}[2] \cong Z_b$ of finite k-groups that takes the Weil pairing to the pairing on Z_b defined in Corollary 2.12.*

Proof of Theorem 4.11. We can again reduce to the case $k = \mathbb{C}$. Fix a choice of Cartan subspace $\mathfrak{c}$, and let $C \subset G^{\mathrm{sc}}$ be the corresponding maximal torus. Choose $y \in \mathfrak{c}^{\mathrm{rs}}$. Let $\gamma_\alpha \in H_1(X_y, \mathbb{F}_2)$ be the element defined in Theorem 4.8. We will derive the theorem from the following statement: fix a root basis $R_{\mathfrak{c}}$ of $\Phi_{\mathfrak{c}}$, and let $\alpha, \beta \in R_{\mathfrak{c}}$ be distinct roots. Then $(\gamma_\alpha, \gamma_\beta) = 1$ if α, β are adjacent in the Dynkin diagram of $\mathfrak{g}$, and $(\gamma_\alpha, \gamma_\beta) = 0$ otherwise. We split the rest of the proof into two cases, according to these possibilities.

Case 1. If α, β are distinct adjacent roots, then we can choose $x \in \mathfrak{c}$ such that the elements of $\Phi_{\mathfrak{c}}$ vanishing on x are exactly the linear combinations of α and β. Let $L = Z_{G^{\mathrm{sc}}}(x)$ and $L^1 = L^{\mathrm{der}}$. Then $L^1 \cong \mathrm{SL}_3$, and the root system $\Phi_{\mathfrak{c}}(x) \subset \Phi_{\mathfrak{c}}$ is spanned by α and β. Moreover, we have by Corollary 4.7 for all $y \in \mathfrak{c}^{\mathrm{rs}}$ sufficiently close to x a commutative diagram

$$\begin{tikzcd} H_1(X_y^1, \mathbb{F}_2) \arrow[r] \arrow[d] & X_*(C(x))/2X_*(C(x)) \arrow[d] \\ H_1(X_y, \mathbb{F}_2) \arrow[r] & X_*(C)/2X_*(C), \end{tikzcd}$$

where $C(x) \subset L^1$ is the maximal torus with Lie algebra $\mathfrak{c} \cap \mathfrak{l}^1$. We know that the horizontal arrows are isomorphisms, and the vertical arrows are injective. The vertical arrows preserve the corresponding pairings.

Now, both of the objects in the top row of the above diagram are 2-dimensional $\mathbb{F}_2$-vector spaces, and their corresponding pairings are nondegenerate. (This is easy to see: the curve X_y^1 is a smooth affine curve of the form $y^2 = x^3 + ax + b$.) There is a unique nondegenerate alternating pairing on any 2-dimensional $\mathbb{F}_2$-vector space, so we deduce that $(\gamma_\alpha, \gamma_\beta) = 1$.

Case 2. If α, β are distinct roots which are not adjacent in the Dynkin diagram of $\mathfrak{g}$, then we can choose $x \in \mathfrak{c}$ such that the roots vanishing on x are exactly the linear combinations of α and β. Let $L = Z_G(x)$ and $L^1 = L^{\mathrm{der}}$. Then $L^1 \cong \mathrm{SL}_2 \times \mathrm{SL}_2$, and X_y has exactly two singularities, each of type A_1. We can choose disjoint open neighborhoods U_1, U_2 of these singularities in X such that for each $y \in \mathfrak{c}^{\mathrm{rs}}$ sufficiently close to x, the map $H_1(U_{1,y} \cup U_{2,y}, \mathbb{F}_2) \to H_1(X_y, \mathbb{F}_2)$ is injective and has image equal to the span of γ_α and γ_β. We see that these homology classes can be represented by cycles contained inside disjoint open sets of X_y. Therefore their intersection pairing is zero, and the theorem follows. $\square$

A parametrization of orbits. We suppose again that k is a general field of characteristic 0. Before stating our last main theorem, we summarize our hypotheses. We fix the following data:

- A split simple adjoint group G over k, of type A_r, D_r, or E_r.

- A stable involution θ of G and a regular nilpotent element $E \in \mathfrak{g}_1$.
- A choice of subregular normal $\mathfrak{sl}_2$-triple (e, h, f).

In terms of these data, we have defined:

- The categorical quotient $B = \mathfrak{g}_1 /\!\!/ G_0$.
- The Kostant section $\kappa \subset \mathfrak{g}_1$.
- A family of reduced connected curves $X \to B$.
- A family of projective curves $Y \to B$ containing X as a fiberwise dense open subset.
- A stabilizer scheme $Z \to B$ whose fiber over $b \in B(k)$ is isomorphic to the stabilizer of any regular element in $\mathfrak{g}_{1,b}$.
- For each $b \in B^{\mathrm{rs}}(k)$, a natural isomorphism $J_{Y_b}[2] \cong Z_b$, that takes the Weil pairing to the nondegenerate alternating pairing on Z_b defined in Corollary 2.12.

Proposition 4.13. *For each $b \in B^{\mathrm{rs}}(k)$, there is a bijection*

$$\mathfrak{g}_{1,b}(k)/G_0(k) \cong \ker\big(H^1(k, J_{Y_b}[2]) \to H^1(k, G_0)\big),$$

which takes the orbit of κ_b to the distinguished element of $H^1(k, J_{Y_b}[2])$.

Proof. Let K be a separable closure of k. We recall that if H is an algebraic group over k which acts on a variety X, and $H(K)$ acts transitively on $X(K)$, then given $x \in X(k)$ there is a bijection

$$X(k)/H(k) \cong \ker\big(H^1(k, Z_H(x)) \to H^1(k, H)\big),$$

under which the $H(k)$-orbit of x is mapped to the distinguished element, by [Gross and Bhargava 2012, Proposition 1]. We apply this with $H = G_0$, $X = \mathfrak{g}_{1,b}$, and base point $x = \kappa_b \in \mathfrak{g}_{1,b}(k)$ induced by the Kostant section. The result follows on using the identification $Z_H(x) \cong J_{Y_b}[2]$ of Theorem 4.10. □

To go further we want to interpret the relative position of the nilpotent elements E and e geometrically.

Lemma 4.14. *There are bijections between the following sets*:

1. *The set of irreducible components of X_0.*
2. *The set of G_0-orbits of regular nilpotent elements in $\mathfrak{g}_1$ containing the G_0-orbit of e in their closure.*
3. *The set of connected components of $Y \setminus X$.*

Proof. The map $\mu_0 \colon G_0 \times X_0 \to \mathcal{N}(\mathfrak{g}_1)$ is flat, and so has open image. This image therefore contains all regular nilpotent G_0-orbits whose closure meets e. On the other hand, one checks using Proposition 2.30 that in each case the number of

regular nilpotent G_0-orbits containing e in their closure is equal to the number of irreducible components of X_0. We can therefore define a bijection between the first two sets by taking an irreducible component of X_0 to the G_0-orbit of any point on its smooth locus.

We write $Y \setminus X = P_1 \cup \cdots \cup P_s$ as a disjoint union of open subschemes, each of which maps isomorphically onto B. By Lemma 4.9, each irreducible component of Y_0 meets a unique section P_i. We define a bijection between the first and third sets above by taking an irreducible component of X_0 to the unique section P_i meeting its closure in Y_0. □

We come now to our main theorem. We choose a section $P \cong B$ inside $Y \setminus X$, and we suppose that E corresponds under the bijection of Lemma 4.14 to the unique component of X_0 whose closure in Y_0 meets P. For each $b \in B^{\mathrm{rs}}(k)$, $P_b \in Y_b(k)$ defines an Abel–Jacobi map $f^{P_b} \colon Y_b \hookrightarrow J_{Y_b}$. (For the definition of this map, see [Milne 1986, §2].)

Theorem 4.15. *For every $b \in B^{\mathrm{rs}}(k)$, there is a commutative diagram, functorial in k, and depending only on e up to $G_0(k)$-conjugacy*:

$$\begin{tikzcd} X_b(k) \arrow[r, "\iota"] \arrow[d, "g"] & \mathfrak{g}_{1,b}(k)/G_0(k) \arrow[d, "\gamma"] \\ J_{Y_b}(k) \arrow[r, "\delta"] & H^1(k, J_{Y_b}[2]). \end{tikzcd}$$

The arrows in this diagram are defined as follows:

- *ι is induced by the inclusion $X_b \hookrightarrow \mathfrak{g}_{1,b}$.*
- *g is the restriction of the Abel–Jacobi map f^{P_b} to $X_b \subset Y_b$.*
- *δ is the usual 2-descent map in Galois cohomology associated to the exact sequence*
$$0 \to J_{Y_b}[2] \to J_{Y_b} \xrightarrow{[2]} J_{Y_b} \to 0.$$
- *γ is the classifying map of Proposition 4.13.*

Proof. We think of the group $H^1(k, J_{Y_b}[2])$ as classifying $J_{Y_b}[2]$-torsors over k. With b as in the theorem, let $E_b = [2]^{-1} f^{P_b}(Y_b) \subset J_{Y_b}$. We write $j_b \colon E_b \to Y_b$ for the natural map. This is a $J_{Y_b}[2]$-torsor over Y_b, and the composite $\delta \circ g$ sends a point $Q \in X_b(k)$ to the class of the torsor $j_b^{-1}(Q) \subset E_b$.

On the other hand, we recall the $J_{Y_b}[2]$-torsor $\Gamma_b \to X_b$ of (4-1), which extends uniquely to a torsor $h_b \colon D_b \to Y_b$, by Theorem 4.10. The composite $\gamma \circ \iota$ sends a point $Q \in X_b(k)$ to the class of $h_b^{-1}(Q)$. It follows from [Milne 1986, Proposition 9.1] that the two covers $D_b \to Y_b$ and $E_b \to Y_b$ become isomorphic as $J_{Y_b}[2]$-torsors after extending scalars to a separable closure of k. To prove the theorem, it therefore

suffices to prove that D_b and E_b are isomorphic as $J_{Y_b}[2]$-torsors over Y_b, before extending scalars. It even suffices to prove that $h_b^{-1}(P_b)$ is always the split torsor, or in other words that $h_b^{-1}(P_b)(k)$ is not the empty set.

Let $\mu\colon G_0 \times \kappa \to \mathfrak{g}_1$ denote the orbit map, and let X' denote the intersection of X with the image of μ. Because of the compatibility between E and P, the subset $X' \cup P$ of the underlying topological space of Y is open; let Y' denote the corresponding open subscheme. Then Y' contains a Zariski open neighborhood of P in Y.

Let $\Gamma' = \mu^{-1}(X')$; this is a Z-torsor over X'. We show that Γ' extends to a Z-torsor over Y'. In fact, there is a commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & H^1_{\text{ét}}(Y', Z) \arrow[r] \arrow[d] & H^1_{\text{ét}}(X', Z) \arrow[r] \arrow[d] & H^0_{\text{ét}}(Y', R^1 j_* Z) \arrow[d] \\
0 \arrow[r] & H^1_{\text{ét}}(Y'_K, Z) \arrow[r] & H^1_{\text{ét}}(X'_K, Z) \arrow[r] & H^0_{\text{ét}}(Y'_K, R^1 j_{K,*} Z),
\end{tikzcd}$$

where $j\colon X' \to Y'$ is the obvious open immersion, and $(\,\cdot\,)_K$ denotes base change to the separable closure K/k. Let $i\colon P \hookrightarrow Y'$ denote the complementary closed immersion. There is an isomorphism $R^1 j_{K,*}(Z) \cong i_{K,*} Z$, and hence

$$H^0_{\text{ét}}(Y'_K, R^1 j_{K,*}(Z)) = H^0_{\text{ét}}(B_K, Z).$$

The group $H^0_{\text{ét}}(B_K, Z)$ is trivial. Indeed, the morphism $Z \to B$ is étale, while the stalk of Z above the origin is trivial. The rightmost vertical arrow in the above diagram is injective, and so the class of Γ' in $H^1_{\text{ét}}(X', Z)$ lifts to $H^1_{\text{ét}}(Y', Z)$. We write $D' \to Y'$ for the corresponding torsor.

Let $F' \to B$ denote the pullback of D' to $B \cong P \hookrightarrow Y'$. We must show that for b as in the theorem, F'_b is the trivial Z-torsor over k. We claim that in fact, F' is trivial. For we can choose a Zariski open neighborhood U_0 of $0 \in B$ and a Galois finite étale cover $U \to U_0$ such that $F' \times_B U$ has a trivialization as a Z-torsor. If U is sufficiently small, then $Z(U) \hookrightarrow Z_0 = 0$ is trivial, so there is a unique such trivialization. By descent, there exists a unique trivialization of F' over U_0. The existence of the contracting $\mathbb{G}_m$-action on $X \to B$ now implies that F' must be globally trivial, as required. This completes the proof of the theorem. □

A conjecture. We hope that the representations studied in this paper will have applications to the study of the average size of the 2-Selmer groups of the Jacobian varieties J_{Y_b}. The first step towards such applications is the following conjecture.

Conjecture 4.16. With assumptions as in Theorem 4.15, there exists a function $\eta\colon J_{Y_b}(k) \to \mathfrak{g}_{1,b}(k)/G_0(k)$, functorial in k, making the diagram

$$\begin{tikzcd} X_b(k) \arrow[r, "\iota"] \arrow[d, "g"] & \mathfrak{g}_{1,b}(k)/G_0(k) \arrow[d, "\gamma"] \\ J_{Y_b}(k) \arrow[r, "\delta"] \arrow[ur, "\eta"] & H^1(k, J_{Y_b}[2]) \end{tikzcd}$$

commute.

The conjecture is true in each case (namely G of type A_2, A_3, or D_4) where the curves Y_b have genus one. The representations we construct in this case are a subset of the ones studied by Bhargava and Ho in their paper on coregular representations associated to genus-one curves [Bhargava and Ho 2013], and in a forthcoming work they apply their representations to the study of the average sizes of Selmer groups [Bhargava and Ho $\geq$ 2014]. In the cases listed above one could also apply the methods of this paper, together with Bhargava's techniques for counting integral points in truncated fundamental domains, to calculate the average size of the 2-Selmer groups of the curves in the corresponding families. Details will appear elsewhere.

Bhargava and Gross [2013] have shown something very close to this conjecture when G is of type A_{2r}. They construct rational orbits using the geometry of the intersection of two quadric hypersurfaces, and apply this to calculate the average size of the 2-Selmer groups of a certain family of hyperelliptic Jacobians. On the other hand, for some other Vinberg representations the work of Gruson, Sam and Weyman [Gruson et al. 2013] gives a relation between the geometric invariant theory and the geometry of the Jacobians of our algebraic curves, and it seems likely that this should extend to an arithmetic relation also.

The proof of Theorem 4.8. In this section we prove Theorem 4.8. Thus G is a simple simply connected group over $k = \mathbb{C}$, θ a stable involution, and $\mathfrak{c} \subset \mathfrak{g}_1$ a Cartan subspace. We fix a normal subregular $\mathfrak{sl}_2$-triple (e, h, f) in $\mathfrak{g}$, and define $S = e + \mathfrak{z}_{\mathfrak{g}}(f)$, $X = e + \mathfrak{z}_{\mathfrak{g}}(f)_1 = S \cap \mathfrak{g}_1$. Let τ denote the automorphism of S induced by $-\theta$; we then have $S^\tau = X$. In what follows we identify all varieties with their complex points.

Lemma 4.17. *Both S^{rs} and X^{rs} are locally trivial fibrations* (*in the analytic topology*) *over $\mathfrak{c}^{\mathrm{rs}}/W$.*

Proof. We combine the Ehresmann fibration theorem and the existence of a good compactification for X^{rs} to see that it is a locally trivial fibration over $\mathfrak{c}^{\mathrm{rs}}/W$. The corresponding result for S follows from the simple relationship between S and X, see Lemma 3.15. □

Corollary 4.18. *The homology groups $H_2(S_b, \mathbb{F}_2)$ and $H_1(X_b, \mathbb{F}_2)$ for $b \in \mathfrak{c}^{\mathrm{rs}}/W$ form local systems $\mathcal{H}_2(S)$ and $\mathcal{H}_1(X)$, and these local systems are canonically isomorphic.*

Proof. Only the second part needs proof. It follows either from a sheaf-theoretic argument, or from the assertion that suspension does not change the monodromy representation of a singularity, at least when one is working modulo 2; see [Arnol'd et al. 1988, Theorem 2.14]. □

Given $y \in \mathfrak{c}$ we write X_y and S_y for the respective fibers over y of the maps $X \times_{\mathfrak{c}/W} \mathfrak{c} \to \mathfrak{c}$ and $S \times_{\mathfrak{c}/W} \mathfrak{c} \to \mathfrak{c}$.

Lemma 4.19. *The local systems $\mathcal{H}_1(X)$ and $\mathcal{H}_2(S)$ become trivial after pullback to $\mathfrak{c}^{\mathrm{rs}}$.*

Proof. In light of Corollary 4.18, it suffices to prove this assertion for $\mathcal{H}_2(S)$. The existence of the Springer resolution implies the existence of a proper morphism $\tilde{S} \to S \times_{\mathfrak{c}/W} \mathfrak{c}$ such that for every $y \in \mathfrak{c}$, the induced map $\tilde{S} \to S_y$ is a minimal resolution of singularities. Moreover, $\tilde{S} \to \mathfrak{c}$ is a locally trivial fiber bundle and $\tilde{S} \times_{\mathfrak{c}} \mathfrak{c}^{\mathrm{rs}} \to S \times_{\mathfrak{c}/W} \mathfrak{c}^{\mathrm{rs}}$ is an isomorphism. See [Slodowy 1980a] for more details. These facts imply the lemma. □

It follows that for any $y, z \in \mathfrak{c}^{\mathrm{rs}}$, the groups $H_1(X_y, \mathbb{F}_2)$ and $H_1(X_z, \mathbb{F}_2)$ are canonically isomorphic.

It is a consequence of Lemma 3.15 that given $b \in \mathfrak{c}/W$, a fiber X_b has a unique nondegenerate critical point if and only if S_b does. Let $\gamma : [0, 1] \to \mathfrak{c}$ be a path such that $\gamma(t)$ is regular semisimple for $0 \le t < 1$, but such that a unique pair of roots $\pm\alpha$ vanishes on $\gamma(1) = x$. Then X_x (or S_x) has a unique nondegenerate critical point, by Corollary 3.16. Let $y = \gamma(0)$. We define a homology class (that we call a vanishing cycle) $[\gamma]_1 \in H_1(X_y, \mathbb{F}_2)$ as follows.

We can find local holomorphic coordinates $z_1, \dots, z_{r+1}$ on X centered at the critical point of X_b and local holomorphic coordinates $u_1, \dots, u_r$ on $\mathfrak{c}/W$ centered at b such that the map $X \to \mathfrak{c}/W$ is locally of the form

$$(z_1, \dots, z_{r+1}) \mapsto (z_1, \dots, z_{r-1}, z_r^2 + z_{r+1}^2).$$

For t close to 1, we can then define a sphere (for a suitable continuous choice of branch of $\sqrt{u_r(t)}$ near $t = 1$):

$$S^1(t) = \left\{ (u_1(t), \dots, u_{r-1}(t), \sqrt{u_r(t)}z_r, \sqrt{u_r(t)}z_{r+1}) \;\middle|\; z_r^2 + z_{r+1}^2 = 1, \Im z_i = 0 \right\}.$$

We define a homology class in $H_1(X_y, \mathbb{F}_2)$ by transporting the class of $S^1(t)$ for t close to 1 along the image of the path γ in $\mathfrak{c}/W$. An entirely analogous procedure defines $[\gamma]_2 \in H_2(S_x, \mathbb{F}_2)$.

Lemma 4.20. *The homology class of the cycle $[\gamma]_1 \in H_1(X_y, \mathbb{F}_2)$ (respectively, $[\gamma]_2 \in H_2(S_y, \mathbb{F}_2)$) is well-defined and depends only on α. Moreover, these classes correspond under the isomorphism $H_1(X_y, \mathbb{F}_2) \cong H_2(S_y, \mathbb{F}_2)$ of Corollary 4.18.*

Proof. It is well known that the $[\gamma]_i$ are well-defined and depend only on the path γ up to homotopy. It follows from Lemma 4.19 that the $[\gamma]_i$ depend only on the endpoint $x = \gamma(1)$ and not on the choice of path. To prove the lemma it suffices to show that $[\gamma]_2$ depends only on α. In fact $[\gamma]_2$ is, by construction, the unique nontrivial element in the kernel of the map

$$H_2(S_y, \mathbb{F}_2) = H_2(\tilde{S}_y, \mathbb{F}_2) \cong H_2(\tilde{S}_x, \mathbb{F}_2) \to H_2(S_x, \mathbb{F}_2).$$

The proof of [Shepherd-Barron 2001, Theorem 3.4] implies that there is an isomorphism of local systems $\mathcal{H}_2(\tilde{S}) \cong X_*(C)/2X_*(C)$ over $\mathfrak{c}$, and that the kernel of the map $H_2(\tilde{S}_x, \mathbb{F}_2) \to H_2(S_x, \mathbb{F}_2)$ corresponds under this isomorphism to the span in $X_*(C)/2X_*(C)$ of $\alpha^\vee$. □

We can therefore define for each $\alpha \in \Phi_{\mathfrak{c}}$ a global section γ_α of the pullback of the local system $\mathcal{H}_1(X)$ to $\mathfrak{c}^{\mathrm{rs}}$, namely the class $[\gamma]_1$ constructed above. Theorem 4.8 now follows from the above facts and the following result.

Lemma 4.21. *Let $R_{\mathfrak{c}} \subset \Phi_{\mathfrak{c}}$ be a choice of root basis, and let $x \in \mathfrak{c}^{\mathrm{rs}}$. Then the set $\{\gamma_\alpha \mid \alpha \in R_{\mathfrak{c}}\}$ is a basis of $H_1(X_x, \mathbb{F}_2)$ as $\mathbb{F}_2$-vector space.*

Proof. This follows immediately from the corresponding fact for the simple coroots $\{\alpha^\vee \mid \alpha \in R_{\mathfrak{c}}\}$. □

Acknowledgements

This paper is a revised version of the author's Harvard Ph.D. thesis, written under the supervision of Benedict H. Gross. I wish to thank him for many useful suggestions and conversations.

References

[Arnol'd et al. 1988] V. I. Arnol'd, V. V. Goryunov, O. V. Lyashko, and V. A. Vasil'ev, "Особенности, I: Покальная и глобальная теория", pp. 5–257 in *Dynamical systems 6*, Itogi Nauki i Tekhniki Sovrem. Probl. Mat. Fund. Napr. **6**, VINITI, Moscow, 1988. Translated as "Singularities, I: Local and global theory" in *Singularity theory I*, Encycl. Math. Sci. **6**, Springer, Berlin, 1998. MR 91h:58010b Zbl 0691.58002

[Bhargava and Gross 2013] M. Bhargava and B. H. Gross, "The average size of the 2-Selmer group of Jacobians of hyperelliptic curves having a rational Weierstrass point", preprint, 2013, available at http://www.math.harvard.edu/~gross/preprints/stable23.pdf.

[Bhargava and Ho 2013] M. Bhargava and W. Ho, "Coregular spaces and genus one curves", preprint, 2013. arXiv 1306.4424

[Bhargava and Ho ≥ 2014] M. Bhargava and W. Ho, "On the average sizes of Selmer groups in families of elliptic curves", In preparation.

[Bhargava and Shankar 2010] M. Bhargava and A. Shankar, "Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves", preprint, 2010. To appear in *Annals of Math.* arXiv 1006.1002

[Bourbaki 1968] N. Bourbaki, *Groupes et algèbres de Lie,* chapitres 4–6, Actualités Scientifiques et Industrielles **1337**, Hermann, Paris, 1968. Translated as *Lie groups and Lie algebras*, Chapters 4–6, Springer, Berlin, 2008. MR 39 #1590 Zbl 0483.22001

[Bourbaki 1975] N. Bourbaki, *Groupes et algèbres de Lie,* chapitres 7 et 8, Actualités Scientifiques et Industrielles **1364**, Hermann, Paris, 1975. Translated in *Lie groups and Lie algebras*, Chapters 7–9, Springer, Berlin, 2008. Zbl 0329.17002

[Brieskorn 1971] E. Brieskorn, "Singular elements of semi-simple algebraic groups", pp. 279–284 in *Actes du Congrès International des Mathématiciens* (Nice, 1970), vol. 2, Gauthier-Villars, Paris, 1971. MR 55 #10720 Zbl 0223.22012

[Collingwood and McGovern 1993] D. H. Collingwood and W. M. McGovern, *Nilpotent orbits in semisimple Lie algebras*, Van Nostrand Reinhold, New York, 1993. MR 94j:17001 Zbl 0972.17008

[Cook 1998] P. R. Cook, "Compactified Jacobians and curves with simple singularities", pp. 37–47 in *Algebraic geometry* (Catania, 1993 and Barcelona, 1994), edited by P. E. Newstead, Lecture Notes in Pure and Appl. Math. **200**, Dekker, New York, 1998. MR 99j:14031 Zbl 0951.14018

[Đoković 2001] D. Ž. Đoković, "The closure diagram for nilpotent orbits of the split real form of E_7", *Represent. Theory* **5** (2001), 284–316. MR 2002g:17019 Zbl 1050.17007

[Đoković and Litvinov 2003] D. Ž. Đoković and M. Litvinov, "The closure ordering of nilpotent orbits of the complex symmetric pair $(\mathrm{SO}_{p+q}, \mathrm{SO}_p \times \mathrm{SO}_q)$", *Canad. J. Math.* **55**:6 (2003), 1155–1190. MR 2004j:22017 Zbl 1060.17002

[Esnault 1980] H. Esnault, "Sur l'identification de singularités apparaissant dans des groupes algébriques complexes", pp. 31–59 in *Seminar on Singularities* (Paris, 1976–1977), edited by D. T. Lê, Publ. Math. Univ. Paris VII **7**, Univ. Paris VII, Paris, 1980. MR 84h:14005

[de Graaf 2011] W. A. de Graaf, "Computing representatives of nilpotent orbits of θ-groups", *J. Symbolic Comput.* **46**:4 (2011), 438–458. MR 2012f:17008 Zbl 1222.17002

[Gross and Bhargava 2012] B. H. Gross and M. Bhargava, "Arithmetic invariant theory", preprint, 2012, available at http://www.math.harvard.edu/~gross/preprints/invariant.pdf.

[Gruson et al. 2013] L. Gruson, S. V. Sam, and J. Weyman, "Moduli of abelian varieties, Vinberg θ-groups, and free resolutions", pp. 419–469 in *Commutative algebra*, edited by I. Peeva, Springer, New York, 2013. Zbl 06153307 arXiv 1203.2575

[Ho 2009] W. Ho, *Orbit parametrizations of curves*, thesis, Princeton University, Princeton, NJ, 2009, available at http://tinyurl.com/hothesis. MR 2713823

[Humphreys 1975] J. E. Humphreys, *Linear algebraic groups*, Graduate Texts in Mathematics **21**, Springer, New York, 1975. MR 53 #633 Zbl 0325.20039

[Kawanaka 1987] N. Kawanaka, "Orbits and stabilizers of nilpotent elements of a graded semisimple Lie algebra", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**:3 (1987), 573–597. MR 89j:17012 Zbl 0651.20046

[Kostant and Rallis 1971] B. Kostant and S. Rallis, "Orbits and representations associated with symmetric spaces", *Amer. J. Math.* **93** (1971), 753–809. MR 47 #399 Zbl 0224.22013

[Levy 2007] P. Levy, "Involutions of reductive Lie algebras in positive characteristic", *Adv. Math.* **210**:2 (2007), 505–559. MR 2008b:17028 Zbl 1173.17019

[Levy 2009] P. Levy, "Vinberg's θ-groups in positive characteristic and Kostant–Weierstrass slices", *Transform. Groups* **14**:2 (2009), 417–461. MR 2010g:17022 Zbl 1209.17016

[Milne 1986] J. S. Milne, "Jacobian varieties", pp. 167–212 in *Arithmetic geometry* (Storrs, CT, 1984), edited by G. Cornell and J. H. Silverman, Springer, New York, 1986. MR 861976 Zbl 0604.14018

[Onishchik and Vinberg 1988] A. L. Onishchik and È. B. Vinberg, *Семинар по группам Ли и алгебраическим группам*, Nauka, Moscow, 1988. Translated as *Lie groups and algebraic groups*, Springer, Berlin, 1990. MR 91g:22001 Zbl 0722.22004

[Panyushev 2005] D. I. Panyushev, "On invariant theory of θ-groups", *J. Algebra* **283**:2 (2005), 655–670. MR 2006g:17022 Zbl 1071.17005

[Poonen 2013] B. Poonen, "Average rank of elliptic curves: after Manjul Bhargava and Arul Shankar", exposé no. 1049 in *Séminaire Bourbaki, volume* 2011/2012, Astérisque **352**, Soc. Math. de France, Paris, 2013.

[Reeder 2011] M. Reeder, "Elliptic centralizers in Weyl groups and their coinvariant representations", *Represent. Theory* **15** (2011), 63–111. MR 2012b:20108 Zbl 1251.20042

[Reeder et al. 2012] M. Reeder, P. Levy, J.-K. Yu, and B. H. Gross, "Gradings of positive rank on simple Lie algebras", *Transform. Groups* **17**:4 (2012), 1123–1190. MR 3000483 Zbl 06136962

[Sekiguchi and Shimizu 1981] J. Sekiguchi and Y. Shimizu, "Simple singularities and infinitesimally symmetric spaces", *Proc. Japan Acad. Ser. A Math. Sci.* **57**:1 (1981), 42–46. MR 82g:17007 Zbl 0481.58009

[Shepherd-Barron 2001] N. I. Shepherd-Barron, "On simple groups and simple singularities", *Israel J. Math.* **123** (2001), 179–188. MR 2002c:14076 Zbl 1007.14001

[Slodowy 1980a] P. Slodowy, "Four lectures on simple groups and singularities", Communications of the Mathematical Institute 11, Rijksuniversiteit Utrecht, 1980. MR 82b:14002 Zbl 0425.22020

[Slodowy 1980b] P. Slodowy, *Simple singularities and simple algebraic groups*, Lecture Notes in Mathematics **815**, Springer, Berlin, 1980. MR 82g:14037 Zbl 0441.14002

[Springer 1977] T. A. Springer, *Invariant theory*, Lecture Notes in Mathematics **585**, Springer, Berlin, 1977. MR 56 #5740 Zbl 0346.20020

[Springer 2009] T. A. Springer, *Linear algebraic groups*, Second ed., Birkhäuser, Boston, 2009. MR 2009i:20089 Zbl 1202.20048

[Steinberg 1974] R. Steinberg, *Conjugacy classes in algebraic groups*, Lecture Notes in Mathematics **366**, Springer, Berlin, 1974. MR 50 #4766 Zbl 0281.20037

[Vinberg 1976] È. B. Vinberg, "Группа Вейля градуированной алгебры Ли", *Izv. Akad. Nauk SSSR Ser. Mat.* **40**:3 (1976), 488–526. Translated as "The Weyl group of a graded Lie algebra" in *Math. USSR-Izvestiya* **10**:3 (1976), 463–495. MR 55 #3175 Zbl 0363.20035

Communicated by Bjorn Poonen
Received 2012-11-08 Revised 2013-02-14 Accepted 2013-03-17

thorne@math.harvard.edu *Department of Mathematics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 7 No. 9 2013