Algebra & Number Theory

Volume 7
2013
No. 3

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 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.507

Ekedahl–Oort strata of hyperelliptic curves in characteristic 2

Arsen Elkin and Rachel Pries

Suppose X is a hyperelliptic curve of genus g defined over an algebraically closed field k of characteristic $p = 2$. We prove that the de Rham cohomology of X decomposes into pieces indexed by the branch points of the hyperelliptic cover. This allows us to compute the isomorphism class of the 2-torsion group scheme $J_X[2]$ of the Jacobian of X in terms of the Ekedahl–Oort type. The interesting feature is that $J_X[2]$ depends only on some discrete invariants of X, namely, on the ramification invariants associated with the branch points. We give a complete classification of the group schemes that occur as the 2-torsion group schemes of Jacobians of hyperelliptic k-curves of arbitrary genus, showing that only relatively few of the possible group schemes actually do occur.

1. Introduction

Suppose k is an algebraically closed field of characteristic $p > 0$. There are several important stratifications of the moduli space $\mathcal{A}_g$ of principally polarized abelian varieties of dimension g defined over k, including the Ekedahl–Oort stratification. The Ekedahl–Oort type characterizes the p-torsion group scheme of the corresponding abelian varieties and, in particular, determines invariants of the group scheme such as the p-rank and a-number. It is defined by the interaction between the Frobenius F and Verschiebung V operators on the p-torsion group scheme. Very little is known about how the Ekedahl–Oort strata intersect the Torelli locus of Jacobians of curves. In particular, one would like to know which group schemes occur as the p-torsion $J_X[p]$ of the Jacobian J_X of a curve X of genus g.

In this paper, we completely answer this question for hyperelliptic k-curves X of arbitrary genus when k has characteristic $p = 2$, a case that is amenable to calculation because of the confluence of hyperelliptic and Artin–Schreier properties. We first prove a decomposition result about the structure of $\mathrm{H}^1_{\mathrm{dR}}(X)$ as a module under the

Elkin is partially supported by the Marie Curie Incoming International Fellowship PIIF-GA-2009-236606. Pries is partially supported by National Science Foundation grant DMS-11-01712.
MSC2010: primary 11G20; secondary 14K15, 14L15, 14H40, 14F40, 11G10.
Keywords: curve, hyperelliptic, Artin–Schreier, Jacobian, p-torsion, a-number, group scheme, de Rham cohomology, Ekedahl–Oort strata.

actions of F and V, where the pieces of the decomposition are indexed by the branch points of the hyperelliptic cover. This is the only decomposition result about the de Rham cohomology of Artin–Schreier curves that we know of, though the action of V on $\mathrm{H}^0(X, \Omega^1)$ and the action of F on $\mathrm{H}^1(X, \mathbb{O})$ have been studied for Artin–Schreier curves under less restrictive hypotheses [Madden 1978; Sullivan 1975].

The second result of this paper is a complete classification of the isomorphism classes of group schemes that occur as the 2-torsion group scheme $J_X[2]$ for a hyperelliptic k-curve X of arbitrary genus when $\mathrm{char}(k) = 2$. The group schemes that occur decompose into pieces indexed by the branch points of the hyperelliptic cover, and we determine the Ekedahl–Oort types of these pieces. In particular, we determine which a-numbers occur for the 2-torsion group schemes of hyperelliptic k-curves of arbitrary genus when $\mathrm{char}(k) = 2$. Before describing the result precisely, we note that it shows that the group scheme $J_X[2]$ depends only on some discrete invariants of X and not on the location of the branch points or the equation of the hyperelliptic cover. This is in sharp contrast to the case of hyperelliptic curves in odd characteristic p, where even the p-rank depends on the location of the branch points [Yui 1978].

Notation 1.1. Suppose k is an algebraically closed field of characteristic $p = 2$. Let X be a k-curve of genus g that is hyperelliptic, in other words, for which there exists a degree two cover $\pi : X \to \mathbb{P}^1$. Let $B \subset \mathbb{P}^1(k)$ denote the set of branch points of π, and let $r := \#B - 1$. After a fractional linear transformation, one may suppose that $0 \in B$ and $\infty \notin B$.

For $\alpha \in B$, the *ramification invariant* d_α is the largest integer for which the higher ramification group of π above α is nontrivial. By [Stichtenoth 2009, Proposition III.7.8], d_α is odd. Let $c_\alpha := (d_\alpha - 1)/2$, and let $x_\alpha := (x - \alpha)^{-1}$.

The cover π is given by an affine equation of the form $y^2 - y = f(x)$ for some nonconstant rational function $f(x) \in k(x)$. After a change of variables of the form $y \mapsto y + \epsilon$, one may suppose the partial fraction decomposition of $f(x)$ has the form

$$f(x) = \sum_{\alpha \in B} f_\alpha(x_\alpha), \tag{1-1}$$

where $f_\alpha(x) \in xk[x^2]$ is a polynomial of degree d_α containing no monomials of even exponent. In particular, the divisor of poles of $f(x)$ on $\mathbb{P}^1$ has the form

$$\mathrm{div}_\infty(f(x)) = \sum_{\alpha \in B} d_\alpha \alpha.$$

By the Riemann–Hurwitz formula [Serre 1968, IV, Proposition 4], the genus g of X satisfies

$$2g + 2 = \sum_{\alpha \in B} (d_\alpha + 1).$$

Recall that the 2-*rank* of (the Jacobian of) the k-curve X is $\dim_{\mathbb{F}_2} \mathrm{Hom}(\mu_2, J_X[2])$, where μ_2 is the kernel of Frobenius on $\mathbb{G}_m$. By the Deuring–Shafarevich formula [Subrao 1975, Theorem 4.2; Crew 1984, Corollary 1.8], the 2-rank of X is r. Note that $g = r + \sum_{\alpha \in B} c_\alpha$. The implication of these formulas is that, for a given genus g (and 2-rank r), there is an additional discrete invariant of the hyperelliptic k-curve X, namely, a partition of $2g+2$ into $r+1$ positive even integers $d_\alpha + 1$. In Section 5a, we show that the Ekedahl–Oort type of X depends only on this discrete invariant.

Theorem 1.2. *Suppose X is a hyperelliptic curve defined over an algebraically closed field k of characteristic 2 with affine equation $y^2 - y = f(x)$, branch locus B, and polynomials f_α for $\alpha \in B$ as described in Notation 1.1. For $\alpha \in B$, consider the Artin–Schreier k-curve Y_α with affine equation $y^2 - y = f_\alpha(x)$. Let E be an ordinary elliptic k-curve. As a module under the actions of Frobenius F and Verschiebung V, the de Rham cohomology of X decomposes as*

$$\mathrm{H}^1_{\mathrm{dR}}(X) \cong \mathrm{H}^1_{\mathrm{dR}}(E)^{\#B-1} \oplus \bigoplus_{\alpha \in B} \mathrm{H}^1_{\mathrm{dR}}(Y_\alpha).$$

As an application of Theorem 1.2, we give a complete classification of the Ekedahl–Oort types that occur for hyperelliptic k-curves. Recall that the 2-torsion group scheme $J_X[2]$ of the Jacobian of a k-curve is a polarized BT_1 group scheme over k (short for polarized Barsotti–Tate truncated level-1 group scheme) and that the isomorphism class of a BT_1 group scheme determines and is determined by its Ekedahl–Oort type; see Section 2 for more details. For $p = 2$ and a natural number c, let G_c be the polarized BT_1 group scheme of rank p^{2c} with Ekedahl–Oort type $[0, 1, 1, 2, 2, \dots, \lfloor c/2 \rfloor]$. For example, G_1 is the 2-torsion group scheme of a supersingular elliptic k-curve. The group scheme G_2 occurs as the 2-torsion of a supersingular nonsuperspecial abelian surface over k. The group scheme G_c is not necessarily indecomposable. More explanation about G_c is given in Sections 2c and 5b.

Before stating the classification result, we note that it also includes a complete description of which a-numbers occur for the Jacobians of hyperelliptic k-curves. Recall that the a-number of X is defined as $a_X := \dim_k \mathrm{Hom}(\alpha_2, J_X[2])$, where α_2 is the kernel of Frobenius on $\mathbb{G}_a$.

Theorem 1.3. *Let X be a hyperelliptic k-curve with affine equation $y^2 - y = f(x)$ defined over an algebraically closed field of characteristic 2 as described in Notation 1.1. Then the 2-torsion group scheme of the Jacobian variety of X is*

$$J_X[2] \simeq (\mathbb{Z}/2 \oplus \mu_2)^r \oplus \bigoplus_{\alpha \in B} G_{c_\alpha},$$

and the a-number of X is $a_X = (g + 1 - \#\{\alpha \in B \mid d_\alpha \equiv 1 \bmod 4\})/2$.

Theorem 1.3 is stated without proof in [van der Geer 1999, 3.2] for the special case when $f(x) \in k[x]$, that is, $r = 0$. There are two interesting things about Theorem 1.3. First, it shows that the Ekedahl–Oort type of $X : y^2 - y = f(x)$ depends only on the orders of the poles of $f(x)$. This is in sharp contrast to the case of hyperelliptic curves in odd characteristic p, where even the p-rank depends on $f(x)$ and the location of the branch points [Yui 1978]. Similarly, it differs from the results of [Bouw 2001; Elkin 2011; Johnston 2007], all of which give bounds for the p-rank and a-number of various kinds of curves that depend strongly on the coefficients of their equations. Likewise, preliminary calculations indicate that it is in contrast to the situation for Artin–Schreier curves in odd characteristic.

Secondly, Theorem 1.3 is interesting because it shows that most of the possibilities for the 2-torsion group scheme of an abelian variety over k do not occur for Jacobians of hyperelliptic k-curves when $\mathrm{char}(k) = 2$. Specifically, there are 2^g possibilities for the 2-torsion group scheme of a g-dimensional abelian variety over k. We determine a subset of these of cardinality equal to the number $P(g+1)$ of partitions of $g+1$ and prove that the group schemes in this subset are exactly those that occur as the 2-torsion $J_X[2]$ for a hyperelliptic k-curve X of genus g. Recall [Hardy and Ramanujan 1918] that $P(g+1)$ grows asymptotically like $e^{\pi\sqrt{2(g+1)/3}}/(4\sqrt{3}(g+1))$ as g goes to infinity. Also, Theorem 1.3 gives the nontrivial bounds $(g-r)/2 \le a_X \le (g+1)/2$ for the a-number.

An earlier nonexistence result of this type is due to Ekedahl [1987], who proved that a curve X of genus $g > p(p-1)/2$ in characteristic $p > 0$ cannot be superspecial and thus $a_X < g$. There are also other recent results about Newton polygons of hyperelliptic (that is, Artin–Schreier) curves in characteristic 2, including several nonexistence results [Blache 2012; Scholten and Zhu 2002]. In addition, there are closed formulas for the number of hyperelliptic curves of genus 3 with given 2-rank over each finite field of characteristic 2 [Nart and Sadornil 2004].

Here is an outline of this paper. Section 2 contains notation and background. Results on $\mathrm{H}^0(X, \Omega^1)$ and the a-number are in Section 3. Theorem 1.2 is with the material on the de Rham cohomology in Section 4. Section 5 contains the results about the Ekedahl–Oort type, including Theorem 1.3.

2. Background

In this paper, all objects are defined over an algebraically closed field k of characteristic $p > 0$, and all curves are smooth, projective, and connected. This section includes background on p-torsion group schemes, Ekedahl–Oort types, the de Rham cohomology, and Frobenius and Verschiebung.

2a. ***The p-torsion group scheme.*** Suppose A is a principally polarized abelian variety of dimension g defined over k. For example, A could be the Jacobian of a

k-curve of genus g. Consider the multiplication-by-p morphism $[p]: A \to A$ that is a finite flat morphism of degree p^{2g}. It factors as $[p] = V \circ F$. Here $F: A \to A^{(p)}$ is the relative Frobenius morphism coming from the p-power map on the structure sheaf; it is purely inseparable of degree p^g. Furthermore, $V: A^{(p)} \to A$ is the Verschiebung morphism.

The *p-torsion group scheme* of A, denoted $A[p]$, is the kernel of $[p]$. It is a finite commutative group scheme annihilated by p, again having morphisms F and V. By [Oort 2001, 9.5], the p-torsion group scheme $A[p]$ is a polarized BT_1 group scheme over k (short for polarized Barsotti–Tate truncated level-1 group scheme) as defined in [Oort 2001, 2.1, 9.2]. The rank of $A[p]$ is p^{2g}.

We now give a brief summary of the classification [Oort 2001, Theorems 9.4 and 12.3] of polarized BT_1 group schemes over k in terms of Dieudonné modules and Ekedahl–Oort type; other useful references are [Kraft 1975] (without polarization) and [Moonen 2001] (for $p \geq 3$).

2b. ***The Dieudonné module and polarizations.*** It is useful to describe the group scheme $A[p]$ using (the modulo p reduction of) the *covariant Dieudonné module* [Oort 2001, 15.3]. This is the dual of the contravariant theory found in [Demazure 1972]. In brief, consider the noncommutative ring $\mathbb{E} = k[F, V]$ generated by semilinear operators F and V with the relations $FV = VF = 0$ and $F\lambda = \lambda^p F$ and $\lambda V = V\lambda^p$ for all $\lambda \in k$. Let $\mathbb{E}(A, B)$ denote the left ideal $\mathbb{E}A + \mathbb{E}B$ of $\mathbb{E}$ generated by A and B. A deep result is that the Dieudonné functor D gives an equivalence of categories between BT_1 group schemes over k (with rank p^{2g}) and finite left $\mathbb{E}$-modules (having dimension $2g$ as a k-vector space). We use the notation $D(\mathbb{G})$ to denote the Dieudonné module of $\mathbb{G}$. For example, the Dieudonné module of the p-torsion group scheme of an ordinary elliptic curve is $D(\mathbb{Z}/p \oplus \mu_p) \simeq \mathbb{E}/\mathbb{E}(F, 1-V) \oplus \mathbb{E}/\mathbb{E}(V, 1-F)$ [Goren 2002, Examples A.5.1 and A.5.3].

The polarization of A induces a symmetry on $A[p]$ as defined in [Oort 2001, 5.1], namely, an antisymmetric isomorphism from $A[p]$ to the Cartier dual group scheme $A[p]^{\mathrm{dual}}$ of $A[p]$. Unfortunately, in characteristic 2, there may be antisymmetric morphisms $A[p] \to A[p]^{\mathrm{dual}}$ that do not come from a polarization. Luckily, this issue can be resolved by defining a polarization on $A[p]$ in terms of a nondegenerate alternating pairing on $D(A[p])$ [Oort 2001, 9.2, 9.5, 12.2].

2c. ***The Ekedahl–Oort type.*** As in [Oort 2001, Sections 5 and 9], the isomorphism type of a BT_1 group scheme $\mathbb{G}$ over k can be encapsulated into combinatorial data. If $\mathbb{G}$ is symmetric with rank p^{2g}, then there is a *final filtration* $N_1 \subset N_2 \subset \cdots \subset N_{2g}$ of $\mathbb{G}$ as a k-vector space that is stable under the action of V and F^{-1} such that $i = \dim(N_i)$ [Oort 2001, 5.4]. If w is a word in V and F^{-1}, then $wD(\mathbb{G})$ is an object in the filtration; in particular, $N_g = VD(\mathbb{G}) = F^{-1}(0)$.

The *Ekedahl–Oort type* of $\mathbb{G}$, also called the *final type*, is $\nu = [\nu_1, \ldots, \nu_g]$, where $\nu_i = \dim(V(N_i))$. The Ekedahl–Oort type of $\mathbb{G}$ does not depend on the choice of a final filtration. There is a restriction $\nu_i \le \nu_{i+1} \le \nu_i + 1$ on the final type. There are 2^g Ekedahl–Oort types of length g since all sequences satisfying this restriction occur. By [Oort 2001, 9.4, 12.3], there are bijections between (i) Ekedahl–Oort types of length g, (ii) polarized BT_1 group schemes over k of rank p^{2g}, and (iii) principal quasipolarized Dieudonné modules of dimension $2g$ over k.

2d. ***The p-rank and a-number.*** Two invariants of (the p-torsion of) an abelian variety are the p-rank and a-number. The *p-rank* of A is $r = \dim_{\mathbb{F}_p} \mathrm{Hom}(\mu_p, A[p])$, where μ_p is the kernel of Frobenius on $\mathbb{G}_m$. Then p^r is the cardinality of $A[p](k)$. The *a-number* of A is $a = \dim_k \mathrm{Hom}(\alpha_p, A[p])$, where α_p is the kernel of Frobenius on $\mathbb{G}_a$. It is well known that $0 \le f \le g$ and $1 \le a + f \le g$. The p-rank of $A[p]$ equals the dimension of $V^g D(\mathbb{G})$. The a-number of $A[p]$ equals $g - \dim(V^2 D(\mathbb{G}))$ [Li and Oort 1998, 5.2.8]. The p-rank equals $\max\{i \mid \nu_i = i\}$, and the a-number equals $g - \nu_g$.

2e. ***The de Rham cohomology.*** Suppose X is a k-curve of genus g, and recall the definition of the noncommutative ring $\mathbb{E} = k[F, V]$ from Section 2b. By [Oda 1969, Section 5], there is an isomorphism of $\mathbb{E}$-modules between the Dieudonné module of the p-torsion group scheme $J_X[p]$ and the de Rham cohomology group $\mathrm{H}^1_{\mathrm{dR}}(X)$. In particular, $\ker(F) = \mathrm{H}^0(X, \Omega^1) = \mathrm{im}(V)$. Recall that $\dim_k \mathrm{H}^1_{\mathrm{dR}}(X) = 2g$.

In [Oda 1969, Section 5], there is the following description of $\mathrm{H}^1_{\mathrm{dR}}(X)$. Let $\mathcal{U} = \{U_i\}$ be a covering of X by affine open subvarieties, and let $U_{ij} := U_i \cap U_j$ and $U_{ijk} := U_i \cap U_j \cap U_k$. For a sheaf $\mathcal{F}$ on X, let

$$\begin{aligned}
\mathrm{C}^0(\mathcal{U}, \mathcal{F}) &:= \{\kappa = (\kappa_i)_i \mid \kappa_i \in \Gamma(U_i, \mathcal{F})\},\\
\mathrm{C}^1(\mathcal{U}, \mathcal{F}) &:= \{\phi = (\phi_{ij})_{i<j} \mid \phi_{ij} \in \Gamma(U_{ij}, \mathcal{F})\},\\
\mathrm{C}^2(\mathcal{U}, \mathcal{F}) &:= \{\psi = (\psi_{ijk})_{i<j<k} \mid \psi_{ijk} \in \Gamma(U_{ijk}, \mathcal{F})\}.
\end{aligned}$$

For convenience, let $\phi_{ii} := 0$ for any $\phi \in \mathrm{C}^1(\mathcal{U}, \mathcal{F})$. There are coboundary operators $\delta : \mathrm{C}^0(\mathcal{U}, \mathcal{F}) \to \mathrm{C}^1(\mathcal{U}, \mathcal{F})$ defined by $(\delta\kappa)_{i<j} = \kappa_i - \kappa_j$ and $\delta : \mathrm{C}^1(\mathcal{U}, \mathcal{F}) \to \mathrm{C}^2(\mathcal{U}, \mathcal{F})$ by $(\delta\phi)_{i<j<k} = \phi_{ij} - \phi_{ik} + \phi_{jk}$. All other maps are applied to $\mathrm{C}^m(\mathcal{U}, \mathcal{F})$ elementwise, for example, $(F\phi)_i := F\phi_i$. As expected, $\delta^2 = 0$.

The de Rham cocycles are defined by

$$\mathrm{Z}^1_{\mathrm{dR}}(\mathcal{U}) := \{(\phi, \omega) \in \mathrm{C}^1(\mathcal{U}, \mathcal{O}) \times \mathrm{C}^0(\mathcal{U}, \Omega^1) \mid \delta\phi = 0,\ d\phi = \delta\omega\},$$

that is, $\phi_{ij} - \phi_{ik} + \phi_{jk} = 0$ and $d\phi_{ij} = \omega_i - \omega_j$ for all indices $i < j < k$. The de Rham coboundaries are defined by

$$\mathrm{B}^1_{\mathrm{dR}}(\mathcal{U}) := \{(\delta\kappa, d\kappa) \in \mathrm{Z}^1_{\mathrm{dR}}(\mathcal{U}) \mid \kappa \in C^0(\mathcal{U}, \mathcal{O})\}.$$

Finally,

$$\mathrm{H}^1_{\mathrm{dR}}(X) \cong \mathrm{H}^1_{\mathrm{dR}}(\mathscr{U}) := \mathrm{Z}^1_{\mathrm{dR}}(\mathscr{U})/\mathrm{B}^1_{\mathrm{dR}}(\mathscr{U}).$$

There is an injective homomorphism $\lambda : \mathrm{H}^0(X, \Omega^1) \to \mathrm{H}^1_{\mathrm{dR}}(X)$ denoted informally by $\omega \mapsto (0, \omega)$, where the second coordinate is defined by $\omega_i = \omega|_{U_i}$. This map is well-defined since $d(0) = \omega|_{U_i} - \omega|_{U_j} = (\delta\omega)_{i<j}$. It is injective because if $(0, \omega_1) \equiv (0, \omega_2) \bmod \mathrm{B}^1_{\mathrm{dR}}(\mathscr{U})$, then $\omega_1 - \omega_2 = d\kappa$, where $\kappa \in C^0(\mathscr{U}, \mathbb{O})$ is such that $\delta\kappa = 0$; thus, $\kappa \in \mathrm{H}^0(\mathscr{U}, \mathbb{O}) \simeq k$ is a constant function on X, and so $\omega_1 - \omega_2 = 0$.

There is another homomorphism $\gamma : \mathrm{H}^1_{\mathrm{dR}}(X) \to \mathrm{H}^1(X, \mathbb{O})$ sending the cohomology class of (ϕ, ω) to the cohomology class of ϕ. The choice of cocycle (ϕ, ω) does not matter since the coboundary conditions on $\mathrm{H}^1_{\mathrm{dR}}(X)$ and $\mathrm{H}^1(X, \mathbb{O})$ are compatible. The homomorphisms λ and γ fit into a short exact sequence

$$0 \to \mathrm{H}^0(X, \Omega^1) \xrightarrow{\lambda} \mathrm{H}^1_{\mathrm{dR}}(X) \xrightarrow{\gamma} \mathrm{H}^1(X, \mathbb{O}) \to 0 \tag{2-1}$$

of k-vector spaces. In Sections 4d and 4f, we construct a suitable section $\sigma : \mathrm{H}^1(X, \mathbb{O}) \to \mathrm{H}^1_{\mathrm{dR}}(X)$ of γ when X is a hyperelliptic k-curve with $\mathrm{char}(k) = 2$.

2f. ***Frobenius and Verschiebung.*** The Cartier operator $\mathscr{C}$ on the sheaf Ω^1 is defined in [Cartier 1957]. Its three principal properties are that it annihilates exact differentials, preserves logarithmic ones, and induces a p^{-1}-linear map on $\mathrm{H}^0(X, \Omega^1)$. The Cartier operator can be computed as follows. Let $x \in k(X)$ be an element that forms a p-basis of $k(X)$ over $k(X)^p$, that is, an element such that every $z \in k(X)$ can be written as

$$z := z_0^p + z_1^p x + \cdots + z_{p-1}^p x^{p-1}$$

for uniquely determined $z_0, \ldots, z_{p-1} \in k(X)$. Then

$$\mathscr{C}(z\, dx/x) := z_0\, dx/x.$$

The Frobenius operator F on the structure sheaf $\mathbb{O}$ of X induces a p-linear map F on $\mathrm{H}^1(X, \mathbb{O})$. By Serre duality, the $k[F]$-module $\mathrm{H}^1(X, \mathbb{O})$ is dual to the $k[\mathscr{C}]$-module $\mathrm{H}^0(X, \Omega^1)$.

The p-linear operator F and the p^{-1}-linear operator V are defined on $\mathrm{H}^1_{\mathrm{dR}}(X)$ as follows. Let $V(\omega) := \mathscr{C}(\omega)$ and $F(\omega) := 0$ for $\omega \in \mathrm{H}^0(X, \Omega^1)$ and $V(f) := 0$ for $f \in \mathrm{H}^1(X, \mathbb{O})$. Then

$$F(f, \omega) := (F(f), F(\omega)) = (f^p, 0) \quad \text{and} \quad V(f, \omega) := (V(f), V(\omega)) = (0, \mathscr{C}(\omega)).$$

With $\mathbb{E} = k[F, V]$ defined in Section 2c, the short exact sequence (2-1) is an exact sequence of $\mathbb{E}$-modules. However, the section σ of (2-1) constructed in Section 4d is not a splitting of $\mathbb{E}$-modules.

3. Results about regular 1-forms and the a-number

We specialize to the case when the algebraically closed field k has characteristic $p = 2$. Consider a hyperelliptic k-curve X with affine equation $y^2 - y = f(x)$ as described in Section 1. For each branch point $\alpha \in B$, recall the definitions of the ramification invariant $d_\alpha = 2c_\alpha + 1$, the function $x_\alpha = (x - \alpha)^{-1}$, and the polynomial $f_\alpha(x_\alpha)$ appearing in the partial fraction decomposition of $f(x)$. Important facts mentioned in Section 1 are that the genus is determined from the ramification invariants by the formula $2g + 2 = \sum_{\alpha \in B}(d_\alpha + 1)$ and that the 2-rank of J_X equals $r = \#B - 1$.

For $\alpha \in B$, let $P_\alpha := \pi^{-1}(\alpha) \in X(k)$ be the ramification point above α, and define the divisor $D_\infty := \pi^{-1}(\infty)$ on X. Recall that $0 \in B$ and $\infty \notin B$, and let $B_\infty := B \cup \{\infty\}$ and $B' := B - \{0\}$.

3a. ***The space*** $\mathbf{H^0(X, \Omega^1)}$. For an integer j and for $\alpha \in B$, consider the 1-forms

$$\omega_{\alpha,j} := x_\alpha^{j-1}\, dx_\alpha \quad \text{on } X.$$

Note that $\omega_{\alpha,j} = -(x-\alpha)^{-j-1}\, dx$ and if $\alpha \in B'$, then $\omega_{\alpha,0} - \omega_{0,0} = -\alpha\, dx/x(x-\alpha)$.

For completeness, we prove the next lemma, a variation of a special case of [Sullivan 1975, Lemma 1(c)].

Lemma 3.1. *A basis for* $\mathrm{H}^0(X, \Omega^1)$ *is given by the* 1*-forms* $\omega_{\alpha,j}$ *for* $\alpha \in B$ *and* $1 \le j \le c_\alpha$ *and* $\omega_{\alpha,0} - \omega_{0,0}$ *for* $\alpha \in B'$.

Proof. For $\alpha \in B$, we can calculate the following divisors on X: $\mathrm{div}(x_\alpha) = D_\infty - 2P_\alpha$,

$$\mathrm{div}(dx_\alpha) = (d_\alpha - 3)P_\alpha + \sum_{\beta \in B - \{\alpha\}} (d_\beta + 1)P_\beta, \tag{3-1}$$

$$\mathrm{div}(\omega_{\alpha,j}) = 2(c_\alpha - j)P_\alpha + (j-1)D_\infty + \sum_{\beta \in B - \{\alpha\}} (d_\beta + 1)P_\beta. \tag{3-2}$$

Thus, $\omega_{\alpha,j}$ is regular for $1 \le j \le c_\alpha$, and $(\omega_{\alpha,0} - \omega_{0,0})$ is regular for $\alpha \in B'$ since

$$\mathrm{div}(\omega_{\alpha,0} - \omega_{0,0}) = 2c_\alpha P_\alpha + 2c_0 P_0 + \sum_{\beta \in B - \{0,\alpha\}} (d_\beta + 1)P_\beta.$$

This set of regular differentials of X is linearly independent because the corresponding set of divisors is linearly independent over $\mathbb{Z}$. It forms a basis since the set has cardinality $r + \sum_{\alpha \in B} c_\alpha = g$. □

Lemma 3.2. *If* $\alpha \in B$, *then*

$$\mathscr{C}(\omega_{\alpha,j}) = \begin{cases} \omega_{\alpha,j/2} & \text{if } j \text{ is even,} \\ 0 & \text{if } j \text{ is odd.} \end{cases}$$

In particular, $\mathscr{C}(\omega_{\alpha,0} - \omega_{0,0}) = \omega_{\alpha,0} - \omega_{0,0}$ *for all* $\alpha \in B'$.

Proof. Using the properties of the Cartier operator found in Section 2f, one computes when j is even that

$$\mathscr{C}(x_\alpha^{j-1}\,dx_\alpha) = x_\alpha^{j/2}\mathscr{C}(dx_\alpha/x_\alpha) = x_\alpha^{j/2-1}\,dx_\alpha$$

and when j is odd that

$$\mathscr{C}(x_\alpha^{j-1}\,dx_\alpha) = x_\alpha^{(j-1)/2}\mathscr{C}(dx_\alpha) = 0. \qquad \square$$

Let $W'_{\alpha,\mathrm{ss}} := \langle \omega_{\alpha,0} - \omega_{0,0}\rangle$ for $\alpha \in B'$, and let $W'_{\alpha,\mathrm{nil}} := \langle \omega_{\alpha,j} \mid 1 \le j \le c_\alpha\rangle$ for $\alpha \in B$, where $\langle\cdot\rangle$ denotes the k-span. These subspaces are invariant under the Cartier operator by Lemma 3.2.

Lemma 3.3. *The subspaces $W'_{\alpha,\mathrm{ss}}$ and $W'_{\alpha,\mathrm{nil}}$ of $\mathrm{H}^0(X, \Omega^1)$ are stable under the action of Verschiebung for each $\alpha \in B$. There is an isomorphism of V-modules*

$$\mathrm{H}^0(X, \Omega^1) \simeq \bigoplus_{\alpha\in B'} W'_{\alpha,\mathrm{ss}} \oplus \bigoplus_{\alpha\in B} W'_{\alpha,\mathrm{nil}}.$$

Proof. This follows immediately from Lemmas 3.1 and 3.2. $\square$

3b. ***Application: The a-number.***

Proposition 3.4. *Let X be a hyperelliptic k-curve with affine equation $y^2 - y = f(x)$ as described in Notation 1.1. If $\mathrm{div}_\infty(f(x)) = \sum_{\alpha\in B} d_\alpha \alpha$ is the divisor of poles of $f(x)$ on $\mathbb{P}^1$, then the a-number of X is*

$$a_X = \frac{g + 1 - \#\{\alpha \in B \mid d_\alpha \equiv 1 \bmod 4\}}{2}.$$

Proof. The a-number of $\mathbb{G} = J_X[2]$ is $a_X = g - \dim(V^2 D(\mathbb{G}))$ [Li and Oort 1998, 5.2.8]. The action of V on $VD(\mathbb{G})$ is the same as the action of the Cartier operator $\mathscr{C}$ on $\mathrm{H}^0(X, \Omega^1)$. So a_X equals the dimension of the kernel of $\mathscr{C}$ on $\mathrm{H}^0(X, \Omega^1)$. By Lemma 3.2, the kernel of $\mathscr{C}$ on $\mathrm{H}^0(X, \Omega^1)$ is spanned by $\omega_{\alpha,j}$ for $\alpha \in B$ and j odd with $1 \le j \le c_\alpha = (d_\alpha - 1)/2$. Thus, the contribution to the a-number from each $\alpha \in B$ is $\lfloor (d_\alpha + 1)/4\rfloor$. In other words, if $d_\alpha \equiv 1 \bmod 4$, the contribution is $(d_\alpha - 1)/4$, and if $d_\alpha \equiv 3 \bmod 4$, the contribution is $(d_\alpha + 1)/4$. Since $g + 1 = \sum_{\alpha\in B}(d_\alpha + 1)/2$, this yields

$$2a_X = (g + 1) - \#\{\alpha \in B \mid d_\alpha \equiv 1 \bmod 4\}. \qquad \square$$

3c. ***Examples with large p-rank.*** Let A be a principally polarized abelian variety over k with dimension g and p-rank r. If $r = g$, then $A[p] \simeq (\mathbb{Z}/p \oplus \mu_p)^g$ and the a-number is $a = 0$. If $r = g - 1$, then $A[p] \simeq (\mathbb{Z}/p \oplus \mu_p)^{g-1} \oplus E[p]$, where E is a supersingular elliptic curve and the a-number is $a = 1$. So the first case where $A[p]$ and a are not determined by the p-rank is when $r = g - 2$.

Example 3.5. Let $g \geq 2$. There are two possibilities for the p-torsion group scheme of a principally polarized abelian variety over k with dimension g and p-rank $g-2$. When $p=2$, both of these occur as the 2-torsion group scheme $J_X[2]$ of the Jacobian of a hyperelliptic k-curve X of genus g.

Proof. If A is a principally polarized abelian variety over k with dimension g and p-rank $g-2$, then $A[p] \simeq (\mu_p \oplus \mathbb{Z}/p)^{g-2} \oplus \mathbb{G}$, where $\mathbb{G}$ is isomorphic to the p-torsion group scheme of an abelian surface Z with p-rank 0. The abelian surface can be superspecial or merely supersingular. In the superspecial case, $\mathbb{G} = (G_1)^2$, where G_1 denotes the p-torsion group scheme of a supersingular elliptic k-curve; in the merely supersingular case, we denote the group scheme G_2; see [Goren 2002, Example A.3.15; Pries 2008, Example 2.3] for a complete description of G_2.

To prove the second claim, consider the two possibilities for a partition of $2g+2$ into $r+1 = g-1$ even integers: (A) $\{2, 2, \ldots, 2, 4, 4\}$ or (B) $\{2, 2, \ldots, 2, 2, 6\}$. In case (A), consider $f(x) \in k(x)$ with $g-1$ poles such that 0 and 1 are poles of order 3 and the other poles are simple. In case (B), consider $f(x) \in k(x)$ with $g-1$ poles such that 0 is a pole of order 5 and the other poles are simple. The kernel of the Cartier operator on $\mathrm{H}^0(X, \Omega^1)$ is spanned by dx/x^2 and $dx/(x-1)^2$ in case (A) and by dx/x^2 in case (B). Thus, the a-number equals 2 in case (A) and equals 1 in case (B). In both cases, this completely determines the group scheme. Namely, the group scheme $J_X[2]$ is isomorphic to $(\mathbb{Z}/2 \oplus \mu_2)^{g-2} \oplus (G_1)^2$ in case (A) and to $(\mathbb{Z}/2 \oplus \mu_2)^{g-2} \oplus G_2$ in case (B). □

For $g \geq 3$ and $r \leq g-3$, the action of V on $\mathrm{H}^0(X, \Omega^1)$ (and, in particular, the value of the a-number) is not sufficient to determine the isomorphism class of the group scheme $J_X[2]$. To determine this group scheme, in the next section we study the $\mathbb{E}$-module structure of $\mathrm{H}^1_{\mathrm{dR}}(X)$.

4. Results on the de Rham cohomology

4a. ***An open covering.*** Let $V' = \mathbb{P}^1 - B_\infty$ and $U' = \pi^{-1}(V') = X - \pi^{-1}(B_\infty)$. For $\alpha \in B_\infty$, let $V_\alpha = V' \cup \{\alpha\}$ and $U_\alpha := U' \cup \{\pi^{-1}(\alpha)\}$. Then the collection $\mathcal{U} := \{U_\alpha \mid \alpha \in B_\infty\}$ is a cover of X by open affine subvarieties. By construction, if $\alpha, \beta \in B_\infty$ are distinct, then $V_{\alpha\beta} := V_\alpha \cap V_\beta = V'$ and $U_{\alpha\beta} := U_\alpha \cap U_\beta = U'$. In particular, the subvarieties $U_{\alpha\beta}$ do not depend on the choice of α and β.

For a sheaf $\mathcal{F}$, let $\mathrm{Z}^1(\mathcal{U}, \mathcal{F})$ and $\mathrm{B}^1(\mathcal{U}, \mathcal{F})$ denote the closed cocycles and coboundaries of $\mathcal{F}$ with respect to $\mathcal{U}$. Recall the definition of the noncommutative ring $\mathbb{E} = k[F, V]$ and the notation about $\mathrm{H}^1_{\mathrm{dR}}(X)$ from Section 2e. In this section, we compute $\mathrm{H}^1(X, \mathcal{O}) \simeq \mathrm{H}^1(\mathcal{U}, \mathcal{O})$ and $\mathrm{H}^1_{\mathrm{dR}}(X) \simeq \mathrm{H}^1_{\mathrm{dR}}(\mathcal{U})$ with respect to the open covering $\mathcal{U}$ of X.

4b. ***Defining components.*** Given a sheaf $\mathcal{F}$ and a cocycle $\phi \in \mathrm{Z}^1(\mathcal{U}, \mathcal{F})$, consider its components $\phi_{\alpha\infty} \in \Gamma(U', \mathcal{F})$ for $\alpha \in B$. We call $\{\phi_{\alpha\infty} \mid \alpha \in B\}$ the set of *defining components* of ϕ. The reason is that the remaining components of ϕ are determined by the coboundary condition $\phi_{\alpha\beta} = \phi_{\alpha\infty} - \phi_{\beta\infty}$. A collection of sections $\{\phi_{\alpha\infty} \in \Gamma(U', \mathcal{F}) \mid \alpha \in B\}$ determines a unique closed cocycle $\phi \in \mathrm{Z}^1(\mathcal{U}, \mathcal{F})$. Thus,

$$\mathrm{Z}^1(\mathcal{U}, \mathcal{F}) \cong \bigoplus_{\alpha \in B} \Gamma(U', \mathcal{F}). \tag{4-1}$$

For $\beta \in B$, consider the natural k-linear map

$$\varphi_\beta : \Gamma(U', \mathcal{O}) \to \mathrm{Z}^1(\mathcal{U}, \mathcal{O})$$

whose defining components for $\alpha \in B$ are

$$(\varphi_\beta(h))_{\alpha\infty} := \begin{cases} h & \text{if } \alpha = \beta, \\ 0 & \text{otherwise.} \end{cases}$$

Also, consider the k-linear map $\varphi_\infty : \Gamma(U', \mathcal{O}) \to \mathrm{Z}^1(\mathcal{U}, \mathcal{O})$ defined by

$$(\varphi_\infty(h))_{\alpha\infty} := -h \quad \text{for all } \alpha \in B.$$

Observe that if $h \in \Gamma(U', \mathcal{O})$, then

$$\sum_{\beta \in B_\infty} \varphi_\beta(h) = 0. \tag{4-2}$$

For $\beta \in B_\infty$, consider the natural k-linear map

$$\psi_\beta : \Gamma(U_\beta, \mathcal{O}) \to \mathrm{C}^0(\mathcal{U}, \mathcal{O})$$

given for $\alpha \in B_\infty$ by

$$(\psi_\beta(h))_\alpha := \begin{cases} h & \text{if } \alpha = \beta, \\ 0 & \text{otherwise.} \end{cases} \tag{4-3}$$

It is straightforward to verify the next lemma.

Lemma 4.1. *Suppose $\beta \in B_\infty$ and $h \in \Gamma(U_\beta, \mathcal{O})$ (that is, h is regular at P_β if $\beta \neq \infty$ and h is regular at the two points in the support of D_∞ if $\beta = \infty$). Then $\varphi_\beta(h|_{U'}) = \delta\psi_\beta(h)$ is a coboundary.*

4c. ***The space*** $\mathbf{H^1(X, \mathcal{O})}$**.** In this section, we find an F-module decomposition of $\mathrm{H}^1(X, \mathcal{O}) \simeq \mathrm{H}^1(\mathcal{U}, \mathcal{O})$. The results could be deduced from Section 3a using the duality between $\mathrm{H}^1(X, \mathcal{O})$ and $\mathrm{H}^0(X, \Omega^1)$. Instead, we take a direct approach because an explicit description of $\mathrm{H}^1(X, \mathcal{O})$ is helpful for studying $\mathrm{H}^1_{\mathrm{dR}}(X)$ in Section 4f.

Lemma 4.2. *Write $D_\infty = P_{\infty,1} + P_{\infty,2}$. Then $\mathrm{ord}_{P_{\infty,1}}(y) = 0$ and $\mathrm{ord}_{P_{\infty,2}}(y) = s$ for some $s \geq 0$ (possibly after reordering). For $\alpha \in B$ and $j \in \mathbb{Z}$, the divisor of poles* *on X of the function $yx_\alpha^{-j} = y(x-\alpha)^j$ satisfies*

$$\operatorname{div}_\infty(y(x-\alpha)^j) \\ = \max(d_\alpha - 2j, 0)P_\alpha + \max(j,0)P_{\infty,1} + \max(j-s,0)P_{\infty,2} + \sum_{\beta\in B-\{\alpha\}} d_\beta P_\beta.$$

Proof. Recall that $\operatorname{div}_\infty(y) = \sum_{\beta\in B} d_\beta P_\beta$. Note that $\operatorname{ord}_{P_{\infty,i}}(y) \geq 0$ for $i = 1, 2$ since $\infty \notin B$. If $\operatorname{ord}_{P_{\infty,2}}(y) > 0$, that is, if y has a zero at $P_{\infty,2}$, then the value of y is one at the Galois conjugate $P_{\infty,1}$ of $P_{\infty,2}$. Thus, y cannot have a zero at both points in the support of D_∞. The second claim follows from the additional fact that $\operatorname{div}(x-\alpha) = 2P_\alpha - D_\infty$ for $\alpha \in B$. □

Lemma 4.2 implies that $y(x-\alpha)^j \in \Gamma(U', \mathbb{O})$ for all $\alpha \in B$ and $j \in \mathbb{Z}$.

Lemma 4.3. *With notation as above,*

(i) $Z^1(\mathcal{U}, \mathbb{O}) = \langle \varphi_\beta((x-\alpha)^j), \varphi_\beta(y(x-\alpha)^j) \mid \alpha, \beta \in B,\ j \in \mathbb{Z}\rangle$, *and*

(ii) *if $\alpha \in B$, then $\langle \varphi_\alpha(y(x-\beta)^j) \mid j \geq 0\rangle = \langle \varphi_\alpha(y(x-\alpha)^j) \mid j \geq 0\rangle$ as subspaces of $Z^1(\mathcal{U}, \mathbb{O})$ for each $\beta \in B$.*

Proof. (i) This is immediate from Equation (4-1) because

$$Z^1(\mathcal{U}, \mathbb{O}) = \bigoplus_{\beta\in B} \langle \varphi_\beta(h) \mid h \in \Gamma(U', \mathbb{O})\rangle.$$

(ii) Both are equal to the subspace $\{\varphi_\alpha(yh(x)) \mid h(x) \in k[x]\}$. □

Lemma 4.4. *Let $\alpha \in B \subset k$ and $j \in \mathbb{Z}$. Then*

(i) $\varphi_\beta((x-\alpha)^j) \in B^1(\mathcal{U}, \mathbb{O})$ *for all $\beta \in B_\infty$,*

(ii) $\varphi_\alpha(y(x-\alpha)^j) \in B^1(\mathcal{U}, \mathbb{O})$ *if $j > c_\alpha$, and*

(iii) $\varphi_\infty(y(x-\alpha)^j) \in B^1(\mathcal{U}, \mathbb{O})$ *if $j \leq 0$.*

Proof. (i) Suppose that $\beta \in B$. If $\beta \neq \alpha$ or if $j \geq 0$, then $(x-\alpha)^j$ is regular at P_β, and so $\varphi_\beta((x-\alpha)^j) \in B^1(\mathcal{U}, \mathbb{O})$ by Lemma 4.1. For $j \geq 0$, it follows from this and Equation (4-2) that the cocycle $\varphi_\infty((x-\alpha)^j) = -\sum_{\beta\in B} \varphi_\beta((x-\alpha)^j)$ is a coboundary. If $j < 0$, then $\varphi_\infty((x-\alpha)^j) \in B^1(\mathcal{U}, \mathbb{O})$ by Lemma 4.1.

Finally, $(x-\alpha)^j \in \Gamma(U_\gamma, \mathbb{O})$ for all $\gamma \in B_\infty - \{\alpha\}$ if $\beta = \alpha \neq \infty$ and $j < 0$. By Equation (4-2),

$$\varphi_\alpha((x-\alpha)^j) = -\sum_{\gamma\in B_\infty-\{\alpha\}} \varphi_\gamma((x-\alpha)^j) = -\sum_{\gamma\in B_\infty-\{\alpha\}} \delta\psi_\gamma((x-\alpha)^j), \tag{4-4}$$

which is a coboundary.

(ii) If $j > c_\alpha$, then $y(x-\alpha)^j \in \Gamma(U_\alpha, \mathbb{O})$ and $\varphi_\alpha(y(x-\alpha)^j) = \delta\psi_\alpha(y(x-\alpha)^j)$.

(iii) If $j \leq 0$, then $y(x-\alpha)^j \in \Gamma(U_\infty, \mathbb{O})$ and $\varphi_\infty(y(x-\alpha)^j) = \delta\psi_\infty(y(x-\alpha)^j)$. □

Consider the cocycles $\phi_{\alpha,j} \in \mathrm{Z}^1(\mathcal{U}, \mathcal{O})$ for $\alpha \in B$ and $j \in \mathbb{Z}$ defined by

$$\phi_{\alpha,j} := \varphi_\alpha(y(x-\alpha)^j).$$

Given $\phi \in \mathrm{Z}^1(\mathcal{U}, \mathcal{O})$, $\tilde{\phi}$ denotes the cohomology class of ϕ in $\mathrm{H}^1(\mathcal{U}, \mathcal{O})$. For $\alpha \in B_\infty$, define the map

$$\tilde{\varphi}_\alpha : \Gamma(U', \mathcal{O}) \to \mathrm{H}^1(\mathcal{U}, \mathcal{O}), \quad f \mapsto \varphi_\alpha(f) \bmod \mathrm{B}^1(\mathcal{U}, \mathcal{O}).$$

We now study $\mathrm{H}^1(\mathcal{U}, \mathcal{O})$; the following lemma is a variant of a special case of [Madden 1978, Lemma 6]:

Lemma 4.5. *A basis for* $\mathrm{H}^1(\mathcal{U}, \mathcal{O})$ *is given by the cohomology classes* $\tilde{\phi}_{\alpha,j}$ *for* $\alpha \in B$ *and* $1 \le j \le c_\alpha$ *and* $\tilde{\phi}_{\alpha,0}$ *for* $\alpha \in B'$.

Proof. The set of cohomology classes $S = \{\tilde{\phi}_{\alpha,j} \mid \alpha \in B,\ 1 \le j \le c_\alpha\} \cup \{\tilde{\phi}_{\alpha,0} \mid \alpha \in B'\}$ has cardinality $r + \sum_{\alpha \in B} c_\alpha = g$. By Lemmas 4.3(i) and 4.4(i), it suffices to show that $\varphi_\beta(y(x-\alpha)^j)$ is in the span of S for $\alpha, \beta \in B$ and $j \in \mathbb{Z}$. By Lemmas 4.3(ii) and 4.4(ii), it suffices to show that the span of S contains $\tilde{\phi}_{0,0}$ and $\tilde{\varphi}_\beta(y(x-\alpha)^{-j})$ for $\alpha, \beta \in B$ and $j > 0$.

The cocycle $\varphi_\infty(y)$ is a coboundary by Lemmas 4.1 and 4.2. Using this and Equation (4-2), one computes in $\mathrm{H}^1(\mathcal{U}, \mathcal{O})$ that

$$\tilde{\phi}_{0,0} = \tilde{\varphi}_0(y) + \tilde{\varphi}_\infty(y) = -\sum_{\beta \in B'} \tilde{\varphi}_\beta(y) = -\sum_{\beta \in B'} \tilde{\phi}_{\beta,0},$$

which is in the span of S.

Now consider $\tilde{\varphi}_\beta(y(x-\alpha)^{-j})$ for $\alpha, \beta \in B$ and $j > 0$. If $0 = r := \#B - 1$, then this cocycle is a coboundary by Equation (4-2) and Lemma 4.4(iii).

Let $r > 0$; first suppose that $\alpha \ne \beta$. Consider the rational function $h = (x-\alpha)^{-j}$, which has no pole at β. Write $h = T + E$, where T is the degree-c_β Taylor polynomial of h at β. Then $\varphi_\beta(yh) = \varphi(yT) + \varphi(yE)$. Note that the function E on $\mathbb{P}^1$ has a zero at β of order at least $c_\beta + 1$. Recall that $\mathrm{ord}_{P_\beta}(x-\beta) = 2$, and observe that $\mathrm{ord}_{P_\beta}(E) \ge 2(c_\beta + 1) = d_\beta + 1$ on X. Since $\mathrm{ord}_{P_\beta}(y) = -d_\beta$, it follows that $yE \in \Gamma(U_\beta, \mathcal{O})$ and thus $\varphi_\beta(yE) \in \mathrm{B}^1(\mathcal{U}, \mathcal{O})$ by Lemma 4.1. The term $\varphi_\beta(yT)$ is, by construction, a linear combination of $\varphi_\beta(y(x-\beta)^j) = \phi_{\beta,j}$ for $0 \le j \le c_\beta$. Thus, $\tilde{\varphi}_\beta(yh)$ is in the span of S, which completes the case when $\alpha \ne \beta$.

If $\alpha = \beta$ and $j > 0$, one can reduce to the previous case by adding the coboundary $\varphi_\infty(y(x-\alpha)^{-j})$ to $\varphi_\alpha(y(x-\alpha)^{-j})$ and using Equation (4-2) to see that

$$\tilde{\varphi}_\alpha(y(x-\alpha)^{-j}) = -\sum_{\gamma \in B - \{\alpha\}} \tilde{\varphi}_\gamma(y(x-\alpha)^{-j}). \qquad \square$$

The next lemma is important for describing the F-module structure of $\mathrm{H}^1(\mathcal{U}, \mathcal{O})$.

Lemma 4.6. *If* $\alpha \in B$ *and* $j \geq 0$, *then*

$$F\tilde{\phi}_{\alpha,j} = \begin{cases} \tilde{\phi}_{\alpha,2j} & \text{if } 2j \leq c_\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Since $(F\phi_{\alpha,j})_{\beta\gamma} = (\phi_{\alpha,j})^2_{\beta\gamma}$, one computes that

$$\begin{aligned}(y(x-\alpha)^j)^2 &= (y+f(x))(x-\alpha)^{2j} \\ &= y(x-\alpha)^{2j} + f(x)(x-\alpha)^{2j}.\end{aligned}$$

The statement follows from the definition of $\tilde{\phi}_{\alpha,j}$ and Lemma 4.4(i). ☐

Now define

$$\begin{aligned} W''_{\alpha,\mathrm{ss}} &:= \langle \tilde{\phi}_{\alpha,0} \rangle && \text{for } \alpha \in B' \text{ and} \\ W''_{\alpha,\mathrm{nil}} &:= \langle \tilde{\phi}_{\alpha,j} \mid 1 \leq j \leq c_\alpha \rangle && \text{for } \alpha \in B. \end{aligned}$$

Lemma 4.7. *The subspaces* $W''_{\alpha,\mathrm{ss}}$ *and* $W''_{\alpha,\mathrm{nil}}$ *of* $\mathrm{H}^1(\mathcal{U}, \mathbb{O})$ *are stable under the action of Frobenius for each* $\alpha \in B$. *There is an isomorphism of* F*-modules*

$$\mathrm{H}^1(\mathcal{U}, \mathbb{O}) \simeq \bigoplus_{\alpha \in B'} W''_{\alpha,\mathrm{ss}} \oplus \bigoplus_{\alpha \in B} W''_{\alpha,\mathrm{nil}}.$$

Proof. This follows immediately from Lemmas 4.5 and 4.6. ☐

4d. ***Auxiliary map.*** The next goal is to define a section $\sigma : \mathrm{H}^1(X, \mathbb{O}) \to \mathrm{H}^1_{\mathrm{dR}}(X)$. To do this, the first step will be to define a homomorphism $\rho : \mathrm{Z}^1(\mathcal{U}, \mathbb{O}) \to \mathrm{C}^0(\mathcal{U}, \Omega^1)$ by defining its components $\rho_\alpha : \mathrm{Z}^1(\mathcal{U}, \mathbb{O}) \to \Gamma(U_\beta, \Omega^1)$ for $\alpha, \beta \in B$. Given $\phi \in \mathrm{Z}^1(\mathcal{U}, \mathbb{O})$ and $\alpha \in B$, the idea is to separate $d\phi$ into two parts: The first part will be regular at P_α and thus belong to $\Gamma(U_\alpha, \Omega^1)$, and the second part will be regular away from P_α and hence belong to $\Gamma(U_\beta, \Omega^1)$ for every $\beta \neq \alpha$.

Notation 4.8. Define the *truncation* operator $\Theta_{\geq i} : k[x, x^{-1}] \to k[x, x^{-1}]$ by

$$\Theta_{\geq i}\Big(\sum_j a_j x^j\Big) := \sum_{j \geq i} a_j x^j.$$

Operators $\Theta_{>i}, \Theta_{\leq i}, \Theta_{<i} : k[x, x^{-1}] \to k[x, x^{-1}]$ can be defined analogously. These operators can also be defined on $k[x_\alpha, x_\alpha^{-1}]$. To clarify some ambiguity in notation, if $m(x_\alpha) \in k[x_\alpha, x_\alpha^{-1}]$, then let $\Theta_{\geq i}(m(x_\alpha))$ denote $\Theta_{\geq i}(m(x))|_{x=x_\alpha}$.

Recall that $x_\alpha := (x-\alpha)^{-1}$, and so $\phi_{\alpha,j} = \varphi_\alpha(y x_\alpha^{-j})$. Then

$$d(y x_\alpha^{-j}) = -j x_\alpha^{-j-1} y \, dx_\alpha + x_\alpha^{-j} \, dy. \tag{4-5}$$

Using partial fractions and the fact that $dy = -d(f(x))$, one sees that

$$dy = -\sum_{\beta \in B} f'_\beta(x_\beta)\, dx_\beta. \tag{4-6}$$

In light of these facts, consider the following definition:

Notation 4.9. For $\alpha \in B$ and $j \geq 0$, define

$$R_{\alpha,j} := \Theta_{\geq 0}(x_\alpha^{-j} f'_\alpha(x_\alpha))\, dx_\alpha \quad \text{and} \quad S_{\alpha,j} := d(y x_\alpha^{-j}) + R_{\alpha,j}.$$

Remark 4.10. Let $a_{\alpha,i} \in k$ be the coefficients of the (odd-power) monomials of the polynomials $f_\alpha(x_\alpha)$ defined in the partial fraction decomposition (1-1):

$$f_\alpha(x_\alpha) = \sum_{i=0}^{c_\alpha} a_{\alpha,i} x_\alpha^{2i+1}.$$

Then

$$R_{\alpha,j} = \sum_{j/2 \leq i \leq c_\alpha} a_{\alpha,i} x_\alpha^{2i-j}\, dx_\alpha = \sum_{j/2 \leq i \leq c_\alpha} a_{\alpha,i}\, \omega_{\alpha,2i-j+1}.$$

Lemma 4.11. *Let $\alpha \in B$ and $j \geq 0$.*

(1) *The differential form $R_{\alpha,j}$ is regular away from P_α, that is, $R_{\alpha,j} \in \Gamma(U_\beta, \Omega^1)$ for all $\beta \in B_\infty - \{\alpha\}$*

(2) *The differential form $S_{\alpha,j}$ is regular at P_α for $0 \leq j \leq c_\alpha$, that is, $S_{\alpha,j} \in \Gamma(U_\alpha, \Omega^1)$.*

Proof. (1) This follows from Remark 4.10 and Equation (3-2).

(2) By Notations 4.8 and 4.9 and Equations (4-5) and (4-6), one sees that

$$S_{\alpha,j} = d(y x_\alpha^{-j}) + \Theta_{\geq 0}(x_\alpha^{-j} f'_\alpha(x_\alpha))\, dx_\alpha \tag{4-7}$$

$$= -j x_\alpha^{-j-1} y\, dx_\alpha - \Theta_{<0}(x_\alpha^{-j} f'_\alpha(x_\alpha))\, dx_\alpha - \sum_{\beta \in B - \{\alpha\}} x_\alpha^{-j} f'_\beta(x_\beta)\, dx_\beta. \tag{4-8}$$

In the first part of Equation (4-8), note that the order of vanishing of $x_\alpha^{-j-1} y\, dx_\alpha$ at P_α is $2d_\alpha - 1 + 2j$ by Lemma 4.2 and Equation (3-1), so this term is regular at P_α.

In the second part of Equation (4-8), note that $\Theta_{<0}(x_\alpha^{-j} f'_\alpha(x_\alpha))$ is contained in $x_\alpha^{-1} k[x_\alpha^{-1}]$. Thus, $\Theta_{<0}(x_\alpha^{-j} f'_\alpha(x_\alpha))$ has a zero of order at least 2 at P_α. As seen in the proof of Lemma 3.1, dx_α has a zero of order $d_\alpha - 3$ at P_α. Thus, $\Theta_{<0}(x_\alpha^{-j} f'_\alpha(x_\alpha))\, dx_\alpha$ is regular at P_α.

The last part of Equation (4-8) is regular at P_α since x_α^{-1} and $f'_\beta(x_\beta)\, dx_\beta$ are regular at P_α. □

4e. *Definition of ρ.* We define a k-linear morphism

$$\rho : \mathrm{Z}^1(\mathcal{U}, \mathcal{O}) \to \mathrm{C}^0(\mathcal{U}, \Omega^1).$$

4e.1. *Definition of ρ on* $\mathrm{B}^1(\mathcal{U},\mathcal{O})$. If $\phi \in \mathrm{B}^1(\mathcal{U},\mathcal{O})$, then for some $\kappa \in \mathrm{C}^0(\mathcal{U},\mathcal{O})$, $\phi = \delta\kappa$. Define

$$\rho(\phi) := d\kappa$$

with differentiation performed component-wise. This map is well-defined since if κ is regular at $P \in X(k)$, then so is $d\kappa$. Moreover, if κ' is another element such that $\phi = \delta\kappa'$, then $\delta(\kappa - \kappa') = 0$, and therefore, $\kappa - \kappa' \in \mathrm{H}^0(\mathcal{U},\mathcal{O})$ is constant and annihilated by d. Let $\rho_\beta(\phi)$ denote $(\rho(\phi))_\beta$.

It follows from the definition that $\mathscr{C}(\rho(\mathrm{B}^1(\mathcal{U},\mathcal{O}))) = 0$ since the Cartier operator annihilates all exact differential forms. Explicitly, the map ρ is computed as follows:

Lemma 4.12. (i) *If $\alpha \in B_\infty$ and $h \in \Gamma(U_\alpha,\mathcal{O})$, then $\rho\varphi_\alpha(h|_{U'}) = d\psi_\alpha(h)$.*

(ii) *If $\alpha \in B$ and $j \leq 0$, then*

$$\rho\varphi_\alpha((x-\alpha)^j) = -\sum_{\gamma \in B_\infty - \{\alpha\}} d\psi_\gamma((x-\alpha)^j).$$

Proof. (i) This is immediate from the definition of the map ρ and Lemma 4.1.

(ii) This follows from part (i), Equation (4-4), and the definition of ρ. □

Example 4.13. We find the value of ρ on the 1-coboundary $\varphi_\alpha(f(x)x_\alpha^{-j})$ if $\alpha \in B$ and $j \geq 0$. Let

$$r_{\alpha,j} := \Theta_{>0}(x_\alpha^{-j} f_\alpha(x_\alpha)) \quad \text{and} \quad s_{\alpha,j} := \Theta_{\leq 0}(x_\alpha^{-j} f_\alpha(x_\alpha)) + \sum_{\beta \neq \alpha} x_\alpha^{-j} f_\beta(x_\beta).$$

Then

$$f(x)x_\alpha^{-j} = r_{\alpha,j} + s_{\alpha,j},$$

and $r_{\alpha,j}$ has a pole at P_α but is regular everywhere else while $s_{\alpha,j}$ is regular at P_α, so

$$\varphi_\alpha(f(x)x_\alpha^{-j}) = \delta\psi_\alpha(s_{\alpha,j}) - \sum_{\beta \in B_\infty - \{\alpha\}} \delta\psi_\beta(r_{\alpha,j}).$$

Therefore, for $\beta \neq \alpha$, by Lemma 4.12, $\rho_\beta\varphi_\alpha(f(x)x_\alpha^{-j}) = -d(r_{\alpha,j})$. Since $f_\alpha(x_\alpha) \in x_\alpha k[x_\alpha^2]$, this simplifies to

$$\rho_\beta\varphi_\alpha(f(x)x_\alpha^{-j}) = \begin{cases} -R_{\alpha,j} & \text{if } j \text{ is even,} \\ 0 & \text{if } j \text{ is odd.} \end{cases} \tag{4-9}$$

Similarly,

$$\rho_\alpha\varphi_\alpha(f(x)x_\alpha^{-j}) = \begin{cases} -S_{\alpha,j} & \text{if } j \text{ is even,} \\ d(f(x)x_\alpha^{-j}) & \text{if } j \text{ is odd.} \end{cases} \tag{4-10}$$

4e.2. *Definition of ρ_β on $\mathrm{Z}^1(\mathcal{U}, \mathcal{O})$.* By Lemma 4.5, $\mathrm{Z}^1(\mathcal{U}, \mathcal{O})$ is generated by $\mathrm{B}^1(\mathcal{U}, \mathcal{O})$ and $\phi_{\alpha,j}$ for $\alpha \in B$ and $0 \le j \le c_\alpha$. For $\alpha, \beta \in B$, define

$$\rho_\beta(\phi_{\alpha,j}) := \begin{cases} R_{\alpha,j} & \text{if } \beta \neq \alpha, \\ S_{\alpha,j} & \text{if } \beta = \alpha, \end{cases}$$

and extend ρ_β to $\mathrm{Z}^1(\mathcal{U}, \mathcal{O})$ linearly. For all $\beta \in B - \{\alpha\}$, note that

$$\rho_\alpha(\phi_{\alpha,j}) = d(yx_\alpha^{-j}) + \rho_\beta(\phi_{\alpha,j}).$$

Lemma 4.14. *There is a well-defined map $\rho : \mathrm{Z}^1(\mathcal{U}, \mathcal{O}) \to \mathrm{C}^0(\mathcal{U}, \Omega^1)$ given by*

$$\rho := \bigoplus_{\beta \in B_\infty} \rho_\beta.$$

Proof. By Section 4e.1 and Lemma 4.11, $\rho_\beta(\mathrm{Z}^1(\mathcal{U}, \mathcal{O})) \subset \Gamma(U_\beta, \Omega^1)$ if $\beta \in B_\infty$. □

Here is an example of a computation of the map ρ.

Lemma 4.15. *Let $\alpha \in B$ and $j \ge 0$. For each $\beta \in B$, in $\Gamma(U_\beta, \Omega^1)$,*

$$\rho_\beta\varphi_\alpha(y^2x_\alpha^{-2j}) = \begin{cases} 0 & \text{if } 0 \le 2j \le c_\alpha, \\ -R_{\alpha,2j} & \text{if } 2j > c_\alpha. \end{cases}$$

In particular, $\rho\varphi_\alpha(y^2x_\alpha^{-2j})$ lies in the subspace $W'_{\alpha,\mathrm{nil}}$ of $\mathrm{H}^0(\mathcal{U}, \Omega^1)$.

Proof. We have $y^2x_\alpha^{-2j} = yx_\alpha^{-2j} + f(x)x_\alpha^{-2j}$, and therefore

$$\varphi_\alpha(y^2x_\alpha^{-2j}) = \phi_{\alpha,2j} + \varphi_\alpha(f(x)x_\alpha^{-2j}).$$

Suppose $0 \le 2j \le c_\alpha$. If $\beta \neq \alpha$, then $\rho_\beta(\phi_{\alpha,2j}) = R_{\alpha,2j} = -\rho_\beta(\varphi_\alpha(f(x)x_\alpha^{-2j}))$ by Equation (4-9). By Equation (4-10), $\rho_\alpha(\phi_{\alpha,2j}) = S_{\alpha,2j} = -\rho_\alpha(\varphi_\alpha(f(x)x_\alpha^{-2j}))$. Thus, $\rho(\phi_{\alpha,2j}) + \rho(\varphi_\alpha(f(x)x_\alpha^{-2j})) = 0$.

Now, suppose that $2j > c_\alpha$. Then yx_α^{-2j} is regular at P_α, and therefore, $\phi_{\alpha,2j}$ is a coboundary with $\rho(\phi_{\alpha,2j}) = d\varphi_\alpha(yx_\alpha^{2j})$. Therefore, for $\beta \neq \alpha$,

$$\rho_\beta(\phi_{\alpha,2j}) + \rho_\beta(\varphi_\alpha(f(x)x_\alpha^{-2j})) = -R_{\alpha,2j},$$

and

$$\rho_\alpha(\phi_{\alpha,2j}) + \rho_\alpha(\varphi_\alpha(f(x)x_\alpha^{-2j})) = d(yx_\alpha^{-2j}) + d(f(x)x_\alpha^{-2j}) - R_{\alpha,2j} = -R_{\alpha,2j}.$$

By Remark 4.10, $R_{\alpha,2j} \in \langle \omega_{\alpha,2i-2j+1} \mid j \le i \le c_\alpha \rangle$. If $2j > c_\alpha$ and $j \le i \le c_\alpha$, then $1 \le 2i - 2j + 1 \le c_\alpha$, and so $R_{\alpha,2j} \in W'_{\alpha,\mathrm{nil}}$. Finally, since $\rho_\beta\varphi_\alpha(y^2x_\alpha^{-2j})$ is independent of the choice of $\beta \in B_\infty$, we have $\rho\varphi_\alpha(y^2x_\alpha^{-2j})$ lies in the kernel $\mathrm{H}^0(\mathcal{U}, \Omega^1)$ of the coboundary map $\delta : C^0(\mathcal{U}, \Omega^1) \to C^1(\mathcal{U}, \Omega^1)$. □

Lemma 4.16. (i) *If $\phi \in \mathrm{Z}^1(\mathcal{U}, \mathcal{O})$, then $\delta\rho(\phi) = d\phi$.*

(ii) *In particular, $\mathscr{C}(\rho_\alpha(\phi)) = \mathscr{C}(\rho_\beta(\phi))$ for all $\alpha, \beta \in B_\infty$.*

(iii) *For all* $\alpha \in B$ *and* $\beta \in B_\infty$, *we have* $\mathscr{C}(\rho_\beta(\phi_{\alpha,j})) = \mathscr{C}(R_{\alpha,j})$.

Proof. (i) The definition of ρ_β implies that $\rho_\alpha(\phi) - \rho_\beta(\phi) = d(\phi)_{\alpha\beta}$ for all $\alpha, \beta \in B_\infty$.

(ii) This follows from part (i) since the Cartier operator annihilates exact differential forms.

(iii) This follows from part (ii) and the definition of ρ_β. □

Remark 4.17. With $a_{\alpha,i}$ defined as in Remark 4.10, one can explicitly compute

$$\mathscr{C}(R_{\alpha,j}) = \begin{cases} \sum_{i=(j+1)/2}^{c_\alpha} \sqrt{a_{\alpha,i}}\,\omega_{\alpha,i-(j-1)/2} & \text{if } j \text{ is odd,} \\ 0 & \text{if } j \text{ is even.} \end{cases}$$

In particular, $\mathscr{C}(R_{\alpha,j}) \in W'_{\alpha,\mathrm{nil}}$.

4f. *The $\mathbb{E}$-module structure of the de Rham cohomology.* Consider the exact sequence of $\mathbb{E}$-modules

$$0 \to \mathrm{H}^0(X, \Omega^1) \xrightarrow{\lambda} \mathrm{H}^1_{\mathrm{dR}}(X) \xrightarrow{\gamma} \mathrm{H}^1(X, \mathbb{O}) \to 0,$$

where $\mathbb{E} = k[F, V]$ is the noncommutative ring defined in Section 2a. Consider the k-linear function

$$\sigma : \mathrm{H}^1(X, \mathbb{O}) \to \mathrm{H}^1_{\mathrm{dR}}(X)$$

defined by $\sigma(\phi) = (\phi, \rho(\phi))$ for $\phi \in \mathrm{Z}^1(\mathscr{U}, \mathbb{O})$.

Lemma 4.18. *The function* σ *is a section of* $\gamma : \mathrm{H}^1_{\mathrm{dR}}(X) \to \mathrm{H}^1(X, \mathbb{O})$.

Proof. The function σ is well-defined because $\sigma(\mathrm{B}^1(\mathscr{U}, \mathbb{O})) \subset \mathrm{B}^1_{\mathrm{dR}}(\mathscr{U})$ by the definition of ρ_β on $\mathrm{B}^1(\mathscr{U}, \mathbb{O})$. It is clearly a section of γ. □

Note that σ is not a splitting of $\mathbb{E}$-modules.
For $\alpha \in B$, let $\lambda_{\alpha,j} := \lambda(\omega_{\alpha,j})$ and $\sigma_{\alpha,j} := \sigma(\tilde{\phi}_{\alpha,j})$.

Proposition 4.19. *For* $0 \le j \le c_\alpha$, *the action of* F *and* V *on* $\mathrm{H}^1_{\mathrm{dR}}(X)$ *is given by*

(i) $F\lambda_{\alpha,j} = 0$,

(ii) $V\lambda_{\alpha,j} = \begin{cases} \lambda_{\alpha,j/2} & \textit{if } j \textit{ is even,} \\ 0 & \textit{if } j \textit{ is odd,} \end{cases}$

(iii) $F\sigma_{\alpha,j} = \begin{cases} \sigma_{\alpha,2j} & \textit{if } j \le c_\alpha/2, \\ \lambda(R_{\alpha,2j}) & \textit{if } j > c_\alpha/2, \end{cases}$

(iv) $V\sigma_{\alpha,j} = \begin{cases} \lambda(\mathscr{C}(R_{\alpha,j})) & \textit{if } j \textit{ is odd,} \\ 0 & \textit{if } j \textit{ is even.} \end{cases}$

Proof. (i) This follows from Section 2f.

(ii) This follows from Lemma 3.2 after applying λ.

(iii) In $\mathrm{Z}^1_{\mathrm{dR}}(\mathscr{U})$,

$$\begin{aligned}F(\sigma_{\alpha,j}) &= (F\phi_{\alpha,j}, 0)\\ &= \big(\varphi_\alpha(y^2x_\alpha^{-2j}), \rho\varphi_\alpha(y^2x_\alpha^{-2j})\big) - (0, \rho\varphi_\alpha(y^2x_\alpha^{-2j}))\\ &= \sigma\varphi_\alpha(y^2x_\alpha^{-2j}) - (0, \rho\varphi_\alpha(y^2x_\alpha^{-2j})).\end{aligned}$$

Since $y^2x_\alpha^{-2j} = yx_\alpha^{-2j} + f(x)x_\alpha^{-2j}$, linearity of σ and φ_α yields that

$$\sigma\varphi_\alpha(y^2x_\alpha^{-2j}) = \sigma\varphi_\alpha(yx_\alpha^{-2j}) + \sigma\varphi_\alpha(f(x)x_\alpha^{-2j}).$$

The term $\sigma\varphi_\alpha(f(x)x_\alpha^{-2j})$ is a coboundary by Lemma 4.4(i), and $\sigma\varphi_\alpha(yx_\alpha^{-2j})$ equals $\sigma_{\alpha,2j}$ if $0 \le 2j \le c_\alpha$ and is a coboundary if $2j > c_\alpha$ by Lemma 4.4(ii). By Lemma 4.15,

$$(0, \rho\varphi_\alpha(y^2x_\alpha^{-2j})) = \begin{cases}0 & \text{if } 0 \le 2j \le c_\alpha,\\ -\lambda(R_{\alpha,2j}) & \text{if } 2j > c_\alpha.\end{cases}$$

(iv) Since $V(\phi, \rho(\phi)) = (0, \mathscr{C}(\rho(\phi)))$, the result follows by Lemma 4.16(iii). □

Consider the subspaces of $\mathrm{H}^1_{\mathrm{dR}}(X)$ given by

$$\begin{aligned}W_{\alpha,\mathrm{ss}} &:= \langle \lambda_{\alpha,0} - \lambda_{0,0}, \sigma_{\alpha,0}\rangle,\\ W_{\alpha,\mathrm{nil}} &:= \langle \lambda_{\alpha,j}, \sigma_{\alpha,j} \mid 1 \le j \le c_\alpha\rangle.\end{aligned}$$

Theorem 4.20. *The subspaces $W_{\alpha,\mathrm{ss}}$ and $W_{\alpha,\mathrm{nil}}$ of $\mathrm{H}^1_{\mathrm{dR}}(X)$ are stable under the action of Frobenius and Verschiebung for each $\alpha \in B$. There is an isomorphism of $\mathbb{E}$-modules*

$$\mathrm{H}^1_{\mathrm{dR}}(X) = \bigoplus_{\alpha\in B'} W_{\alpha,\mathrm{ss}} \oplus \bigoplus_{\alpha\in B} W_{\alpha,\mathrm{nil}}.$$

Proof. The stability is immediate by Proposition 4.19, Remark 4.10, and Lemma 4.15. The decomposition follows from Lemmas 4.18, 3.3, and 4.7. □

Theorem 1.2 is immediate from Theorem 4.20.

5. Results on the Ekedahl–Oort type

For a natural number c, let G_c be the unique symmetric BT_1 group scheme of rank p^{2c} with Ekedahl–Oort type $[0, 1, 1, 2, 2, \dots, \lfloor c/2\rfloor]$. In other words, this means that there is a final filtration $N_1 \subset N_2 \subset \cdots \subset N_{2c}$ of $D(G_c)$ as a k-vector space, which is stable under the action of V and F^{-1} and with $i = \dim(N_i)$, such that $\dim(V(N_i)) = \lfloor i/2\rfloor$. In Section 5a, we prove that group schemes of the form G_c appear in the decomposition of $J_X[2]$ when X is a hyperelliptic k-curve. In Section 5b, we describe the Dieudonné module of G_c for arbitrary c and give examples.

5a. *The final filtration for hyperelliptic curves in characteristic* 2. Suppose X is a hyperelliptic k-curve with affine equation $y^2 - y = f(x)$ as described in Notation 1.1. For $\alpha \in B$, recall that $c_\alpha = (d_\alpha - 1)/2$, where d_α is the ramification invariant of X above α. Recall the subspaces $W_{\alpha,\mathrm{nil}}$ of $\mathrm{H}^1_{\mathrm{dR}}(X)$ from Section 4f. Define subspaces $N_{\alpha,i}$ of $W_{\alpha,\mathrm{nil}}$ for $0 \le i \le 2c_\alpha$ as follows: $N_{\alpha,0} := \{0\}$ and

$$N_{\alpha,i} := \begin{cases} \langle \lambda_{\alpha,j} \mid 1 \le j \le i \rangle & \text{if } 1 \le i \le c_\alpha, \\ N_{\alpha,c_\alpha} \oplus \langle \sigma_{\alpha,j} \mid 1 \le j \le i \rangle & \text{if } c_\alpha + 1 \le i \le 2c_\alpha. \end{cases}$$

Proposition 5.1. *The filtration $N_{\alpha,0} \subset N_{\alpha,1} \subset N_{\alpha,2} \subset \cdots \subset N_{\alpha,2c_\alpha}$ is a final filtration of $W_{\alpha,\mathrm{nil}}$ for each $\alpha \in B$. Furthermore, $V(N_{\alpha,i}) = N_{\alpha,\lfloor i/2 \rfloor}$.*

Proof. Let $0 \le i \le 2c_\alpha$. Then $\dim(N_{\alpha,i}) = i$. By Proposition 4.19, $V(N_{\alpha,i}) = N_{\alpha,\lfloor i/2 \rfloor}$, and $F^{-1}(N_{\alpha,i}) = N_{\alpha,c_\alpha + \lceil i/2 \rceil}$. Thus, the filtration $N_{\alpha,0} \subset N_{\alpha,1} \subset N_{\alpha,2} \subset \cdots \subset N_{\alpha,2c_\alpha}$ is stable under the action of V and F^{-1}. □

Theorem 5.2. *Let k be an algebraically closed field of characteristic $p = 2$. Let X be a hyperelliptic k-curve with affine equation $y^2 - y = f(x)$ as described in Notation 1.1. Then the 2-torsion group scheme of X decomposes as*

$$J_X[2] \simeq (\mathbb{Z}/2 \oplus \mu_2)^r \oplus \bigoplus_{\alpha \in B} G_{c_\alpha},$$

and the a-number of X is

$$a_X = (g + 1 - \#\{\alpha \in B \mid d_\alpha \equiv 1 \bmod 4\})/2.$$

Proof. By [Oda 1969, Section 5], there is an isomorphism of $\mathbb{E}$-modules between the Dieudonné module $D(J_X[2])$ and the de Rham cohomology $\mathrm{H}^1_{\mathrm{dR}}(X)$. By Theorem 4.20, there is an isomorphism of $\mathbb{E}$-modules

$$\mathrm{H}^1_{\mathrm{dR}}(X) = \bigoplus_{\alpha \in B'} W_{\alpha,\mathrm{ss}} \oplus \bigoplus_{\alpha \in B} W_{\alpha,\mathrm{nil}}.$$

If $\alpha \in B'$, then $W_{\alpha,\mathrm{ss}}$ is isomorphic to $\mathbb{E}/\mathbb{E}(F, 1-V) \oplus \mathbb{E}/\mathbb{E}(V, 1-F) \simeq D(\mathbb{Z}/2 \oplus \mu_2)$. Finally, Proposition 5.1 shows $W_{\alpha,\mathrm{nil}} \simeq D(G_{c_\alpha})$, which completes the proof of the statement about $J_X[2]$. The statement about a_X can be found in Proposition 3.4. □

As a corollary, we highlight the special case when $r = 0$ (for example, when $f(x) \in k[x]$). Corollary 5.3 is stated without proof in [van der Geer 1999, 3.2].

Corollary 5.3. *Let k be an algebraically closed field of characteristic $p = 2$. Suppose X is a hyperelliptic k-curve of genus g and p-rank $r = 0$. Then the Ekedahl–Oort type of $J_X[2]$ is* $[0, 1, 1, 2, 2, \ldots, \lfloor g/2 \rfloor]$, *and the a-number $a_X = \lfloor (g+1)/2 \rfloor$.*

Proof. This is a special case of Theorem 5.2 where $\#B = 1$. □

The next immediate corollary of Theorem 5.2 is included to emphasize that Theorem 5.2 gives a complete classification of the 2-torsion group schemes that occur as $J_X[2]$ when X is a hyperelliptic k-curve.

Corollary 5.4. *Let k be an algebraically closed field of characteristic $p = 2$. Let G be a polarized BT_1 group scheme over k of rank p^{2g}. Let $0 \le r \le g$. Then $G \simeq J_X[2]$ for some hyperelliptic k-curve X of genus g and p-rank r if and only if there exist nonnegative integers $c_1, \dots, c_{r+1}$ such that $\sum_{i=1}^{r+1} c_i = g - r$ and such that*

$$G \simeq (\mathbb{Z}/2 \oplus \mu_2)^r \oplus \bigoplus_{\alpha \in B} G_{c_\alpha}.$$

Remark 5.5. For fixed g, the number of isomorphism classes of polarized BT_1 group schemes of rank p^{2g} that occur as $J_X[2]$ for some hyperelliptic k-curve X of genus g equals the number of partitions of $g + 1$. To see this, note that the isomorphism class of $J_X[2]$ is determined by the multiset $\{d_1, \dots, d_{r+1}\}$, where $d_i = 2c_i + 1$ and $\sum_{i=1}^{r+1}(d_i + 1) = 2g + 2$. So the number of isomorphism classes equals the number of partitions of $2g + 2$ into positive even integers.

Remark 5.6. The examples in Section 5b show that the factors G_c appearing in the decomposition of $J_X[2]$ in Theorem 5.2 may not be indecomposable as polarized BT_1 group schemes.

5b. ***Description of a particular Ekedahl–Oort type.*** Recall that G_c is the unique polarized BT_1 group scheme over k of rank p^{2c} that has Ekedahl–Oort type $[0, 1, 1, 2, 2, \dots, \lfloor c/2 \rfloor]$. Recall that $\mathbb{E} = k[F, V]$ is the noncommutative ring defined in Section 2b. In this section, we describe the Dieudonné module $D(G_c)$. We start with some examples to motivate the notation. These show that G_c is sometimes indecomposable and sometimes decomposes into polarized BT_1 group schemes of smaller rank. The first four examples were found using preexisting tables.

Example 5.7. (i) For $c = 1$, the Ekedahl–Oort type is $[0]$. This Ekedahl–Oort type occurs for the p-torsion group scheme of a supersingular elliptic curve. See [Goren 2002, Example A.3.14; Pries 2008, Example 2.3] for a description of G_1. It has Dieudonné module $\mathbb{E}/\mathbb{E}(F + V)$.

(ii) For $c = 2$, the Ekedahl–Oort type is $[0, 1]$. This Ekedahl–Oort type occurs for the p-torsion group scheme of a supersingular abelian surface that is not superspecial. See [Goren 2002, Example A.3.15; Pries 2008, Example 2.3] for a description of G_2. It has Dieudonné module $\mathbb{E}/\mathbb{E}(F^2 + V^2)$.

(iii) For $c = 3$, the Ekedahl–Oort type is $[0, 1, 1]$. This Ekedahl–Oort type occurs for an abelian threefold with p-rank 0 and a-number 2 whose p-torsion is indecomposable as a polarized BT_1 group scheme. By [Pries 2008, Lemma 3.4], G_3 has Dieudonné module $\mathbb{E}/\mathbb{E}(F^2 + V) \oplus \mathbb{E}/\mathbb{E}(V^2 + F)$.

(iv) For $c = 4$, the Ekedahl–Oort type is [0, 1, 1, 2]. This Ekedahl–Oort type occurs for an abelian fourfold with p-rank 0 and a-number 2 whose p-torsion decomposes as a direct sum of polarized BT_1 group schemes of rank p^2 and p^6. By [Pries 2008, Table 4.4], G_4 has Dieudonné module $\mathbb{E}/\mathbb{E}(F+V) \oplus \mathbb{E}/\mathbb{E}(F^3+V^3)$.

We now provide an algorithm to determine the Dieudonné module $D(G_c)$ for all positive integers $c \in \mathbb{N}$ following the method of [Oort 2001, Section 9.1].

Proposition 5.8. *The Dieudonné module $D(G_c)$ is the $\mathbb{E}$-module generated as a k-vector space by $\{X_1, \ldots, X_c, Y_1, \ldots, Y_c\}$ with the actions of F and V given by*

(i) $F(Y_j) = 0$,

(ii) $V(Y_j) = \begin{cases} Y_{2j} & \text{if } j \le c/2, \\ 0 & \text{if } j > c/2, \end{cases}$

(iii) $F(X_i) = \begin{cases} X_{j/2} & \text{if } j \text{ is even}, \\ Y_{c-(j-1)/2} & \text{if } j \text{ is odd}, \end{cases}$

(iv) $V(X_j) = \begin{cases} 0 & \text{if } j \le (c-1)/2, \\ -Y_{2c-2j+1} & \text{if } j > (c-1)/2. \end{cases}$

Proof. By definition of G_c, there is a final filtration $N_1 \subset N_2 \subset \cdots \subset N_{2c}$ of $D(G_c)$ as a k-vector space, which is stable under the action of V and F^{-1} and with $i = \dim(N_i)$, such that $\nu_i := \dim(V(N_i)) = \lfloor i/2 \rfloor$. This implies that $\nu_i = \nu_{i-1}$ if and only if i is odd. In the notation of [Oort 2001, Section 9.1], this yields $m_i = 2i$ and $n_i = 2g - 2i + 1$ for $1 \le i \le g$; also, let

$$Z_i := \begin{cases} X_{i/2} & \text{if } i \text{ is even}, \\ Y_{c-(i-1)/2} & \text{if } i \text{ is odd}. \end{cases}$$

By [Oort 2001, Section 9.1], for $1 \le i \le g$, the action of F is given by $F(Y_i) = 0$ and $F(X_i) = Z_i$ and the action of V by $V(Z_i) = 0$ and $V(Z_{2g-i+1}) = (-1)^{i-1} Y_i$. $\square$

More notation is needed to give an explicit description of $D(G_c)$.

Notation 5.9. Let $c \in \mathbb{N}$ be fixed. Let $I := \{j \in \mathbb{N} \mid \lceil (c+1)/2 \rceil \le j \le c\}$, which is a set of cardinality $\lfloor (c+1)/2 \rfloor$. For $j \in I$, let $\ell(j)$ be the odd part of j, and let $e(j)$ be the nonnegative integer such that $j = 2^{e(j)} \ell(j)$. Let $s(j) := c - (\ell(j) - 1)/2$. One can check that $\{s(j) \mid j \in I\} = I$. Also, let $m(j) := 2c - 2j + 1$, and let $\epsilon(j)$ be the nonnegative integer such that $t(j) := 2^{\epsilon(j)} m(j) \in I$. One can check that $\{t(j) \mid j \in I\} = I$. Thus, there is a unique bijection $\iota : I \to I$ such that $t(\iota(j)) = s(j)$ for each $j \in I$.

Proposition 5.10. *Recall Notation 5.9. For $c \in \mathbb{N}$, the set $\{X_j \mid j \in I\}$ generates the Dieudonné module $D(G_c)$ as an $\mathbb{E}$-module subject to the relations, for $j \in I$, $F^{e(j)+1}(X_j) + V^{\epsilon(\iota(j))+1}(X_{\iota(j)}) = 0$. Also, $\{X_j \mid j \in I\}$ is a basis for the quotient of $D(G_c)$ by the left ideal $D(G_c)(F, V)$.*

Proof. Proposition 5.8 implies that $F^{e(j)}(X_j) = X_{\ell(j)}$ and $F(X_{\ell(j)}) = Y_{s(j)}$. Also, $V(X_j) = -Y_{m(j)}$, and so $V^{\epsilon(j)+1}(X_j) = -Y_{t(j)}$. This yields the stated relations. To complete the first claim, it suffices to show that the span of $\{X_j \mid j \in I\}$ under the action of F and V contains the k-module generators of $D(C_c)$ listed in Proposition 5.8. This follows from the observations that $X_i = F(X_{2i})$ if $1 \le i \le \lfloor c/2 \rfloor$, that $Y_i = V(Y_{i/2})$ if i is even, and that $Y_i = V(-X_{c-(i-1)/2})$ if i is odd. By [Li and Oort 1998, 5.2.8], the dimension of $D(G_c)$ modulo $D(G_c)(F, V)$ equals the a-number. Since $a = |I|$ by Corollary 5.3, it follows that the set $|I|$ of generators of $D(G_c)$ is linearly independent modulo $D(G_c)(F, V)$. □

Here are some more examples. The columns of the table below list the value of c, the generators of $D(G_c)$ as an $\mathbb{E}$-module ($X_{i_1} - X_{i_2}$ denotes $\{X_i \mid i_1 \le i \le i_2\}$), and the relations among these generators. The last column is the number of summands of $D(G_c)$ in its decomposition as an $\mathbb{E}$-module (as opposed to as a polarized $\mathbb{E}$-module). The table can be verified in two ways: first, by checking it with Proposition 5.10 and second, by computing the action of F and V on a k-basis for $D(G_c)$, using this to construct a final filtration of $D(G_c)$ stable under V and F^{-1}, and then checking that it matches the Ekedahl–Oort type of G_c. In Example 5.11, we illustrate the second method.

c	generators	relations	# summands
5	$X_3 - X_5$	$FX_3 + V^3X_5,\ F^3X_4 + VX_3,\ FX_5 + VX_4$	1
6	$X_4 - X_6$	$F^3X_4 + V^2X_5,\ FX_5 + V^3X_6,\ F^2X_6 + VX_4$	1
7	$X_4 - X_7$	$F^3X_4 + VX_4,\ FX_5 + VX_5,\ F^2X_6 + V^2X_6,\ FX_7 + V^3X_7$	4
8	$X_5 - X_8$	$FX_5 + V^2X_7,\ F^2X_6 + VX_5,\ FX_7 + VX_6,\ F^4X_8 + V^4X_8$	2
9	$X_5 - X_9$	$FX_5 + VX_6,\ F^2X_6 + V^4X_9,\ FX_7 + V^2X_8,\ F^4X_8 + VX_5,\ FX_9 + VX_7$	1
10	$X_6 - X_{10}$	$F^2X_6 + VX_6,\ FX_7 + VX_7,\ F^4X_8 + V^2X_8,\ FX_9 + V^2X_9,\ F^2X_{10} + V^4X_{10}$	5

Example 5.11. For $c = 7$, the group scheme G_7 that has Ekedahl–Oort type $[0, 1, 1, 2, 2, 3, 3]$ is isomorphic to a direct sum of polarized BT_1 group schemes of ranks p^2, p^4, and p^8 and has Dieudonné module

$$\mathbf{M} := \mathbb{E}/\mathbb{E}(F+V) \oplus \mathbb{E}/\mathbb{E}(F^2+V^2) \oplus \mathbb{E}/\mathbb{E}(V+F^3) \oplus \mathbb{E}/\mathbb{E}(F^3+V).$$

Proof. Let $\{1_A, V_A\}$ be the basis of the submodule $A = \mathbb{E}/\mathbb{E}(F+V)$ of $\mathbf{M}$, $\{1_B, V_B, V_B^2, F_B^2\}$ the basis of the submodule $B = \mathbb{E}/\mathbb{E}(F^2+V^2)$, $\{1_C, V_C, V_C^2, V_C^3\}$ the basis of the submodule $C = \mathbb{E}/\mathbb{E}(F+V^3)$, and $\{1_{C'}, F_{C'}, F_{C'}^2, F_{C'}^3\}$ the basis of the submodule $C' = \mathbb{E}/\mathbb{E}(F^3 + V)$. The action of Frobenius and Verschiebung on the elements of these bases is

x	1_A	V_A	1_B	V_B	V_B^2	F_B	1_C	V_C	V_C^2	V_C^3	$1_{C'}$	$F_{C'}$	$F_{C'}^2$	$F_{C'}^3$
Vx	V_A	0	V_B	V_B^2	0	0	V_C	V_C^2	V_C^3	0	$F_{C'}^3$	0	0	0
Fx	V_A	0	F_B	0	0	V_B^2	V_C^3	0	0	0	$F_{C'}$	$F_{C'}^2$	$F_{C'}^3$	0

To verify the proposition, one can repeatedly apply V and F^{-1} to construct a filtration $N_1 \subset N_2 \subset \cdots \subset N_{14}$ of $\mathbf{M}$ as a k-vector space that is stable under the action of V and F^{-1} such that $i = \dim(N_i)$. To save space, we summarize the calculation by listing a generator t_i for N_i/N_{i-1}:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
t_i	V_C^3	V_C^2	V_B^2	V_C	V_A	$F_{C'}^3$	V_B	1_C	$F_{C'}^2$	1_A	F_B	$F_{C'}$	$1_{C'}$	1_B

Then one can check that $V(N_i) = N_{\lfloor i/2 \rfloor}$ and $F^{-1}(N_i) = N_{7+\lceil i/2 \rceil}$, which verifies that the Ekedahl–Oort type of $\mathbf{M}$ is [0, 1, 1, 2, 2, 3, 3]. □

Remark 5.12. One could ask when $D(G_c)$ decomposes as much as numerically possible, in other words, when the a-number equals the number of summands of $D(G_c)$ in its decomposition as an $\mathbb{E}$-module. For example, $D(G_c)$ has this property when $c \in \{1, 2, 3, 4, 7, 10\}$ but not when $c \in \{5, 6, 8, 9\}$. This phenomenon occurs if and only if the bijection ι from Notation 5.9 is the identity.

Remark 5.13. The group scheme G_8 decomposes as the direct sum of two indecomposable polarized BT_1 group schemes, one whose Ekedahl–Oort type is [0, 0, 1, 1] and the other whose covariant Dieudonné module is $\mathbb{E}/\mathbb{E}(F^4 + V^4)$. We take this opportunity to note that there is a mistake in [Pries 2008, Example in Section 3.3]. The covariant Dieudonné module of $I_{4,3} = [0, 0, 1, 1]$ is stated incorrectly. To fix it, consider the method of [Oort 2001, Section 9.1]. Consider the k-vector space of dimension 8 generated by $X_1, \ldots, X_4$ and $Y_1, \ldots, Y_4$. Consider the operation F defined by $F(Y_i) = 0$ for $1 \le i \le 4$,

$$F(X_1) = Y_4, \quad F(X_2) = Y_3, \quad F(X_3) = X_1, \quad F(X_4) = Y_2.$$

Consider the operation V defined by

$$\begin{aligned} &V(X_1) = 0, \quad V(X_2) = -Y_4, \quad V(X_3) = -Y_2, \quad V(X_4) = -Y_1, \\ &V(Y_1) = Y_3, \quad V(Y_2) = 0, \quad V(Y_3) = 0, \quad V(Y_4) = 0. \end{aligned}$$

Thus, $D(I_{4,3})$ is generated by X_2, X_3, and X_4 modulo the three relations

$$F X_2 + V^2 X_4, \quad F^2 X_3 + V X_2, \quad V X_3 + F X_4.$$

5c. ***Newton polygons.*** There are several results in characteristic 2 about the Newton polygons of hyperelliptic (for example, Artin–Schreier) curves X of genus g and 2-rank 0. For example, [Blache 2012, Remark 3.6] states that if $2^{n-1}-1 \le g \le 2^n-2$, then the generic first slope of the Newton polygon of an Artin–Schreier curve of genus g and 2-rank 0 is $1/n$. This statement is generalized to odd primes p in [Blache 2012, Proposition 3.5]. See also earlier work in [Scholten and Zhu 2002, Theorem 1.1(III)].

The Ekedahl–Oort type of $J_X[2]$ gives information about the Newton polygon of X but does not determine it completely. Using Corollary 5.3 and [Harashita 2007, Section 3.1 and Theorem 4.1], one can show that the first slope of the Newton polygon of X is at least $1/n$. Since this is weaker than [Blache 2012, Theorem 4.3], we do not include the details.

More generally, one could consider the case that X is a hyperelliptic k-curve of genus g and arbitrary p-rank. One could use Theorem 5.2 to give partial information (namely, a lower bound) for the Newton polygon of X.

Acknowledgments

We would like to thank Jeff Achter and the anonymous referee for helpful comments.

References

[Blache 2012] R. Blache, "Valuation of exponential sums and the generic first slope for Artin–Schreier curves", *J. Number Theory* **132**:10 (2012), 2336–2352. MR 2944758 Zbl 06074389

[Bouw 2001] I. I. Bouw, "The p-rank of ramified covers of curves", *Compositio Math.* **126**:3 (2001), 295–322. MR 2002e:14045 Zbl 1056.14036

[Cartier 1957] P. Cartier, "Une nouvelle opération sur les formes différentielles", *C. R. Acad. Sci. Paris* **244** (1957), 426–428. MR 18,870b Zbl 0077.04502

[Crew 1984] R. M. Crew, "Etale p-covers in characteristic p", *Compositio Math.* **52**:1 (1984), 31–45. MR 85f:14011 Zbl 0558.14009

[Demazure 1972] M. Demazure, *Lectures on p-divisible groups*, Lecture Notes in Mathematics **302**, Springer, Berlin, 1972. MR 88a:14049 Zbl 0247.14010

[Ekedahl 1987] T. Ekedahl, "On supersingular curves and abelian varieties", *Math. Scand.* **60**:2 (1987), 151–178. MR 88g:14037 Zbl 0641.14007

[Elkin 2011] A. Elkin, "The rank of the Cartier operator on cyclic covers of the projective line", *J. Algebra* **327** (2011), 1–12. MR 2011m:14071 Zbl 1216.14026

[Goren 2002] E. Z. Goren, *Lectures on Hilbert modular varieties and modular forms*, CRM Monograph Series **14**, American Mathematical Society, Providence, RI, 2002. MR 2003c:11038 Zbl 0986.11037

[Harashita 2007] S. Harashita, "Ekedahl–Oort strata and the first Newton slope strata", *J. Algebraic Geom.* **16**:1 (2007), 171–199. MR 2007k:14043 Zbl 1113.14031

[Hardy and Ramanujan 1918] G. H. Hardy and S. Ramanujan, "Asymptotic formulæ in combinatory analysis", *Proc. Lond. Math. Soc. (2)* **17** (1918), 75–115. Reprinted as pp. 276–309 in *Collected papers of Srinivasa Ramanujan*, AMS Chelsea Publ., Providence, RI, 2000. JFM 46.0198.02

[Johnston 2007] O. Johnston, "A note on the a-numbers and p-ranks of Kummer covers", preprint, 2007. arXiv 0710.2120

[Kraft 1975] H. Kraft, "Kommutative algebraische p-Gruppen (mit Anwendungen auf p-divisible Gruppen und abelsche Varietäten)", manuscript, Sonderforsch. Bereich Bonn, 1975.

[Li and Oort 1998] K.-Z. Li and F. Oort, *Moduli of supersingular abelian varieties*, Lecture Notes in Mathematics **1680**, Springer, Berlin, 1998. MR 99e:14052 Zbl 0920.14021

[Madden 1978] D. J. Madden, "Arithmetic in generalized Artin–Schreier extensions of $k(x)$", *J. Number Theory* **10**:3 (1978), 303–323. MR 80d:12009 Zbl 0384.12008

[Moonen 2001] B. Moonen, "Group schemes with additional structures and Weyl group cosets", pp. 255–298 in *Moduli of abelian varieties* (Texel Island, 1999), edited by C. Faber et al., Progr. Math. **195**, Birkhäuser, Basel, 2001. MR 2002c:14074 Zbl 1084.14523

[Nart and Sadornil 2004] E. Nart and D. Sadornil, "Hyperelliptic curves of genus three over finite fields of even characteristic", *Finite Fields Appl.* **10**:2 (2004), 198–220. MR 2005a:11090 Zbl 1049.11063

[Oda 1969] T. Oda, "The first de Rham cohomology group and Dieudonné modules", *Ann. Sci. École Norm. Sup.* (4) **2** (1969), 63–135. MR 39 #2775 Zbl 0175.47901

[Oort 2001] F. Oort, "A stratification of a moduli space of abelian varieties", pp. 345–416 in *Moduli of abelian varieties* (Texel Island, 1999), edited by C. Faber et al., Progr. Math. **195**, Birkhäuser, Basel, 2001. MR 2002b:14055 Zbl 1052.14047

[Pries 2008] R. Pries, "A short guide to p-torsion of abelian varieties in characteristic p", pp. 121–129 in *Computational arithmetic geometry* (San Francisco, CA, 2006), edited by K. E. Lauter and K. A. Ribet, Contemp. Math. **463**, American Mathematical Society, Providence, RI, 2008. MR 2009m:11085 Zbl 1151.14328

[Scholten and Zhu 2002] J. Scholten and H. J. Zhu, "Hyperelliptic curves in characteristic 2", *Int. Math. Res. Not.* **2002**:17 (2002), 905–917. MR 2003d:11089 Zbl 1034.14013

[Serre 1968] J.-P. Serre, *Corps locaux*, 2nd ed., Actualités scientifiques et industrielles **1296**, Hermann, Paris, 1968. MR 50 #7096 Zbl 0423.12017

[Stichtenoth 2009] H. Stichtenoth, *Algebraic function fields and codes*, 2nd ed., Graduate Texts in Mathematics **254**, Springer, Berlin, 2009. MR 2010d:14034 Zbl 1155.14022

[Subrao 1975] D. Subrao, "The p-rank of Artin–Schreier curves", *Manuscripta Math.* **16**:2 (1975), 169–193. MR 51 #12868 Zbl 0321.14017

[Sullivan 1975] F. J. Sullivan, "p-torsion in the class group of curves with too many automorphisms", *Arch. Math. (Basel)* **26** (1975), 253–261. MR 52 #13846 Zbl 0331.14016

[van der Geer 1999] G. van der Geer, "Cycles on the moduli space of abelian varieties", pp. 65–89 in *Moduli of curves and abelian varieties: The Dutch intercity seminar on moduli*, edited by C. Faber and E. Looijenga, Aspects Math. **E33**, Vieweg, Braunschweig, 1999. MR 2001b:14009 Zbl 0974.14031

[Yui 1978] N. Yui, "On the Jacobian varieties of hyperelliptic curves over fields of characteristic $p > 2$", *J. Algebra* **52**:2 (1978), 378–410. MR 58 #10920 Zbl 0404.14008

Communicated by Richard Taylor
Received 2010-07-07 Revised 2012-04-11 Accepted 2012-04-16

A.Elkin@warwick.ac.uk *Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom*

pries@math.colostate.edu *Department of Mathematics, Colorado State University, Fort Collins, CO, 80523, United States*

ALGEBRA AND NUMBER THEORY 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.533

Cycle classes and the syntomic regulator

Bruno Chiarellotto, Alice Ciccioni and Nicola Mazzari

Let $\mathcal{V} = \operatorname{Spec}(R)$ and R be a complete discrete valuation ring of mixed characteristic $(0, p)$. For any flat R-scheme $\mathcal{X}$, we prove the compatibility of the de Rham fundamental class of the generic fiber and the rigid fundamental class of the special fiber. We use this result to construct a syntomic regulator map $\mathrm{reg}_{\mathrm{syn}} : CH^i(\mathcal{X}/\mathcal{V}, 2i-n) \to H^n_{\mathrm{syn}}(\mathcal{X}, i)$ when $\mathcal{X}$ is smooth over R with values in the syntomic cohomology defined by A. Besser. Motivated by the previous result, we also prove some of the Bloch–Ogus axioms for the syntomic cohomology theory but viewed as an absolute cohomology theory.

Introduction

Let $\mathcal{V} = \operatorname{Spec}(R)$ with R a complete discrete valuation ring of mixed characteristic and perfect residue field. Given $\mathcal{X}$, an algebraic $\mathcal{V}$-scheme, one can consider the de Rham cohomology of its generic fiber $\mathcal{X}_K$ and the rigid cohomology of its special fiber $\mathcal{X}_k$. These two cohomology groups are related by a canonical *cospecialization* map $cosp : H^n_{\mathrm{rig},c}(\mathcal{X}_k) \to H^n_{\mathrm{dR},c}(\mathcal{X}_K)$ (in general not an isomorphism) [Baldassarri et al. 2004, Section 6]. There is also the notion of rigid and de Rham cycle class. The starting result of this paper is the compatibility of these cycle classes with respect to the cospecialization map (see Theorem 1.4.1 for the precise statement).

In the case $\mathcal{X}$ is smooth (possibly nonproper) over $\mathcal{V}$, we get the following corollary (see Corollary 1.5.1). Let $sp_{CH} : CH^*(\mathcal{X}_K) \otimes \mathbb{Q} \to CH^*(\mathcal{X}_k) \otimes \mathbb{Q}$ be the specialization of Chow rings constructed by Grothendieck in [Berthelot et al. 1971, Appendix]. Then the diagram

$$\begin{tikzcd} CH^q(\mathcal{X}_K/K) \otimes \mathbb{Q} \arrow[r, "\eta_{\mathrm{dR}}"] \arrow[d, "sp_{CH}"'] & H^{2q}_{\mathrm{dR}}(\mathcal{X}_K) \arrow[d, "sp"] \\ CH^q(\mathcal{X}_k/k) \otimes \mathbb{Q} \arrow[r, "\eta_{\mathrm{rig}}"'] & H^{2q}_{\mathrm{rig}}(\mathcal{X}_k/K) \end{tikzcd}$$

The three authors were supported by the Cariparo Eccellenza Grant “Differential Methods in Arithmetic, Geometry and Algebra”.
MSC2010: primary 14F43; secondary 14F30, 19F27.
Keywords: syntomic cohomology, cycles, regulator map, rigid cohomology, de Rham cohomology.

is commutative, where η_{dR} and η_{rig} are the de Rham and cycle class maps, respectively, and *sp* is the Poincaré dual of *cosp*. In the proof, we use the main results of [Baldassarri et al. 2004; Bosch et al. 1995; Petrequin 2003].

This result can be viewed as a generalization of a theorem of Messing [1987, Theorem B3.1], in which he further assumes $\mathscr{X}$ to be proper (not only smooth) over $\mathscr{V}$. In that case, rigid and crystalline coincide, and the map *sp* is an isomorphism [Berthelot 1997b].

This compatibility result is the motivation for an alternative construction of the regulator map (see Proposition 1.6.6)

$$\mathrm{reg}_{\mathrm{syn}} : CH^i(\mathscr{X}/\mathscr{V}, 2i-n) \to H^n_{\mathrm{syn}}(\mathscr{X}, i)$$

with values in the syntomic cohomology group defined by Besser [2000] (for $\mathscr{X}$ smooth over $\mathscr{V}$). For this proof, we use an argument of Bloch [1986] and the existence of a syntomic cycle class (see Proposition 1.6.2).

The aforementioned results motivated us to investigate further the properties of syntomic cohomology. We are not able to formulate even the basic Bloch–Ogus axioms using Besser's framework. Thus, we have followed the interpretation of syntomic cohomology of Bannai [2002] as an absolute one. To this end, we define a triangulated category of p-adic Hodge complexes, *pHD* (see Definition 2.0.11). An object M of *pHD* can be represented by a diagram of the form $M_{\mathrm{rig}} \to M_K \leftarrow M_{\mathrm{dR}}$, where $M_?$ is a complex of K-vector spaces endowed with a Frobenius automorphism when $? = \mathrm{rig}$ and with a filtration when $? = \mathrm{dR}$. In *pHD*, there is a naturally defined tensor product, and $\mathbb{K}$ denotes the unit object of *pHD*. The main difference with respect to [Bannai 2002] is that the maps in the diagram are not necessarily quasi-isomorphisms.

From [Besser 2000], we get (in Proposition 5.3.1) that there are functorial p-adic Hodge complexes $R\Gamma(\mathscr{X})$ satisfying

$$R\Gamma_{\mathrm{rig}}(\mathscr{X}) \to R\Gamma_K(\mathscr{X}) \leftarrow R\Gamma_{\mathrm{dR}}(\mathscr{X})$$

and inducing the specialization map in cohomology (that is, taking the cohomology of each element of the diagram). Meanwhile, we show how the constructions made by Besser may be obtained using the theory of generalized Godement resolution (also called the bar resolution). In particular, we use the results of [van der Put and Schneider 1995] in order to have enough points for rigid analytic spaces. Further, we consider the twisted version $R\Gamma(\mathscr{X})(i)$, which is given by the same complexes but with the Frobenius (respectively the filtration) twisted by i (see Remark 2.2.1).

We then prove (see Proposition 5.3.4) that the syntomic cohomology groups $H^n_{\mathrm{syn}}(\mathscr{X}, i)$ of [Besser 2000] are isomorphic to the (absolute cohomology) groups $H^n_{\mathrm{abs}}(\mathscr{X}, i) := \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma(\mathscr{X})(i)[n])$. This result generalizes that of [Bannai

2002] (which was given only for $\mathcal{V}$-schemes with good compactification) to any smooth algebraic $\mathcal{V}$-scheme.

For this absolute cohomology, we can prove some of the Bloch–Ogus axioms. In fact, we construct a p-adic Hodge complex $R\Gamma_c(\mathcal{X})(i)$ related to rigid and de Rham cohomology with compact support. Therefore, we can define an absolute cohomology with compact support functorial with respect to proper maps $H^n_{\mathrm{abs},c}(\mathcal{X}, i) := \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma_c(\mathcal{X})(i)[n])$ and an absolute homology theory $H_n^{\mathrm{abs}}(\mathcal{X}, i) := \mathrm{Hom}_{pHD}(R\Gamma_c(\mathcal{X})(i)[n], \mathbb{K})$.

We wish to point out that the constructions above are essentially consequences of the work done by Besser and Bannai, but it seems hard to prove the following results without the formalism of Godement resolutions that we develop in Section 3. Let $\mathcal{X}$ be a smooth scheme over $\mathcal{V}$. Then

(i) there is a cup product pairing

$$H^n_{\mathrm{abs}}(\mathcal{X}, i) \otimes H^m_{\mathrm{abs},c}(\mathcal{X}, j) \to H^{n+m}_{\mathrm{abs},c}(\mathcal{X}, i+j)$$

induced by the natural pairings defined on the cohomology of the generic and the special fiber (see Corollary 5.4.4),

(ii) there is a Poincaré duality isomorphism (see Proposition 5.4.5) and

(iii) there is a Gysin map; that is, given a proper morphism $f : \mathcal{X} \to \mathcal{Y}$ of smooth algebraic $\mathcal{V}$-schemes of relative dimension d and e, respectively, then there is a canonical map

$$f_* : H^n_{\mathrm{abs}}(\mathcal{X}, i) \to H^{n+2c}_{\mathrm{abs}}(\mathcal{Y}, i+c),$$

where $c = e - d$ (see Corollary 5.4.7).

Notation. In this paper, R is a complete discrete valuation ring with fraction field K and residue field k with k perfect. We assume $\mathrm{char}(K) = 0$ and $\mathrm{char}(k) = p > 0$.

The ring of Witt vectors of k is denoted by R_0, and K_0 is its field of fractions. The Frobenius of K_0 is denoted by σ. The category of bounded complexes of K-vector spaces is denoted by $C^b(K)$.

If V is a K-vector space, then $V^\vee$ is the dual vector space.

We use $X, Y, \ldots$ for schemes over k or K; $\mathcal{X}, \mathcal{Y}, \ldots$ for K-analytic spaces; P, Q for formal K-schemes; $\mathsf{P}_K, \mathsf{Q}_K$ for the associated Raynaud fibers and $\mathcal{X}, \mathcal{Y}$ for (algebraic) $\mathcal{V}$-schemes, where $\mathcal{V} = \mathrm{Spec}(R)$. Finally, $\widehat{\mathcal{X}}$ denotes the p-adic completion of a $\mathcal{V}$-scheme $\mathcal{X}$.

1. Cycle classes

1.1. ***Higher Chow groups.*** Following Bloch [1986], we recall the definition of the higher Chow groups $CH^*(X/K, *)$ of a K-scheme X. For any $n \geq 0$, let

$\Delta^n := \operatorname{Spec}(\mathbb{Z}[t_0, \dots, t_n]/(\sum_i t_i - 1))$ with face maps $\partial_i(n) : \Delta^n \to \Delta^{n+1}$, which in coordinates are given by $\partial_i(n)(t_0, \dots, t_n) := (t_0, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_n)$. Let X be a smooth K-scheme of relative dimension d (this hypothesis is not necessary, but we deal only with smooth schemes). Let $z^q(X/K, 0)$ be the free abelian group generated by the irreducible and closed subschemes of X of codimension q. Let $z^q(X/K, n)$ denote the free abelian group generated by the closed subschemes $W \subset \Delta^n_X := X \times \Delta^n$ such that $W \in z^q(\Delta^n_X/K, 0)$ and meets all faces properly; that is, if $F \subset \Delta^n_X$ is a face of codimension c, then the codimension of each irreducible component of the intersection $F \cap W$ is greater than or equal to $c+q$ on Δ^n_X.

Using the differential $\sum_i (-1)^i \partial_i^*(n) : z^q(X/K, n+1) \to z^q(X/K, n)$, one obtains a complex of abelian groups $z^q(X/K, *)$. We set $\Gamma^i_X(q) := z^q(X/K, 2q-i)$ and

$$CH^q(X/K, 2q-i) := H^i(\Gamma_X(q)).$$

These groups are in fact isomorphic to the Voevodsky–Suslin motivic cohomology $H^i_{mot}(X/K, q)$ of the generic fiber X [Mazza et al. 2006, Theorem 19.1].

1.2. ***Relative cycles.*** Let $\mathscr{X}$ be an algebraic and flat $\mathscr{V}$-scheme. By the theory of relative cycles [Suslin and Voevodsky 2000], one can define the group $z^q(\mathscr{X}/\mathscr{V}, 0)$ to be the free abelian group generated by *universally integral relative cycles* of codimension q (we can use the codimension because $\mathscr{X}$ is assumed to be equidimensional over $\mathscr{V}$). By [Ivorra 2005, Part I, Lemma 1.2.6], $z^q(\mathscr{X}/\mathscr{V}, 0)$ is the free abelian group generated by the closed subschemes $\mathscr{W} \subset \mathscr{X}$ that are integral, of codimension q and flat over $\mathscr{V}$. Then we can define the group $z^q(\mathscr{X}/\mathscr{V}, n)$ as the free abelian group generated by the integral and flat $\mathscr{V}$-schemes $\mathscr{W} \in z^q(\Delta^n_{\mathscr{X}}, 0)$ meeting all faces properly and such that $\partial_i(n-1)^*\mathscr{W}$ is flat over $\mathscr{V}$ for all i. Thus, we can form a complex $z^q(\mathscr{X}/\mathscr{V}, *)$ with the same boundary maps as $z^q(X/K, *)$.

Definition 1.2.1. With the notation above, we define the higher Chow groups of $\mathscr{X}$ over $\mathscr{V}$ to be

$$CH^q(\mathscr{X}/\mathscr{V}, 2q-i) := H^i(\Gamma_X(q)),$$

where $\Gamma^i_{\mathscr{X}/\mathscr{V}}(q) := z^q(\mathscr{X}/\mathscr{V}, 2q-i)$.

Remark 1.2.2. (i) Recall that by Lemma 5.1.1, any closed and flat subscheme of $\mathscr{X}$ is completely determined by its generic fiber. Then $z^q(\mathscr{X}/\mathscr{V}, *)$ is a subcomplex of $z^q(\mathscr{X}_K/K, *)$ inducing a canonical map in (co)homology

$$\gamma : CH^q(\mathscr{X}/\mathscr{V}, 2q-i) := H^i(\Gamma_{\mathscr{X}/\mathscr{V}}(q)) \to CH^q(\mathscr{X}_K/K, 2q-i).$$

(ii) It follows easily that, for $i = 2q$, the map $\gamma : CH^q(\mathscr{X}/\mathscr{V}, 0) \to CH^q(\mathscr{X}_K/K, 0)$ is surjective by the snake lemma. In the general case, we don't know whether or not this map is injective or surjective.

1.3. ***De Rham and rigid fundamental/cycle classes.*** In the following, we refer to [Hartshorne 1975; Baldassarri et al. 2004; Petrequin 2003] for the definitions and the properties of the (algebraic) de Rham and the rigid cohomology theory. Let W be an integral scheme of dimension r over K (resp. over k). Then we can associate to W its de Rham (resp. rigid) fundamental class, which is an element of the dual of the top de Rham (resp. rigid) cohomology with compact support

$$[W]_{\mathrm{dR}} = \mathrm{tr}_{\mathrm{dR}} \in H^{2r}_{\mathrm{dR},c}(W)^\vee \quad (\text{resp. } [W]_{\mathrm{rig}} = \mathrm{tr}_{\mathrm{rig}} \in H^{2r}_{\mathrm{rig},c}(W)^\vee).$$

For the de Rham case, this class is first defined in [Hartshorne 1975, Section 7] as an element of the de Rham homology; by Poincaré duality [Baldassarri et al. 2004, Theorem 3.4], it corresponds to the trace map. The rigid case is treated in [Petrequin 2003, Section 2.1, Section 6].

Now let X be a K-scheme (resp. k-scheme) of dimension d and $w = \sum_i n_i W_i \in z^q(X/K, 0)$ (resp. $\in z^q(X/k, 0)$) be a dimension-r cycle on X. The cohomology with compact support is functorial with respect to proper maps; hence, there is a canonical map

$$f : \bigoplus_i H^{2r}_{?,c}(W_i)^\vee \to H^{2r}_{?,c}(|w|)^\vee,$$

where $|w| = \bigcup_i W_i$ is the support of w and ? is dR or rig according to the choice of the base field. With the notation above, we define the de Rham (resp. rigid) cycle class of w as

$$[w]_? := f\Big(\sum_i [W_i]_?\Big) \in H^{2r}_{?,c}(|w|)^\vee, \quad \text{where } ? = \mathrm{dR}, \mathrm{rig}\,.$$

Again by functoriality, this defines an element of $H^{2r}_{?,c}(X)^\vee$.

1.4. ***Compatibility.*** Let $\mathscr{X}$ be a flat $\mathscr{V}$-scheme, and let $w \in z^q(\mathscr{X}/\mathscr{V}, 0)$ be a relative cycle. We can write $w = \sum_i n_i \mathscr{W}_i$, where $\mathscr{W}_i$ is an integral flat $\mathscr{V}$-scheme closed in $\mathscr{X}$ and of codimension q. Then w defines a cycle $w_K \in z^q(\mathscr{X}_K/K, 0)$ (resp. $w_k \in z^q(\mathscr{X}_k/k, 0)$); on the generic fiber, we get simply $w_K = \sum_i n_i (\mathscr{W}_i)_K$. However, on the special fiber, we must write the irreducible decomposition of each $(\mathscr{W}_i)^{\mathrm{red}}_k$, say $W_{i,1} \cup \cdots \cup W_{i,r_i}$, and then consider the multiplicities, that is,

$$w_k = \sum_i \Big(n_i \sum_j m_{i,j} W_{i,j}\Big),$$

where $m_{i,j} := \mathrm{length}(\mathscr{O}_{\mathscr{W}_k, W_{i,j}})$.

We can consider the de Rham and the rigid cycle classes of w, that is,

$$[w_K]_{\mathrm{dR}} \in H^{2r}_{\mathrm{dR},c}(|w_K|)^\vee \quad \text{and} \quad [w_k]_{\mathrm{rig}} \in H^{2r}_{\mathrm{rig},c}(|w_k|)^\vee.$$

The rigid homology groups $H_{2r}^{\mathrm{rig}}(|w_k|)$ are defined as the dual of $H^{2r}_{\mathrm{rig},c}(|w_k|)$; see [Petrequin 2003, Section 2]. We will prove that these cycle classes are compatible under (co)specialization and induce a well-defined syntomic cohomology class.

Theorem 1.4.1. *Let $\mathscr{X}$ be a flat $\mathscr{V}$-scheme of relative dimension d, and let $w \in z^q(\mathscr{X}/\mathscr{V}, 0)$ be a relative cycle of codimension q (and relative dimension $r := d-q$). Then $cosp([w_k]_{\mathrm{rig}}) = [w_K]_{\mathrm{dR}}$.*

Proof. First of all, consider the basic case: $w = \mathscr{W}$ is an integral closed subscheme of $\mathscr{X}$ smooth over $\mathscr{V}$. By [Baldassarri et al. 2004, proof of Theorem 6.9], we have a commutative diagram

$$\begin{tikzcd} H^{2r}_{\mathrm{rig},c}(\mathscr{W}_k/K) \arrow[rr, "cosp"] \arrow[dr, "{[w_k]_{\mathrm{rig}}}"'] & & H^{2r}_{\mathrm{dR},c}(\mathscr{W}_K) \arrow[dl, "{[w_K]_{\mathrm{dR}}}"] \\ & K & \end{tikzcd}$$

where the rigid (resp. de Rham) trace map $[w_k]_{\mathrm{rig}}$ (resp. $[w_K]_{\mathrm{dR}}$) is an isomorphism of K-vector spaces.

Given a general relative cycle w, we can reduce to the basic case by arguing as follows. First by linearity and the functoriality of the specialization map, we can restrict to the case $w = \mathscr{W}$ with $\mathscr{W}$ integral. Then the generic fiber $\mathscr{W}_K$ is integral, and by [Grothendieck 1967, Proposition 17.15.12], there exists a closed K-subscheme T such that $\mathscr{W}_K \setminus T$ is smooth over K. Let $\mathscr{T}$ be the flat extension of T (see Lemma 5.1.1); then $\mathscr{T}$ is of codimension at least 1 in $\mathscr{W}$, and the complement $\mathscr{W} \setminus \mathscr{T}$ is a flat $\mathscr{V}$-scheme of relative dimension r. Consider the long exact sequence

$$\cdots H^{2r-1}_{\mathrm{dR},c}(T) \to H^{2r}_{\mathrm{dR},c}(\mathscr{W}_K \setminus \mathscr{T}_K) \to H^{2r}_{\mathrm{dR},c}(\mathscr{W}_K) \to H^{2r}_{\mathrm{dR},c}(T) \cdots .$$

Note that here the first and last terms vanish for dimensional reasons. The same happens (*mutatis mutandis*) for the rigid cohomology of the special fiber.

Hence, from now on, we can assume that $\mathscr{W}$ is integral and that its generic fiber $\mathscr{W}_K$ is smooth. In this setting, we apply the reduced fiber theorem for schemes [Bosch et al. 1995, Theorem 2.1]; that is, there exist a finite field extension K'/K and a finite morphism $f : \mathscr{Y} \to \mathscr{W} \times_{\mathscr{V}} \mathscr{V}'$ of $\mathscr{V}'$-schemes such that

(i) $f_{K'} : \mathscr{Y}_{K'} \to \mathscr{W}_{K'}$ is an isomorphism and

(ii) $\mathscr{Y}/\mathscr{V}'$ is flat and has reduced geometric fibers.

Recall that the cospecialization map commutes with finite field extension, and the same holds for both the rigid and de Rham trace maps. By [Petrequin 2003, proof of Proposition 6.4], the rigid fundamental/cycle class is preserved by finite morphisms, that is,

$$[\mathscr{Y}_{k'}]_{\mathrm{rig}} \circ f_{k'}^* = [\mathscr{W}_{k'}]_{\mathrm{rig}} \quad \text{where} \quad f_{k'}^* : H^{2r}_{\mathrm{rig},c}(\mathscr{W}_{k'}) \to H^{2r}_{\mathrm{rig},c}(\mathscr{Y}_{k'}).$$

From the discussion above, there is no loss of generality in assuming that $\mathcal{W}$ has reduced geometric fibers and smooth generic fiber $\mathcal{W}_K$. Now let S be the singular locus of the special fiber $\mathcal{W}_k$. Again by [Grothendieck 1967, Proposition 17.15.12], the complement $\mathcal{W}_k \setminus S$ is an open and dense subscheme of $\mathcal{W}_k$, and it is smooth over k. The scheme S has codimension at least 1 in $\mathcal{W}_k$; hence, $H^{2r}_{\mathrm{rig},c}(\mathcal{W}_k \setminus S) \to H^{2r}_{\mathrm{rig},c}(\mathcal{W}_k)$ is an isomorphism. From this, it follows that we can assume $\mathcal{W}$ to be smooth over $\mathcal{V}$, where we know that the claim is true. $\square$

1.5. ***The smooth case.*** From now on, assume $\mathcal{X}$ to be smooth over $\mathcal{V}$. By the compatibility of (co)specialization with Poincaré duality [Baldassarri et al. 2004, Theorem 6.9], the (de Rham or rigid) cycle class map defines an element $\eta_{\mathrm{rig}}(w_k) \in H^{2q}_{\mathrm{rig},|w_k|}(\mathcal{X}_k/K)$ (resp. $\eta_{\mathrm{dR}}(w_K) \in H^{2q}_{\mathrm{dR},|w_K|}(\mathcal{X}_K)$) compatible with respect to the specialization morphism

$$sp(\eta_{\mathrm{dR}}(w_K)) = \eta_{\mathrm{rig}}(w_k). \tag{1}$$

Before stating the next corollary, we need to introduce some further notation. Let X be a smooth scheme over K (resp. over k); then the de Rham (resp. rigid) cycle class map factors through the Chow groups, inducing a map

$$\eta_{\mathrm{dR}} : CH^q(X/K) \to H^{2q}_{\mathrm{dR}}(X) \quad (\text{resp. } \eta_{\mathrm{rig}} : CH^q(X/k) \to H^{2q}_{\mathrm{rig}}(X/K)),$$

where, by abuse of notation, $\eta_{\mathrm{dR}}(W)$ (resp. $\eta_{\mathrm{rig}}(W)$) is viewed as an element of $H^{2q}_{\mathrm{dR}}(X)$ (resp. $H^{2q}_{\mathrm{rig}}(X/K)$) via the canonical map $H^{2q}_{\mathrm{dR},W}(X) \to H^{2q}_{\mathrm{dR}}(X)$ (resp. $H^{2q}_{\mathrm{rig},W}(X/K) \to H^{2q}_{\mathrm{rig}}(X/K)$) for any W integral subscheme of codimension q [Hartshorne 1975, Proposition 7.8.1, page 60; Petrequin 2003, Corollary 7.6].

Also, we recall that in [Berthelot et al. 1971, Exp. X, Appendix], a specialization map for the classical Chow ring is constructed:

$$sp_{CH} : CH^*(\mathcal{X}_K/K) \otimes \mathbb{Q} \to CH^*(\mathcal{X}_k/k) \otimes \mathbb{Q}.$$

Explicitly, the map is given as follows. Let $W \subset \mathcal{X}_K$ be an integral scheme of codimension q, and let $\mathcal{W}$ denote its Zariski closure in $\mathcal{X}$. Then the specialization of $sp_{CH}[W]$ is the class representing the subscheme $\mathcal{W}_k$.

Corollary 1.5.1. *With the notation above, the diagram*

$$\begin{tikzcd} CH^q(\mathcal{X}_K/K) \otimes \mathbb{Q} \arrow[r, "\eta_{\mathrm{dR}}"] \arrow[d, "sp_{CH}"'] & H^{2q}_{\mathrm{dR}}(\mathcal{X}_K) \arrow[d, "sp"] \\ CH^q(\mathcal{X}_k/k) \otimes \mathbb{Q} \arrow[r, "\eta_{\mathrm{rig}}"'] & H^{2q}_{\mathrm{rig}}(\mathcal{X}_k/K) \end{tikzcd}$$

commutes (we tensored each term by $\mathbb{Q}$ to guarantee the existence of sp_{CH}).

1.6. ***Syntomic cohomology.*** For any smooth $\mathscr{V}$-scheme $\mathscr{X}$, Besser [2000, Definition 6.1] defined the (rigid) syntomic cohomology groups $H^n_{\mathrm{syn}}(\mathscr{X}, i)$. We will be rather sketchy on the definitions because we will give another construction later. For such a cohomology, there is a long exact sequence

$$\longrightarrow H^n_{\mathrm{syn}}(\mathscr{X}, i) \longrightarrow H^n_{\mathrm{rig}}(\mathscr{X}_k/K_0) \oplus F^i H^n_{\mathrm{dR}}(\mathscr{X}_K) \xrightarrow{h} H^n_{\mathrm{rig}}(\mathscr{X}_k/K_0) \oplus H^n_{\mathrm{rig}}(\mathscr{X}_k/K) \xrightarrow{+} \tag{2}$$

where $h(x_0, x_{\mathrm{dR}}) = (\phi(x_0^\sigma) - p^i x_0, x_0 \otimes 1_K - sp(x_{\mathrm{dR}}))$.

Roughly, these groups are defined as $H^n_{\mathrm{syn}}(\mathscr{X}, i) := H^n(\mathbb{R}\Gamma_{Bes}(\mathscr{X}, i))$, where

$$\mathbb{R}\Gamma_{\mathrm{Bes}}(\mathscr{X}, i) := \mathrm{Cone}\big(\mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus Fil^i\, \mathbb{R}\Gamma_{\mathrm{dR}}(\mathscr{X}) \to \mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus \mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K)\big)[-1]$$

is a complex of abelian groups functorial in $\mathscr{X}$ and such that

$$H^n(\mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0)) = H^n_{\mathrm{rig}}(\mathscr{X}_k/K_0) \quad \text{and} \quad H^n(Fil^i\, \mathbb{R}\Gamma_{\mathrm{dR}}(\mathscr{X})) = F^i H^n_{\mathrm{dR}}(\mathscr{X}_K).$$

The functoriality of $\mathbb{R}\Gamma_{\mathrm{Bes}}(\,\cdot\,, i)$ allows us to give the following definition:

Definition 1.6.1. Let $\mathscr{X}$ be a smooth $\mathscr{V}$-scheme. Let $\mathfrak{X} \subset \mathscr{X}$ be a closed subscheme of $\mathscr{X}$. We define the syntomic complex with support in $\mathfrak{X}$ using the complexes defined by Besser as

$$\mathbb{R}\Gamma_{\mathrm{Bes},\mathfrak{X}}(\mathscr{X}, i) := \mathrm{Cone}\big(\mathbb{R}\Gamma_{\mathrm{Bes}}(\mathscr{X}, i) \to \mathbb{R}\Gamma_{\mathrm{Bes}}(\mathscr{X} \setminus \mathfrak{X}, i)\big)[-1].$$

This is a complex of abelian groups functorial with respect to cartesian squares. This fact will be used in the proof of Proposition 1.6.6. The cohomology of this complex is the *syntomic cohomology with support* in $\mathfrak{X}$ denoted by

$$H^n_{\mathrm{syn},\mathfrak{X}}(\mathscr{X}, i) := H^n(\mathbb{R}\Gamma_{\mathrm{Bes},\mathfrak{X}}(\mathscr{X}, i))$$

so that we get a long exact sequence

$$\cdots H^n_{\mathrm{syn},\mathfrak{X}}(\mathscr{X}, i) \to H^n_{\mathrm{syn}}(\mathscr{X}, i) \to H^n_{\mathrm{syn}}(\mathscr{X} \setminus \mathfrak{X}, i) \to \cdots .$$

Proposition 1.6.2 (syntomic cycle class). *Let $\mathscr{X}$ be a smooth $\mathscr{V}$-scheme, and let $w \in z^q(\mathscr{X}/\mathscr{V}, 0)$ be a relative cycle of codimension q. Then the canonical mapping*

$$\psi : H^{2q}_{\mathrm{syn},|w|}(\mathscr{X}, q) \to H^{2q}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K_0) \oplus F^q H^{2q}_{\mathrm{dR},|w_K|}(\mathscr{X}_K)$$

is injective, and there exists a unique element $\eta_{\mathrm{syn}}(w) \in H^{2q}_{\mathrm{syn},|w|}(\mathscr{X}, q)$ such that $\psi(\eta_{\mathrm{syn}}(w)) = (\eta_{\mathrm{rig}}(w_k), \eta_{\mathrm{dR}}(w_K))$.

Proof. By the definition of syntomic cohomology with support, there is a long exact sequence similar to (2):

$$\to H^{2q-1}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K_0)\oplus H^{2q-1}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K) \to H^{2q}_{\mathrm{syn},|w|}(X,q) \to$$
$$H^{2q}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K_0)\oplus F^q H^{2q}_{\mathrm{dR},|w_K|}(\mathscr{X}_K) \to H^{2q}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K_0)\oplus H^{2q}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K) \to .$$

The last term on the left vanishes because of weak purity in rigid cohomology [Berthelot 1997b, Corollary 5.7]. It follows that $H^{2q}_{\mathrm{syn},|w|}(\mathscr{X},q)$ consists of the pairs $(x,y)\in H^{2q}_{\mathrm{rig},|w_k|}(\mathscr{X}_k/K_0)\oplus F^q H^{2q}_{\mathrm{dR},|w_K|}(\mathscr{X}_K)$ such that $\phi(x^\sigma)=p^q x$ and $sp(y)=x\otimes 1_K$. By Hodge theory, we have $F^q H^{2q}_{\mathrm{dR},|w_K|}(\mathscr{X}_K)=H^{2q}_{\mathrm{dR},|w_K|}(\mathscr{X}_K)$. Moreover, the Frobenius acts on $[w_k]_{\mathrm{rig}}$ as multiplication by p^q [Petrequin 2003, Proposition 7.13]. Hence, in view of (1), we can easily conclude the proof. □

Lemma 1.6.3 (functoriality). *Let $f:\mathscr{X}'\to\mathscr{X}$ be a closed immersion of smooth $\mathscr{V}$-schemes, and let $w\in z^q(\mathscr{X}/\mathscr{V},0)$ be a relative cycle of codimension q. Assume that the preimage $f^{-1}w$ lies in $z^q(\mathscr{X}'/\mathscr{V},0)$; then $f^*\eta_{\mathrm{syn}}(w)=\eta_{\mathrm{syn}}(f^{-1}w)$.*

Proof. It is not restrictive to assume that $w=\mathscr{W}$ is an integral subscheme of $\mathscr{X}$ flat over $\mathscr{V}$.

We first show that it is sufficient to prove that $f^*\eta_{\mathrm{dR}}(w_K)=\eta_{\mathrm{dR}}(f^{-1}w_K)$. In fact, the syntomic cycle class can be viewed as an element $(\eta_{\mathrm{rig}}(w_k),\eta_{\mathrm{dR}}(w_K))$ in the direct sum of rigid and de Rham cohomology. Then note that $sp(\eta_{\mathrm{dR}}(w_K))=\eta_{\mathrm{rig}}(w_k)\otimes 1_K$ and that the specialization map is functorial.

To prove that $f^*\eta_{\mathrm{dR}}(\mathscr{W}_K)=\eta_{\mathrm{dR}}(f^{-1}\mathscr{W}_K)$, we first reduce by excision to the case where $\mathscr{W}_K$ is smooth over K (just remove from $\mathscr{W}_K$, $\mathscr{X}_K$ and $\mathscr{X}'_K$ the singular points of $\mathscr{W}_K$). In the same way, we can further assume $f^{-1}\mathscr{W}_K$ to be smooth over K. Now we can use the same proof as [Petrequin 2003, Proposition 7.1] to conclude. □

Lemma 1.6.4 (homotopy). *The (rigid) syntomic cohomology is homotopy invariant:*

$$H^n_{\mathrm{syn}}(\mathscr{X}\times\mathbb{A}^1_{\mathscr{V}},q)\cong H^n_{\mathrm{syn}}(\mathscr{X},q).$$

Proof. Just consider the long exact sequences of syntomic cohomology, and note that the de Rham cohomology (of smooth schemes) is homotopy invariant by [Hartshorne 1975, Proposition 7.9.1]. The same holds for rigid cohomology, for instance, using the Künneth formula [Berthelot 1997a]. □

Lemma 1.6.5 (weak purity). *Let $\mathscr{X}$ be a smooth $\mathscr{V}$-scheme. Let $\mathscr{Z}\subset\mathscr{X}$ be a closed subscheme of $\mathscr{X}$ of codimension q. Then*

$$H^n_{\mathrm{syn},\mathscr{Z}}(\mathscr{X},i)=0\quad \textit{for all } n<2q.$$

Proof. This follows directly from the long exact sequence of syntomic cohomology and the weak purity in de Rham and rigid cohomology [Hartshorne 1975, Section 7.2; Petrequin 2003, Section 1]. □

Proposition 1.6.6. *Let $\mathscr{X}$ be a smooth $\mathscr{V}$-scheme. The syntomic cycle class map induces a group homomorphism* $\mathrm{reg}_{\mathrm{syn}} : CH^i(\mathscr{X}/\mathscr{V}, 2i-n) \to H^n_{\mathrm{syn}}(\mathscr{X}, i)$.

Proof. The construction is analogous to that provided in [Bloch 1986]. Consider the cohomological double complex $\mathbb{R}\Gamma_{\mathrm{Bes}}(\Delta^{-n}_{\mathscr{X}}, q)^m$ nonzero for $m \geq 0$ and $n \leq 0$; the differential in n is induced by $\partial_i^{(-n)}$ in the usual way. Similarly define the double complex

$$\mathbb{R}\Gamma_{\mathrm{Bes,supp}}(\Delta^{-n}_{\mathscr{X}}, q)^m := \operatorname*{colim}_{w \in z^q(\mathscr{X}, -n)} \mathbb{R}\Gamma_{\mathrm{Bes},|w|}(\Delta^{-n}_{\mathscr{X}}, q)^m.$$

For technical reasons, we truncate these complexes (nontrivially):

$$A^{n,m}_{?} = \tau_{n \geq -N} \mathbb{R}\Gamma_{\mathrm{Bes},?}(\Delta^{-n}_{\mathscr{X}}, q)^m, \quad \text{where } ? = \varnothing, \text{supp},$$

for N even and $N \gg 0$.

Consider the spectral sequence

$$E_1^{n,m} := H^m(A^{\bullet,n}) \Rightarrow H^{n+m}(sA^{\bullet,*}),$$

where s denotes the associated simple complex of a double complex. By homotopy invariance (Lemma 1.6.4), $E_1^{n,m} := H^m_{\mathrm{syn}}(\Delta^{-n}_{\mathscr{X}}, q)$ is isomorphic to $H^m_{\mathrm{syn}}(\mathscr{X}, q)$ for $-N \leq n \leq 0$ and $m \geq 0$; otherwise, it is 0. Moreover, $d_1^{n,m} = 0$ except for n even, $-N \leq n < 0$ and $m \geq 0$, in which case $d_1^{n,m} = \mathrm{id}$. This gives an isomorphism $H^i(sA^{n,m}) \cong H^i_{\mathrm{syn}}(\mathscr{X}, q)$.

In the spectral sequence

$$E^{n,m}_{1,\mathrm{supp}} := H^m(A^{\bullet,n}_{\mathrm{supp}}) \Rightarrow H^{n+m}(sA^{\bullet,*}_{\mathrm{supp}}),$$

we have

$$E^{n,m}_{1,\mathrm{supp}} = \operatorname*{colim}_{w \in z^q(X,-n)} H^m_{\mathrm{syn},|w|}(\Delta^{-n}_{\mathscr{X}}, q) \quad \text{for } -N \leq n \leq 0 \text{ and } m \geq 0$$

and $E^{n,m}_{1,\mathrm{supp}} = 0$ otherwise. Applying Lemma 1.6.3 to the face morphisms, it is easy to prove that the syntomic cycle class induces a natural map of complexes $\Gamma^\bullet_{\mathscr{X}/\mathscr{V}}(q) \to E^{\bullet-2q,2q}_{1,\mathrm{supp}}$, and hence, for all i, a map $CH^q(\mathscr{X}/\mathscr{V}, 2q-i) \to E^{i-2q,2q}_{2,\mathrm{supp}}$. The groups $E^{n,m}_{r,\mathrm{supp}}$ are zero for $m < 2q$ and $r \geq 1$ due to weak purity. Hence, there are natural maps $E^{i-2q,2q}_{2,\mathrm{supp}} \to E^{i-2q,2q}_{\infty,\mathrm{supp}} \to H^i(sA^{\bullet,*}_{\mathrm{supp}})$. By construction, there is a map $H^i(sA^{\bullet,*}_{\mathrm{supp}}) \to H^i(sA^{\bullet,*})$. Composing all these maps, we obtain the expected map $CH^q(\mathscr{X}, 2q-i) \to H^i_{\mathrm{syn}}(\mathscr{X}, q)$. □

Corollary 1.6.7. *With the notation above, there is a commutative diagram*

$$\begin{tikzcd} CH^q(\mathcal{X}/\mathcal{V})\otimes\mathbb{Q} \arrow[r, "\gamma\otimes 1_{\mathbb{Q}}"] \arrow[d, "\mathrm{reg}_{\mathrm{syn}}"'] & CH^q(\mathcal{X}_K/K)\otimes\mathbb{Q} \arrow[r, "sp_{CH}"] \arrow[d, "\eta_{\mathrm{dR}}"'] & CH^q(\mathcal{X}_k/k)\otimes\mathbb{Q} \arrow[d, "\eta_{\mathrm{rig}}"'] \\ H^{2q}_{\mathrm{syn}}(\mathcal{X},q) \arrow[r, "\pi"'] & F^q H^{2q}_{\mathrm{dR}}(\mathcal{X}_K) \arrow[r, "sp"'] & H^{2q}_{\mathrm{rig}}(\mathcal{X}_k/K), \end{tikzcd}$$

where π is the composition

$$H^n_{\mathrm{syn}}(\mathcal{X},i)\longrightarrow H^n_{\mathrm{rig}}(\mathcal{X}_k/K_0)\oplus F^iH^n_{\mathrm{dR}}(\mathcal{X}_K)\xrightarrow{\mathrm{pr}_2} F^iH^n_{\mathrm{dR}}(\mathcal{X}_K)$$

and γ is the map described in Remark 1.2.2.

Proof. Just note that in this case, $\mathrm{reg}_{\mathrm{syn}}$ is the map induced by the syntomic cycle class in the usual way. □

Remark 1.6.8. Via Chern classes, Besser [2000, Theorem 7.5] obtained a regulator $c_i^{2i-n}: K_{2i-n}(X)\to H^n_{\mathrm{syn}}(\mathcal{X},i)$. At present, we cannot compare it with the regulator of Proposition 1.6.6. This is because we don't know how to relate K-theory with the higher Chow groups we have defined. Nevertheless, we expect that there exists a map $CH^i(\mathcal{X}/\mathcal{V},2i-n)_{\mathbb{Q}}\to K_{2i-n}(X)_{\mathbb{Q}}$. This issue will be treated in a future work.

2. *p*-adic Hodge complexes

Having defined a regulator map with values in the syntomic cohomology, it is tempting to check (some of) the Bloch–Ogus axioms for this theory. We address this problem by viewing the syntomic cohomology as an absolute cohomology theory.

Thus, in this section, we define a triangulated category of *p-adic Hodge complexes* similar to that of [Bannai 2002]. See also [Beĭlinson 1986; Huber 1995; Levine 1998, Chapter V, Section 2.3].

The syntomic cohomology will be computed by Hom groups in this category.

Definition 2.0.9 (see [Bannai 2002, Section 2]). Let $C^b_{\mathrm{rig}}(K)$ be the category of pairs $(M^\bullet,\phi)$, where

(i) $M^\bullet = M_0^\bullet\otimes_{K_0}K$ and $M_0^\bullet$ is a complex in $C^b(K_0)$;

(ii) (Frobenius structure) if $(M_0^\bullet)^\sigma := M_0^\bullet\otimes_\sigma K_0$, then $\phi:(M_0^\bullet)^\sigma\to M_0^\bullet$ is a K_0-linear morphism.

The morphisms in this category are morphisms in $C^b(K_0)$ compatible with respect to the Frobenius structure. In this way, we get an abelian category.

Let Filt_K be the category of K-vector spaces with a descending, exhaustive and separated filtration. We write $C^b_{\mathrm{dR}}(K)=C^b(\mathrm{Filt}_K)$, and we write the objects of this category as pairs $(M^\bullet, F)$, where

(i) $M^\bullet$ is a complex in $C^b(K)$ and

(ii) (Hodge filtration) F is a (separated and exhaustive) filtration on $M^\bullet$.

Remark 2.0.10 (strictness). We review some technical facts about filtered categories. For a full discussion, see [Huber 1995, Sections 2 and 3].

The category Filt_K (and also $C^b_{\mathrm{dR}}(K)$) is additive but not abelian. It is an exact category when one takes for short exact sequences those that are exact as sequences of K-vector spaces and are such that the morphisms are strict with respect to the filtrations; recall that a morphism $f : (M, F) \to (N, F)$ is *strict* if $f(F^i M) = F^i N \cap \mathrm{Im}(f)$.

An object $(M^\bullet, F) \in C^b_{\mathrm{dR}}(K)$ is a *strict complex* if its differentials are strict as morphism of filtered vector spaces. Strict complexes can be characterized also by the fact that the canonical spectral sequence $H^q(F^p) \Rightarrow H^{p+q}(M^\bullet)$ degenerates at E_1.

One can define canonical truncation functors on $C^b_{\mathrm{dR}}(K)$: For $M^\bullet \in C^b_{\mathrm{dR}}(K)$, let

$$\tau_{\le n}(M^\bullet, F)^i := \begin{cases} M^i & \text{if } i < n, \\ \mathrm{Ker}(d^n) & \text{if } i = n, \\ 0 & \text{if } i > n, \end{cases} \qquad \tau_{\ge n}(M^\bullet, F)^i := \begin{cases} 0 & \text{if } i < n-1, \\ \mathrm{Coim}(d^n) & \text{if } i = n-1, \\ M^i & \text{if } i \ge n. \end{cases}$$

It is important to note that the naive cohomology object $\tau_{\le n}\tau_{\ge n}(M^\bullet, F)$ of a strict complex $(M^\bullet, F)$ agrees with the cohomology $H^n(M^\bullet)$ of the complex of K-vector spaces underlying $(M^\bullet, F)$ [Huber 1995, Proposition 2.1.3 and Section 3].

Definition 2.0.11 (see [Bannai 2002, Definition 2.2]). From the discussion above, there is an exact functor $\Phi_{\mathrm{rig}} : C^b_{\mathrm{rig}}(K) \to C^b(K)$ (resp. $\Phi_{\mathrm{dR}} : C^b_{\mathrm{dR}}(K) \to C^b(K)$) induced by $(M^\bullet, \phi) \mapsto M^\bullet$ (resp. $(M^\bullet, F) \mapsto M^\bullet$). We define the category *pHC* of *p-adic Hodge complexes* whose objects are systems $M = (M^\bullet_{\mathrm{rig}}, M^\bullet_{\mathrm{dR}}, M^\bullet_K, c, s)$, where

(i) $(M^\bullet_{\mathrm{rig}}, \phi)$ is an object of $C^b_{\mathrm{rig}}(K)$ and $H^*(M^\bullet_{\mathrm{rig}})$ is finitely generated over K,

(ii) $(M^\bullet_{\mathrm{dR}}, F)$ is an object of $C^b_{\mathrm{dR}}(K)$ and $H^*(M^\bullet_{\mathrm{dR}})$ is finitely generated over K and

(iii) $M^\bullet_K$ is an object of $C^b(K)$ and $c : M^\bullet_{\mathrm{rig}} \to M^\bullet_K$ (resp. $s : M^\bullet_{\mathrm{dR}} \to M^\bullet_K$) is a morphism in $C^b(K)$. Hence, c and s give a diagram in C^b_K

$$M^\bullet_{\mathrm{rig}} \xrightarrow{c} M^\bullet_K \xleftarrow{s} M^\bullet_{\mathrm{dR}}.$$

A morphism in *pHC* is given by a system $f := (f_{\mathrm{rig}}, f_{\mathrm{dR}}, f_K)$, where $f_? : M^\bullet_? \to N^\bullet_?$ is a morphism in $C^b_{\mathrm{rig}}(K)$, $C^b_{\mathrm{dR}}(K)$ or $C^b(K)$ for $? = \mathrm{rig}$, dR or K, respectively, and such that they are compatible with respect to the diagram in (iii) above.

2.1. ***Derived version.*** A *homotopy* in *pHC* is a system of homotopies h_i compatible with the comparison maps c and s. We define the category *pHK* to be the category *pHC* modulo morphisms homotopic to 0. We say that a morphism $f = (f_{\mathrm{rig}}, f_{\mathrm{dR}}, f_K)$ in *pHC* (or *pHK*) is a *quasi-isomorphism* if $f_?$ is a quasi-isomorphism for $? = \mathrm{rig}$ (or K) and f_{dR} is a filtered quasi-isomorphism, that is, $\mathrm{gr}_F(f_{\mathrm{dR}})$ is a quasi-isomorphism. Finally, we say that $M \in pHC$ (or *pHK*) is *acyclic* if $M_? = 0$ is acyclic for any $? = \mathrm{rig}, \mathrm{dR}, K$.

Lemma 2.1.1. (i) *The category pHK is a triangulated category.*

(ii) *The localization of pHK with respect to the class of quasi-isomorphisms exists. This category*, *denoted pHD*, *is a triangulated category.*

(iii) *On the category pHD*, *it is possible to define a nondegenerate t-structure* (*resp. truncation functors*) *compatible with the standard t-structure* (*resp. truncation functors*) *defined on* $C^b(K)$, $C^b_{\mathrm{rig}}(K)$ *and* $C^b_{\mathrm{dR}}(K)$.

Proof. The proof is the same as [Bannai 2002, Proposition 2.6]. See also [Huber 1995, Section 2] for a survey on how to derive exact categories. □

Remark 2.1.2. In the terminology of Huber [1995, Section 4], the category *pHC* is the *glued exact category* of C^b_{rig} and C^b_{dR} via C^b_K. The foregoing definition is inspired by Bannai [2002], who constructs a *rigid* glued exact category C^b_{MF}, that is, the comparison maps are all quasi-isomorphisms. For our purposes, we cannot impose this strong assumption. This is motivated by the fact that the (co)specialization is not an isomorphism for a general smooth $\mathcal{V}$-scheme.

2.2. ***Derived*** **Hom.** Let $M^\bullet$ and $M'^\bullet$ be two objects of *pHC*. Consider the diagram $\mathcal{H}(M^\bullet, M'^\bullet)$ (of complexes of abelian groups)

$$\begin{tikzcd}
\mathrm{Hom}^\bullet_{K_0}((M^\bullet_0)^\sigma, M'^\bullet_0) & & \mathrm{Hom}^\bullet_K(M^\bullet_{\mathrm{rig}}, M'^\bullet_K) & & \mathrm{Hom}^\bullet_K(M^\bullet_{\mathrm{dR}}, M'^\bullet_K) \\
\mathrm{Hom}^\bullet_{K_0}(M^\bullet_0, M'^\bullet_0) \arrow[u, "h_0"] \arrow[urr, "h_1"] & & \mathrm{Hom}^\bullet_K(M^\bullet_K, M'^\bullet_K) \arrow[u, "h_2"] \arrow[urr, "h_3"] & & \mathrm{Hom}^{\bullet,F}_K(M^\bullet_{\mathrm{dR}}, M'^\bullet_{\mathrm{dR}}) \arrow[u, "h_4"]
\end{tikzcd}$$

where $h_0(x_0) = x_0 \circ \phi - \phi' \circ x_0^\sigma$; $h_1(x_0) = c' \circ (x_0 \otimes \mathrm{id}_K)$; $h_2(x_K) = x_K \circ c$; $h_3(x_K) = x_K \circ s$ and $h_4(x_{\mathrm{dR}}) = s' \circ x_{\mathrm{dR}}$; $\mathrm{Hom}^{\bullet,F}_K(M^\bullet_{\mathrm{dR}}, M'^\bullet_{\mathrm{dR}})$ is the complex of morphisms compatible with respect to the filtrations. Then define the two complexes of abelian groups $\Gamma_0(M^\bullet, M'^\bullet) :=$ (direct sum of the bottom row) and $\Gamma_1(M^\bullet, M'^\bullet) :=$ (direct sum of the top row). Finally, consider the cone

$$\Gamma(M^\bullet, M'^\bullet) := \mathrm{Cone}\big(\psi_{M^\bullet, M'^\bullet} : \Gamma_0(M^\bullet, M'^\bullet) \to \Gamma_1(M^\bullet, M'^\bullet)\big)[-1],$$

where

$$\psi_{M^\bullet, M'^\bullet} : (x_0, x_K, x_{\mathrm{dR}}) \mapsto (-h_0(x_0), h_1(x_0) - h_2(x_K), h_3(x_K) - h_4(x_{\mathrm{dR}})).$$

Remark 2.2.1. (i) Let $\mathbb{K}(-n)$ be the Tate twisted p-adic Hodge complex: i.e., $\mathbb{K}(-n)_{\mathrm{rig}}$ (resp. $\mathbb{K}(-n)_{\mathrm{dR}}$, $\mathbb{K}(-n)$) is equal to K concentrated in degree 0; the Frobenius is $\phi(\lambda) := p^n\sigma(\lambda)$; the filtration is $F^i = K$ for $i \leq n$ and 0 otherwise.

(ii) Given two p-adic Hodge complexes $M^\bullet$ and $M'^\bullet$, we define their tensor product $M^\bullet \otimes M'^\bullet$ component-wise, that is, $(M^\bullet_{\mathrm{rig}} \otimes M'^\bullet_{\mathrm{rig}}, M^\bullet_{\mathrm{dR}} \otimes M'^\bullet_{\mathrm{dR}}, M^\bullet_K \otimes M'^\bullet_K, c \otimes c', s \otimes s')$. The complex $M^\bullet \otimes \mathbb{K}(n)$ is denoted by $M^\bullet(n)$.

(iii) The complex $\Gamma(\mathbb{K}, M^\bullet(n))$ is quasi-isomorphic to

$$\mathrm{Cone}(M_0^\bullet \oplus F^n M^\bullet_{\mathrm{dR}} \xrightarrow{\eta} M_0^\bullet \oplus M_K^\bullet)[-1],$$
$$\eta(x_0, x_{\mathrm{dR}}) = (p^{-n}\phi(x_0^\sigma) - x_0, c(x_0 \otimes \mathrm{id}_K) - s(x_{\mathrm{dR}})),$$

where $x_0 \in M_0^\bullet$ and $x_{\mathrm{dR}} \in F^n M^\bullet_{\mathrm{dR}}$.

If c is a quasi-isomorphism, letting sp denote the composition of

$$H^q(F^n M^\bullet_{\mathrm{dR}}) \xrightarrow{s^*} H^q(M_K^\bullet) \underset{\cong}{\xleftarrow{c^*}} H^q(M^\bullet_{\mathrm{rig}}),$$

we obtain a long exact sequence

$$\to H^q(\Gamma(\mathbb{K}, M^\bullet(n))) \to H^q(M_0^\bullet) \oplus H^q(F^n M^\bullet_{\mathrm{dR}}) \xrightarrow{\eta'} H^q(M_0^\bullet) \oplus H^q(M^\bullet_{\mathrm{rig}}) \to,$$

where $\eta'(x_0, x_{\mathrm{dR}}) = (p^{-n}\phi(x_0^\sigma) - x_0, x_0 \otimes 1_K - sp(x_{\mathrm{dR}}))$.

If s is a quasi-isomorphism, letting $cosp$ denote the composition of

$$H^q(M^\bullet_{\mathrm{rig}}) \xrightarrow{c^*} H^q(M_K^\bullet) \underset{\cong}{\xleftarrow{s^*}} H^q(M^\bullet_{\mathrm{dR}}),$$

we obtain a long exact sequence

$$\to H^q(\Gamma(\mathbb{K}, M^\bullet(n))) \to H^q(M_0^\bullet) \oplus H^q(F^n M^\bullet_{\mathrm{dR}}) \xrightarrow{\eta''} H^q(M_0^\bullet) \oplus H^q(M^\bullet_{\mathrm{dR}}) \to,$$

where $\eta''(x_0, x_{\mathrm{dR}}) = (p^{-n}\phi(x_0^\sigma) - x_0, cosp(x_0 \otimes 1_K) - x_{\mathrm{dR}})$.

Proposition 2.2.2 (extension formula). *With the notation above, there is a canonical morphism of abelian groups*

$$\mathrm{Hom}_{pHD}(M^\bullet, M'^\bullet[n]) \cong H^n(\Gamma(M^\bullet, M'^\bullet)).$$

In particular, if $M^\bullet = M$ and $M'^\bullet = M'$ are concentrated in degree 0, *then* $H^n(\Gamma(M, M')) = 0$ *for $n \geq 2$ and $n < 0$.*

Proof. By the octahedron axiom, we have the following triangle in $D^b(Ab)$:

$$\mathrm{Ker}\,\psi_{M^\bullet, M'^\bullet} \to \Gamma(M^\bullet, M'^\bullet) \to \mathrm{Coker}\,\psi_{M^\bullet, M'^\bullet}[-1] \xrightarrow{+}.$$

Its cohomological long exact sequence is

$$\xrightarrow{\partial} H^n(\mathrm{Ker}\,\psi_{M^\bullet, M'}) \to H^n(\Gamma(M, M'^\bullet)) \to H^n(\mathrm{Coker}\,\psi_{M^\bullet, M'^\bullet}[-1]) \xrightarrow{\partial}.$$

Note that by construction, $H^n(\operatorname{Ker}\psi_{M^\bullet,M'^\bullet}) = \operatorname{Hom}_{pHK}(M^\bullet, M'^\bullet[n])$. Also, we have

$$\operatorname{Hom}_{pHD}(M^\bullet, M'^\bullet[n]) = \operatorname*{colim}_I \operatorname{Hom}_{pHK}(M^\bullet, M''^\bullet[n]),\ I = \{\text{quis } g : M'^\bullet \to M''^\bullet\}.$$

Thus, the result is proved if we show that

(i) $H^n(\Gamma(M^\bullet, M'^\bullet)) \cong H^n(\Gamma(M^\bullet, M''^\bullet))$ holds given any $g : M'^\bullet \to M''^\bullet$ quasi-isomorphism and

(ii) $\operatorname{colim}_I H^n(\operatorname{Coker}\psi_{M^\bullet,M''^\bullet}[-1]) = 0$.

The first claim follows from the exactness of $\Gamma(M^\bullet, \cdot)$, and the second is proved in [Beĭlinson 1986, 1.7, 1.8] (and with more details in [Huber 1995, Lemma 4.2.8; Bannai 2002, Lemma 2.15]) with the assumption that all the gluing maps are quasi-isomorphisms, but this hypothesis is not necessary. □

Lemma 2.2.3 (tensor product). *Let $M^\bullet$, $M'^\bullet$ and $I^\bullet$ be p-adic Hodge complexes. For any $\alpha \in K$, there is a morphism of complexes*

$$\cup_\alpha : \Gamma(I^\bullet, M^\bullet) \otimes \Gamma(I^\bullet, M'^\bullet) \to \Gamma(I^\bullet, M^\bullet \otimes M'^\bullet).$$

All such $\cup_\alpha$ are equivalent up to homotopy.

Proof. See [Beĭlinson 1986, 1.11]. □

Remark 2.2.4 (enlarging the diagram). We recall some results from [Levine 1998, Chapter V, 2.3.3]. Let $M_1^\bullet \xrightarrow{f} M_2^\bullet \xleftarrow{g} M_3^\bullet$ (resp. $M_1^\bullet \xleftarrow{f} M_2^\bullet \xrightarrow{g} M_3^\bullet$) be a diagram of complexes in $C^b(K)$. Let $P^\bullet = \operatorname{Cone}(f - g : M_1^\bullet \oplus M_3^\bullet \to M_2^\bullet)[-1]$ be the *quasipullback* complex (resp. $Q^\bullet = \operatorname{Cone}((f, -g) : M_2^\bullet \to M_1^\bullet \oplus M_3^\bullet)$ be the *quasipushout*). Assume that f is a quasi-isomorphism. Then the diagrams

$$\begin{tikzcd} P^\bullet \arrow[r, "h"] \arrow[d] & M_3^\bullet \arrow[d, "g"] \\ M_1^\bullet \arrow[r, "f"'] & M_2^\bullet \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} M_2^\bullet \arrow[r, "f"] \arrow[d, "g"'] & M_1^\bullet \arrow[d] \\ M_3^\bullet \arrow[r, "k"'] & P^\bullet \end{tikzcd}$$

are commutative up to homotopy and are such that h and k are quasi-isomorphisms.

Now let pHC' be a category of systems $(M_{\mathrm{rig}}^\bullet, M_{\mathrm{dR}}^\bullet, M_1^\bullet, M_2^\bullet, M_3^\bullet, c, s, f, g)$ similar to Definition 2.0.11 and such that there is a diagram

$$M_{\mathrm{rig}}^\bullet \xrightarrow{c} M_1^\bullet \xleftarrow{f} M_2^\bullet \xrightarrow{g} M_3^\bullet \xleftarrow{s} M_{\mathrm{dR}}^\bullet.$$

Then the quasipushout induces a functor from the category pHC' to the category pHC. This functor is compatible with tensor product after passing to the categories pHK' and pHK.

3. Godement resolution

Here we recall some facts about the *generalized* Godement resolution, also called the bar-resolution. We refer to [Ivorra 2005]; see also [Weibel 1994, Section 8.6].

3.1. ***General construction.*** Let $u : P \to X$ be a morphism of Grothendieck topologies so that $P^\sim$ (resp. $X^\sim$) is the category of abelian sheaves on P (resp. X). Then we have a pair of adjoint functors (u^*, u_*), where $u^* : X^\sim \to P^\sim$ and $u_* : P^\sim \to X^\sim$. For any object $\mathscr{F}$ of $X^\sim$, we can define a cosimplicial object $B^*(\mathscr{F}) : \Delta \to X^\sim$ in the following way. First let $\eta : \mathrm{id}_{X^\sim} \to u_*u^*$ and $\epsilon : u^*u_* \to \mathrm{id}_{P^\sim}$ be the natural transformations induced by adjunction.

Endow $B^n(\mathscr{F}) := (u_*u^*)^{n+1}(\mathscr{F})$ with codegeneracy maps

$$\sigma_i^n := (u_*u^*)^i u_*\epsilon u^*(u_*u^*)^{n-1-i} : B^n(\mathscr{F}) \to B^{n-1}(\mathscr{F}) \quad \text{for } i = 0, \dots, n-1$$

and cofaces

$$\delta_i^{n-1} := (u_*u^*)^i \eta (u_*u^*)^{n-i} : B^{n-1}(\mathscr{F}) \to B^n(\mathscr{F}) \quad \text{for } i = 0, \dots, n.$$

Lemma 3.1.1. *With the notation above, let $sB^*(\mathscr{F})$ be the associated complex of objects of $X^\sim$. Then there is a canonical map $b_{\mathscr{F}} : \mathscr{F} \to sB^*(\mathscr{F})$ such that $u^*(b_{\mathscr{F}})$ is a quasi-isomorphism. Moreover, if u^* is exact and conservative, then $b_{\mathscr{F}}$ is a quasi-isomorphism.*

Proof. See [Ivorra 2005, Chapter III, Lemma 3.4.1]. □

Thus, for any sheaf $\mathscr{F} \in X^\sim$ (or complex of sheaves), we can define a functorial map $b_{\mathscr{F}} : \mathscr{F} \to sB^*(\mathscr{F})$ with $sB^n(\mathscr{F}) := (u_*u^*)^{n+1}\mathscr{F}$. We will denote this complex of sheaves $\mathrm{Gd}_P(\mathscr{F})$. In the case u^* is exact and conservative, $\mathrm{Gd}_P(\mathscr{F})$ is a canonical resolution of $\mathscr{F}$. If $\mathscr{F}^\bullet$ is a complex of sheaves on X, $\mathrm{Gd}_P(\mathscr{F}^\bullet)$ denotes the simple complex $s(\mathrm{Gd}_P(\mathscr{F}^i)^j)$. Often, we will need to iterate this process, and we will write $\mathrm{Gd}_P^2(\mathscr{F}) := \mathrm{Gd}_P(\mathrm{Gd}_P(\mathscr{F}))$.

Now suppose there is a commutative diagram of sites

$$\begin{tikzcd} P \arrow[r, "g"] \arrow[d, "u"'] & Q \arrow[d, "v"] \\ X \arrow[r, "f"'] & Y \end{tikzcd}$$

and a morphism of sheaves $a : \mathscr{G} \to f_*\mathscr{F}$, where $\mathscr{F}$ (resp. $\mathscr{G}$) is a sheaf on X (resp. Y).

Lemma 3.1.2. *There is a canonical map* $\mathrm{Gd}_Q(\mathscr{G}) \to f_* \mathrm{Gd}_P(\mathscr{F})$ *compatible with* $b_{\mathscr{F}}$ *and* $b_{\mathscr{G}}$.

Proof. We need only show that there is a canonical map $v_*v^*\mathscr{G} \to f_*u_*u^*\mathscr{F}$ lifting a. First consider the composition $\mathscr{G} \to f_*\mathscr{F} \to f_*u_*u^*\mathscr{F}$. Then we get a map $\mathcal{G} \to v_* g_* u^* \mathcal{F}$ because $v_* g_* = f_* u_*$. By adjunction, this gives $v^* \mathcal{G} \to g_* u^* \mathcal{F}$. Then we apply v_* and use $v_* g_* = f_* u_*$ to obtain the desired map. □

Remark 3.1.3 (tensor product). The Godement resolution is compatible with tensor products; that is, for any pair of sheaves $\mathcal{F}$ and $\mathcal{G}$ on X, there is a canonical quasi-isomorphism $\mathrm{Gd}_P(\mathcal{F}) \otimes \mathrm{Gd}_P(\mathcal{G}) \to \mathrm{Gd}_P(\mathcal{F} \otimes \mathcal{G})$. The same holds for complexes that are bounded below [Friedlander and Suslin 2002, Appendix A].

3.2. *Points of sites/topoi.*

Definition 3.2.1 (see [Artin et al. 1972a, Example IV, Section 6]). Let X be a site and $Sh(X)$ be the associated topos of sheaves of sets. A *point* of X is a morphism of topoi $\pi : Set \to Sh(X)$, that is, a pair of adjoint functors (π^*, π_*) such that π^* is left exact.

Example 3.2.2. Let X be a scheme. Then any point x of the topological space underlying X gives a point π_x of the Zariski site of X. We call them Zariski points.

Now let x be a geometric point of X; then it induces a point π_x for the étale site of X. We call them étale points of X.

Let $\mathcal{F}$ be a Zariski (resp. étale) sheaf X and P be the set of Zariski (resp. étale) points of X. Then the functor $\mathcal{F} \mapsto \bigsqcup_{\pi \in P} \mathcal{F}_\pi := \pi^* \mathcal{F}$ is exact and conservative. In other words, the Zariski (resp. étale) site of X has enough points.

Example 3.2.3 (points on rigid analytic spaces [van der Put and Schneider 1995]). Let $\mathcal{X}$ be a rigid analytic space over K. We recall that a *filter* f on $\mathcal{X}$ is a collection $(\mathcal{U}_\alpha)_\alpha$ of admissible open subsets of X satisfying

(i) $\mathcal{X} \in f$ and $\varnothing \notin f$,

(ii) if $\mathcal{U}_\alpha, \mathcal{U}_\beta \in f$, then $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \in f$ and

(iii) if $\mathcal{U}_\alpha \in f$ and the admissible open $\mathcal{V}$ contains $\mathcal{U}_\alpha$, then $\mathcal{V} \in f$.

A *prime filter* on $\mathcal{X}$ is a filter p satisfying moreover

(iv) if $\mathcal{U} \in p$ and $\mathcal{U} = \bigcup_{i \in I} \mathcal{U}'_i$ is an admissible covering of $\mathcal{U}$, then $\mathcal{U}'_{i_0} \in p$ for some $i_0 \in I$.

Let $P(\mathcal{X})$ be the set of all prime filters of $\mathcal{X}$. The filters on $\mathcal{X}$ are ordered with respect to inclusion. We can give to $P(\mathcal{X})$ a topology and define a morphism of sites $\sigma : P(\mathcal{X}) \to \mathcal{X}$. Also, we let $Pt(\mathcal{X})$ denote the set of prime filters with the discrete topology. Let $i : Pt(\mathcal{X}) \to P(X)$ be the canonical inclusion and $\xi = \sigma \circ i$.

Remark 3.2.4. Let $p = (\mathcal{U}_\alpha)_\alpha$ be a prime filter on $\mathcal{X}$ as above. Then p is a point of the site $\mathcal{X}$ (see Definition 3.2.1). Using the construction of the continuous map σ of [van der Put and Schneider 1995], we get that the morphism of topoi $\pi : Set \to Sh(\mathcal{X})$, associated to p, is defined in the following way. For any sheaf (of sets) $\mathcal{F}$ on $\mathcal{X}$, let $\pi^*(\mathcal{F}) = \operatorname{colim}_\alpha \mathcal{F}(\mathcal{U}_\alpha)$; for any set $\mathcal{S}$ and $\mathcal{V}$ admissible open in $\mathcal{X}$, let

$$\pi_*\mathcal{S}(\mathcal{V}) = \begin{cases} \mathcal{S} & \text{if } \mathcal{V} = \mathcal{U}_\alpha \text{ for some } \alpha, \\ 0 & \text{otherwise,} \end{cases}$$

where 0 denotes the final object in the category *Set*. In fact, with the notation above, we easily get the adjunction

$$\operatorname{Hom}_{Set}(\pi^*(\mathcal{F}), \mathcal{S}) = \lim_{\alpha} \operatorname{Hom}_{Set}(\mathcal{F}(\mathcal{U}_\alpha), \mathcal{S}) = \operatorname{Hom}_{Sh(\mathcal{X})}(\mathcal{F}, \pi_*\mathcal{S}).$$

Lemma 3.2.5. *With the notation above, the functor $\xi^{-1} : Sh(\mathcal{X}) \to Sh(Pt(\mathcal{X}))$ is exact and conservative. In other words, for any $p \in Pt(\mathcal{X})$, $\mathcal{F} = 0$ if all $\mathcal{F}_p = 0$ and the functors $Sh(\mathcal{X}) \ni \mathcal{F} \mapsto \mathcal{F}_p$ are exact.*

Proof. See [van der Put and Schneider 1995, Section 4] after the proof of Theorem 1. □

4. Rigid and de Rham complexes

We begin this section by recalling the construction of the rigid complexes of [Besser 2000, Section 4]. Instead of the techniques of [Artin et al. 1972b, Exposé V.bis], we use the machinery of generalized Godement resolution as developed in Section 3. This alternative approach was also mentioned by Besser in the introduction of his paper. We then recall the construction of the de Rham complexes.

4.1. ***Rigid complexes.*** We define a *rigid triple* to be a system $(X, \overline{X}, \mathsf{P})$, where X is an algebraic k-scheme, $j : X \to \overline{X}$ is an open embedding into a proper k-scheme and $\overline{X} \to \mathsf{P}$ is a closed embedding into a p-adic formal $\mathcal{V}$-scheme P that is smooth in a neighborhood of X.

Definition 4.1.1 [Besser 2000, 4.2, 4.4]. Let $(X, \overline{X}, \mathsf{P})$ and $(Y, \overline{Y}, \mathsf{Q})$ be two rigid triples, and let $f : X \to Y$ be a morphism of k-schemes. Let $\mathcal{U} \subset]\overline{X}[_{\mathsf{P}}$ be a strict neighborhood of $]X[_{\mathsf{P}}$ and $F : \mathcal{U} \to \mathsf{Q}_K$ be a morphism of K-rigid spaces. We say that F is *compatible* with f if it induces the following commutative diagram:

$$\begin{array}{ccc}]X[_{\mathsf{P}} & \xrightarrow{F} &]Y[_{\mathsf{Q}} \\ \downarrow{\scriptstyle sp} & & \downarrow{\scriptstyle sp} \\ X & \xrightarrow[f]{} & Y \end{array}$$

We write $\operatorname{Hom}_f(\mathcal{U}, \mathsf{Q}_K)$ for the collection of morphisms compatible with f.

The collection of rigid triples forms a category RT with the set of pairs (f, F) written as $\operatorname{Hom}((X, \overline{X}, \mathsf{P}), (Y, \overline{Y}, \mathsf{Q}))$, where $f : X \to Y$ is a k-morphism and $F \in \operatorname{colim}_{\mathcal{U}} \operatorname{Hom}_f(\mathcal{U}, \mathsf{Q})$.

Proposition 4.1.2. (i) *There is a functor*

$$(\mathrm{Sch}/k)^\circ \to C(K_0), \quad X \mapsto R\Gamma_{\mathrm{rig}}(X/K_0)$$

from the category of algebraic k-schemes with proper morphisms Sch/k, *such that* $H^i(R\Gamma_{\mathrm{rig}}(X/K_0)) \cong H^i_{\mathrm{rig}}(X/K_0)$. *Moreover, a canonical σ-linear endomorphism of* $R\Gamma_{\mathrm{rig}}(X/K_0)$ *exists inducing the Frobenius on cohomology.*

(ii) *There are two functors* $\mathrm{RT} \to C(K)$

$$\widetilde{R\Gamma}_{\mathrm{rig}}(X)_{\overline{X},\mathsf{P}} \quad \text{and} \quad R\Gamma_{\mathrm{rig}}(X/K)_{\overline{X},\mathsf{P}}$$

and functorial quasi-isomorphisms with respect to maps of rigid triples

$$R\Gamma_{\mathrm{rig}}(X/K) \leftarrow \widetilde{R\Gamma}_{\mathrm{rig}}(X)_{\overline{X},\mathsf{P}} \to R\Gamma_{\mathrm{rig}}(X)_{\overline{X},\mathsf{P}}.$$

Proof. See [Besser 2000, 4.9, 4.21, 4.22]. □

Remark 4.1.3. The building block of the construction is the functor $R\Gamma_{\mathrm{rig}}(X/K)_{\overline{X},\mathsf{P}}$. That complex is constructed with a system of compatible resolution of the overconvergent de Rham complexes $j_X^\dagger \Omega^\bullet_{\mathcal{U}}$, where $\mathcal{U}$ runs over all strict neighborhoods of the tube of X. Using Godement resolution, we can explicitly define

$$R\Gamma_{\mathrm{rig}}(X/K)_{\overline{X},\mathsf{P}} := \operatorname*{colim}_{\mathcal{U}} \Gamma(\mathcal{U}, \mathrm{Gd}_{\mathrm{an}}\, j_X^\dagger\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathcal{U}}),$$

where $\mathrm{Gd}_{\mathrm{an}} = \mathrm{Gd}_{Pt(\mathcal{U})}$. This will be an essential ingredient for achieving the main results of the paper.

All the proofs of [Besser 2000, Section 4] work using this *Godement* complex. We recall that

$$R\Gamma_{\mathrm{rig}}(X/K) := \operatorname*{colim}_{A \in \mathrm{SET}^0_X} R\Gamma_{\mathrm{rig}}(X/K)_{\overline{X}_A,\mathsf{P}_A},$$

$$\widetilde{R\Gamma}_{\mathrm{rig}}(X/K) := \operatorname*{colim}_{A \in \mathrm{SET}^0_{(X,\overline{X},\mathsf{P})}} R\Gamma_{\mathrm{rig}}(X/K)_{\overline{X}_A,\mathsf{P}_A},$$

where SET^0_X and $\mathrm{SET}^0_{(X,\overline{X},\mathsf{P})}$ are filtered categories of indexes.

With some modifications, we can provide a compact support version of the functors above. We just need to be careful in the choice of morphisms of rigid triples.

Definition 4.1.4. Let $(X, \overline{X}, \mathsf{P})$ and $(Y, \overline{Y}, \mathsf{Q})$ be two rigid triples, and let $f : X \to Y$ be a morphism of k-schemes. Let $F : \mathcal{U} \to \mathsf{Q}_K$ be compatible with f (as in Definition 4.1.1). We say that F is *strict* if there is a commutative diagram

$$\begin{tikzcd}]X[_{\mathsf{P}} \arrow[r] \arrow[d, "F"] & \mathcal{U} \arrow[d, "F"] & \mathcal{U} \setminus]X[_{\mathsf{P}} \arrow[l] \arrow[d, "F"] \\ ]Y[_{\mathsf{Q}} \arrow[r] & \mathcal{V} & \mathcal{V} \setminus]Y[_{\mathsf{Q}}, \arrow[l] \end{tikzcd}$$

where $\mathcal{V}$ is a strict neighborhood of $]Y[_{\mathsf{Q}}$ in $]\bar{Y}[_{\mathsf{Q}}$.

It is easy to show that strict morphisms are composable. We let RT_c denote the category of *rigid triples with proper morphisms*, which is the (not full) subcategory of RT with the same objects and morphisms pairs (f, F), where f is proper and F is a germ of a strict compatible morphism.

Lemma 4.1.5. (i) *Let $(X, \bar{X}, \mathsf{P})$ be a rigid triple, and let $\mathcal{U}$ be a strict neighborhood of $]X[_{\mathsf{P}}$. Then*

$$H^i(\Gamma(\mathcal{U}, \mathrm{Gd}_{\mathrm{an}}\, \underline{\Gamma}_{]X[_{\mathsf{P}}}\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathcal{U}})) = H^i_{\mathrm{rig},c}(X).$$

(ii) *Let $(Y, \bar{Y}, \mathsf{Q})$ be another rigid triple, $f : X \to Y$ be a proper k-morphism and $F : \mathcal{U} \to \mathsf{Q}_K$ be a morphism of K-analytic space compatible with f and strict. Then there is a canonical map*

$$F^* : \mathrm{Gd}_{\mathrm{an}}\, \underline{\Gamma}_{]Y[_{\mathsf{Q}}}\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathsf{Q}_K} \to F_* \,\mathrm{Gd}_{\mathrm{an}}\, \underline{\Gamma}_{]X[_{\mathsf{P}}}\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathcal{U}}.$$

Proof. (i) It is sufficient to note that $\mathrm{Gd}_{\mathrm{an}}(\Omega^\bullet_{\mathcal{U}})$ is a complex of flasque sheaves and that a flasque sheaf is acyclic for $\underline{\Gamma}_{]X[_{\mathsf{P}}}$. Let F be a flasque sheaf on the rigid analytic space $\mathcal{U}$. By definition, $\underline{\Gamma}_{]X[_{\mathsf{P}}} = \mathrm{Ker}(a : F \to i_* i^* F)$. It is easy to check that $\mathbb{R}^q i_* i^* F = 0$ for $q \geq 1$. Hence, $\mathbb{R}\underline{\Gamma}_{]X[_{\mathsf{P}}} F \cong \mathrm{Cone}(a : F \to i_* i^* F)[-1]$. But by hypothesis, the map a is surjective, so $\mathrm{Cone}(a : F \to i_* i^* F)[-1] \cong \mathrm{Ker}(a) = \mathbb{R}^0\underline{\Gamma}_{]X[_{\mathsf{P}}}$.
(ii) First consider the canonical pullback of differential forms $F^* : \Omega_{\mathsf{Q}_K} \to F_*\Omega_{\mathcal{U}}$. Then by Lemma 3.1.2, we get a map $\mathrm{Gd}_{Pt(\mathsf{Q}_K)}\, \Omega^\bullet_{\mathsf{Q}_K} \to F_*\, \mathrm{Gd}_{Pt(\mathcal{U})}\, \Omega_{\mathcal{U}}$. Applying the functor $\underline{\Gamma}_{]X[_{\mathsf{P}}}$ to the adjoint map, we get

$$\underline{\Gamma}_{]X[_{\mathsf{P}}} F^{-1}\, \mathrm{Gd}_{Pt(\mathsf{Q}_K)}\, \Omega_{\mathsf{Q}_K} \to \underline{\Gamma}_{]X[_{\mathsf{P}}}\, \mathrm{Gd}_{Pt(\mathcal{U})}\, \Omega_{\mathcal{U}}.$$

But the strictness of F implies that there is a canonical map $F^{-1}\underline{\Gamma}_{]Y[_{\mathsf{Q}}} \to \underline{\Gamma}_{]X[_{\mathsf{P}}} F^{-1}$ [Le Stum 2007, proof of Proposition 5.2.17]. Hence, we have a map

$$F^{-1}\underline{\Gamma}_{]Y[_{\mathsf{Q}}}\, \mathrm{Gd}_{Pt(\mathsf{Q}_K)}\, \Omega_{\mathsf{Q}_K} \to \underline{\Gamma}_{]X[_{\mathsf{P}}}\, \mathrm{Gd}_{Pt(\mathcal{U})}\, \Omega_{\mathcal{U}}.$$

We can conclude the proof by taking the adjoint of this map and again applying Lemma 3.1.2. ☐

Proposition 4.1.6. (i) *There is a functor*

$$(\mathrm{Sch}_c/k)^\circ \to C(K_0), \qquad X \mapsto R\Gamma_{\mathrm{rig},c}(X/K_0)$$

from the category of algebraic k-schemes with proper morphisms Sch_c/k, such that $H^i(R\Gamma_{\mathrm{rig},c}(X/K_0)) \cong H^i_{\mathrm{rig},c}(X/K_0)$. Moreover, there exists a canonical σ-linear endomorphism of $R\Gamma_{\mathrm{rig},c}(X/K_0)$ inducing the Frobenius on cohomology.

(ii) *There are two functors* $\mathrm{RT}_c \to C(K)$

$$\widetilde{R\Gamma}_{\mathrm{rig},c}(X)_{\overline{X},\mathsf{P}} \quad \textit{and} \quad R\Gamma_{\mathrm{rig},c}(X/K)_{\overline{X},\mathsf{P}}$$

and functorial quasi-isomorphisms with respect to maps of rigid triples

$$R\Gamma_{\mathrm{rig},c}(X/K) \leftarrow \widetilde{R\Gamma}_{\mathrm{rig},c}(X)_{\overline{X},\mathsf{P}} \to R\Gamma_{\mathrm{rig},c}(X)_{\overline{X},\mathsf{P}}.$$

Proof. In view of Lemma 4.1.5, it suffices to mimic the construction given in [Besser 2000, 4.9, 4.21, 4.22] but using only proper morphisms of k-schemes and strict compatible maps. In this case, the functors used in the construction are

$$\begin{aligned} R\Gamma_{\mathrm{rig},c}(X/K)_{\overline{X},\mathsf{P}} &:= \operatorname*{colim}_{\mathcal{U}} \Gamma(\mathcal{U}, \mathrm{Gd}_{\mathrm{an}}\, \underline{\Gamma}_{]X[_{\mathsf{P}}}\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathcal{U}}),\\ R\Gamma_{\mathrm{rig},c}(X/K) &:= \operatorname*{colim}_{A\in \mathrm{SET}^0_X} R\Gamma_{\mathrm{rig},c}(X/K)_{\overline{X}_A,\mathsf{P}_A},\\ \widetilde{R\Gamma}_{\mathrm{rig},c}(X/K) &:= \operatorname*{colim}_{A\in \mathrm{SET}^0_{(X,\overline{X},\mathsf{P})}} R\Gamma_{\mathrm{rig},c}(X/K)_{\overline{X}_A,\mathsf{P}_A}. \end{aligned}$$ □

4.2. ***de Rham complexes.*** Now we focus on de Rham complexes, and we deal with smooth K-algebraic schemes. Let X be a smooth algebraic K-scheme. The (algebraic) de Rham cohomology of X is the hypercohomology of its complex of Kähler differentials $H^i_{\mathrm{dR}}(X/K) := H^i(X, \Omega^\bullet_{X/K})$ [Grothendieck 1966b]. We can also define the de Rham cohomology with compact support [Baldassarri et al. 2004, Section 1] as the hypercohomology groups $H^i_{dR,c}(X/K) := H^i(\overline{X}, \lim_n J^n\Omega^\bullet_{\overline{X}/K})$, where $X \to \overline{X}$ is a smooth compactification and J is the sheaf of ideals associated to the complement $\overline{X} \setminus X$ (this definition does not depend on the choice of $\overline{X}$ [Baldassarri et al. 2004, Theorem 1.8]). In order to consider the Hodge filtration on the de Rham cohomology groups, we fix a normal crossings compactification $g : X \to Y$ and let $D := Y \setminus X$ be the complement divisor (this is possible by the Nagata compactification theorem and the Hironaka resolution theorem [Deligne 1971, Section 3.2.1]). We let $\Omega^\bullet_Y\langle D\rangle$ denote the de Rham complex of Y with logarithmic poles along D (in the Zariski topology) [Jannsen 1990, 3.3]. Let $I \subset \mathcal{O}_Y$ be the defining sheaf of ideals of D.

Proposition 4.2.1. *With the notation above,*

(i) *there is a canonical isomorphism*

$$H^i_{\mathrm{dR}}(X) \cong H^i(Y, \Omega^\bullet_Y\langle D\rangle) \quad (\textit{resp. } H^i_{\mathrm{dR},c}(X) \cong H^i(Y, I\Omega^\bullet_Y\langle D\rangle)),$$

(ii) *the spectral sequence*

$$E_1^{p,q} = H^q(Y, \Omega^p) \Rightarrow H^{p+q}(Y, \Omega^\bullet)$$

degenerates at 1 *for* $\Omega^\bullet = \Omega^\bullet_Y\langle D\rangle$ *and* $I\Omega^\bullet_Y\langle D\rangle$ *and*

(iii) *the filtration induced by this spectral sequence on* $H^i_{\mathrm{dR}}(X)$ (*resp.* $H^i_{\mathrm{dR},c}(X)$) *is independent of the choice of* Y. *Namely*,

$$F^j H^i_{\mathrm{dR}}(X) := H^i(Y, \sigma^{\geq j}\Omega^\bullet_Y\langle D\rangle) \quad (\textit{resp. } H^i_{\mathrm{dR},c}(X) := H^i(Y, \sigma^{\geq j} I\Omega^\bullet_Y\langle D\rangle)),$$

where $\sigma^{\geq j}$ *is the stupid filtration.*

Proof. Using the argument of [Deligne 1971, 3.2.11], we get the independence of the choice of Y. The same holds for $H^i_{\mathrm{dR},c}(X)$.

Since our base field K is of characteristic 0, we can find an embedding $\tau : K \to \mathbb{C}$. Then by [GAGA 1955–1956], we get these isomorphisms of filtered vector spaces:

$$H^i(Y, \Omega^\bullet_Y\langle D\rangle) \otimes_K \mathbb{C} \cong H^i(Y_h, \Omega^\bullet_{Y_h}\langle D_h\rangle)$$
$$(\text{resp. } H^i(Y, I\Omega^\bullet_Y\langle D\rangle) \otimes_K \mathbb{C} \cong H^i(Y_h, I_h\Omega^\bullet_{Y_h}\langle D_h\rangle)),$$

where $(\,\cdot\,)_h$ is the complex analytification functor and I_h is the defining sheaf of D_h. Thus, we conclude by [Deligne 1971, Section 3] (resp. [Peters and Steenbrink 2008, Part II, Example 7.25] for the compact support case). □

Remark 4.2.2. The degeneracy of the spectral sequence in (ii) of the proposition above can be proved algebraically [Deligne and Illusie 1987]. We don't know an algebraic proof of the isomorphism in (i).

In the sequel, a morphism of pairs (X, Y) as above is a commutative square

$$\begin{array}{ccc} X & \xrightarrow{g} & Y \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle v} \\ X' & \xrightarrow[g']{} & Y'. \end{array}$$

We say that the morphism is strict if the square is cartesian.

The complex $\Omega^\bullet_Y\langle D\rangle$ (resp. $I\Omega^\bullet_Y\langle D\rangle$) is a complex of Zariski sheaves over Y functorial with respect to the pair (X, Y) (resp. strict morphisms of pairs). We can construct two different (generalized) Godement resolutions (see Section 3): one using Zariski points and the other via the K-analytic space associated to Y.

We will write $Pt(Y) = Pt(Y_{\mathrm{zar}})$ for the set of Zariski points of Y with the discrete topology and $Pt(Y_{\mathrm{an}})$ for the discrete site of rigid points (Example 3.2.3) of Y_{an}; $Pt(Y_{\mathrm{an}}) \sqcup Pt(Y)$ is the direct sum in the category of sites.

Proposition 4.2.3. *With the notation of Proposition 4.2.1, let* $w : Y_{\mathrm{an}} \to Y_{\mathrm{zar}}$ *be the canonical map from the rigid analytic site to the Zariski site of* Y. *Then for any Zariski sheaf* Ω *on* Y, *there is a diagram*

$$\mathrm{Gd}_{Pt(Y)}(\Omega) \leftarrow \mathrm{Gd}_{Pt(Y_{\mathrm{an}})\sqcup Pt(Y)}(\Omega) \to w_*\,\mathrm{Gd}_{Pt(Y_{\mathrm{an}})}(w^*\Omega).$$

If we further consider $\Omega = \Omega^\bullet_Y\langle D\rangle$ (*resp.* $\Omega = I\Omega^\bullet_Y\langle D\rangle$), *then the diagram is functorial with respect to the pair* (X, Y) (*resp.* (X, Y) *and strict morphisms*). *The same holds true with* $\mathrm{Gd}^2_?$ *instead of* $\mathrm{Gd}_?$.

Proof. The first claim follows from Lemma 3.1.2 applied to the commutative diagram of sites

$$\begin{tikzcd} Pt(Y_{\mathrm{an}}) \arrow[r] \arrow[d] & Pt(Y_{\mathrm{an}}) \sqcup Pt(Y) \arrow[d] & Pt(Y) \arrow[l] \arrow[d] \\ Y_{\mathrm{an}} \arrow[r, "w"'] & Y_{\mathrm{zar}} & Y_{\mathrm{zar}} \arrow[l, "\mathrm{id}"] \end{tikzcd}$$

with respect to the canonical map $\Omega \to w_* w^* \Omega$. The second claim follows from the functoriality of the complex $\Omega^\bullet_Y\langle D\rangle$ (resp. $I\Omega^\bullet_Y\langle D\rangle$). □

Proposition 4.2.4. (i) *Let* $g : X \to Y$ *be a normal crossings compactification as in Section 4.2. Then there is a quasi-isomorphism of complexes of sheaves*

$$\Omega^\bullet_Y\langle D\rangle \to \mathrm{Gd}^2_{Pt(Y)}(\Omega^\bullet_Y\langle D\rangle) \quad (\textit{resp. } I\Omega^\bullet_Y\langle D\rangle \to \mathrm{Gd}^2_{Pt(Y)}\, I\Omega^\bullet_Y\langle D\rangle),$$

and the stupid filtration on $\Omega^\bullet_Y\langle D\rangle$ (*resp.* $I\Omega^\bullet_Y\langle D\rangle$) *induces a filtration on the right term of the morphism.*

(ii) *Let* Sm/K (*resp.* Sm_c/K) *be the category of algebraic and smooth* K*-schemes* (*resp. with proper morphisms*). *Let* $D^b_{\mathrm{dR}}(K)$ *be the derived category of the exact category of filtered vector spaces. Then there exist two functors*

$$R\Gamma_{\mathrm{dR}}(\,\cdot\,) : (\mathrm{Sm}/K)^\circ \to D^b_{\mathrm{dR}}(K) \quad \textit{and} \quad R\Gamma_{\mathrm{dR},c}(\,\cdot\,) : (\mathrm{Sm}_c/K)^\circ \to D^b_{\mathrm{dR}}(K)$$

such that with the same notation as (i), $R\Gamma_{\mathrm{dR}}(X) = \Gamma(Y, \mathrm{Gd}^2_{Pt(Y)}(\Omega^\bullet_Y\langle D\rangle))$ *and* $R\Gamma_{\mathrm{dR},c}(X) = \Gamma(Y, \mathrm{Gd}^2_{Pt(Y)} I\Omega^\bullet_Y\langle D\rangle)$.

(iii) *The filtered complexes* $R\Gamma_{\mathrm{dR}}(X)$ *and* $R\Gamma_{\mathrm{dR},c}(X)$ *are strict* (see Remark 2.0.10).

Proof. (i) This follows directly from the definition of Godement resolution.

(ii) This follows from the functoriality of the Godement resolution with respect to morphism of pairs and [Deligne 1971, 3.2.11] or [Huber 1995, Lemma 15.2.3] for the compact support case.

(iii) This follows by [Peters and Steenbrink 2008, Part II, Section 4.3, Section 7.3.1]. □

5. Syntomic cohomology

In this section, we construct the p-adic Hodge complexes needed to define the rigid syntomic cohomology groups (also with compact support) for a smooth algebraic $\mathscr{V}$-scheme. The functoriality will be a direct consequence of the construction.

5.1. ***Compactifications.***

Lemma 5.1.1 [Grothendieck 1966a, 2.8.5]. *Let $f : \mathscr{X} \to \mathscr{V}$ be a morphism of schemes, and let $Z \subset \mathscr{X}_K$ be a closed subscheme of the generic fiber of $\mathscr{X}$. Then there exists a unique closed subscheme $\mathscr{Z} \subset \mathscr{X}$ that is flat over $\mathscr{V}$ and satisfies $\mathscr{Z}_K = Z$. Thus, $\mathscr{Z}$ is the schematic closure of Z in $\mathscr{X}$.*

Proposition 5.1.2. *Let $\mathscr{X}$ be a smooth scheme over $\mathscr{V}$. Then there exists a generic* normal crossings compactification, *that is, an open embedding $g : \mathscr{X} \to \mathscr{Y}$ such that*

(i) *$\mathscr{Y}$ is proper over $\mathscr{V}$,*

(ii) *$\mathscr{Y}_K$ is smooth over K and*

(iii) *$\mathscr{D}_K \subset \mathscr{Y}_K$ is a normal crossings divisor, where $\mathscr{D} = \mathscr{Y} \setminus \mathscr{X}$.*

Proof. First by Nagata [Conrad 2007], there exists an open embedding $\mathscr{X}_K \to Y$, where Y is a proper K-scheme. By the Hironaka resolution theorem, we can assume that Y is a smooth compactification of $\mathscr{X}_K$ with complement a normal crossings divisor. Now we can define $\mathscr{Y}'$ to be the gluing of $\mathscr{X}$ and Y along the common open subscheme $\mathscr{X}_K$. These schemes are all of finite type over $\mathscr{V}$. It follows from the construction that $\mathscr{Y}'$ is a scheme, separated and of finite type over $\mathscr{V}$, whose generic fiber is Y.

The Nagata compactification theorem works also in a relative setting, namely for a separated and finite type morphism; hence, we can find a $\mathscr{V}$-scheme $\mathscr{Y}$ that is a compactification of $\mathscr{Y}'$ over $\mathscr{V}$. Thus, we get an open and dense embedding $h : \mathscr{Y}'_K = Y \to \mathscr{Y}_K$ of proper K-schemes, so h is the identity, and the statement is proven. □

Remark 5.1.3. We can also give another proof of the previous proposition assuming an embedding in a smooth $\mathscr{V}$-scheme. First by Nagata [Conrad 2007], there exists an open embedding $\mathscr{X} \to \overline{\mathscr{X}}$. Now assume that $\overline{\mathscr{X}}$ is embeddable in a smooth $\mathscr{V}$-scheme $\mathscr{W}$; then by [Włodarczyk 2005, Theorem 1.0.2], we can get a resolution of the K-scheme $\overline{\mathscr{X}}_K$ by making a sequence of blowups with respect to a family of closed subsets in good position with respect to the regular locus of $\overline{\mathscr{X}}_K$, in particular $Z_i \cap \mathscr{X}_K = \varnothing$. One can perform the same construction directly over $\mathscr{V}$, replacing the closed $Z_i \subset \mathscr{W}_K$ with their Zariski closure $\mathscr{Z}_i$ in $\mathscr{W}$. By hypothesis, $Z_i \subset \mathscr{W}_K \setminus \mathscr{X}_K$; hence, its closure $\mathscr{Z}_i$ is contained in $\mathscr{W} \setminus \mathscr{X}$, and $(\mathscr{Z}_i)_K = Z_i$ by Lemma 5.1.1. The construction doesn't affect what happens on the generic fiber because the blowup construction is local and behaves well with respect to open immersions [Hartshorne 1977, Chapter II, 7.15]. This will give $\mathscr{Y}$ as in the proposition.

5.2. ***Connecting maps.*** From now on, we keep the notation of Proposition 5.1.2 with $g : \mathscr{X} \to \mathscr{Y}$ being fixed. To simplify the notation, $\mathcal{X}$ (resp. $\mathcal{Y}$) denotes the rigid analytic space associated to the generic fiber of $\mathscr{X}$ (resp. $\mathscr{Y}$), usually denoted $\mathscr{X}_K^{\mathrm{an}}$. Let $w : \mathcal{Y} \to (\mathscr{Y}_K)_{\mathrm{zar}}$ be the canonical map of sites.

Sometimes we will simply write $\mathrm{Gd}_{\mathrm{an}} = \mathrm{Gd}_{Pt(\mathcal{U})}$ if the rigid space $\mathcal{U}$ is clear from the context. Similarly, we write $\mathrm{Gd}_{\mathrm{zar}} = \mathrm{Gd}_{Pt(X)}$ to denote the Godement resolution with respect to Zariski points of a K-scheme X.

Lemma 5.2.1. *With the notation above, we have the following morphisms of complexes of K-vector spaces*:

$$\Gamma(\mathcal{Y}, \mathrm{Gd}_{\mathrm{an}}\, j^\dagger\, \mathrm{Gd}_{\mathrm{an}}\, w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle) \to \Gamma(\mathcal{X}, \mathrm{Gd}_{\mathrm{an}}\, j^\dagger\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathcal{X}}) \to R\Gamma_{\mathrm{rig}}(\mathcal{X}_k/K)_{\mathcal{Y}_k,\widehat{\mathcal{Y}}},$$

$$\Gamma(\mathcal{Y}, \mathrm{Gd}_{\mathrm{an}}\, \Gamma_{]\mathcal{X}_k[}\, \mathrm{Gd}_{\mathrm{an}}\, w^* I\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle) \\ \to \Gamma(\mathcal{X}, \mathrm{Gd}_{\mathrm{an}}\, \Gamma_{]\mathcal{X}_k[}\, \mathrm{Gd}_{\mathrm{an}}\, \Omega^\bullet_{\mathcal{X}}) \to R\Gamma_{\mathrm{rig},c}(\mathcal{X}_k/K)_{\mathcal{Y}_k,\widehat{\mathcal{Y}}}.$$

All the maps are quasi-isomorphisms. We let a and b denote the composition of the maps in the first and second diagrams, *respectively.*

Proof. By construction, $R\Gamma_{\mathrm{rig}}(\mathcal{X}_k/K)_{\mathcal{Y}_k,\widehat{\mathcal{Y}}}$ (respectively $R\Gamma_{\mathrm{rig},c}(\mathcal{X}_k/K)_{\mathcal{Y}_k,\widehat{Y}}$) is a direct limit of complexes indexed over the strict neighborhoods of $]\mathcal{X}_k[_{\hat{\mathcal{Y}}}$, and $\mathcal{X}$ is one of them (see Remark 4.1.3 and Proposition 4.1.6). Hence, the map on the right (of both diagrams) comes from the universal property of the direct limit.

For the map on the left, consider first the canonical inclusion of algebraic differential forms with log poles into the analytic ones $w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle \to g^{\mathrm{an}}_*\Omega^\bullet_{\mathcal{X}}$. By Lemma 3.1.2, we get a map

$$\mathrm{Gd}_{Pt(\mathcal{Y})}\, w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle \to g^{\mathrm{an}}_*\, \mathrm{Gd}_{Pt(\mathcal{X})}\, \Omega^\bullet_{\mathcal{X}}.$$

Then applying the $j^\dagger$ functor and noting that $j^\dagger g^{\mathrm{an}}_* = g^{\mathrm{an}}_* j^\dagger$ [Le Stum 2007, 5.1.14], one obtains a morphism $j^\dagger\, \mathrm{Gd}_{Pt(\mathcal{Y})}\, w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle \to g^{\mathrm{an}}_* j^\dagger\, \mathrm{Gd}_{Pt(\mathcal{X})}\, \Omega^\bullet_{\mathcal{X}}$. This is what we need to apply Lemma 3.1.2 again and conclude the proof for the first diagram.

For the second diagram, repeat the argument using [Le Stum 2007, 5.2.15]. □

Lemma 5.2.2. *With the notation above*, *we have the following morphisms of complexes of K-vector spaces*, *which we denote by a' and b'*, *respectively*:

$$\Gamma(\mathcal{Y}, \mathrm{Gd}^2_{\mathrm{an}}\, w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle) \to \Gamma(\mathcal{Y}, \mathrm{Gd}_{\mathrm{an}}\, j^\dagger\, \mathrm{Gd}_{\mathrm{an}}\, w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle),$$

$$\Gamma(\mathcal{Y}, \mathrm{Gd}_{\mathrm{an}}\, \Gamma_{]\mathcal{X}_k[}\, \mathrm{Gd}_{\mathrm{an}}\, w^* I\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle) \to \Gamma(\mathcal{Y}, \mathrm{Gd}^2_{\mathrm{an}}\, w^* I\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle).$$

Proof. The maps a' and b' are induced by the canonical maps $\Omega \to j^\dagger\Omega$ and $\Gamma_{]\mathcal{X}_k[}\Omega \to \Omega$, respectively, where Ω is an abelian sheaf on $\mathcal{Y}$. In particular, we consider $\Omega = \mathrm{Gd}_{\mathrm{an}}\, w^*\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle$ (respectively $\Omega = \mathrm{Gd}_{\mathrm{an}}\, w^* I\Omega^\bullet_{\mathcal{Y}_K}\langle \mathcal{D}_K\rangle$). To conclude the proof, apply the functor $\mathrm{Gd}_{\mathrm{an}}$ again and take global sections. □

5.3. ***Syntomic complexes.*** Now we put together all we have done, getting a diagram, say $R\Gamma'(\mathcal{X})$, of complexes of K-vector spaces:

$$\begin{tikzcd}
R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \arrow[r, "\alpha_1"] & R\Gamma_{\mathrm{rig}}(\mathscr{X}/K) \\
\widetilde{R\Gamma}_{\mathrm{rig}}(\mathscr{X}/K)_{\mathscr{Y}_k,\hat{\mathscr{Y}}} \arrow[ur, "\alpha_2"] \arrow[r, "\alpha_3"] & R\Gamma_{\mathrm{rig}}(\mathscr{X}/K)_{\mathscr{Y}_k,\hat{\mathscr{Y}}} \\
\Gamma(\mathscr{Y}, \mathrm{Gd}^2_{\mathrm{an}}\, w^*\Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \arrow[ur, "\alpha_4"] \arrow[r, "\alpha_5"] & \Gamma(\mathscr{Y}, \mathrm{Gd}^2_{\mathrm{an}}\, w^*\Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \\
\Gamma(\mathscr{Y}_K, \mathrm{Gd}^2_{\mathrm{an+zar}}\, \Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \arrow[ur, "\alpha_6"] \arrow[r, "\alpha_7"] & \Gamma(\mathscr{Y}_K, \mathrm{Gd}^2_{\mathrm{zar}}\, \Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \\
\Gamma(\mathscr{Y}_K, \mathrm{Gd}^2_{\mathrm{zar}}\, \Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \arrow[ur, "\alpha_8"] &
\end{tikzcd}$$

where α_1, α_5, α_8 are the identity maps; α_2 and α_3 are the maps of Proposition 4.1.2; α_4 is the composition of $a \circ a'$ (see Lemma 5.2.1 and Lemma 5.2.2) and α_6 and α_7 are defined in Proposition 4.2.3. By repeatedly applying the quasipushout construction, we obtain a diagram of the shape

$$R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \to R\Gamma_K(\mathscr{X}) \leftarrow R\Gamma_{\mathrm{dR}}(\mathscr{X}). \tag{3}$$

It represents an object of *pHC* that we denote $R\Gamma(\mathscr{X})$.

Similarly, we can construct a p-adic Hodge complex $R\Gamma_c(\mathscr{X})$ associated to the diagram $R\Gamma'_c(\mathscr{X})$ defined as

$$\begin{tikzcd}
R\Gamma_{\mathrm{rig},c}(\mathscr{X}/K_0) \arrow[r, "\beta_1"] & R\Gamma_{\mathrm{rig},c}(\mathscr{X}/K) \\
\widetilde{R\Gamma}_{\mathrm{rig},c}(\mathscr{X}/K)_{\mathscr{Y}_k,\hat{\mathscr{Y}}} \arrow[ur, "\beta_2"] \arrow[r, "\beta_3"] & R\Gamma_{\mathrm{rig},c}(\mathscr{X}/K)_{\mathscr{Y}_k,\hat{\mathscr{Y}}} \\
\Gamma(\mathscr{Y}, \mathrm{Gd}_{\mathrm{an}}\, \Gamma_{]X[}\, \mathrm{Gd}_{\mathrm{an}}\, w^* I\Omega_{Y_K}\langle D\rangle) \arrow[ur, "\beta_4"] \arrow[r, "\beta_5"] & \Gamma(\mathscr{Y}, \mathrm{Gd}^2_{\mathrm{an}}\, w^* I\Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \\
\Gamma(\mathscr{Y}_K, \mathrm{Gd}^2_{\mathrm{an+zar}}\, I\Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \arrow[ur, "\beta_6"] \arrow[r, "\beta_7"] & \Gamma(\mathscr{Y}_K, \mathrm{Gd}^2_{\mathrm{zar}}\, I\Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \\
\Gamma(\mathscr{Y}_K, \mathrm{Gd}^2_{\mathrm{zar}}\, I\Omega_{\mathscr{Y}_K}\langle\mathscr{D}_K\rangle) \arrow[ur, "\beta_8"] &
\end{tikzcd}$$

where β_1 and β_8 are the identity maps; β_2 and β_3 are the maps of Proposition 4.1.2; β_4 is the map b of Lemma 5.2.1; $\beta_5 = b'$ of Lemma 5.2.2 and β_6 and β_7 are defined in Proposition 4.2.3. Note that β_6 is a quasi-isomorphism by GAGA.

Proposition 5.3.1. *Let Sm/R (resp. Sm_c/R) be the category of algebraic and smooth R-schemes (resp. with proper morphisms). The previous construction induces the functors*

$$R\Gamma(\,\cdot\,) : (\mathrm{Sm}/\mathscr{V})^\circ \to pHD \quad \text{and} \quad R\Gamma_c(\,\cdot\,) : (\mathrm{Sm}_c/\mathscr{V})^\circ \to pHD\,.$$

Proof. Let $f : \mathscr{X} \to \mathscr{X}'$ be a morphism of smooth $\mathscr{V}$-schemes. To get the functoriality, we just have to show that can find two gncd compactifications $g : \mathscr{X} \to \mathscr{Y}$ and $g' : \mathscr{X}' \to \mathscr{Y}'$ and a map $h : \mathscr{Y} \to \mathscr{Y}'$ extending f, that is, $hg = g'f$. We argue as in [Deligne 1971, Section 3.2.11]. Fix two gncd compactifications $g' : \mathscr{X}' \to \mathscr{Y}'$ and

$l : \mathscr{X} \to \mathscr{X}$. Then consider the canonical map $\mathscr{X} \to \mathscr{X} \times \mathscr{Y}'$ induced by l and $g' f$. Let $\overline{\mathscr{X}}$ be the closure of $\mathscr{X}$ in $\mathscr{X} \times \mathscr{Y}'$ and use the same argument as in the proof of Proposition 5.1.2 to get $\mathscr{Y}$, which is generically a resolution of the singularities of $\overline{\mathscr{X}}$. Then by Proposition 4.2.3 and Proposition 4.1.2, we get the functoriality of $R\Gamma(\,\cdot\,)$.

If we further assume f to be proper, then we can apply the same argument in order to get the commutative square $hg = g' f$ as above. Then by properness, this square is also cartesian by [Huber 1995, Lemma 15.2.3]. From this fact and Propositions 4.1.6 and 4.2.3, we obtain the functoriality of $R\Gamma_c(\,\cdot\,)$ with respect to proper maps. □

Definition 5.3.2. Let $\mathscr{X}$ be a smooth algebraic scheme over $\mathscr{V}$. For any integers n and i, we define the *absolute cohomology* groups of $\mathscr{X}$ as

$$H^n_{\mathrm{abs}}(\mathscr{X}, i) := \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma(\mathscr{X})(i)[n]) = H^n(\Gamma(\mathbb{K}, R\Gamma(\mathscr{X})(i))) \tag{4}$$

and the *absolute cohomology with compact support* groups of $\mathscr{X}$ as

$$H^n_{\mathrm{abs},c}(\mathscr{X}, i) := \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma_c(\mathscr{X})(i)[n]) = H^n(\Gamma(\mathbb{K}, R\Gamma_c(\mathscr{X})(i))). \tag{5}$$

A direct consequence of the definition is the existence of the following long exact sequence, which should be considered to be a p-adic analog of the corresponding sequence for Deligne–Beilinson cohomology [Beĭlinson 1985, Introduction].

Proposition 5.3.3. *With the notation above, we have the long exact sequences*

$$\to H^n_{\mathrm{abs}}(\mathscr{X}, i) \to H^n_{\mathrm{rig}}(\mathscr{X}_k/K_0) \oplus F^i H^n_{\mathrm{dR}}(\mathscr{X}_K) \xrightarrow{h} H^n_{\mathrm{rig}}(\mathscr{X}_k/K_0) \oplus H^n_{\mathrm{rig}}(\mathscr{X}_k/K) \xrightarrow{+},$$

where $h(x_0, x_{\mathrm{dR}}) := (\phi(x_0^\sigma) - p^i x_0, x_0 \otimes 1_K - sp(x_{\mathrm{dR}}))$ *and*

$$\to H^n_{\mathrm{abs},c}(\mathscr{X}, i) \to H^n_{\mathrm{rig},c}(\mathscr{X}_k/K_0) \oplus F^i H^n_{\mathrm{dR},c}(\mathscr{X}_K) \xrightarrow{h_c} H^n_{\mathrm{rig},c}(\mathscr{X}_k/K_0) \oplus H^n_{\mathrm{dR},c}(\mathscr{X}_K) \xrightarrow{+},$$

where $h_c(x_0, x_{\mathrm{dR}}) := (\phi_c(x_0\sigma) - p^i x_0, cosp(x_0 \otimes 1_K) - x_{\mathrm{dR}})$.

Proof. By Proposition 2.2.2, the absolute cohomology is the cohomology of a mapping cone, namely $\Gamma(\mathbb{K}, R\Gamma(\mathscr{X}))$ (or $\Gamma_c(\mathbb{K}, R\Gamma(\mathscr{X}))$ for the compact support case). The long exact sequences above are easily induced from the distinguished triangle defined by the term of the mapping cone (see Remark 2.2.1(iii)). Indeed, one need only note that the map $c : R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \otimes K \to R\Gamma_K(\mathscr{X})$ (resp. $s : R\Gamma_{\mathrm{dR}}(\mathscr{X}) \to R\Gamma_K(\mathscr{X})$) is a quasi-isomorphism. □

Proposition 5.3.4. *With the notation above, there is a canonical isomorphism between the absolute cohomology we have defined and the* (*rigid*) *syntomic cohomology of Besser*:

$$H^n_{\mathrm{syn}}(\mathscr{X}, i) \cong H^n_{\mathrm{abs}}(\mathscr{X}, i).$$

Proof. Besser defines a complex

$$\mathbb{R}\Gamma_{\mathrm{Bes}}(\mathscr{X}, i) := \mathrm{Cone}\big(\mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus Fil^i\, \mathbb{R}\Gamma_{\mathrm{dR}}(\mathscr{X}) \to \mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus \mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K)\big)[-1]$$

and the syntomic cohomology groups (of degree n and twisted i) $H^n(\mathbb{R}\Gamma_{\mathrm{Bes}}(\mathscr{X}, i))$ (see [Besser 2000, proof of Proposition 6.3]). Note that, modulo the choice of the flasque resolution, $\mathbb{R}\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) = R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0)$ (the left-hand side is the notation used by Besser with the bold $\mathbb{R}$); the complex $Fil^i\, \mathbb{R}\Gamma_{\mathrm{dR}}(\mathscr{X})$ is a direct limit over all the normal crossing compactifications of $\mathscr{X}_K$ of complexes of the form $\Gamma(Y, \mathrm{Gd}^2_{Pt(Y)}\, \sigma^{\geq i}\Omega_Y\langle D\rangle)$ so that our $F^i R\Gamma_{\mathrm{dR}}(\mathscr{X})$ is an element of this direct limit. To conclude the proof, recall that by Remark 2.2.1(iii), we obtain

$$H^n_{\mathrm{abs}}(\mathscr{X}, i) \cong H^n(\mathrm{Cone}(R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus F^i R\Gamma_{\mathrm{dR}}(\mathscr{X}) \to R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_{\mathrm{rig}}(\mathscr{X}/K))[-1]),$$

and it is easy to check that the maps in the two mapping cones are defined in the same way. □

Remark 5.3.5. The isomorphism $H^n(R\Gamma_{\mathrm{syn}}(\mathscr{X}, i)) \cong \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma(\mathscr{X})(i)[n])$ can be viewed as a generalization of the result of Bannai [2002], who considers only smooth schemes $\mathscr{X}$ with a fixed compactification $\mathscr{Y}$ and such that $\mathscr{Y} \setminus \mathscr{X} = \mathscr{D}$ is a relative normal crossings divisor over $\mathscr{V}$. Moreover, Bannai's construction is not functorial with respect to $\mathscr{X}$; functoriality holds only with respect to a so-called syntomic datum. We should point out that the category defined by Bannai is endowed with a t-structure whose heart is the category MF^f_K of (weakly) admissible filtered Frobenius modules. In our case, we don't have such a nice picture.

Remark 5.3.6. The category of p-adic Hodge complexes is not endowed with internal Hom. This is due to the fact that the Frobenius is only a quasi-isomorphism; hence, we cannot invert it. In particular, this happens for the complex $R\Gamma_c(\mathscr{X})$; thus, we cannot define the dual $R\Gamma_c(\mathscr{X})^\vee := \mathbb{R}\mathscr{H}om(R\Gamma_c(\mathscr{X}), \mathbb{K})$ and an absolute homology theory as

$$H_n^{\mathrm{abs}}(\mathscr{X}, i) := \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma_c(\mathscr{X})^\vee(-i)[-n]);$$

see [Huber 1995, Section 15.3]. Nevertheless, the usual adjunction between Hom and $\otimes$ should give a natural isomorphism

$$\mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma_c(\mathscr{X})^\vee) \cong \mathrm{Hom}_{pHD}(R\Gamma_c(\mathscr{X}), \mathbb{K}),$$

and the right term does makes sense in our setting. This motivates the following definition:

Definition 5.3.7. With the notation of Definition 5.3.2, we define the *absolute homology* groups of $\mathscr{X}$ as

$$H_n^{\mathrm{abs}}(\mathscr{X}, i) := \mathrm{Hom}_{pHD}(R\Gamma_c(\mathscr{X}), \mathbb{K}(-i)[-n]) = H^{-n}(\Gamma(R\Gamma_c(\mathscr{X}), \mathbb{K}(-i))). \quad (6)$$

5.4. ***Cup product and Gysin map.*** We are going to prove that there is a morphism

$$R\Gamma(\mathscr{X}) \otimes R\Gamma_c(\mathscr{X}) \to R\Gamma_c(\mathscr{X})$$

of p-adic Hodge complexes. This induces a pairing at the level of the complexes computing absolute cohomology. The key point is the compatibility of the de Rham and rigid pairings with respect to the specialization and cospecialization maps.

Let us start by fixing some notation. Let $\mathscr{X}$ be a smooth algebraic $\mathscr{V}$-scheme, $g : \mathscr{X} \to \mathscr{Y}$ be a gncd compactification and $\mathscr{D} = \mathscr{Y} \setminus \mathscr{X}$ be the complement; $\mathcal{Y}$ is the rigid analytic space associated to the K-scheme $\mathscr{Y}_K$; as before, we let $w : \mathcal{Y} \to \mathscr{Y}_{K,\mathrm{zar}}$ denote the canonical morphism of sites (see the notation after Proposition 5.1.2). For any $\mathbb{O}_{\mathscr{Y}_K}$-module F, w^* denotes its pullback.

Remark 5.4.1.

(i) The wedge product of algebraic differentials induces the pairing

$$p_{\mathrm{dR}} : \Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \otimes I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \to I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle.$$

(ii) The analytification of p_{dR} gives a pairing

$$w^*\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \otimes w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \to w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle.$$

Hence, by [Berthelot 1997a, Lemma 2.1], we get the pairing

$$p_{\mathrm{rig}} : j^\dagger w^*\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \otimes \underline{\Gamma}_{]\mathscr{X}_k[}w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \to \underline{\Gamma}_{]\mathscr{X}_k[}w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle.$$

Lemma 5.4.2 (sheaves level). *The diagram*

$$\begin{tikzcd}
\begin{array}{c}\mathrm{Gd_{an}}^2(w^*\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle) \\ \otimes \mathrm{Gd}^2_{\mathrm{an}}(w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle)\end{array} \arrow[r] & \mathrm{Gd}^2_{\mathrm{an}}(w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle) \\
\begin{array}{c}\mathrm{Gd}^2_{\mathrm{an}}(w^*\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle) \\ \otimes \mathrm{Gd_{an}}\,\underline{\Gamma}_{]\mathscr{X}_k[}\,\mathrm{Gd_{an}}(w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle)\end{array} \arrow[u] \arrow[d] \arrow[r, "m"] & \mathrm{Gd_{an}}\,\underline{\Gamma}_{]\mathscr{X}_k[}\,\mathrm{Gd_{an}}(w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle) \arrow[u] \arrow[d, "1"] \\
\begin{array}{c}\mathrm{Gd_{an}}\, j^\dagger\, \mathrm{Gd_{an}}(w^*\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle) \\ \otimes \mathrm{Gd_{an}}\,\underline{\Gamma}_{]\mathscr{X}_k[}\,\mathrm{Gd_{an}}(w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle)\end{array} \arrow[r] & \mathrm{Gd_{an}}\,\underline{\Gamma}_{]\mathscr{X}_k[}\,\mathrm{Gd_{an}}(w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle)
\end{tikzcd}$$

commutes, where $m := p_{\mathrm{rig}} \circ (j^\dagger \otimes 1)$ by definition.

Proof. The bottom square commutes by construction, and we get $p_{\mathrm{rig}} \circ (j^\dagger \otimes 1) = w^*p_{\mathrm{dR}}$ restricted to $w^*\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle \otimes \underline{\Gamma}_{]\mathscr{X}_k[}w^*I\Omega_{\mathscr{Y}_K}\langle \mathscr{D}_K\rangle$. $\square$

Proposition 5.4.3. *Let $\mathscr{X}$ be a smooth $\mathscr{V}$-scheme. Then there exists a morphism*

$$\pi : R\Gamma(\mathscr{X}) \otimes R\Gamma_c(\mathscr{X}) \to R\Gamma_c(\mathscr{X})$$

of p-adic Hodge complexes, which is functorial with respect to $\mathscr{X}$ (as a morphism in pHD). Moreover, taking the cohomology of this map, we get the compatibility

$$\begin{tikzcd}
H^n_{\mathrm{dR}}(\mathscr{X}_K) \otimes H^m_{\mathrm{dR},c}(\mathscr{X}_K) \arrow[r] & H^{n+m}_{\mathrm{dR},c}(\mathscr{X}_K) \\
H^n_{\mathrm{dR}}(\mathscr{X}_K) \otimes H^m_{\mathrm{rig},c}(\mathscr{X}_k) \arrow[r] \arrow[u, "\mathrm{id}\otimes cosp"] \arrow[d, "sp\otimes\mathrm{id}"'] & H^{n+m}_{\mathrm{rig},c}(\mathscr{X}_k) \arrow[u, "cosp"'] \arrow[d, "\mathrm{id}"] \\
H^n_{\mathrm{rig}}(\mathscr{X}_k) \otimes H^m_{\mathrm{rig},c}(\mathscr{X}_k) \arrow[r] & H^{n+m}_{\mathrm{rig},c}(\mathscr{X}_k)
\end{tikzcd}$$

Proof. It is sufficient to provide a pairing of the enlarged diagrams (see Remark 2.2.4). Thus, we have to define a morphism of diagrams $\pi' : R\Gamma'(\mathscr{X}) \otimes R\Gamma'_c(\mathscr{X}) \to R\Gamma'_c(\mathscr{X})$ (notation as in Section 5.3). It is easy to construct π' using the previous lemma and the compatibility of the Godement resolution with the tensor product. □

Corollary 5.4.4. *There is a functorial pairing (induced by π of Proposition 5.4.3)*

$$H^n_{\mathrm{abs}}(\mathscr{X}, i) \otimes H^m_{\mathrm{abs},c}(\mathscr{X}, j) \to H^{n+m}_{\mathrm{abs},c}(\mathscr{X}, i+j).$$

Proof. Consider the pairing of Proposition 5.4.3, which induces a morphism $R\Gamma(\mathscr{X})(i) \otimes R\Gamma_c(\mathscr{X})(j) \to R\Gamma_c(\mathscr{X})(i+j)$. We then get the corollary by Lemma 2.2.3 and Definition 5.3.2. □

Proposition 5.4.5 (Poincaré duality). *Let $\mathscr{X}$ be a smooth and algebraic $\mathscr{V}$-scheme of dimension d. Then there is a canonical isomorphism*

$$H^n_{\mathrm{abs}}(\mathscr{X}, i) \cong H^{\mathrm{abs}}_{2d-n}(\mathscr{X}, d-i).$$

Proof. By definition (Equations (4) and (6)), it is sufficient to prove that the complex $\Gamma(\mathbb{K}, R\Gamma(\mathscr{X}))$ is quasi-isomorphic to the diagram $\Gamma(R\Gamma_c(\mathscr{X}), \mathbb{K}[-2d](-d))$.

First recall that $\Gamma(\mathbb{K}, R\Gamma(\mathscr{X}))$ is defined as

$$\mathrm{Cone}\big(R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_K(\mathscr{X}) \oplus F^0 R\Gamma_{\mathrm{dR}}(\mathscr{X}) \\ \xrightarrow{\psi} R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_K(\mathscr{X}) \oplus R\Gamma_K(\mathscr{X})\big)[-1],$$

where $\psi(x_0, x_K, x_{\mathrm{dR}}) := (\phi(x_0) - x_0, c(x_0 \otimes \mathrm{id}_K) - x_K, x_K - s(x_{\mathrm{dR}}))$. To define the desired map, we need to modify this complex, replacing $R\Gamma_K(\mathscr{X})$ with $R\Gamma_{\mathrm{dR}}(\mathscr{X})$ as

$$\mathrm{Cone}\big(R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_{\mathrm{dR}}(\mathscr{X}) \oplus F^0 R\Gamma_{\mathrm{dR}}(\mathscr{X}) \\ \xrightarrow{\psi'} R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_K(\mathscr{X}) \oplus R\Gamma_{\mathrm{dR}}(\mathscr{X})\big)[-1],$$

where $\psi'(x_0, x'_{\mathrm{dR}}, x_{\mathrm{dR}}) := (\phi(x_0) - x_0, c(x_0 \otimes \mathrm{id}_K) - s(x'_{\mathrm{dR}}), x'_{\mathrm{dR}} - x_{\mathrm{dR}})$. It is easy to see that this new complex, call it $M^\bullet$, is quasi-isomorphic to $\Gamma(\mathbb{K}, R\Gamma(\mathscr{X}))$.

Because the filtered complex $R\Gamma_{\mathrm{dR},c}(\mathscr{X})$ is strict, the truncation $\tau_{\geq 2d} R\Gamma_{\mathrm{dR},c}(\mathscr{X})$ is the usual truncation of complexes of K-vector spaces (see Lemma 2.1.1(iii) and Remark 2.0.10). Then the cup product induces a morphism of complexes $M^\bullet \to \Gamma(R\Gamma_c(\mathscr{X}), \tau_{\geq 2d} R\Gamma_c(\mathscr{X}))$ that is a quasi-isomorphism by the Poincaré duality theorems for rigid and de Rham cohomology [Berthelot 1997a; Huber 1995]. Explicitly, this map is induced by the commutative diagram

$$\begin{array}{ccc}
\begin{array}{c} R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_{\mathrm{dR}}(\mathscr{X}) \\ \oplus F^0 R\Gamma_{\mathrm{dR}}(\mathscr{X}) \end{array} & \xrightarrow{\alpha} & \begin{array}{c} \mathrm{Hom}^\bullet_{K_0}(N_0, \tau_{\geq 2d} N_0) \oplus \mathrm{Hom}^\bullet_K(N_K, \tau_{\geq 2d} N_K) \\ \oplus \mathrm{Hom}^{\bullet, F}_K(N_{\mathrm{dR}}, \tau_{\geq 2d} N_{\mathrm{dR}}) \end{array} \\
\Big\downarrow{\scriptstyle \psi'} & & \Big\downarrow{\scriptstyle \xi} \\
\begin{array}{c} R\Gamma_{\mathrm{rig}}(\mathscr{X}/K_0) \oplus R\Gamma_{\mathrm{rig}}(\mathscr{X}/K) \\ \oplus R\Gamma_{\mathrm{dR}}(\mathscr{X}) \end{array} & \xrightarrow[\beta]{} & \begin{array}{c} \mathrm{Hom}^\bullet_{K_0}(N_0^\sigma, \tau_{\geq 2d} N_0) \oplus \mathrm{Hom}^\bullet_K(N_{\mathrm{rig}}, \tau_{\geq 2d} N_K) \\ \oplus \mathrm{Hom}^\bullet_K(N_{\mathrm{dR}}, \tau_{\geq 2d} N_K), \end{array}
\end{array}$$

where $N := R\Gamma_c(\mathscr{X})$,

$$\xi(f_0, f_K, \mathrm{dR}) := (\phi_c \circ f_0^\sigma - f_0 \circ \phi_c, c \circ (f_0 \otimes \mathrm{id}_K) - f_K \circ c, f_K \circ s - s \circ x_{\mathrm{dR}}),$$

$$\alpha(x_0, x'_{\mathrm{dR}}, x_{\mathrm{dR}}) : (y_0, y_K, y_{\mathrm{dR}}) \mapsto (x_0 \cup y_0, s(x'_{\mathrm{dR}}) \cup y_K, x_{\mathrm{dR}} \cup y_{\mathrm{dR}}),$$

$$\beta(x_0, x_{\mathrm{rig}}, x_{\mathrm{dR}}) : (y_0, y_{\mathrm{rig}}, y_{\mathrm{dR}}) \mapsto (x_0 \cup \phi_c(y_0), x_{\mathrm{rig}} \cup y_{\mathrm{rig}}, x_{\mathrm{dR}} \cup y_{\mathrm{dR}}).$$

To conclude the proof, it is sufficient to apply the exact functor $\Gamma(R\Gamma_c(\mathscr{X}), \cdot)$ to the quasi-isomorphisms

$$\tau_{\geq 2d} R\Gamma_c(\mathscr{X}) \leftarrow H^{2d}(R\Gamma_c(\mathscr{X}))[-2d] \to \mathbb{K}(-d)[-2d]. \qquad \square$$

Remark 5.4.6. We would like to point out some technical issues regarding Poincaré duality in syntomic cohomology.

(i) If it were possible to define an internal Hom in the category *pHC* of p-adic Hodge complexes (see Remark 5.3.6), then by Proposition 5.4.3, one would obtain the natural isomorphism

$$R\Gamma(\mathscr{X})(d)[2d] \cong R\Gamma_c(\mathscr{X})^\vee$$

in the triangulated category *pHD*, where $R\Gamma_c(\mathscr{X})^\vee := \mathbb{R}\mathscr{H}om_{pHC}(R\Gamma_c(\mathscr{X}), \mathbb{K})$ is the dual of $R\Gamma_c(\mathscr{X})$. Then one would get by adjunction the duality

$$\begin{aligned} \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma(X)(i)[n]) &\cong \mathrm{Hom}_{pHD}(\mathbb{K}, R\Gamma_c(X)^\vee(i-d)[n-2d]) \\ &\cong \mathrm{Hom}_{pHD}(R\Gamma_c(X)(d-i)[2d-n], \mathbb{K}). \end{aligned}$$

(ii) The Grothendieck–Leray spectral sequence for absolute homology is

$$E_2^{p,q} = \mathrm{Hom}_{pHD}(H^{-q}(R\Gamma_c(\mathscr{X})(i))[-p], \mathbb{K}),$$

and it degenerates to the short exact sequences

$$0 \to \mathrm{Hom}_{pHD}(H^n(R\Gamma_c(\mathscr{X})(i)), \mathbb{K}) \to H_n^{\mathrm{syn}}(\mathscr{X}, i) \\ \to \mathrm{Hom}_{pHD}(H^{n+1}(R\Gamma_c(\mathscr{X})(i)), \mathbb{K}[1]) \to 0.$$

Directly by Proposition 2.2.2, the group $\mathrm{Hom}_{pHD}(H^n(R\Gamma_c(\mathscr{X})(i)), \mathbb{K})$ is

$$\Big\{ (x_0, x_{\mathrm{dR}}) \in H^n_{\mathrm{rig},c}(\mathscr{X}_k/K_0)^\vee \oplus (H^n_{\mathrm{dR},c}(\mathscr{X}_K)/F^{i+1})^\vee \\ : (x_0 \otimes 1_K) = x_{\mathrm{dR}} \circ cosp, x_0 \circ \phi_c = p^i x_0^\sigma \Big\}.$$

In cohomology, the Frobenius is an invertible; hence, the equation $x_0 \circ \phi_c = p^i x_0^\sigma$ is equivalent to $x_0 = p^i x_0^\sigma \phi_c^{-1}$, where the latter is the formula for the Frobenius of internal $\mathscr{H}om(H^n_{\mathrm{rig},c}(\mathscr{X}_k), K(-i))$ in the category of isocrystals (that is, modules with Frobenius). Hence, by Poincaré duality, we get $\phi((x_0^\vee)^\sigma) = p^{d-i} x_0^\vee$; $x^\vee \in H^{2d-n}_{\mathrm{rig}}(\mathscr{X}_k/K_0)$ is the cohomology class corresponding to x_0, where ϕ is the Frobenius of $H^{2d-n}_{\mathrm{rig}}(\mathscr{X}_k/K_0)$.

(iii) The absolute homology is defined via the complex $R\Gamma_c(\mathscr{X})$, but it is not clear how to relate it to the dual of the absolute cohomology with compact support.

Corollary 5.4.7 (Gysin map). *Let $f : \mathscr{X} \to \mathscr{Y}$ be a proper morphism of smooth algebraic $\mathscr{V}$-schemes of relative dimension d and e, respectively. Then there is a canonical map*

$$f_* : H^n_{\mathrm{abs}}(\mathscr{X}, i) \to H^{n+2c}_{\mathrm{abs}}(\mathscr{Y}, i + c),$$

where $c = e - d$.

Proof. This is a direct consequence of the previous proposition and the functoriality of $R\Gamma_c(\mathscr{X})$ with respect to proper morphisms. □

Remark 5.4.8. With the notation above, we get a morphism of spectral sequences

$$g : E_2^{p,q}(\mathscr{X}) := \mathrm{Hom}_{pHD}(\mathbb{K}, H^q(\mathscr{X})(i)[p]) \to E_2^{p,q+2c}(\mathscr{Y}) \\ := \mathrm{Hom}_{pHD}(\mathbb{K}, H^{q+2c}(\mathscr{Y})(i)[p])$$

compatible with f_*. The map g is induced by the Gysin morphism in de Rham and rigid cohomology.

We conclude by saying that it is natural to expect that the Gysin map for syntomic cohomology is compatible with the K-theory pushforward under the regulator defined by Besser. We plan to address this open problem in a future work.

Acknowledgment

We acknowledge the help of and useful discussion with K. Bannai, A. Besser, F. Ivorra and Q. Liu.

References

[Artin et al. 1972a] M. Artin, A. Grothendieck, J. Verdier, P. Deligne, and B. Saint-Donat (editors), *Théorie des topos et cohomologie étale des schémas. Tome 1: Théorie des topos*, Lecture Notes in Mathematics **269**, Springer, Berlin, 1972. Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964. MR 50 #7130 Zbl 0234.00007

[Artin et al. 1972b] M. Artin, A. Grothendieck, J. Verdier, P. Deligne, and B. Saint-Donat (editors), *Théorie des topos et cohomologie étale des schémas. Tome 2*, Lecture Notes in Mathematics **270**, Springer, Berlin, 1972. Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964. MR 50 #7131 Zbl 0237.00012

[Baldassarri et al. 2004] F. Baldassarri, M. Cailotto, and L. Fiorot, "Poincaré duality for algebraic de Rham cohomology", *Manuscripta Math.* **114**:1 (2004), 61–116. MR 2006b:14035 Zbl 1071.14023

[Bannai 2002] K. Bannai, "Syntomic cohomology as a p-adic absolute Hodge cohomology", *Math. Z.* **242**:3 (2002), 443–480. MR 2004e:14037 Zbl 1053.14018

[Beĭlinson 1985] A. A. Beĭlinson, "Higher regulators and values of L-functions", *J. Sov. Math.* **30** (1985), 2036–2070. Zbl 0588.14013

[Beĭlinson 1986] A. A. Beĭlinson, "Notes on absolute Hodge cohomology", pp. 35–68 in *Applications of algebraic K-theory to algebraic geometry and number theory, Part I* (Boulder, CO, 1983), edited by S. J. Bloch et al., Contemp. Math. **55**, Amer. Math. Soc., Providence, RI, 1986. MR 87m:14019 Zbl 0621.14011

[Berthelot 1997a] P. Berthelot, "Dualité de Poincaré et formule de Künneth en cohomologie rigide", *C. R. Acad. Sci. Paris Sér. I Math.* **325**:5 (1997), 493–498. MR 2000c:14023 Zbl 0908.14006

[Berthelot 1997b] P. Berthelot, "Finitude et pureté cohomologique en cohomologie rigide", *Invent. Math.* **128**:2 (1997), 329–377. MR 98j:14023 Zbl 0908.14005

[Berthelot et al. 1971] P. Berthelot, A. Grothendieck, and L. Illusie (editors), *Théorie des intersections et théorème de Riemann–Roch* (*SGA 6*), Lecture Notes in Mathematics **225**, Springer, Berlin, 1971. MR 50 #7133 Zbl 0218.14001

[Besser 2000] A. Besser, "Syntomic regulators and p-adic integration, I: Rigid syntomic regulators", *Israel J. Math.* **120** (2000), 291–334. MR 2002c:14035 Zbl 1001.19003

[Bloch 1986] S. Bloch, "Algebraic cycles and higher K-theory", *Adv. in Math.* **61**:3 (1986), 267–304. MR 88f:18010 Zbl 0608.14004

[Bosch et al. 1995] S. Bosch, W. Lütkebohmert, and M. Raynaud, "Formal and rigid geometry, IV: The reduced fibre theorem", *Invent. Math.* **119**:2 (1995), 361–398. MR 97e:11075 Zbl 0839.14014

[Conrad 2007] B. Conrad, "Deligne's notes on Nagata compactifications", *J. Ramanujan Math. Soc.* **22**:3 (2007), 205–257. MR 2009d:14002 Zbl 1142.14001

[Deligne 1971] P. Deligne, "Théorie de Hodge, II", *Inst. Hautes Études Sci. Publ. Math.* 40 (1971), 5–57. MR 58 #16653a Zbl 0219.14007

[Deligne and Illusie 1987] P. Deligne and L. Illusie, "Relèvements modulo p^2 et décomposition du complexe de de Rham", *Invent. Math.* **89**:2 (1987), 247–270. MR 88j:14029 Zbl 0632.14017

[Friedlander and Suslin 2002] E. M. Friedlander and A. Suslin, "The spectral sequence relating algebraic K-theory to motivic cohomology", *Ann. Sci. École Norm. Sup.* (4) **35**:6 (2002), 773–875. MR 2004b:19006 Zbl 1047.14011

[GAGA 1955–1956] J.-P. Serre, "Géométrie algébrique et géométrie analytique", *Ann. Inst. Fourier, Grenoble* **6** (1955–1956), 1–42. MR 18,511a Zbl 0075.30401

[Gillet and Messing 1987] H. Gillet and W. Messing, "Cycle classes and Riemann–Roch for crystalline cohomology", *Duke Math. J.* **55**:3 (1987), 501–538. MR 89c:14025 Zbl 0651.14014

[Grothendieck 1966a] A. Grothendieck, "Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* 28 (1966), 255. MR 36 #178 Zbl 0144.19904

[Grothendieck 1966b] A. Grothendieck, "On the de Rham cohomology of algebraic varieties", *Inst. Hautes Études Sci. Publ. Math.* 29 (1966), 95–103. MR 33 #7343 Zbl 0145.17602

[Grothendieck 1967] A. Grothendieck, "Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* 32 (1967), 361. MR 39 #220 Zbl 0153.22301

[Hartshorne 1975] R. Hartshorne, "On the De Rham cohomology of algebraic varieties", *Inst. Hautes Études Sci. Publ. Math.* 45 (1975), 5–99. MR 55 #5633 Zbl 0326.14004

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Huber 1995] A. Huber, *Mixed motives and their realization in derived categories*, Lecture Notes in Mathematics **1604**, Springer, Berlin, 1995. MR 98d:14030 Zbl 0938.14008

[Ivorra 2005] F. Ivorra, *Réalisation l-adique des motifs mixtes*, thesis, Université Paris, 6, 2005, Available at http://www.uni-due.de/~hm0035/These.pdf.

[Jannsen 1990] U. Jannsen, *Mixed motives and algebraic K-theory*, Lecture Notes in Mathematics **1400**, Springer, Berlin, 1990. MR 91g:14008 Zbl 0691.14001

[Le Stum 2007] B. Le Stum, *Rigid cohomology*, Cambridge Tracts in Mathematics **172**, Cambridge University Press, 2007. MR 2009c:14029 Zbl 1131.14001

[Levine 1998] M. Levine, *Mixed motives*, Mathematical Surveys and Monographs **57**, American Mathematical Society, Providence, RI, 1998. MR 99i:14025 Zbl 0902.14003

[Mazza et al. 2006] C. Mazza, V. Voevodsky, and C. Weibel, *Lecture notes on motivic cohomology*, Clay Mathematics Monographs **2**, American Mathematical Society, Providence, RI, 2006. MR 2007e:14035 Zbl 1115.14010

[Peters and Steenbrink 2008] C. A. M. Peters and J. H. M. Steenbrink, *Mixed Hodge structures*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **52**, Springer, Berlin, 2008. MR 2009c:14018 Zbl 1138.14002

[Petrequin 2003] D. Petrequin, "Classes de Chern et classes de cycles en cohomologie rigide", *Bull. Soc. Math. France* **131**:1 (2003), 59–121. MR 2004b:14030 Zbl 1083.14505

[van der Put and Schneider 1995] M. van der Put and P. Schneider, "Points and topologies in rigid geometry", *Math. Ann.* **302**:1 (1995), 81–103. MR 96k:32070 Zbl 0867.11049

[Suslin and Voevodsky 2000] A. Suslin and V. Voevodsky, "Relative cycles and Chow sheaves", pp. 10–86 in *Cycles, transfers, and motivic homology theories*, Ann. of Math. Stud. **143**, Princeton Univ. Press, 2000. MR 1764199 Zbl 1019.14004

[Weibel 1994] C. A. Weibel, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics **38**, Cambridge University Press, 1994. MR 95f:18001 Zbl 0797.18001

[Włodarczyk 2005] J. Włodarczyk, "Simple Hironaka resolution in characteristic zero", *J. Amer. Math. Soc.* **18**:4 (2005), 779–822. MR 2006f:14014 Zbl 1084.14018

Communicated by Hélène Esnault
Received 2010-10-12 Revised 2011-12-22 Accepted 2012-05-03

chiarbru@math.unipd.it *Università degli Studi di Padova, Via Trieste 63, 35100 Padova, Italy*

ciccioni@math.unipd.it *Università degli Studi di Padova, Via Trieste 63, 35100 Padova, Italy*

nicola.mazzari@math.u-bordeaux1.fr *IMB, Université Bordeaux 1 351, cours de la Libération, 33405 Talence cedex, France*

ALGEBRA AND NUMBER THEORY 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.567

Zeros of real irreducible characters of finite groups

Selena Marinelli and Pham Huu Tiep

We prove that if all real-valued irreducible characters of a finite group G with Frobenius–Schur indicator 1 are nonzero at all 2-elements of G, then G has a normal Sylow 2-subgroup. This result generalizes the celebrated Ito–Michler theorem (for the prime 2 and real, absolutely irreducible, representations), as well as several recent results on nonvanishing elements of finite groups.

1. Introduction

Suppose that G is a finite group. Let $\mathrm{Irr}(G)$ be the set of the irreducible complex characters of G, and let $\mathbb{F}$ be a subfield of $\mathbb{C}$. Write $\mathrm{Irr}_{\mathbb{F}}(G)$ for the set of those $\chi \in \mathrm{Irr}(G)$ such that $\chi(g) \in \mathbb{F}$ for all $g \in G$. Hence $\mathrm{Irr}_{\mathbb{R}}(G)$ is the set of *real-valued* (or *real*) irreducible characters of G.

As shown in recent papers [Dolfi et al. 2008; Navarro et al. 2009; Navarro and Tiep 2010], several fundamental results on characters of finite groups admit a version in which $\mathrm{Irr}(G)$ is replaced by $\mathrm{Irr}_{\mathbb{F}}(G)$ for a suitable field $\mathbb{F}$. For instance, S. Dolfi, G. Navarro and P. H. Tiep proved in [Dolfi et al. 2008] that if all $\chi \in \mathrm{Irr}_{\mathbb{R}}(G)$ have odd degree, then a Sylow 2-subgroup of G is normal in G (therefore, providing a strong version of the celebrated Ito–Michler theorem for the prime $p = 2$).

In this paper, we turn our attention to the *nonvanishing* elements of a finite group G. These elements, introduced by M. Isaacs, G. Navarro and T. R. Wolf in [Isaacs et al. 1999], are the $x \in G$ such that $\chi(x) \neq 0$ for all $\chi \in \mathrm{Irr}(G)$. Since their definition, there has been an increasing interest in the set of the nonvanishing elements of finite groups. See for instance [Dolfi et al. 2009; Dolfi et al. 2010c; Dolfi et al. 2010d; Dolfi et al. 2010a; Dolfi et al. 2010b]. One of most relevant results in this area was obtained by S. Dolfi, E. Pacifici, L. Sanus and P. Spiga in [Dolfi et al. 2009], where they proved that if all the p-elements of a finite group G are nonvanishing, then G has a normal Sylow p-subgroup. Since characters of degree not divisible by p cannot vanish on any p-element (by an elementary

Tiep is supported by NSF grants DMS-0901241 and DMS-1201374.
MSC2010: primary 20C15; secondary 20C33.
Keywords: real irreducible character, nonvanishing element, Frobenius–Schur indicator.

argument involving roots of unity — see for instance Lemma 5.1), this result is again an extension of the Ito–Michler theorem.

Recall that the Frobenius–Schur indicator of $\chi \in \mathrm{Irr}(G)$ is 0 if χ is nonreal, ± 1 if χ is real; moreover it is 1 precisely when χ is afforded by a real representation of G.

Our main result in this paper is the following.

Theorem A. *Let G be a finite group. If $\chi(x) \neq 0$ for all real-valued irreducible characters χ of G with Frobenius–Schur indicator* 1 *and all* 2*-elements $x \in G$, then G has a normal Sylow* 2*-subgroup.*

Since odd degree characters do not vanish on 2-elements, Theorem A above provides at the same time a generalization of [Dolfi et al. 2008, Theorem A] and of the $p = 2$ case of [Dolfi et al. 2009, Theorem A]. As an immediate consequence of Theorem A, we obtain the following refinement of the Ito–Michler theorem for the prime 2 and real, absolutely irreducible, representations:

Theorem B. *Let G be a finite group. If $\chi(1)$ is odd for all real-valued irreducible characters χ of G with Frobenius–Schur indicator* 1*, then G has a normal Sylow* 2*-subgroup.*

A few remarks are in order here. First of all, the hypotheses of our Theorem A here are strictly more general than those of [Dolfi et al. 2008, Theorem A]. In Section 5 below, we will describe an interesting family of examples of groups G, having real irreducible characters of even degree, such that all 2-elements of G are nonvanishing. We also mention that in order to obtain the solvable part of Theorem A, we will prove a result guaranteeing the existence of real 2-defect zero characters, which might be of independent interest; see Theorem 2.4.

2. Regular orbits and characters of 2-defect zero

We will need the following result, showing that real characters are remarkably well-behaved across odd sections. As usual, if N is a normal subgroup of a group G and $\theta \in \mathrm{Irr}(N)$, we denote by $I_G(\theta)$ the inertia subgroup of θ in G and by $\mathrm{Irr}(G|\theta)$ the set of the irreducible characters of G that lie over θ. For brevity, we call $\chi \in \mathrm{Irr}(G)$ *strongly real* if the Frobenius–Schur indicator of χ equals 1, and let $\mathrm{Irr}_+(G)$ denote the set of all strongly real irreducible characters of G. Certainly, if $H \leq G$ and $\chi = \lambda^G \in \mathrm{Irr}(G)$ for some $\lambda \in \mathrm{Irr}_+(H)$, then $\chi \in \mathrm{Irr}_+(G)$.

Lemma 2.1. *Let G be a finite group and let $N \lhd G$ with G/N of odd order.*

(i) *If $\theta \in \mathrm{Irr}_{\mathbb{R}}(N)$, then there exists a unique $\chi \in \mathrm{Irr}_{\mathbb{R}}(G|\theta)$.*

(ii) *If $\theta \in \mathrm{Irr}_+(N)$, then there exists a unique $\chi \in \mathrm{Irr}_+(G|\theta)$.*

Proof. Part (i) is [Navarro and Tiep 2008, Corollary 2.2].

For (ii), let $T = I_G(\theta)$. Since $|T/N|$ is odd, θ extends to a real character λ of T by [Navarro and Tiep 2008, Lemma 2.1]. As $\lambda_N = \theta$ is strongly real, the same holds for λ. Now $\chi = \lambda^G$ is irreducible and strongly real. The uniqueness of χ follows from (i). □

Lemma 2.2. *Let $G = N\langle j\rangle$ be a split extension of a normal subgroup N by a subgroup $\langle j\rangle$ of order 2. Suppose that $\alpha \in \operatorname{Irr}(N)$ is of odd degree, and that $\alpha^j = \bar{\alpha} \neq \alpha$. Then α^G is irreducible and strongly real.*

Proof. The irreducibility of α^G is obvious. Let α be afforded by a representation $\Phi : N \to \mathrm{GL}_n(\mathbb{C})$, so that $n = \alpha(1)$ is odd. Then the representations $\Phi^j : x \mapsto \Phi(jxj^{-1})$ and $\Phi^* : x \mapsto {}^t\Phi(x^{-1})$ afford the same character $\bar{\alpha}$, whence $\Phi(jxj^{-1}) = A\,{}^t\Phi(x^{-1})A^{-1}$ for some $A \in \mathrm{GL}_n(\mathbb{C})$. Conjugating by j once more, we see that $A \cdot {}^tA^{-1}$ commutes with $\Phi(x)$ for all $x \in N$. By Schur's lemma, ${}^tA = \kappa A$ for some $\kappa \in \mathbb{C}$. Transposing once more, we get $\kappa^2 = 1$. But $A \in \mathrm{GL}_n(\mathbb{C})$ and n is odd, so $\kappa = 1$, that is, $A = {}^tA$. Now we define $\Psi : G \to \mathrm{GL}_{2n}(\mathbb{C})$ by

$$\Psi(x) = \begin{pmatrix} \Phi(x) & 0 \\ 0 & A\,{}^t\Phi(x^{-1})A^{-1} \end{pmatrix}, \qquad \Psi(xj) = \Psi(x) \cdot \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$$

for all $x \in N$; in particular, $\Psi(xj) = \Psi(j)\Psi(jxj^{-1})$. It is straightforward to check that Ψ is a group homomorphism, and that ${}^t\Psi(g) \cdot M\Psi(g) = M$ for all $g \in G$ and with

$$M := \begin{pmatrix} 0 & A^{-1} \\ {}^tA^{-1} & 0 \end{pmatrix}.$$

Thus the $\mathbb{C}G$-module corresponding to Ψ supports a G-invariant symmetric bilinear form (with Gram matrix M) and affords the character α^G, whence α^G is strongly real. □

Note that the examples with $(G, N, \alpha(1)) = (2\mathsf{S}_7, 2\mathsf{A}_7, 4)$ and with $(G, N, \alpha(1)) = (Q_8, C_4, 1)$ show that one cannot remove any of the assumptions of Lemma 2.2.

A character $\chi \in \operatorname{Irr}(G)$ is said to be of *p-defect zero* for a given prime p if p does not divide $|G|/\chi(1)$. By a fundamental result of R. Brauer [Isaacs 1976, Theorem 8.17], if $\chi \in \operatorname{Irr}(G)$ is a character of p-defect zero, then $\chi(g) = 0$ for every element $g \in G$ such that p divides the order $o(g)$ of g. Next we recall the following result of G. R. Robinson:

Lemma 2.3 [Robinson 1989, Remark 2, p. 254]. *Let G be a finite group and let $\chi \in \operatorname{Irr}(G)$ be a real character of 2-defect zero. Then χ is strongly real.*

Let G be a finite group and let U be a faithful G-module. We recall that a G-orbit $\{u^g \mid g \in G\}$ of G on U is a *regular orbit* if its cardinality is equal to $|G|$ or, equivalently, if $\mathbf{C}_G(u) = 1$.

Theorem 2.4. *Let G be a finite group. Assume that $\mathbf{O}_2(G) = 1$ and that G has a nilpotent normal 2-complement M. Let P be a Sylow 2-subgroup of G and assume that whenever U is a faithful $\mathbb{F}_q[P]$-module, P has a regular orbit on U, where q is a prime dividing $|M|$. Then there exists a strongly real irreducible character $\chi \in \mathrm{Irr}_+(G)$ of 2-defect zero.*

If P is an abelian 2-group, then P has a regular orbit in every faithful action on a module of coprime characteristic. In fact, this is an application of Brodkey's theorem [1963].

We observe, for completeness, that a 2-group P acting faithfully on a module U of characteristic $q \neq 2$ has no regular orbit on U only if q is either a Mersenne or Fermat prime, and P involves a section isomorphic to the dihedral group D_8. This follows from [Manz and Wolf 1993, Theorems 4.4 and 4.8], using Maschke's theorem and standard arguments for passing from irreducible to completely reducible modules.

Theorem 2.4 will be derived from the following result, whose somewhat more technical statement will be needed in the proof of Theorem A.

Theorem 2.5. *Let G be a finite group with a nontrivial Sylow 2-subgroup P. Assume that $\mathbf{O}_2(G) = 1$ and that G has a nilpotent normal 2-complement M. Assume in addition that, whenever U is a faithful $\mathbb{F}_q[P]$-module, P has a regular orbit on U, where q is a prime dividing $|M|$. Then there exist a character $\theta \in \mathrm{Irr}(M)$ and an element $z \in P$, such that $\theta^G \in \mathrm{Irr}(G)$ and $\theta^z = \bar{\theta}$.*

Proof. Let $P \in \mathrm{Syl}_2(G)$. Since $\mathbf{O}_2(G) = 1$, P acts faithfully on M. By coprimality, P acts faithfully on $M/\Phi(M)$, as well. So, by factoring out $\Phi(M)$, we can assume that

$$M = L_1 \times L_2 \times \cdots \times L_k,$$

where each L_i is an irreducible $\mathbb{F}_{q_i}[P]$-module for some prime $q_i \neq 2$. We define $W_i = \mathbf{C}_P(L_i)$ for any $i = 1, \ldots, k$. Observe that W_i is a normal subgroup of P for each i, and that $\bigcap_{i=1}^k W_i = 1$, since P acts faithfully on M.

Now, let $\mathcal{B}$ be a subset of $\{W_1, \ldots, W_k\}$ minimal such that

$$\bigcap_{W \in \mathcal{B}} W = 1.$$

We can assume that $\mathcal{B} = \{W_1, \ldots, W_n\}$ for some $n \leq k$. Thus P acts faithfully on $U = L_1 \times \cdots \times L_n$.

By assumption, for all $i \in \{1, \ldots, n\}$, there exists an element $u_i \in L_i$ such that $\mathbf{C}_P(u_i) = W_i$. So, if we set $u = (u_1, \ldots, u_n) \in U$, it follows that $\mathbf{C}_P(u) = 1$. Now, we consider the dual group $\hat{U} = \mathrm{Irr}(U)$. Since $|U|$ is odd, by [Isaacs 1976, Theorem 13.24], U and $\hat{U}$ are isomorphic as P-modules. Hence there exists $\mu \in \hat{U}$

such that $I_P(\mu)=1$, where $I_P(\mu)$ is the inertia group of μ in P. Consider now, for $1\le j\le n$, the subgroup

$$H_j=\bigcap_{\substack{1\le t\le n\\ t\ne j}} W_t.$$

Note that H_j is a normal subgroup of P and that H_j is not contained in W_j, by the minimality of $\mathcal{B}$. Furthermore, $H_j\cap W_j=1$ for each j. Now, let $z_j\in \mathbf{Z}(P)\cap H_j$ be an involution; such an element certainly exists, as H_j is a nontrivial normal subgroup of P. So, $\mathbf{C}_{L_j}(z_j)$ is a P-submodule of L_j and $\mathbf{C}_{L_j}(z_j)<L_j$ as $z_j\notin W_j$. As L_j is irreducible, it follows that $\mathbf{C}_{L_j}(z_j)=1$. Hence z_j inverts every element of L_j; see, for instance, [Huppert 1998, Theorem 16.9(e)]. Moreover, as $z_j\in H_j$, z_j centralizes L_i for every $i\ne j$, $1\le i, j\le n$.

Let $z=z_1\cdots z_n$, and observe that z inverts every element of U. By the isomorphism of P-modules $U\cong \hat{U}$, then z inverts every irreducible character of U. In particular, $\mu^z=\mu^{-1}=\bar{\mu}$. Now, we can write $M=U\times N$, where N is P-invariant. Let $\theta=\mu\times 1_N\in \operatorname{Irr}(M)$. Then, we have $\theta^z=\overline{\theta}$ and $I_P(\theta)=I_P(\mu)=1$. Thus, by Clifford theory $\theta^G\in\operatorname{Irr}(G)$ and the proof is complete. □

Proof of Theorem 2.4. Clearly, we may assume $P\ne 1$. So, by Theorem 2.5 there exists a character $\theta\in\operatorname{Irr}(M)$ such that $\chi=\theta^G\in\operatorname{Irr}(G)$ and an element $z\in P$ such that $\theta^z=\overline{\theta}$. Hence,

$$\overline{\chi}=\overline{\theta^G}=\overline{\theta}^G=(\theta^z)^G=\theta^G=\chi,$$

so $\chi\in\operatorname{Irr}_{\mathbb{R}}(G)$. Next, since $\chi(1)=|G:M|\,\theta(1)=|P|\,\theta(1)$, χ is a character of 2-defect zero of G. Hence χ is strongly real by Lemma 2.3. □

3. Proof of Theorem A

We will need the following deep result concerning the existence of suitable strongly real characters of almost simple groups. We state it here and prove it in Section 4.

Theorem 3.1. *Let S be any finite nonabelian simple group. For any H with $S\le H\le \operatorname{Aut}(S)$, there exist a character $\theta\in\operatorname{Irr}(S)$ and a 2-element $x\in S$, such that both the following conditions apply*:

(i) $\theta(x^\sigma)=0$ *for all* $\sigma\in\operatorname{Aut}(S)$.

(ii) *There exists a subgroup J with $I_H(\theta)\le J\le H$ and a strongly real character $\alpha\in\operatorname{Irr}(J|\theta)$.*

We can now proceed to prove Theorem A, which we restate below.

Theorem 3.2. *Let G be a finite group and $P\in\operatorname{Syl}_2(G)$. Suppose that $\chi(x)\ne 0$ for all $\chi\in\operatorname{Irr}_+(G)$ and for all 2-elements $x\in G$. Then $P\lhd G$.*

Proof. Let G be a minimal counterexample to the statement; in particular, $P \neq 1$. Let $M \neq 1$ be a minimal normal subgroup of G.

1. Observe that $\mathrm{Irr}_+(G/M) \subseteq \mathrm{Irr}_+(G)$ and 2-elements of G/M lift to 2-elements of G. Hence, $PM \lhd G$ by minimality of G. If M_1 is another minimal normal subgroup of G with $M_1 \neq M$, then $G/M \times G/M_1$ has a normal Sylow 2-subgroup, as both G/M and G/M_1 do. Since $M \cap M_1 = 1$, G is isomorphic to a subgroup of $G/M \times G/M_1$ and hence G has a normal Sylow 2-subgroup, a contradiction. So, M is the only minimal normal subgroup of G.

2. Suppose first that 2 divides $|M|$. If M is solvable, then M is a 2-group and so $P = PM \lhd G$, a contradiction. Hence M is not solvable. Thus $M = S_1 \times \cdots \times S_t$, where $S_i \cong S$, a nonabelian simple group. Write $S := S_1$, $H := \mathbf{N}_G(S)$ and $C := \mathbf{C}_G(S)$. Thus, H/C is isomorphic to a subgroup $\overline{H}$ of $\mathrm{Aut}(S)$, with $S \leq \overline{H} \leq \mathrm{Aut}(S)$. By Theorem 3.1, there exists a character $\theta \in \mathrm{Irr}(S)$ and a 2-element $x \in S$ such that $\theta(x^\sigma) = 0$ for all $\sigma \in \mathrm{Aut}(S)$. Moreover, there exists a subgroup J with $I_{\overline{H}}(\theta) \leq J \leq \overline{H}$ and a strongly real character $\alpha \in \mathrm{Irr}(J|\theta)$. By the Clifford correspondence [Isaacs 1976, Theorem 6.11], $\alpha = \lambda^J$ for a suitable character $\lambda \in \mathrm{Irr}(I_{\overline{H}}(\theta)|\theta)$. Therefore, $\beta := \lambda^{\overline{H}}$ is an irreducible character of $\overline{H}$. Furthermore, β is strongly real as $\beta = (\lambda^J)^{\overline{H}} = \alpha^{\overline{H}}$, and β lies over θ.

Let now $\psi := \theta \times 1_S \times \cdots \times 1_S \in \mathrm{Irr}(M)$. Note that $C \lhd I_G(\psi) \leq H$ and that $I_G(\psi)/C$ is isomorphic to $I_{\overline{H}}(\theta)$. Hence, by lifting characters from the corresponding factor groups modulo C, we can view $\lambda \in \mathrm{Irr}(I_G(\psi)|\psi)$ and $\lambda^H = \beta \in \mathrm{Irr}_+(H)$.

Define $\chi = \lambda^G$. By the Clifford correspondence, χ is an irreducible character of G and, since $\chi = \beta^G$, we have $\chi \in \mathrm{Irr}_+(G)$. We will show that χ vanishes on the 2-element $g = (x, x, \ldots, x) \in M$. In fact, χ lies over ψ and hence the restriction χ_M is a sum of conjugates ψ^y, with $y \in G$. Now, each conjugate ψ^y is of the form

$$\psi^y = 1_S \times \cdots \times 1_S \times \theta^\sigma \times 1_S \times \cdots \times 1_S,$$

for a suitable $\sigma \in \mathrm{Aut}(S)$. Thus $\psi^y(g) = \theta(x^{\sigma^{-1}}) = 0$ for all $y \in G$. It follows that $\chi(g) = 0$, against our assumptions.

3. We have shown that M is an elementary abelian q-group for some prime $q \neq 2$.

Let $Z := \Omega_1(\mathbf{Z}(P))$ so that $Z \neq 1$. Since $|M|$ is odd, $ZM/M = \Omega_1(\mathbf{Z}(PM/M))$ and so $ZM \lhd G$. Observe also that M is a normal nilpotent 2-complement of ZM and that Z is a Sylow 2-subgroup of ZM. Moreover, $\mathbf{O}_2(ZM) = 1$, as $\mathbf{O}_2(ZM)$ is normal in G and M is the unique minimal normal subgroup of G. Finally, since Z is abelian, Z has a regular orbit on every faithful Z-module of odd characteristic. Thus, by Theorem 2.5 there exist $\theta \in \mathrm{Irr}(M)$ and $z \in Z$, such that $\theta^z = \overline{\theta}$ and $\theta^{ZM} \in \mathrm{Irr}(ZM)$. Since $Z \neq 1$ and $q \neq 2$, we must have that $\overline{\theta} \neq \theta$ and $z \neq 1$; in fact z is an involution.

Let $T = I_G(\theta) \cap PM = I_{PM}(\theta)$. Since $q \neq 2$, θ has a *canonical extension* $\gamma \in \mathrm{Irr}(T)$; see [Isaacs 1976, Corollary 8.16]. By uniqueness of the canonical extension of $\overline{\theta}$, it follows that $\gamma^z = \overline{\gamma}$. Let $\delta = \gamma^{PM}$. By Clifford theory δ is irreducible. Since $\theta^z = \overline{\theta} \neq \theta$, we see that $z \notin T$ but z normalizes T. It follows that $K := T\langle j \rangle$ is a split extension of T by $\langle j \rangle$. We have already shown that $\gamma^z = \overline{\gamma}$. Also, $\gamma_M = \theta$ is nonreal. Hence γ is nonreal and has degree 1. Applying Lemma 2.2 to the character γ of T, we see that γ^K is strongly real. Consequently, $\delta = (\gamma^K)^{PM}$ is strongly real.

Recalling that PM is a normal subgroup of odd index in G, by Lemma 2.1(ii) there exists a character $\chi \in \mathrm{Irr}_+(G)$ that lies over δ.

Now, we show that $\delta(g) = 0$ for every $g \in ZM \smallsetminus M$. In fact, as $\theta^{ZM} \in \mathrm{Irr}(ZM)$, by Clifford theory $T \cap ZM = I_{ZM}(\theta) = M$. As both M and ZM are normal in G, we get that for all $x \in G$, $T^x \cap ZM = (T \cap ZM)^x = M^x = M$. So,

$$ZM \cap \Big(\bigcup_{x \in G} T^x\Big) = M.$$

As $\delta = \gamma^{PM}$ with $\gamma \in \mathrm{Irr}(T)$, the formula of character induction yields that $\delta(g) = 0$ for all $g \in ZM \smallsetminus M$.

Note now that, because $ZM > M$, there exists a 2-element $g_0 \in ZM \smallsetminus M$. Finally, observe that χ_{PM} is a sum of conjugates δ^y of δ in G and that $\delta^y(g_0) = \delta(g_0^{y^{-1}}) = 0$, since $g_0^{y^{-1}} \in ZM \smallsetminus M$ for all $y \in G$. Therefore, we conclude that $\chi(g_0) = 0$, with $\chi \in \mathrm{Irr}_+(G)$ and g_0 a 2-element of G, the final contradiction. □

4. Almost simple groups

This section is devoted to proving Theorem 3.1. First we handle some easy cases:

Lemma 4.1. *Theorem 3.1 holds if S is an alternating group, a sporadic simple group, or a simple group of Lie type in characteristic* 2.

Proof. The cases of A_5, A_6, and all the sporadic groups can be verified directly using [Conway et al. 1985]. Assume $S = \mathsf{A}_n$ with $n \geq 7$; in particular $\mathrm{Aut}(S) \cong \mathsf{S}_n$. As shown in [Dolfi et al. 2009, Proposition 2.4], one can find a character θ satisfying the conditions described in Theorem 3.1, which extends to a strongly real character α of $\mathrm{Aut}(S)$.

If $S = {}^2F_4(2)'$, then we can choose $J = H$ and $\theta \in \mathrm{Irr}(H)$ of degree 2048 if $H = S$ and of degree 4096 if $H = \mathrm{Aut}(S)$ (and $x \neq 1$ any 2-element in S); see [Conway et al. 1985]. The case $\mathrm{Sp}_4(2)' \cong \mathsf{A}_6$ has been considered above. For all other simple groups of Lie type in characteristic 2, we choose θ to be the Steinberg character St and $1 \neq x \in S$ to be any 2-element: it is well-known [Feit 1993] that St vanishes at any 2-singular element and extends to the character of a rational representation of H. □

Lemma 4.2. *Let G be a finite group with a normal subgroup N, and $\chi \in \mathrm{Irr}(G)$. Then χ_N is irreducible (over N) if and only if the characters $\chi\alpha$, where $\alpha \in \mathrm{Irr}(G/N)$, are all irreducible and pairwise distinct. Moreover, in this case the irreducible characters of G that lie above χ_N are precisely the characters $\chi\alpha$, where $\alpha \in \mathrm{Irr}(G/N)$.*

Proof. The "only if" direction is Gallagher's theorem [Isaacs 1976, Theorem 6.17]. For the "if" direction, observe that the hypothesis implies

$$(\chi_N)^G = \chi \cdot (1_N)^G = \sum_{\alpha \in \mathrm{Irr}(G/N)} \alpha(1)\chi\alpha$$

contains χ with multiplicity 1, and so $[\chi_N, \chi_N]_N = [\chi, (\chi_N)^G]_G = 1$, as stated. □

In the rest of this section, let S be a simple group of Lie type in characteristic $p > 2$. We can find a simple algebraic group $\mathcal{G}$ of adjoint type defined over a field of characteristic p and a Frobenius morphism $F : \mathcal{G} \to \mathcal{G}$ such that $S = [G, G]$ for $G := \mathcal{G}^F$. We refer to [Carter 1985; Digne and Michel 1991] for basic facts on the Deligne–Lusztig theory of complex representations of finite groups of Lie type. In particular, irreducible characters of G are partitioned into (rational) Lusztig series that are labeled by conjugacy classes of semisimple elements s in the dual group L, where the pair $(\mathcal{L}, F^*)$ is dual to $(\mathcal{G}, F)$ and $L = \mathcal{L}^{F^*}$. Since $\mathcal{L}$ is simply connected, $\mathbf{C}_{\mathcal{L}}(s)$ is connected for any semisimple element $s \in \mathcal{L}$; see [Carter 1985, Theorem 3.5.6]. Hence the L-conjugacy class s^L corresponds to a (unique) irreducible (semisimple) character χ_s of G of degree $[L : \mathbf{C}_L(s)]_{p'}$; see [Digne and Michel 1991, §14]. Since χ_s belongs to the Lusztig series defined by s^L, two semisimple characters χ_s and χ_t are equal precisely when s and t are conjugate in L.

The structure of $\mathrm{Aut}(S)$ is described in [Gorenstein et al. 1994, Theorem 2.5.12]; in particular, it is a split extension of G by an abelian group (of field and graph automorphisms), which we denote by $A(S)$.

Our arguments will rely on the following proposition, which is of independent interest:

Proposition 4.3. *In the notation above, assume that $s \in L$ is a semisimple element of order coprime to $|\mathbf{Z}(L)|$. Then the following statements hold.*

(i) *If s is real in L then χ_s is strongly real.*

(ii) *Let σ be a bijective morphism of the algebraic group $\mathcal{G}$ commuting with F and let σ^* be the corresponding morphism of $\mathcal{L}$. Assume that χ_s is σ-invariant. Then s and $\sigma^*(s)$ are L-conjugate. Moreover, if σ is a Frobenius morphism, then s is $\mathcal{L}$-conjugate to some element in $\mathcal{L}^{\sigma^*}$; in particular, $|s|$ divides $|\mathcal{L}^{\sigma^*}|$.*

(iii) *$\theta := (\chi_s)_S$ is irreducible (over S).*

(iv) *Let σ be a bijective morphism of $\mathcal{G}$ commuting with F (and so fixing G and S). Suppose that σ fixes θ. Then σ fixes χ_s.*

Proof. Part (ii) and the statement about χ_s being real in (i) can be proved exactly in the same way as [Dolfi et al. 2008, Lemma 2.5] using [Navarro et al. 2008, Corollary 2.5]. Assume now that s is real. Since $\mathcal{G}$ is of adjoint type, $\mathbf{Z}(\mathcal{G}) = 1$; in particular, it is connected, and $\mathbf{Z}(G) = \mathbf{Z}(\mathcal{G})^F = 1$ by [Carter 1985, Proposition 3.6.8]. Hence, by [Vinroot 2010, Theorem 4.2], the Frobenius–Schur indicator of χ_s is 1, as stated in (i).

For (iii), by [Digne and Michel 1991, Proposition 13.30] and its proof, the characters $\alpha \in \mathrm{Irr}(G/S)$ are exactly the semisimple characters χ_z with $z \in \mathbf{Z}(L)$; moreover, $\chi_{sz} = \chi_s \chi_z$. Observe that sz and st are not L-conjugate if $z, t \in \mathbf{Z}(L)$ are distinct. (Indeed, suppose $g(sz)g^{-1} = st$ for some $g \in L$. Then since $|s|$ is coprime to $|\mathbf{Z}(L)|$, we have

$$|s| = |gsg^{-1}| = |s \cdot (tz^{-1})| = |s| \cdot |tz^{-1}|,$$

and so $z = t$.) It follows that the characters χ_{sz} are all irreducible and pairwise distinct. By Lemma 4.2, $\theta = (\chi_s)_S$ is irreducible, and $\mathrm{Irr}(G|\theta) = \{\chi_{sz} \mid z \in \mathbf{Z}(L)\}$.

Suppose now that σ fixes θ as in (iv). Since σ fixes G, it now fixes $\mathrm{Irr}(G|\theta)$ and so it sends χ_s to χ_{sz} for some $z \in \mathbf{Z}(L)$. Let σ^* be the morphism of $\mathcal{L}$ corresponding to σ. By [Navarro et al. 2008, Corollary 2.5], sz and $\sigma^*(s)$ are L-conjugate. In particular, $|s| = |\sigma^*(s)| = |sz| = |s| \cdot |z|$, and so $z = 1$ as stated. □

Proposition 4.4. *Theorem 3.1 holds if S is one of the following simple groups in characteristic $p > 2$: $G_2(q)$, ${}^2G_2(q)$, ${}^3D_4(q)$, $F_4(q)$, or $E_8(q)$, where $q = p^f$.*

Proof. Notice that in each of these cases, $\mathrm{Out}(S) = A(S)$ is cyclic, of order $2f$ if $S = G_2(q)$ and $p = 3$, of order $3f$ if $S = {}^3D_4(q)$, and of order f otherwise; see for instance [Gorenstein et al. 1994, Theorem 2.5.12]. Furthermore, $S = G \cong L$; see [Carter 1985, p. 120]. Choose the integer m to be 6, 12, 12, or 30, if $S = {}^2G_2(q)$, ${}^3D_4(q)$, $F_4(q)$, or $E_8(q)$, respectively. If $S = G_2(q)$, we choose $m = 3$ if $q = 3^f$ with f odd, and $m = 6$ otherwise. By [Zsigmondy 1892], there exists a *primitive prime divisor* (p.p.d.) $r = r(p, mf)$ of $p^{mf} - 1$, that is, a prime divisor of $p^{mf} - 1$ that does not divide $\prod_{i=1}^{mf-1}(p^i - 1)$. According to [Moretó and Tiep 2008, Lemma 2.3], L contains a semisimple element s of order r for which $\mathbf{C}_L(s)$ is a torus of order dividing $\Phi_m(q)$ if $\Phi_m(t)$ is the m-th cyclotomic polynomial in t; in particular, s is regular. It is well-known [Tiep and Zalesski 2005, Proposition 3.1] that every semisimple element $s \in L$ is real. It then follows by Proposition 4.3(i) that $\theta := \chi_s$ is strongly real.

We claim that χ_s is not stable under any nontrivial outer automorphism σ of S. Indeed, since $\mathrm{Aut}(S)$ is a split extension of S by the cyclic group $\mathrm{Out}(S)$ in the cases under consideration [Gorenstein et al. 1994, Theorem 2.5.12] we can choose σ to be a power σ_0^e of some canonical outer automorphism σ_0 of S.

(a) If $S = {}^2G_2(q)$, $G_2(q)$ (with $p \neq 3$), $F_4(q)$, or $E_8(q)$, then σ_0 is induced by the field automorphism $x \mapsto x^p$, and so we can choose e such that $1 \leq e < f$ and $e \mid f$. In this case, $|\mathscr{L}^{\sigma^*}|$ is equal to the order of $|L|$, but with q replaced by p^e, and hence is not divisible by r by the choice of r.

(b) Now suppose that $S = G_2(q)$ and $p = 3$. Then $1 \leq e < 2f$, $e \mid 2f$, and $|\mathscr{L}^{\sigma^*}|$ equals $|G_2(p^{e/2})|$ if $2 \mid e$ and $|{}^2G_2(p^e)|$ if e is odd. In either case, r is coprime to $|\mathscr{L}^{\sigma^*}|$ by the choice of r.

(c) Finally, consider the case $S = {}^3D_4(q)$. Then we can choose e so that $1 \leq e < 3f$ and $e \mid 3f$. Now $|\mathscr{L}^{\sigma^*}|$ equals $|D_4(p^e)|$ and so r is coprime to $|\mathscr{L}^{\sigma^*}|$ by the choice of r.

We have shown that $I_{\mathrm{Aut}(S)}(\chi_s) = S$, whence $I_H(\chi_s) = S$. Since $\chi_s(1) = [L : \mathbf{C}_L(s)]_{p'}$ and $|\mathbf{C}_L(s)|$, being a divisor of $\Phi_m(q)$, is odd, we see that χ_s has 2-defect zero and so χ_s vanishes at any 2-element $1 \neq x \in S$. Hence we are done by taking $J = S$ and $\alpha = \chi_s$. □

Proposition 4.5. *Theorem 3.1 holds if S is any of the following simple groups of Lie type in characteristic $p > 2$: $\mathrm{PSL}_2(q)$ with $q \geq 5$; $\mathrm{PSp}_{2n}(q)$ with $n \geq 2$; $\Omega_{2n+1}(q)$ with $n \geq 3$; $\mathrm{P}\Omega^{\epsilon}_{2n}(q)$ with $2 \mid n$, $n \geq 6$ for $\epsilon = +$, and $n \geq 4$ for $\epsilon = -$; or $E_7(q)$.*

Proof. **1.** Recall that $L = \mathrm{SL}_2(q)$, respectively $\mathrm{Spin}_{2n+1}(q)$, $\mathrm{Sp}_{2n}(q)$, $\mathrm{Spin}^{\epsilon}_{2n}(q)$, or $E_7(q)_{sc}$ in the described cases; in particular, $\mathbf{Z}(L)$ is a 2-group. We write $q = p^f$ as usual. By [Tiep and Zalesski 2005, Proposition 3.1], any semisimple element in L is real. Now we choose a semisimple element $s \in L$ of (odd) order r, where r is selected as follows.

(i) If $L = \mathrm{SL}_2(q)$, then $r = (q+\epsilon)/2$ if $\epsilon = \pm 1$ is chosen so that $q \equiv \epsilon \pmod 4$.

(ii) Next, $r = r(p, 2nf)$ is a p.p.d. of $p^{2nf} - 1$ in the other classical cases, unless $L = \mathrm{Spin}^+_{2n}(q)$.

(iii) In the case $L = E_7(q)$, $r = r(p, 18f)$.

(iv) In the remaining case, $L = \mathrm{Spin}^+_{2n}(q)$ contains a central product
$$C = \mathrm{Spin}^-_4(q) * \mathrm{Spin}^-_{2n-4}(q),$$
and we choose $s = s_1 s_2 \in C$ where $s_1 \in \mathrm{Spin}^-_4(q) \cong \mathrm{SL}_2(q^2)$ has order $(q^2+1)/2$ and $s_2 \in \mathrm{Spin}^-_{2n-4}(q)$ has order $(q^{n-2}+1)/2$. More precisely, if β and γ denote some elements in $\bar{\mathbb{F}}_q^\times$ of orders $(q^2+1)/2$ and $(q^{n-2}+1)/2$, respectively, then we can choose s to act on the natural $\mathscr{L}$-module $\bar{\mathbb{F}}_q^{2n}$ with spectrum $\{\beta^i, \beta^{-i} \mid i = 1, q\} \cup \{\gamma^{q^j}, \gamma^{-q^j} \mid 0 \leq j \leq n-3\}$.

In these cases, it is straightforward to check (see for instance [Moretó and Tiep 2008, Lemmas 2.3 and 2.4]) that s is a regular semisimple element, and $T^* := \mathbf{C}_L(s)$ is a maximal torus of order $q+\epsilon$, q^n+1, $\Phi_2(q)\Phi_{18}(q)$, or $(q^2+1)(q^{n-2}+1)$, respectively. Hence, χ_s is a strongly real irreducible character of G, and in fact $\chi_s = \pm R^G_{T,\vartheta}$ is a Deligne–Lusztig character corresponding to some maximal torus T of G in duality with T^*; in particular, $|T| = |T^*|$. (Indeed, since $\mathcal{T}^* := \mathbf{C}_{\mathcal{L}}(s)$ is a torus, this is the unique torus containing s. Now if $(\mathcal{T}, \vartheta)$ is in duality with $(\mathcal{T}^*, s)$, then $T = \mathcal{T}^F$ and $\chi_s = \pm R^G_{T,\vartheta}$.) Now the character formula [Carter 1985, Theorem 7.2.8] shows that $\chi_s(x) = 0$ for any semisimple element $x \in G$ with $|\mathbf{C}_G(x)|$ *not* divisible by $|T|$.

2. By Proposition 4.3(iii), $\theta := (\chi_s)_S$ is irreducible and strongly real. Furthermore, when $S = \mathrm{PSL}_2(q)$, we have $\theta(1) = q-\epsilon$ and so θ has 2-defect 0, whence it vanishes at any nontrivial 2-element $x \in S$. In the remaining cases, we will find an involution $x \in S$ such that $|\mathbf{C}_G(x)|$ is not divisible by $|T|$. If $S = \mathrm{PSp}_{2n}(q)$, choose x to be an involution with centralizer of type $\mathrm{Sp}_2(q)\times\mathrm{Sp}_{2n-2}(q)$ (in $\mathrm{Sp}_{2n}(q)$). If $S = \Omega_{2n+1}(q)$, choose x to be an involution with centralizer of type $\mathrm{GO}^+_4(q) \times \mathrm{GO}_{2n-3}(q)$ (in $\mathrm{GO}_{2n+1}(q)$). For $S = \mathrm{P}\Omega^\epsilon_{2n}(q)$, choose x to be an involution with centralizer of type $\mathrm{GO}^+_4(q) \times \mathrm{GO}^\epsilon_{2n-4}(q)$ (in $\mathrm{GO}^\epsilon_{2n}(q)$). Finally, for S of type $E_7(q)$, choose x to be an involution with centralizer of type $\mathrm{SL}_2(q) * \mathrm{Spin}_{16}(q)$; see [Gorenstein et al. 1994, Table 4.5.1]. It is straightforward to check that $|\mathbf{C}_G(x)|$ is not divisible by $|T|$ for the chosen element x. Then for any $\sigma \in \mathrm{Aut}(S)$, $|\mathbf{C}_G(x^\sigma)| = |\mathbf{C}_G(x)|$ (as $G \lhd \mathrm{Aut}(S)$), whence $\theta(x^\sigma) = 0$.

3. Next we show that any automorphism $\sigma \in \mathrm{Aut}(S)$ that fixes θ must belong to G. Since $\mathrm{Aut}(S) = G : A(S)$ and G fixes $\theta = (\chi_s)_S$, we may assume $\sigma \in A(S)$. Recall that $|A(S)| = 2f$ if $S = \mathrm{P}\Omega^\epsilon_{2n}(q)$ and $|A(S)| = f$ otherwise. Let $\sigma_0 \in A(S)$ denote the automorphism of S (and of G, $\mathcal{G}$) induced by the field automorphism $y \mapsto y^p$. If $S = \mathrm{P}\Omega^\epsilon_{2n}(q)$, we denote by $\tau \in A(S)$ the nontrivial graph automorphism commuting with F (otherwise set $\tau = 1_S$). Notice that $G = \mathcal{G}^F$ with $F = \sigma_0^f$, unless $S = \mathrm{P}\Omega^-_{2n}(q)$ in which case $F = \tau\sigma_0^f$. Then $A(S)$ is generated by σ_0 and τ. It follows that σ can be extended to a Frobenius morphism of $\mathcal{G}$, which commutes with F, unless $\sigma = \tau$ and $S = \mathrm{P}\Omega^+_{2n}(q)$. Replacing σ by $\tau\sigma_0^f$ in the latter case, we again see that σ extends to a Frobenius morphism of $\mathcal{G}$ that commutes with F. Since σ fixes θ, σ fixes χ_s by Proposition 4.3(iv), which in turn implies that $|s|$ divides $|\mathcal{L}^{\sigma^*}|$ by Proposition 4.3(ii).

3a. First consider the case $\sigma = \sigma_0^e$. Then $|\mathcal{L}^{\sigma^*}|$ equals the order of L but with q replaced by p^e; denote it by $|L(p^e)|$. Suppose $S = \mathrm{PSL}_2(q)$; in particular $A(S) = \langle\sigma_0\rangle \cong C_f$ and so we may choose $e \mid f$. If $q \equiv 1 \pmod 4$, we get $|s| = (p^f+1)/2$ divides $p^{2e}-1$, which is possible only when $e = f$. If $q \equiv -1 \pmod 4$, then f is odd; hence $|s| = (p^f-1)/2$ can divide $p^{2e}-1$ only when

$e = f$. Next suppose $S = \mathrm{PSp}_{2n}(q)$ or $\Omega_{2n+1}(q)$. Then $|s| = r(p, 2nf)$ can divide $|L(p^e)|$ only when $f \mid e$. If S is of type $E_7(q)$, then $|s| = r(p, 18f)$ can divide $|L(p^e)|$ only when $f \mid e$. In all these cases, $A(S) = \langle \sigma_0 \rangle \cong C_f$, so we conclude that $\sigma_S = 1_S$. Consider the case $S = \mathrm{P}\Omega^-_{2n}(q)$. Since $|s| = r(p, 2nf)$ divides $|L(p^e)|$, we get $f \mid e$. Recall that $A(S) = \langle \sigma_0 \rangle \cong C_{2f}$ for $S = \mathrm{P}\Omega^-_{2n}(q)$, so we may assume $e \mid (2f)$. Now if $2f \mid e$ then $\sigma_S = 1_S$. On the other hand, if $f = e$, then $|s| = r(p, 2nf)$ does not divide $|L(p^e)|$.

Assume now that $S = \mathrm{P}\Omega^+_{2n}(q)$. Since the order of $(\sigma_0)_S$ is f, we may assume that $0 \le e \le f/2$. Observe that $|s|$ is divisible by some p.p.d. $r_1 = r(p, (2n-4)f)$. Now since r_1 divides $|L(p^e)|$, we get $(2n-4)f \mid je$ for some j, $1 \le j \le 2n-2$. But then $je \le (n-1)f < (2n-4)f$, so $e = 0$ and $\sigma_S = 1_S$.

3b. It remains to consider the case σ is not contained in $\langle \sigma_0 \rangle$. This can happen only when $S = \mathrm{P}\Omega^+_{2n}(q)$. Since $(\sigma_0)^f$ acts trivially on S, we can write $\sigma = \tau\sigma_0^e$ with $1 \le e \le f$. Moreover, replacing σ by $\sigma^{-1} = \tau\sigma_0^{f-e}$ (while acting on S), we may in fact assume that $1 \le e \le f/2$ or $e = f$. Now $r_1 = r(p, (2n-4)f)$ divides $|s|$ and $|s|$ divides $|\mathscr{L}^{\sigma^*}|$, and so we get $(2n-4)f \mid 2je$ for some j with $1 \le j \le n$. It follows that $e \ge (n-2)f/n > f/2$ as $n \ge 6$. We have shown that $e = f$, and so by Proposition 4.3(ii), s is $\mathscr{L}$-conjugate to some element $s' \in \mathscr{L}^{\sigma^*} = \mathrm{Spin}^-_{2n}(q)$. Certainly, $|s'| = |s|$ is divisible by $r_1 = r(p, (2n-4)f)$. Observe that the r_1-part of s' has centralizer of type $\mathrm{GO}^+_4(q) \times \mathrm{GO}^-_{2n-4}(q)$ (in $\mathrm{GO}^-_{2n}(q)$). Hence, the action of s' on the natural $\mathscr{L}$-module $V = \overline{\mathbb{F}}_q^{2n}$ is induced by $\mathrm{diag}(A, B)$ with $A \in \mathrm{GO}^+_4(q)$ and $B \in \mathrm{GO}^-_{2n-4}(q)$. But in this case, the spectrum of s' and s on V cannot have the shape indicated in (iv) above.

We have shown that $I_{\mathrm{Aut}(S)}(\theta) = G$; in particular, if $S \le H \le \mathrm{Aut}(S)$, we have $I_H(\theta) = G \cap H$. Choosing $J = G \cap H$ and $\alpha = (\chi_s)_J$, we are done. $\square$

Lemma 4.6. *Let $\mathscr{L}$ be a simple simply connected algebraic group of type A_n with $n \ge 2$, D_n with $n \ge 3$ odd, or E_6, $F : \mathscr{L} \to \mathscr{L}$ a Frobenius morphism, and let $L := \mathscr{L}^F$. Let $\varphi \in \mathrm{Aut}(L)$ be a (nontrivial) graph automorphism of L (modulo inner-diagonal automorphisms). Then $\varphi(s)$ and s^{-1} are conjugate in L for any semisimple element $s \in L$.*

Proof. It is well-known [Steinberg 1968, §10] that such an automorphism φ lifts to an automorphism $\varphi = \psi\tau$ of $\mathscr{L}$, where ψ is inner: $\psi(x) = gxg^{-1}$ for some $g \in \mathscr{L}$, and τ acts as the inversion $t \mapsto t^{-1}$ on some maximal torus $\mathscr{T}$ of $\mathscr{L}$. Since s is semisimple, $s = hth^{-1}$ for some $t \in \mathscr{T}$ and $h \in \mathscr{L}$. Thus

$$\varphi(s) = \psi\tau(hth^{-1}) = g\tau(h)t^{-1}\tau(h)^{-1}g^{-1} = zs^{-1}z^{-1},$$

where $z := g\tau(h)h^{-1} \in \mathscr{L}$. Since s and $\varphi(s)$ are F-stable, we see $z^{-1}F(z) \in \mathbf{C}_{\mathscr{L}}(s)$. But $\mathscr{L}$ is simply connected; hence $\mathbf{C}_{\mathscr{L}}(s)$ is connected and F-stable. Therefore, by the Lang–Steinberg theorem, there is $c \in \mathbf{C}_{\mathscr{L}}(s)$ such that $z^{-1}F(z) = c^{-1}F(c)$, that is, $u := zc^{-1} \in L$. It follows that $\varphi(s) = zs^{-1}z^{-1} = us^{-1}u^{-1}$ is L-conjugate to s^{-1}. $\square$

Proposition 4.7. *Theorem 3.1 holds if S is any of the following simple groups of Lie type in characteristic $p > 2$: $\mathrm{PSL}_n(q)$ with $n \geq 3$, $\mathrm{PSU}_n(q)$ with $n \geq 5$ odd, $\mathrm{P}\Omega^{\epsilon}_{2n}(q)$ with $n \geq 5$ odd, $E_6(q)$, or ${}^2E_6(q)$.*

Proof. Recall that $L = \mathrm{SL}_n(q)$, $\mathrm{SU}_n(q)$, $\mathrm{Spin}^{\epsilon}_{2n}(q)$, $E_6(q)_{sc}$, or ${}^2E_6(q)_{sc}$ in the described cases, respectively, and we write $q = p^f$ as usual. In all these cases, S, G, and L have an outer automorphism that lifts to an involutive graph automorphism τ of $\mathcal{L}$ mentioned in the proof of Lemma 4.6. In particular, $\tau(X) = {}^tX^{-1}$ in the SL and SU cases, and τ acts on $S = \mathrm{P}\Omega^{\epsilon}_{2n}(q)$ as a conjugation by some element $X \in \mathrm{GO}^{\epsilon}_{2n}(q) \smallsetminus \mathrm{SO}^{\epsilon}_{2n}(q)$. Also recall that $G \lhd \mathrm{Aut}(S)$ and τG generates a subgroup of order 2 in $\mathrm{Aut}(S)/G \cong A(S)$. Our proof will be divided in two cases according to whether the subgroup GH of $\mathrm{Aut}(S)$ contains $\langle G, \tau\rangle$ or not. In the former case, we will choose $\theta = \chi_S$ with $\chi \in \mathrm{Irr}(G)$ being nonreal and use τ to produce a real character for some subgroup $J > I_H(\theta)$. In the latter case, we choose $\theta = \chi_S$ with $\chi \in \mathrm{Irr}(G)$ being real and with $I_H(\theta) \leq G$. In fact, we also consider $\mathrm{PSU}_n(q)$ with $n \geq 4$ even in all parts of this proof, except in part 6 below. Moreover, even though the case of $\mathrm{PSU}_n(q)$ with $n \geq 5$ odd is also handled in Proposition 4.8 (below) using a different method, we also treat it here, since the character χ constructed here in this case will be used in some of our other works.

Case I (GH does not contain $\langle G, \tau\rangle$). **1.** We will construct $\chi \in \mathrm{Irr}(G)$, $\theta = \chi_S$, and $x \in S$ as follows.

Case Ia. Suppose $S = \mathrm{PSL}_3(q)$ and f is odd. Then $A(S) \cong C_{2f}$ contains a unique involution τ and $|H/(H \cap G)|$, $|(H \cap G)/S|$ are odd. In this case, we choose χ to be the unipotent (Weil) character of G of degree $q(q+1)$ and $x \in S$ to be *any* element of order $(q^2-1)_2$. Note that $\chi + 1_G$ is just the permutation character of G acting on the 1-spaces of the natural $\mathrm{GL}_3(q)$-module $\mathbb{F}_q^3$. It follows that χ is strongly real, $\theta = \chi_S$ is irreducible, and $\chi(x) = 0$. By Lemma 2.1(ii), θ extends to a strongly real character of $J := I_H(\theta)$.

Now we may assume that we are not in the case (Ia), and choose a semisimple element $s \in L$ of order r as follows.

Case Ib. Suppose that $S = \mathrm{PSL}^{\epsilon}_n(q)$, where either $n \geq 4$, or $(n, \epsilon) = (3, +)$ and $2|f$. Choose $m \in \{n, n-1\}$ to be *even* and $r = r(p, mf)$ a p.p.d. of $p^{mf}-1$. Note that our hypothesis on n and f guarantees that r exists, and furthermore, r is coprime to $|\mathbf{Z}(L)| = \gcd(n,\ q-\epsilon)$. Embed $\mathrm{Sp}_m(q)$ in $L = \mathrm{SL}^{\epsilon}_n(q)$ and choose $s \in \mathrm{Sp}_m(q)$ of order r. One can check that $|\mathbf{C}_L(s)| = (q^{m/2}+1)^2(q-\epsilon)^{n-m-1}$ if $\epsilon = -$ and $m \equiv 2 \pmod 4$, and $|\mathbf{C}_L(s)| = (q^m-1)(q-\epsilon)^{n-m-1}$ otherwise.

Case Ic. Next suppose that $S = \mathrm{P}\Omega_{2n}^{\epsilon}(q)$. Then choose $r = r(p, (2n-2)f) > 2$, a p.p.d. of $p^{(2n-2)f} - 1$, and $s \in \mathrm{Spin}_{2n-2}^{-}(q) < L$ of order r. By [Moretó and Tiep 2008, Lemma 2.4], $|\mathbf{C}_L(s)| = (q^{n-1}+1)(q+\epsilon)$.

Case Id. In the case $L = E_6^{\epsilon}(q)_{sc}$ (where $\epsilon = +$ for E_6 and $\epsilon = -$ for ${}^2E_6(q)$), we choose $s \in F_4(q) < L$ of order $r = r(p, 12f) \geq 13$. By [Moretó and Tiep 2008, Lemma 2.3], $|\mathbf{C}_L(s)| = \Phi_{12}(q) \cdot (q^2 + q\epsilon + 1)$.

In Cases Ib–Id, it is straightforward to check that s is a regular semisimple element; furthermore, s is real by [Tiep and Zalesski 2005, Proposition 3.1]. Hence, $\chi = \chi_s$ is a strongly real irreducible character of G, and, arguing as in part (1) of the proof of Proposition 4.5, we see that $\chi_s(x^{\sigma}) = 0$ for all $\sigma \in \mathrm{Aut}(S)$, whenever $x \in G$ is any semisimple element with $|\mathbf{C}_G(x)|$ *not* divisible by $|T| = |\mathbf{C}_L(s)|$. Also, by Proposition 4.3(iii), $\theta := (\chi_s)_S$ is irreducible.

2. Observe that, when $L = E_6^{\epsilon}(q)$, χ and θ have 2-defect 0, whence they vanish at any 2-element $1 \neq x \in S$. In the remaining cases, we now find a 2-element $x \in S$ such that $|\mathbf{C}_G(x)|$ is not divisible by $|T|$. If $L = \mathrm{SL}_n^{\epsilon}(q)$, we choose x represented by $\mathrm{diag}(x_1, I_{n-2}) \in \mathrm{SL}_2(q) \times \mathrm{SL}_{n-2}^{\epsilon}(q)$ with $|x_1| = 4$. One can then check that $|\mathbf{C}_G(x)| = |\mathrm{GL}_{n-2}^{\epsilon}(q)| \cdot (q - \alpha)$ with $\alpha = \pm 1$ chosen such that $4 \mid (q-\alpha)$, whence $|\mathbf{C}_G(x)|$ is not divisible by $|T|$. Finally, if $L = \mathrm{Spin}_{2n}^{\epsilon}(q)$, then we choose x to be an involution with centralizer of type $\mathrm{GO}_4^{+}(q) \times \mathrm{GO}_{2n-4}^{\epsilon}(q)$ (in $\mathrm{GO}_{2n}^{\epsilon}(q)$). It is easy to see that $|\mathbf{C}_G(x)|$ is not divisible by r for the chosen element x. Thus for all $\sigma \in \mathrm{Aut}(S)$, $\theta(x^{\sigma}) = 0$, as required in Theorem 3.1(i).

3. It remains to show that $I_H(\theta) \leq G$ and so θ extends to the strongly real character $\alpha = \chi_J$ of $J = I_H(\theta) = G \cap H$. Since G fixes $\theta = \chi_S$ and GH does not contain $\langle G, \tau \rangle$, it suffices to show that $I_{\mathrm{Aut}(S)}(\theta) = \langle G, \tau \rangle$. Consider any automorphism $\sigma \in \mathrm{Aut}(S)$ that fixes θ. Since $\mathrm{Aut}(S) = G : A(S)$, we may assume $\sigma \in A(S)$, and so in the notation of the proof of Proposition 4.5, we may write $\sigma = \tau^i (\sigma_0)^e$ with $i, e \geq 0$. Since σ fixes θ, σ fixes χ_s by Proposition 4.3(iv), which in turn implies that s and $\sigma^*(s)$ are L-conjugate by Proposition 4.3(ii). But s is real, and $\tau(s)$ is L-conjugate to s^{-1} by Lemma 4.6. Hence, replacing σ by σ^{-1} if necessary, we may assume that $\sigma = (\sigma_0)^e$, where $0 \leq e \leq f/2$ in the (untwisted) cases of SL, Spin^{+}, and E_6. In the (twisted) cases of SU, Spin^{-}, and ${}^2E_6$, since τ acts on S as σ_0^f, replacing σ by σ^{-1} we may assume that $\sigma = (\sigma_0)^e$ with $0 \leq e \leq 2f/3$. Also, $r = |s|$ divides $|\mathcal{L}^{\sigma^*}|$ by Proposition 4.3(ii). In either case, we can now check that this can happen only when $e = 0$, that is, $\sigma \in \langle G, \tau \rangle$.

Case II (GH contains $\langle G, \tau \rangle$). **4.** In this case, we will choose a semisimple element $s \in L$ of order r as follows.

Case IIa. Suppose that $S = \mathrm{PSL}_n(q)$. Choose $m \in \{n, n-1\}$ to be *odd* (so $m \geq 3$) and $r_1 = r(p, mf)$ a p.p.d. of $p^{mf} - 1$. Furthermore, choose $r_2 = 1$ if f is odd, and $r_2 = r(p, mf/2)$ a p.p.d. of $p^{mf/2} - 1$ if $2 \mid f$. Then $r = r_1 r_2$ is coprime to $|\mathbf{Z}(L)| = \gcd(n,\ q-1)$. Since $L \geq \mathrm{SL}_m(q)$ contains a cyclic subgroup of order $(q^m - 1)/(q-1)$, we can find a semisimple element $s \in L$ of order r. One can check that $|\mathbf{C}_L(s)| = (q^m - 1)(q-1)^{n-m-1}$.

Case IIb. Suppose that $S = \mathrm{PSU}_n(q)$ with $n \geq 3$. Choose $m \in \{n, n-1\}$ to be *odd* (so $m \geq 3$) and $r = r(p, 2mf)$ a p.p.d. of $p^{2mf} - 1$; in particular, r is coprime to $|\mathbf{Z}(L)| = \gcd(n,\ q+1)$. Now we can find a semisimple element $s \in L$ of order r, with $|\mathbf{C}_L(s)| = (q^m + 1)(q+1)^{n-m-1}$.

Case IIc. Suppose that $S = \mathrm{P}\Omega^+_{2n}(q)$. Choose $r_1 = r(p, nf)$ to be a p.p.d. of $p^{nf} - 1$. Furthermore, choose $r_2 = 1$ if f is odd, and $r_2 = r(p, nf/2)$ a p.p.d. of $p^{nf/2} - 1$ if $2 \mid f$, and set $r = r_1 r_2$. Since $\mathrm{SO}^+_{2n}(q) > \mathrm{GL}_n(q)$ contains a cyclic subgroup of order $q^n - 1$, we can find a semisimple element $s \in L$ of (odd) order r. One can check that $|\mathbf{C}_L(s)| = q^n - 1$.

Case IId. Assume now that $S = \mathrm{P}\Omega^-_{2n}(q)$. Then choose $r = r(p, 2nf)$ to be a p.p.d. of $p^{2nf} - 1$. Since n is odd, $\mathrm{GO}^-_{2n}(q) > \mathrm{GU}_n(q)$ contains a cyclic subgroup of order $q^n + 1$, and so we can find a semisimple element $s \in L$ of order r, with $|\mathbf{C}_L(s)| = q^n + 1$.

Case IIe. Next suppose that $L = E_6(q)_{sc}$. Then choose $r_1 = r(p, 9f)$, a p.p.d. of $p^{9f} - 1$, and choose $r_2 = 1$ if f is odd, and $r_2 = r(p, 9f/2)$, a p.p.d. of $p^{9f/2} - 1$ if $2 \mid f$. Then $r = r_1 r_2$ is coprime to $|\mathbf{Z}(L)| = (3, q-1)$. We claim that there is a regular semisimple element $s \in L$ with $T^* = \mathbf{C}_L(s)$ of order $\Phi_9(q)$. Indeed, by [Moretó and Tiep 2008, Lemma 2.3], there is a regular semisimple element $s_1 \in L$ of order r_1 with $T^* := \mathbf{C}_L(s_1)$ of order $\Phi_9(q)$. If f is odd, set $s = s_1$. Assume $2 \mid f$. Then $\Phi_9(q) = (q^9 - 1)/\Phi_1(q)\Phi_3(q)$ is divisible by r_2, so T^* contains an element s_2 of order r_2. Now set $s = s_1 s_2$.

Case IIf. In the case $L = {}^2E_6(q)_{sc}$, we choose $s \in L$ of order $r = r(p, 18f) \geq 19$. By [Moretó and Tiep 2008, Lemma 2.3], we have $|\mathbf{C}_L(s)| = \Phi_{18}(q)$.

In all these cases, it is straightforward to check that s is a regular semisimple element. Hence, as above, $\chi = \chi_s \in \mathrm{Irr}(G)$, and $\chi_s(x^\sigma) = 0$ for all $\sigma \in \mathrm{Aut}(S)$, whenever $x \in G$ is any semisimple element with $|\mathbf{C}_G(x)|$ *not* divisible by $|T| = |\mathbf{C}_L(s)|$. Also, by Proposition 4.3(iii), $\theta := (\chi_s)_S$ is irreducible.

5. Observe that, when $L = E_6^\epsilon(q)$, both χ and θ have 2-defect 0, whence they vanish at any 2-element $1 \neq x \in S$. In the remaining cases, one easily checks that the 2-element $x \in S$ constructed in part 2 of this proof has the property that $|\mathbf{C}_G(x)|$ is not divisible by $|T^*|$. Thus $\theta(x^\sigma) = 0$ for all $\sigma \in \mathrm{Aut}(S)$, as required in Theorem 3.1(i).

Next we claim that s is *not* real in L. Assume the contrary: $gsg^{-1} = s^{-1}$ for some $g \in L$. Then g normalizes $T^* = \mathbf{C}_L(s)$ and $g^2 \in T^*$. But (using for instance

[Fleischmann et al. 1998, §5 and Theorem 5.7]) one can see that $N_L(T^*)/T^*$ has odd order (indeed it is C_m in IIa and IIb, C_n in IIc and IId, and C_9 in IIe and IIf). It follows that $g \in T^*$ and so $s^2 = 1$, a contradiction.

Now we show that $I_{\mathrm{Aut}(S)}(\chi) = G$. Consider any automorphism $\sigma \in \mathrm{Aut}(S)$ that fixes χ. As in 3), we may write $\sigma = \tau^i(\sigma_0)^e$ with $e \geq 0$ and $i = 0, 1$. Moreover, $\sigma \notin G\tau$ (otherwise $\chi^\sigma = (\chi_s)^\tau = \chi_{s^{-1}} \neq \chi$ as s is not real), so $e > 0$ if $i = 1$. Hence, $r = |s|$ must divide $|\mathscr{L}^{\sigma^*}|$ by Proposition 4.3(ii).

First we consider the twisted cases: $L = \mathrm{SU}_n(q)$, $\mathrm{Spin}_{2n}^-(q)$, or ${}^2E_6(q)_{sc}$. Then $A(S) = \langle \sigma_0 \rangle \cong C_{2f}$ and $(\sigma_0)^f = \tau$ on S. Replacing σ by σ^{-1} if necessary, we may assume that $0 \leq e < f$ and $i = 0$. The condition $r = |s|$ divides $|\mathscr{L}^{\sigma^*}|$ now implies that $e = 0$.

Finally we consider the untwisted cases: $L = \mathrm{SL}_n(q)$, $\mathrm{Spin}_{2n}^+(q)$, or $E_6(q)_{sc}$. Then $A(S) = \langle \sigma_0 \rangle \times \langle \tau \rangle \cong C_f \times C_2$. Replacing σ by σ^{-1} if necessary, we may assume that $0 \leq e \leq f/2$. If $i = 0$, then the condition $r = |s|$ divides $|\mathscr{L}^{\sigma^*}|$ now implies that $e = 0$, that is, $\sigma \in G$. Next assume that $i = 1$ (and so $0 < e \leq f/2$), and $L = \mathrm{SL}_n(q)$ for instance. Then r divides $|\mathscr{L}^{\sigma^*}| = |\,\mathrm{SU}_n(p^e)|$, and so $r_1 = r(p, mf)$ divides $p^{je} - (-1)^j$ for some j, $1 \leq j \leq n$. If j is even, then

$$(n-1)f \leq mf \mid je \leq nf/2,$$

a contradiction as $n \geq 3$. Hence j is odd. Recall that $m \in \{n, n-1\}$ is chosen to be odd and $1 \leq j \leq n$, so $j \leq m$. Now $mf \mid 2je \leq mf$ implies that $e = f/2$. In this case we have that $r_2 = r(p, mf/2)$ divides $p^{ke} - (-1)^k$ for some k, $1 \leq k \leq n$. In particular, $me \mid 2ke$ and so $m \mid 2k$, which implies $m \mid k$ because m is odd. Since $1 \leq k \leq n$ and $m \geq n-1$, we obtain that $k = m$ and so k is odd. But in this case $r_2 = r(p, ke)$ cannot divide $p^{ke} + 1$, a contradiction. The same argument shows that $r = |s|$ cannot divide $|\mathscr{L}^{\sigma^*}|$ if $i = 1$ and $L = \mathrm{Spin}_{2n}^+(q)$ or $E_6(q)_{sc}$.

6. We have shown that $I_{\mathrm{Aut}(S)}(\chi) = G$. Hence, $I_H(\theta) = H \cap G$ by Proposition 4.3(iv). Since

$$H/(G \cap H) \cong GH/G \geq \langle G, \tau \rangle / G \cong C_2$$

by the main hypothesis in Case II, we can find $\varphi \in H \smallsetminus G$ such that φ induces τ modulo G and $\varphi^2 \in G \cap H$. Now set $J = \langle G \cap H, \varphi \rangle$ and $\alpha = (\chi_{G \cap H})^J$. Then by Lemma 4.6 and [Navarro et al. 2008, Corollary 2.5] we have

$$\chi^\varphi = \chi^\tau = (\chi_s)^\tau = \chi_{\tau(s)} = \chi_{s^{-1}} = \bar{\chi},$$

in particular, $\theta^\varphi = \bar{\theta}$, but $\theta^\varphi \neq \theta$ as $\varphi \notin G \cap H = I_H(\theta)$. Since $S \lhd J$, this implies that $\alpha \in \mathrm{Irr}(J|\theta)$. Also, α equals $\chi + \bar{\chi}$ on $G \cap H$ and 0 on $J \smallsetminus (G \cap H)$, whence it is real.

Under the extra assumption that $S \not\cong \mathrm{PSU}_n(q)$ with $n \geq 4$ *even*, we now show that α is strongly real. Indeed, setting $K := \langle G, \varphi \rangle = \langle G, \tau \rangle$ and $\vartheta := \chi^K$ (as

$K \cap H = J$), we see that $\vartheta_J = \alpha$, so $\vartheta \in \mathrm{Irr}(K)$. Also, ϑ equals $\chi + \bar{\chi}$ on G and 0 on $K \smallsetminus G$, so ϑ is real.

- Now, if $S = \mathrm{PSL}_n(q)$, then K is a quotient of $\langle \mathrm{GL}_n(q), \tau \rangle$, and so ϑ is strongly real by [Gow 1983, Theorem 2].
- Suppose $S = \mathrm{P}\Omega^{\epsilon}_{2n}(q)$. Then $\langle S, \tau \rangle \leq R := \mathrm{PGO}^{\epsilon}_{2n}(q) < K$. Since $\theta^{\tau} = \theta^{\varphi} = \bar{\theta} \neq \theta$ and $\vartheta_S = \theta + \bar{\theta}$, we see that ϑ_R is irreducible. By the main result of [Gow 1985], ϑ_R, as an irreducible character of $\mathrm{GO}^{\epsilon}_{2n}(q)$, is strongly real. In turn, this implies that ϑ is strongly real.
- Suppose that $S = \mathrm{PSU}_n(q)$ with $n \geq 3$ *odd* or $S = E^{\epsilon}_6(q)$. Then $|\mathbf{C}_L(s)| = (q^n+1)/(q+1)$, $\Phi_9(q)$ or $\Phi_{18}(q)$, respectively, and is odd; hence χ and ϑ are of 2-defect zero. Since ϑ is real of 2-defect 0, it is strongly real by Lemma 2.3.

Thus in all cases ϑ is strongly real, and so is $\alpha = \vartheta_J$, as claimed. □

Proposition 4.8. *Theorem 3.1 holds if* $S = \mathrm{PSU}_n(q)$ *where* $n \geq 3$ *and* q *is odd.*

Proof. Keep all the notation of the proof of Proposition 4.7.

1. First we consider the case $S = \mathrm{PSU}_3(q)$. When $q = 3$, one can check using [Conway et al. 1985] that $\mathrm{Irr}(S)$ contains a character θ of degree 14, which extends to a strongly real character of $\mathrm{Aut}(S)$ and vanishes at all elements of order 8 in S. Furthermore, the case where GH contains $\langle G, \tau \rangle$ has already been considered in Case II of the proof of Proposition 4.7. So we may assume that $q \geq 5$ and that GH does not contain $\langle G, \tau \rangle$.

In the notation of [Geck 1990, Table 3.1], consider the irreducible character $\theta = \chi^{(u)}_{q^3+1}$ of degree q^3+1 of $L = \mathrm{SU}_3(q)$, with $u := q+1$. Since θ is trivial at $\mathbf{Z}(L)$, we will view it as an irreducible character θ of $S = L/\mathbf{Z}(L)$. Using the character values listed in [Geck 1990, Table 3.1], one checks that θ is real and τ-invariant (indeed,

$$\chi^{(u)}_{q^3+1} = \chi^{(-u)}_{q^3+1} = \chi^{(uq)}_{q^3+1}$$

by our choice of u). Next, the largest degree of irreducible characters of $G = \mathrm{PGU}_3(q)$ is $(q+1)(q^2-1)$, which is less than $3\theta(1)$. Since G/S has order 1 or 3, it follows that θ is G-invariant. Hence, by [Navarro and Tiep 2008, Lemma 2.1], θ extends to a unique $\chi \in \mathrm{Irr}_{\mathbb{R}}(G)$. Viewing χ as a real irreducible character of $\mathrm{GU}_3(q)$, we conclude by [Ohmori 1981, Theorem 7(ii)] that χ is strongly real.

Next we show that $I_{\mathrm{Aut}(S)}(\theta) = \langle G, \tau \rangle$. Since θ is invariant under G and τ and $\mathrm{Aut}(G) = G : A(S)$, it suffices to show that the only nontrivial element $\sigma = (\sigma_0)^e \in A(S)$ that fixes θ is $\tau = (\sigma_0)^f$. So assume that $1 \leq e \leq f$ for such a σ. Consider an element y belonging to the conjugacy class $C^{(1)}_7$ in [Geck 1990, Table 1.1], so that y^{σ} belongs to the class $C^{(p^e)}_7$. Then the condition $\theta(y) = \theta(y^{\sigma})$ implies that

$$\delta + \delta^{-1} = \delta^{p^e} + \delta^{-p^e}$$

for a fixed $(q-1)$-st primitive root δ of unity in $\mathbb{C}$. Since $1 \le e \le f$ and $q \ge 5$, it follows that $e = f$, as claimed.

We have shown that θ extends to the strongly real character $\alpha = \chi_J$ of $J = H \cap G = I_H(\theta)$. It remains to find an element x satisfying the condition (i) of Theorem 3.1. Suppose first that $q \equiv 3 \pmod 4$. Then we choose $x \in S$ to be any element of order 4 that affords eigenvalue 1 on the natural module $\mathbb{F}_{q^2}^3$ for L. Observe that $x^{\mathrm{Aut}(S)}$ is just the conjugacy class $C_6^{(0,(q+1)/4,3(q+1)/4)}$ in [Geck 1990, Table 1.1], and so $\theta(x) = 0$; cf. [ibid., Table 3.1].

Assume now that $q \equiv 1 \pmod 4$. Then we choose $x \in S$ to be any element of order 8. Any $\mathrm{Aut}(S)$-conjugate x^σ of such an x belongs to the conjugacy class $C_7^{k(q^2-1)/8}$ in [ibid., Table 1.1] for some odd integer k. Hence

$$\theta(x^\sigma) = \delta^{k(q^2-1)/8} + \delta^{-k(q^2-1)/8} = 0$$

since k is odd and $|\delta| = q-1$.

2. From now on we may assume that $S = \mathrm{PSU}_n(q)$ with $n \ge 4$. Then it was shown in parts 2. and 5. of the proof of [Dolfi et al. 2008, Theorem 2.1] that there is a permutation character ρ of $\mathrm{Aut}(S)$ such that $\rho_S = 1_S + \varphi + \psi$ is the sum of three irreducible (unipotent) characters of S, all of distinct degrees, and with exactly one, call it θ, of even degree. In fact, ρ_S is just the permutation character of the action of S on the singular 1-spaces of the natural L-module $V = \mathbb{F}_{q^2}^n$, and

$$\varphi(1) = \frac{(q^n - (-1)^n)(q^{n-1} + (-1)^n q^2)}{(q+1)(q^2-1)}, \qquad \psi(1) = \frac{(q^n + (-1)^n q)(q^n - (-1)^n q^2)}{(q+1)(q^2-1)}.$$

Since $S \lhd \mathrm{Aut}(S)$, it follows that the same is true for ρ, and so θ extends to a strongly real character of even degree of $\mathrm{Aut}(S)$. Note that $\theta = \varphi$ if $n \equiv 0, 3 \pmod 4$ and $\theta = \psi$ if $n \equiv 1, 2 \pmod 4$.

3. It remains to find a 2-element $h \in S$ such that $\theta(h^\sigma) = 0$ for all $\sigma \in \mathrm{Aut}(S)$. It suffices to show that $\theta(h) = 0$ since θ is $\mathrm{Aut}(S)$-invariant. To this end, we will use the technique of *dual pairs*; see for instance [Liebeck et al. 2010; Tiep 2010]. We consider the dual pair $X * Y$ inside $\Gamma := \mathrm{GU}_{2n}(q)$, where $X = \mathrm{GU}_2(q)$ and $Y = \mathrm{GU}_n(q)$. More precisely, we view X as $\mathrm{GU}(U)$, where $U = \mathbb{F}_{q^2}^2$ is endowed with a nondegenerate Hermitian form $(\cdot,\cdot)_U$, and Y is meant to be $\mathrm{GU}(V)$, where $V = \mathbb{F}_{q^2}^n$ is endowed with a nondegenerate Hermitian form $(\cdot,\cdot)_V$. Now we consider $W = U \otimes_{\mathbb{F}_{q^2}} V$ with the Hermitian form $(\cdot,\cdot)$ defined via $(u \otimes v, u' \otimes v') = (u,u')_U \cdot (v,v')_V$ for $u \in U$ and $v \in V$. The action of $X \times Y$ on V induces a homomorphism $X \times Y \to \Gamma := \mathrm{GU}(W)$. Recall (see [Tiep and Zalesskii 1997, §4]) that for any $m \ge 1$, the class function

$$\zeta_{m,q}(g) = (-1)^m (-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g-1)} \tag{4-1}$$

is a (reducible) *Weil character* of $\mathrm{GU}_m(q)$ of degree q^m, where $\mathrm{Ker}(g-1)$ is the fixed point subspace of $g \in \mathrm{GU}_m(q)$ on the natural module $(\overline{\mathbb{F}}_{q^2})^m$ for $\mathrm{GU}_m(q)$. By [Liebeck et al. 2010, Proposition 6.3], the restriction of $\zeta := \zeta_{2n,q}$ to $X \times Y$ decomposes as

$$\zeta_{X\times Y} = \sum_{\alpha \in \mathrm{Irr}(X)} \alpha \otimes \mathsf{D}_\alpha, \tag{4-2}$$

where the Y-characters $\mathsf{D}^\circ_\alpha := \mathsf{D}_\alpha - k_\alpha \cdot 1_Y$ are all irreducible and distinct, for some $k_\alpha \in \{0, 1\}$. Furthermore, $k_\alpha = 1$ precisely when $\alpha = 1_X$ or α is the Steinberg character St of X. Also, D_α can be computed explicitly using the formula

$$\mathsf{D}_\alpha(g) = \frac{1}{|X|}\sum_{x\in X} \overline{\alpha(x)}\zeta(xg). \tag{4-3}$$

In particular, one can show (see [Liebeck et al. 2010, Table III]) that $\mathsf{D}^\circ_{1_X}$ is the only irreducible constituent of ζ_Y of degree $\varphi(1)$, and $\mathsf{D}^\circ_{\mathsf{St}}$ is the only irreducible constituent of ζ_Y of degree $\psi(1)$. On the other hand, (4-1) and (4-2) imply that

$$\zeta_Y = \sum_{\alpha \in \mathrm{Irr}(X)} \alpha(1)\cdot \mathsf{D}^\circ_\alpha + (q+1)\cdot 1_Y$$

is just the permutation character of Y on the points of the vector space V, whence ζ_Y contains ρ_Y, the inflation of $\rho_{\mathrm{PGU}_n(q)}$ to $Y = \mathrm{GU}_n(q)$. It follows that

$$\varphi = (\mathsf{D}^\circ_{1_X})_S - 1_S,\;\; \psi = (\mathsf{D}^\circ_{\mathsf{St}})_S - 1_S.$$

Together with (4-1) and (4-3), this will allow us to find the desired element h.

4. Among the irreducible characters of $X = \mathrm{GU}_2(q)$, there are $q + 1$ distinct characters ζ_2^i, where $0 \le i \le q$, which are known as (irreducible) *Weil characters* of X. They are computed explicitly in [Tiep and Zalesskii 1997, Lemma 4.1]; furthermore, $\zeta_{2,q} = \sum_{i=0}^{q} \zeta_2^i$ and $\zeta_2^0 = \mathsf{St}$. In particular,

$$[\zeta_{2,q}, \mathsf{St}]_X = 1. \tag{4-4}$$

Let $\mu_{q+1} := \{c \in \mathbb{F}_{q^2} \mid c^{q+1} = 1\}$. Note that, for any $c \in \mu_{q+1}$, $x \mapsto \zeta_{2,q}(cx)$ is a class function on X. Moreover, using the well-known character table of X (see for instance [Ennola 1963]), we can check that

$$\frac{1}{|X|}\sum_{x\in X} \zeta_{2,q}(cx)\zeta_{2,q}(dx)\overline{\mathsf{St}(x)} = 1 \tag{4-5}$$

for any $c, d \in \mu_{q+1}$, and

$$\frac{1}{|X|}\sum_{x\in X} \zeta_{2,q}(cx)^2\zeta_{2,q}(dx) = 1 \tag{4-6}$$

whenever $c, d \in \mu_{q+1}$ and $c \ne d$.

5. Now we are ready to find the desired element h. This will be done according to $n \pmod 8$. Let $N = 2(q^2-1)_2$ denote the 2-part $(q^4-1)_2$ of (q^4-1), and let γ be a fixed N-th primitive root of unity in $\overline{\mathbb{F}}_{q^2}$. Observe that $\mathrm{GU}_4(q)$ has a cyclic maximal torus T of order q^4-1 that contains an element g_4 conjugate to $\mathrm{diag}(\gamma, \gamma^{-q}, \gamma^{q^2}, \gamma^{-q^3})$ over $\overline{\mathbb{F}}_{q^2}$, and set

$$g_8 := \mathrm{diag}(g_4, g_4^{-1}) \in \mathrm{SU}_8(q).$$

Note that no eigenvalue of g_4 and g_8 belongs to $\mathbb{F}_{q^2}$ by the choice of γ. On the other hand, any eigenvalue of any $x \in X = \mathrm{GU}_2(q)$ belongs to $\mathbb{F}_{q^2}^\times$.

Let $8 \mid n$, and choose $h = \mathrm{diag}(g_8, \dots, g_8) \in \mathrm{SU}_n(q)$. Then, for any $x \in X$, no eigenvalue of xh can be equal to 1, whence $\zeta(xh) = 1$ by (4-1). Hence, by (4-3) we have

$$\theta(h) = \mathsf{D}^\circ_{1_X}(h) = \frac{1}{|X|}\sum_{x\in X}\zeta(xh) - 1 = [1_X, 1_X] - 1 = 0.$$

Next, for $n \equiv 1 \pmod 8$ we choose $h = \mathrm{diag}(g_8, \dots, g_8, 1) \in \mathrm{SU}_n(q)$. Then, for any $x \in X$, no eigenvalue of xg_8 can be equal to 1, whence $\zeta(xh) = \zeta_{2,q}(x)$ by (4-1). It then follows by (4-3) and (4-4) that

$$\theta(h) = \mathsf{D}^\circ_{\mathsf{St}}(h) = \frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(x)\overline{\mathsf{St}(x)} - 1 = [\zeta_{2,q}, \mathsf{St}] - 1 = 0.$$

For $n \equiv 2 \pmod 8$ we choose $h = \mathrm{diag}(g_8, \dots, g_8, 1, 1) \in \mathrm{SU}_n(q)$. Then, $\zeta(xh) = \zeta_{2,q}(x)^2$ for any $x \in X$ by (4-1). By (4-3) and (4-5) applied to $c = d = 1$ we have

$$\theta(h) = \mathsf{D}^\circ_{\mathsf{St}}(h) = \frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(x)^2\overline{\mathsf{St}(x)} - 1 = 0.$$

For $n \equiv 3 \pmod 8$ we choose $h = \mathrm{diag}(g_8, \dots, g_8, 1, -1, -1) \in \mathrm{SU}_n(q)$. Again, $\zeta(xh) = \zeta_{2,q}(x)\zeta_{2,q}(-x)^2$ for all $x \in X$. By (4-3) and (4-6) applied to $(c, d) = (-1, 1)$ we have

$$\theta(h) = \mathsf{D}^\circ_{1_X}(h) = \frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(-x)^2\zeta_{2,q}(x) - 1 = 0.$$

Assume that $n \equiv 5 \pmod 8$. Note that $1 \neq c_1 := \det(g_4^{-1}) = \gamma^{(q^4-1)/(q+1)} \in \mu_{q+1}$. Let $h = \mathrm{diag}(g_8, \dots, g_8, g_4, c_1) \in \mathrm{SU}_n(q)$. Then, $\zeta(xh) = \zeta_{2,q}(c_1x)$ for any $x \in X$ by (4-1), and $\mathsf{St}(c_1x) = \mathsf{St}(x)$ since St is trivial at $\mathbf{Z}(X)$. Hence, by (4-3) and (4-4) we have

$$\begin{aligned}\mathsf{D}^{\circ}_{\mathsf{St}}(h) &= \frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(c_1x)\overline{\mathsf{St}(x)}-1\\ &= \frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(c_1x)\overline{\mathsf{St}(c_1x)}-1=[\zeta_{2,q},\mathsf{St}]_X-1=0.\end{aligned}$$

For $n\equiv 6 \pmod 8$ we choose $h=\mathrm{diag}(g_8,\ldots,g_8,g_4,1,c_1)\in \mathrm{SU}_n(q)$. Then, $\zeta(xh)=\zeta_{2,q}(c_1x)\zeta_{2,q}(x)$ for any $x\in X$ by (4-1). By (4-3) and (4-5) applied to $(c,d)=(c_1,1)$ we have

$$\theta(h)=\mathsf{D}^{\circ}_{\mathsf{St}}(h)=\frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(c_1x)\zeta_{2,q}(x)\overline{\mathsf{St}(x)}-1=0.$$

For $n\equiv 7 \pmod 8$ we choose $h=\mathrm{diag}(g_8,\ldots,g_8,g_4,1,1,c_1)\in \mathrm{SU}_n(q)$. Then, $\zeta(xh)=\zeta_{2,q}(x)^2\zeta_{2,q}(c_1x)$ for any $x\in X$ by (4-1). By (4-3) and (4-6) applied to $(c,d)=(1,c_1)$ we have

$$\theta(h)=\mathsf{D}^{\circ}_{1_X}(h)=\frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(x)^2\zeta_{2,q}(c_1x)-1=0.$$

Finally, assume that $n\equiv 4 \pmod 8$. Then we choose $h_4\in \mathrm{SU}_4(q)$ to be conjugate to $\mathrm{diag}(\gamma^2,\gamma^{-2q},1,\gamma^{2q-2})$ over $\overline{\mathbb{F}}_{q^2}$, and set $h=\mathrm{diag}(g_8,\ldots,g_8,h_4)\in \mathrm{SU}_n(q)$. Note that $\gamma^2,\gamma^{-2q}\in\mathbb{F}^{\times}_{q^2}$ and $1\neq c_2:=\gamma^{2q-2}\in\mu_{q+1}$. Now, for any $x\in X$, we have

$$\zeta(xh)=\zeta_{2,q}(x)\zeta_{2,q}(c_2x)\zeta_{2,q}(\gamma^2x)\zeta_{2,q}(\gamma^{-2q}x)$$

by (4-1). Direct computation using the character table of $\mathrm{GU}_2(q)$ shows that

$$\theta(h)=\mathsf{D}^{\circ}_{1_X}(h)=\frac{1}{|X|}\sum_{x\in X}\zeta_{2,q}(x)\zeta_{2,q}(c_2x)\zeta_{2,q}(\gamma^2x)\zeta_{2,q}(\gamma^{-2q}x)-1=0,$$

and so we are done. □

To complete the proof of Theorem 3.1, we handle the case $S=\mathrm{P}\Omega_8^+(q)$:

Proposition 4.9. *Theorem 3.1 holds in the case $S=\mathrm{P}\Omega_8^+(q)$ with q odd.*

Proof. **1.** Suppose that $q=3$. Then, according to [Conway et al. 1985], S has a unique irreducible character θ of degree 300, which is strongly real, and a unique conjugacy class ($4E$ in the notation of [ibid.]) of elements x of order 4 with $|\mathbf{C}_S(x)|=1536$ and $\theta(x)=0$. It follows that $\theta(x^\sigma)=0$ for all $\sigma\in\mathrm{Aut}(S)$. Furthermore, one can show (directly, or using [GAP 2004]) that θ extends to a rational character of $\mathrm{Aut}(S)=S\cdot\mathsf{S}_4$. From now on we may assume that $q=p^f\geq 5$.

2. Choose $\epsilon = \pm 1$ such that $q \equiv \epsilon \pmod 4$. Also view S as $L/\mathbf{Z}(L)$, where $L = \mathrm{Spin}_8^+(q)$ and $\mathbf{Z}(L) \cong C_2 \times C_2$. Fix an orthonormal basis $(e_1, \ldots, e_4)$ of $\mathbb{R}^4$ and realize the simple roots of the algebraic group $\mathscr{L} = \mathrm{Spin}_8(\bar{\mathbb{F}}_q)$ as

$$\alpha_1 = e_1 - e_2, \quad \alpha_2 = e_2 - e_3, \quad \alpha_3 = e_3 - e_4, \quad \alpha_4 = e_3 + e_4$$

as usual. Then the four fundamental weights of $\mathscr{L}$ are given by

$$\varpi_1 = e_1, \quad \varpi_2 = e_1 + e_2, \quad \varpi_3 = \frac{e_1 + e_2 + e_3 - e_4}{2}, \quad \varpi_4 = \frac{e_1 + e_2 + e_3 + e_4}{2}.$$

Let $\Gamma \cong \mathrm{S}_3$ denote the subgroup of $A(S)$ consisting of graph automorphisms. Then Γ permutes the 3 fundamental weights ϖ_1, ϖ_3, ϖ_4 transitively and faithfully, and fixes ϖ_2. Consider the corresponding $\mathscr{L}$-modules $V_i = V(\varpi_i)$ with highest weight ϖ_i, $i = 1, 3, 4$. Then the set of weights of V_i is

$$\begin{array}{ll} \{\pm e_j \mid 1 \le j \le 4\}, & \text{when } i = 1, \\ \left\{\frac{1}{2}\sum_{j=1}^4 a_j e_j \mid a_j = \pm 1,\ \prod_{j=1}^4 a_j = -1\right\}, & \text{when } i = 3 \\ \left\{\frac{1}{2}\sum_{j=1}^4 a_j e_j \mid a_j = \pm 1,\ \prod_{j=1}^4 a_j = 1\right\}, & \text{when } i = 4. \end{array}$$

We can think of V_1 as the natural module for $K := \Omega_8^+(q)$.

3. We will use the description above to show that L contains a regular semisimple element s of (odd) order $N := (q^3 + \epsilon)/2$ if $q \ge 9$ and $N := (q^2 + 1)/2$ if $q = 5, 7$ such that s^σ is not L-conjugate to s for any nontrivial $\sigma \in \Gamma$.

Indeed, assume that $q \ge 9$. Then fix $\delta \in \bar{\mathbb{F}}_q^\times$ of order $(q^3 + \epsilon)/2$ and choose $s \in L$ to be the unique inverse image of odd order of $\bar{s} \in \Omega_2^{-\epsilon}(q) \times \Omega_6^{-\epsilon}(q) < K$ with spectrum $\mathrm{Spec}(s, V_1) = \{\delta^j \mid j \in J_1\}$, where

$$J_1 = \{\pm 1, \pm r, \pm r^2, \pm 2(r^2 + r + 1)\}$$

and $r := -\epsilon q$. Thus we may assume that

$$e_1(s) = \delta, \quad e_2(s) = \delta^r, \quad e_3(s) = \delta^{r^2}, \quad e_4(s) = \delta^{2(r^2+r+1)}.$$

Hence $\mathrm{Spec}(s, V_i) = \{\delta^{(N+j)/2} \mid j \in J_i\}$ for $i = 3, 4$, where

$$J_3 = \{\pm(3r^2 + 3r + 1), \pm(3r^2 + r + 3), \pm(r^2 + 3r + 3), \pm(r^2 + r + 1)\},$$
$$J_4 = \{\pm(3r^2 + r + 1), \pm(r^2 + 3r + 1), \pm(r^2 + r + 3), \pm 3(r^2 + r + 1)\}.$$

Recall that $|\delta| = N \ge 5(r^2 + r + 1)$ and $|r| \ge 9$ since $q \ge 9$. Hence δ belongs to $\mathrm{Spec}(s, V_1)$ but neither to $\mathrm{Spec}(s, V_3)$ nor $\mathrm{Spec}(s, V_4)$, and similarly $\delta^{(N+r^2+r+3)/2}$ belongs to $\mathrm{Spec}(s, V_4)$ but not to $\mathrm{Spec}(s, V_3)$. Thus s has pairwise different spectra on the three modules V_1, V_3 and V_4 permuted faithfully by Γ, whence s and s^σ cannot be L-conjugate for any $1 \ne \sigma \in \Gamma$. Arguing as in the proof of [Moretó and

Tiep 2008, Lemma 2.3], we can view s as an element of $\mathrm{SO}_8^+(q)$ to calculate the order of its centralizer and find that $T^* = \mathbf{C}_L(s)$ is a torus of order $(q+\epsilon)(q^3+\epsilon)$; in particular, s is regular.

Suppose now that $q = 5$ or 7. Then fix $\delta \in \bar{\mathbb{F}}_q^\times$ of order $(q^2+1)/2$ and choose $s \in L$ to be the unique inverse image of odd order of $\bar{s} \in \Omega_4^-(q) \times \Omega_4^-(q) < K$ with spectrum $\mathrm{Spec}(s, V_1) = \{\delta^j \mid j \in J_1\}$, where

$$J_1 = \{\pm 1, \pm q, \pm 2, \pm 2q\}.$$

Thus we may assume that

$$e_1(s) = \delta, \quad e_2(s) = \delta^q, \quad e_3(s) = \delta^2, \quad e_4(s) = \delta^{2q}.$$

Hence $\mathrm{Spec}(s, V_i) = \{\delta^{j/2} \mid j \in J_i\}$ for $i = 3, 4$, where

$$\begin{aligned} J_3 &= \{\pm(3q+1), \pm(q+3), \pm(3q-1), \pm(q-3)\}, \\ J_4 &= \{\pm(q+1), \pm(3q-3), \pm(q-1), \pm(3q+3)\}. \end{aligned}$$

One can again check that s has pairwise different spectra on the three modules V_1, V_3 and V_4, and so s and s^σ cannot be L-conjugate for any $1 \neq \sigma \in \Gamma$. Furthermore, s is regular and $T^* = \mathbf{C}_L(s)$ is a torus of order $(q^2+1)^2$.

4. By [Tiep and Zalesski 2005, Proposition 3.1], s is real. It now follows by Proposition 4.3 that χ_s is a strongly real irreducible character of G, and $\theta := (\chi_s)_S$ is irreducible. We claim that $I_H(\theta) \leq G$. Once this is completed, we can take $J = G \cap H$ and $\alpha = (\chi_s)_J$ as usual. As in the proof of Proposition 4.7, it suffices to show that if $\sigma \in A(S) \cong C_f \times \mathrm{S}_3$ fixes χ_s then σ is trivial. Write $\sigma = \tau(\sigma_0)^e$ for some $\tau \in \Gamma$ and $0 \leq e < f$ (and σ_0 is induced by the field automorphism $y \mapsto y^p$ as usual). By the results of 3) we may assume $0 < e < f$; in particular, $f \geq 2$ and so $q \geq 9$. By Proposition 4.3(ii), s^L is σ^*-stable and $N = |s|$ divides $|\mathscr{L}^{\sigma^*}|$.

First assume that $|\tau| = 3$, that is, τ is a triality graph automorphism. Then $N = (q^3+\epsilon)/2$ divides $|\mathscr{L}^{\sigma^*}| = |{}^3D_4(p^e)|$. Using a suitable p.p.d. of N one can now show that $3f \mid 12e$ and so $f \mid 4e$. It follows that s^L is stable under $\sigma^4 = \tau$, contrary to the results of 3).

Now we may assume that $|\tau| = 1$ or 2, and so $N = (q^3+\epsilon)/2$ divides $|\mathscr{L}^{\sigma^*}| = |\mathrm{Spin}_8^\alpha(p^e)|$ with $\alpha = +$ or $-$, respectively. Using a suitable p.p.d. of N we now see that $3f \mid 6e$ or $3f \mid 8e$. If f is odd, then we get that $f \mid e$, a contradiction as $0 < e < f$. Hence f is even, $\epsilon = +$, and $N = (q^3+1)/2$ is divisible by $r = r(p, 6f)$, a p.p.d. of $p^{6f} - 1$. Since r divides $|\mathrm{Spin}_8^\alpha(p^e)|$, we must have $6f \mid 6e$ or $6f \mid 8e$. In the former case we again have $f \mid e$, a contradiction. So $6f \mid 8e$, and $e = 3f/4$ as $0 < e < f$. In this case, s^L is stable under $\sigma^2 = (\sigma_0)^{f/2}$. Repeating the argument above, we see that $r = r(p, 6f)$ divides $|\mathrm{Spin}_8^+(p^{f/2})|$, which is impossible.

5. We have shown that $I_H(\theta) = G \cap H$ and obviously θ extends to the strongly real character $(\chi_s)_{G\cap H}$. It remains to find a 2-element $x \in S$ such that $\theta(x^\sigma) = 0$ for all $\sigma \in \mathrm{Aut}(S)$. Since s is regular, we have $\chi_s = \pm R^G_{T,\vartheta}$ for some maximal torus T of order $|T| = |T^*|$. Recall that $q \equiv \epsilon \pmod 4$, hence we can choose $x \in S$ to be represented by $\mathrm{diag}(-I_2, I_6) \in \Omega_2^\epsilon(q) \times \Omega_6^\epsilon(q) < \Omega_8^+(q)$ with centralizer $\mathrm{GO}_2^\epsilon(q) \times \mathrm{GO}_6^\epsilon(q)$ (in $\mathrm{GO}_8^+(q)$). It is easy to see that $|\mathbf{C}_G(x)|$ is not divisible by $|T|$. Thus, $\theta(x^\sigma) = 0$ for any $\sigma \in \mathrm{Aut}(S)$. □

5. Final remarks

We start with a well-known lemma; see, for instance, [Bubboloni et al. 2009, Lemma 2.1]. We provide a proof for the sake of completeness.

Lemma 5.1. *Let $\chi \in \mathrm{Irr}(G)$ and let g be a p-element of the group G, p a prime. If $\chi(g) = 0$, then p divides $\chi(1)$.*

Proof. Let ω be a primitive p^a-th root of unity, where $p^a = o(g)$, and write $n = \chi(1)$. Then $\chi(g) = \sum_{i=1}^n \omega^{k_i} = 0$ for suitable integers $0 \le k_i \le p^a$, and ω is a root of the polynomial $q(x) = \sum_{i=1}^n x^{k_i}$. Hence, the p^a-th cyclotomic polynomial $\Phi(x)$ divides $q(x)$ (over $\mathbb{Q}$, hence also over $\mathbb{Z}$ by Gauss's lemma). In particular, $\Phi(1) = p$ divides $q(1) = \chi(1)$, as required. □

Using Lemma 5.1, from Theorem A we immediately obtain Theorem B, which in turn implies the following.

Corollary 5.2 [Dolfi et al. 2008, Theorem A]. *Let G be a finite group. If every $\chi \in \mathrm{Irr}_{\mathbb{R}}(G)$ has odd degree, then G has a normal Sylow 2-subgroup.*

However, the following class of examples shows that it is not possible to deduce our Theorem A from [Dolfi et al. 2008, Theorem A] (even if we require $\chi(x) \neq 0$ for all $\chi \in \mathrm{Irr}_{\mathbb{R}}(G)$ and all 2-elements $x \in G$).

Example 5.3. For every Mersenne prime $q > 7$ there exists a Frobenius $\{2, q\}$-group G such that

(a) $\chi(g) \neq 0$ for all $\chi \in \mathrm{Irr}_{\mathbb{R}}(G)$ and every 2-element $g \in G$, and

(b) there exists a $\chi_0 \in \mathrm{Irr}_{\mathbb{R}}(G)$ with $\chi(1)$ even.

Let $q = 2^t - 1$ be a Mersenne prime, with $t > 3$ a prime. Write $n = t - 1$. As shown in [Isaacs 1989, Section 4] and in [Riedl 1999], one can construct a remarkable class of 2-groups (or, in general, p-groups for any prime p) $P_n(2, t)$ as subgroups of the group of units of suitable skew-polynomial rings. We recall that the same class of groups has also been considered in [Hanaki and Okuyama 1997], where they are given as matrix groups.

We mention (see [Riedl 1999]) that the group $P = P_n(2, t)$ has order 2^{tn}, that the upper central series of P coincides with the lower central series of P and that all its factors are elementary abelian groups of order 2^t. Moreover, P has a fixed point free group of automorphisms Q of order q. Hence, the semidirect product $G = PQ$ is a Frobenius $\{2, q\}$-group.

As proved in [Bubboloni et al. 2009, Example 1] (see also [Isaacs et al. 1999, Theorem 5.1]), $\chi(g) \neq 0$ for every $\chi \in \mathrm{Irr}(G)$ and for every element $g \in G$ of 2-power order. So, in particular, (a) is satisfied.

To prove (b), we denote by $\mathrm{Cl}_{\mathbb{R}}(P)$ the set of the P-conjugacy classes of real elements of P. (Observe that they are precisely the classes where every irreducible character of P assumes a real value). As an application of Brauer permutation lemma [Isaacs 1976, (6.32)], we know that $|\mathrm{Irr}_{\mathbb{R}}(P)| = |\mathrm{Cl}_{\mathbb{R}}(P)|$. Let $W = \mathbf{Z}_2(P)$ be the second term of the (upper) central series of G. By [Riedl 1999, part (i) of Corollary 2.12 and Lemma 6.1], we see that $|W| = 2^{2t}$ and that W is elementary abelian, because $t > 3$. Since every involution is a real element of P, it follows that $|\mathrm{Cl}_{\mathbb{R}}(P)| > |\mathbf{Z}(P)| = 2^t$. Therefore, as P has $|P/P'| = 2^t$ linear characters, we conclude that there exists a nonlinear $\psi \in \mathrm{Irr}_{\mathbb{R}}(P)$. So, $\chi = \psi^G \in \mathrm{Irr}_{\mathbb{R}}(G)$ is a real character of even degree of G.

Acknowledgments

We thank S. Dolfi and G. Navarro for many conversations on the subject of the paper. We are also grateful to the referee for helpful comments on the paper and especially for suggesting that we look at characters with Frobenius–Schur indicator 1, which led to a stronger version of our main result.

References

[Brodkey 1963] J. S. Brodkey, "A note on finite groups with an abelian Sylow group", *Proc. Amer. Math. Soc.* **14** (1963), 132–133. MR 26 #200 Zbl 0112.26005

[Bubboloni et al. 2009] D. Bubboloni, S. Dolfi, and P. Spiga, "Finite groups whose irreducible characters vanish only on p-elements", *J. Pure Appl. Algebra* **213**:3 (2009), 370–376. MR 2009k:20012 Zbl 1162.20004

[Carter 1985] R. W. Carter, *Finite groups of Lie type: Conjugacy classes and complex characters*, John Wiley & Sons, New York, 1985. MR 87d:20060 Zbl 0567.20023

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups: Maximal subgroups and ordinary characters for simple groups*, Oxford University Press, Eynsham, 1985. MR 88g:20025 Zbl 0568.20001

[Digne and Michel 1991] F. Digne and J. Michel, *Representations of finite groups of Lie type*, London Mathematical Society Student Texts **21**, Cambridge University Press, 1991. MR 92g:20063 Zbl 0815.20014

[Dolfi et al. 2008] S. Dolfi, G. Navarro, and P. H. Tiep, "Primes dividing the degrees of the real characters", *Math. Z.* **259**:4 (2008), 755–774. MR 2009f:20006 Zbl 1149.20006

[Dolfi et al. 2009] S. Dolfi, E. Pacifici, L. Sanus, and P. Spiga, "On the orders of zeros of irreducible characters", *J. Algebra* **321**:1 (2009), 345–352. MR 2009k:20013 Zbl 1162.20005

[Dolfi et al. 2010a] S. Dolfi, G. Navarro, E. Pacifici, L. Sanus, and P. H. Tiep, "Non-vanishing elements of finite groups", *J. Algebra* **323**:2 (2010), 540–545. MR 2010k:20015 Zbl 1192.20006

[Dolfi et al. 2010b] S. Dolfi, E. Pacifici, and L. Sanus, "Groups whose vanishing class sizes are not divisible by a given prime", *Arch. Math.* (*Basel*) **94**:4 (2010), 311–317. MR 2011e:20007 Zbl 1202.20010

[Dolfi et al. 2010c] S. Dolfi, E. Pacifici, L. Sanus, and P. Spiga, "On the vanishing prime graph of finite groups", *J. Lond. Math. Soc.* (2) **82**:1 (2010), 167–183. MR 2011i:20010 Zbl 1203.20024

[Dolfi et al. 2010d] S. Dolfi, E. Pacifici, L. Sanus, and P. Spiga, "On the vanishing prime graph of solvable groups", *J. Group Theory* **13**:2 (2010), 189–206. MR 2011i:20023 Zbl 1196.20029

[Ennola 1963] V. Ennola, "On the characters of the finite unitary groups", *Ann. Acad. Sci. Fenn. Ser. A I No.* **323** (1963), 35. MR 28 #143 Zbl 0109.26001

[Feit 1993] W. Feit, "Extending Steinberg characters", pp. 1–9 in *Linear algebraic groups and their representations* (Los Angeles, CA, 1992), edited by R. S. Elman et al., Contemp. Math. **153**, Amer. Math. Soc., Providence, RI, 1993. MR 94k:20012 Zbl 0823.20013

[Fleischmann et al. 1998] P. Fleischmann, W. Lempken, and P. H. Tiep, "The p-intersection subgroups in quasi-simple and almost simple finite groups", *J. Algebra* **207**:1 (1998), 1–42. MR 99j:20021 Zbl 0911.20015

[GAP 2004] *GAP – groups, algorithms, and programming*, Version 4.4, 2004, available at http://www.gap-system.org.

[Geck 1990] M. Geck, "Irreducible Brauer characters of the 3-dimensional special unitary groups in nondefining characteristic", *Comm. Algebra* **18**:2 (1990), 563–584. MR 91b:20016

[Gorenstein et al. 1994] D. Gorenstein, R. Lyons, and R. Solomon, *The classification of the finite simple groups*, Mathematical Surveys and Monographs **40**, American Mathematical Society, Providence, RI, 1994. MR 95m:20014 Zbl 0816.20016

[Gow 1983] R. Gow, "Properties of the characters of the finite general linear group related to the transpose-inverse involution", *Proc. London Math. Soc.* (3) **47**:3 (1983), 493–506. MR 85k:20130 Zbl 0551.20023

[Gow 1985] R. Gow, "Real representations of the finite orthogonal and symplectic groups of odd characteristic", *J. Algebra* **96**:1 (1985), 249–274. MR 87b:20015 Zbl 0576.20026

[Hanaki and Okuyama 1997] A. Hanaki and T. Okuyama, "Groups with some combinatorial properties", *Osaka J. Math.* **34**:2 (1997), 337–356. MR 98m:05200 Zbl 0894.20006

[Huppert 1998] B. Huppert, *Character theory of finite groups*, de Gruyter Expositions in Mathematics **25**, Walter de Gruyter & Co., Berlin, 1998. MR 99j:20011 Zbl 0932.20007

[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

[Isaacs 1989] I. M. Isaacs, "Coprime group actions fixing all nonlinear irreducible characters", *Canad. J. Math.* **41**:1 (1989), 68–82. MR 90j:20038 Zbl 0686.20002

[Isaacs et al. 1999] I. M. Isaacs, G. Navarro, and T. R. Wolf, "Finite group elements where no irreducible character vanishes", *J. Algebra* **222**:2 (1999), 413–423. MR 2000k:20008 Zbl 0959.20009

[Liebeck et al. 2010] M. W. Liebeck, E. A. O'Brien, A. Shalev, and P. H. Tiep, "The Ore conjecture", *J. Eur. Math. Soc.* (*JEMS*) **12**:4 (2010), 939–1008. MR 2011e:20016 Zbl 1205.20011

[Manz and Wolf 1993] O. Manz and T. R. Wolf, *Representations of solvable groups*, London Mathematical Society Lecture Note Series **185**, Cambridge University Press, 1993. MR 95c:20013 Zbl 0928.20008

[Moretó and Tiep 2008] A. Moretó and P. H. Tiep, "Prime divisors of character degrees", *J. Group Theory* **11**:3 (2008), 341–356. MR 2009f:20007 Zbl 1149.20008

[Navarro and Tiep 2008] G. Navarro and P. H. Tiep, "Rational irreducible characters and rational conjugacy classes in finite groups", *Trans. Amer. Math. Soc.* **360**:5 (2008), 2443–2465. MR 2008k:20014 Zbl 1137.20009

[Navarro and Tiep 2010] G. Navarro and P. H. Tiep, "Degrees of rational characters of finite groups", *Adv. Math.* **224**:3 (2010), 1121–1142. MR 2011h:20015 Zbl 1200.20006

[Navarro et al. 2008] G. Navarro, P. H. Tiep, and A. Turull, "Brauer characters with cyclotomic field of values", *J. Pure Appl. Algebra* **212**:3 (2008), 628–635. MR 2009b:20018 Zbl 1173.20007

[Navarro et al. 2009] G. Navarro, L. Sanus, and P. H. Tiep, "Real characters and degrees", *Israel J. Math.* **171** (2009), 157–173. MR 2010d:20013 Zbl 1186.20010

[Ohmori 1981] Z. Ohmori, "On the Schur indices of reductive groups, II", *Quart. J. Math. Oxford Ser.* (2) **32**:128 (1981), 443–452. MR 83f:20028 Zbl 0474.20022

[Riedl 1999] J. M. Riedl, "Character degrees, class sizes, and normal subgroups of a certain class of p-groups", *J. Algebra* **218**:1 (1999), 190–215. MR 2000f:20025 Zbl 0932.20015

[Robinson 1989] G. R. Robinson, "The Frobenius–Schur indicator and projective modules", *J. Algebra* **126**:1 (1989), 252–257. MR 90k:20019 Zbl 0742.20013

[Steinberg 1968] R. Steinberg, *Lectures on Chevalley groups*, Yale University, New Haven, Conn., 1968. MR 57 #6215 Zbl 1196.22001

[Tiep 2010] P. H. Tiep, "Dual pairs of finite classical groups in cross characteristics", pp. 161–179 in *Character theory of finite groups*, edited by M. L. Lewis et al., Contemp. Math. **524**, Amer. Math. Soc., Providence, RI, 2010. MR 2012a:20025 Zbl 1227.20005

[Tiep and Zalesski 2005] P. H. Tiep and A. E. Zalesski, "Real conjugacy classes in algebraic groups and finite groups of Lie type", *J. Group Theory* **8**:3 (2005), 291–315. MR 2006b:20010 Zbl 1076.20033

[Tiep and Zalesskii 1997] P. H. Tiep and A. E. Zalesskii, "Some characterizations of the Weil representations of the symplectic and unitary groups", *J. Algebra* **192**:1 (1997), 130–165. MR 99d:20074

[Vinroot 2010] C. R. Vinroot, "Duality, central characters, and real-valued characters of finite groups of Lie type", *Osaka J. Math.* **47**:2 (2010), 523–534. MR 2012c:20030 Zbl 1209.20010

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

Communicated by Ronald Mark Solomon
Received 2011-03-21 Revised 2012-02-02 Accepted 2012-03-16

marinelli@mail.dm.unipi.it *Dipartimento di Matematica L. Tonelli, Università di Pisa, Largo Bruno Pontecorvo 5, I-56127 Pisa, Italy*

tiep@math.arizona.edu *Department of Mathematics, University of Arizona, 617 N. Santa Rita Ave., P.O. Box 210089, Tucson AZ 85721-0089, United States*
http://math.arizona.edu/~tiep/

ALGEBRA AND NUMBER THEORY 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.595

The biHecke monoid of a finite Coxeter group and its representations

Florent Hivert, Anne Schilling and Nicolas Thiéry

For any finite Coxeter group W, we introduce two new objects: its cutting poset and its biHecke monoid. The cutting poset, constructed using a generalization of the notion of blocks in permutation matrices, almost forms a lattice on W. The construction of the biHecke monoid relies on the usual combinatorial model for the 0-Hecke algebra $H_0(W)$, that is, for the symmetric group, the algebra (or monoid) generated by the elementary bubble sort operators. The authors previously introduced the Hecke group algebra, constructed as the algebra generated simultaneously by the bubble sort and antisort operators, and described its representation theory. In this paper, we consider instead the *monoid* generated by these operators. We prove that it admits $|W|$ simple and projective modules. In order to construct the simple modules, we introduce for each $w \in W$ a combinatorial module T_w whose support is the interval $[1, w]_R$ in right weak order. This module yields an algebra, whose representation theory generalizes that of the Hecke group algebra, with the combinatorics of descents replaced by that of blocks and of the cutting poset.

1. Introduction

In this paper we introduce two novel objects for any finite Coxeter group W: its *cutting poset* and its *biHecke monoid*. The cutting poset is constructed using a generalization of blocks in permutation matrices to any Coxeter group and is almost a lattice. The biHecke monoid is generated simultaneously by the sorting and antisorting operators associated to the combinatorial model of the 0-Hecke algebra $H_0(W)$. It turns out that the representation theory of the biHecke monoid, and in particular the construction of its simple modules, is closely tied to the cutting poset.

Hivert was partly supported by ANR grant 06-BLAN-0380. Schilling was in part supported by NSF grants DMS–0652641, DMS–0652652, and DMS–1001256. Thiérly was in part supported by NSF grants DMS–0652641 and DMS–0652652.

MSC2010: primary 20M30, 20F55; secondary 06D75, 16G99, 20C08.

Keywords: Coxeter groups, Hecke algebras, representation theory, blocks of permutation matrices.

The study of these objects combines methods from and impacts several areas of mathematics: Coxeter group theory, monoid theory, representation theory, combinatorics (posets, permutations, descent sets), as well as computer algebra. The guiding principle is the use of representation theory, combined with computer exploration, to extract combinatorial structures from an algebra, and in particular a monoid algebra, often in the form of posets or lattices. This includes the structures associated to monoid theory (such as for example Green's relations), but also goes beyond. For example, we find connections between the classical orders of Coxeter groups (left, right, and left-right weak order and Bruhat order) and Green's relations on our monoids ($\mathcal{R}$, $\mathcal{L}$, $\mathcal{J}$, and $\mathcal{H}$-order and ordered monoids), and these orders play a crucial role in the combinatorics and representation theory of the biHecke monoid.

The usual combinatorial model for the 0-Hecke algebra $H_0(\mathfrak{S}_n)$ of the symmetric group is the algebra (or monoid) generated by the (anti) bubble sort operators $\pi_1, \dots, \pi_{n-1}$, where π_i acts on words of length n and sorts the letters in positions i and $i+1$ decreasingly. By symmetry, one can also construct the bubble sort operators $\bar{\pi}_1, \dots, \bar{\pi}_{n-1}$, where $\bar{\pi}_i$ acts by sorting increasingly, and this gives an isomorphic construction $\overline{H}_0$ of the 0-Hecke algebra. This construction generalizes naturally to any finite Coxeter group W. Furthermore, when W is a Weyl group, and hence can be affinized, there is an additional operator π_0 projecting along the highest root.

In [Hivert and Thiéry 2009] the first and last author constructed the *Hecke group algebra* $\mathcal{H}W$ by gluing together the 0-Hecke algebra and the group algebra of W along their right regular representation. Alternatively, $\mathcal{H}W$ can be constructed as the *biHecke algebra* of W, by gluing together the two realizations $H_0(W)$ and $\overline{H}_0(W)$ of the 0-Hecke algebra. $\mathcal{H}W$ admits a more conceptual description as the algebra of all operators on $\mathbb{K}W$ preserving left antisymmetries; the representation theory of $\mathcal{H}W$ follows, governed by the combinatorics of descents. In [Hivert et al. 2009], the authors further proved that, when W is a Weyl group, $\mathcal{H}W$ is a natural quotient of the affine Hecke algebra.

In this paper, following a suggestion of Alain Lascoux, we study the *biHecke monoid* $M(W)$, obtained by gluing together the two 0-Hecke *monoids*. This involves the combinatorics of the usual poset structures on W (left, right, left-right, Bruhat order), as well as the new cutting poset. Building upon the extensive study of the representation theory of the 0-Hecke algebra [Norton 1979; Carter 1986; Denton 2010; 2011], we explore the representation theory of the biHecke monoid. In the process, we prove that the biHecke monoid is aperiodic and its Borel submonoid fixing the identity is $\mathcal{J}$-trivial. This sparked our interest in the representation theory of $\mathcal{J}$-trivial and aperiodic monoids, and the general results we found along the way are presented in [Denton et al. 2010/11].

We further prove that the simple and projective modules of M are indexed by the elements of W. In order to construct the simple modules, we introduce for each $w \in W$ a combinatorial module T_w whose support is the interval $[1, w]_R$ in right weak order. This module yields an algebra, whose representation theory generalizes that of the Hecke group algebra, with the combinatorics of descents replaced by that of blocks and of the cutting poset.

Let us finish by giving some additional motivation for the study of the biHecke monoid. In type A, the tower of algebras $(\mathbb{K}M(\mathfrak{S}_n))_{n\in\mathbb{N}}$ possesses long sought-after properties. Indeed, it is well known that several combinatorial Hopf algebras arise as Grothendieck rings of towers of algebras. The prototypical example is the tower of algebras of the symmetric groups that gives rise to the Hopf algebra Sym of symmetric functions, on the Schur basis [Macdonald 1995; Zelevinsky 1981]. Another example, due to Krob and Thibon [1997], is the tower of the 0-Hecke algebras of the symmetric groups that gives rise to the Hopf algebra QSym of quasisymmetric functions of [Gessel 1984], on the F_I basis. The product rule on the F_I is naturally lifted through the descent map to a product on permutations, leading to the Hopf algebra FQSym of free quasisymmetric functions [Duchamp et al. 2002]. This calls for the existence of a tower of algebras $(A_n)_{n\in\mathbb{N}}$, such that each A_n contains $H_0(\mathfrak{S}_n)$ and has its simple modules indexed by the elements of $\mathfrak{S}_n$. The biHecke monoids $M(\mathfrak{S}_n)$, and their Borel submonoids $M_1(\mathfrak{S}_n)$ and $M_{w_0}(\mathfrak{S}_n)$, satisfy these properties, and are therefore expected to yield new representation theoretical interpretations of the bases of FQSym.

In the remainder of this introduction, we briefly review Coxeter groups and their 0-Hecke monoids, introduce the biHecke monoid, which is our main object of study, and outline the rest of the paper.

1a. ***Coxeter groups.*** Let (W, S) be a Coxeter group, that is, a group W with a presentation

$$W = \langle S \mid (ss')^{m(s,s')} \quad \text{for all } s, s' \in S\rangle, \tag{1-1}$$

with $m(s, s') \in \{1, 2, \dots, \infty\}$ and $m(s, s) = 1$. The elements $s \in S$ are called *simple reflections*, and the relations can be rewritten as

$$\begin{aligned} s^2 &= 1 && \text{for all } s \in S, \\ \underbrace{ss'ss's\cdots}_{m(s,s')} &= \underbrace{s'ss'ss'\cdots}_{m(s,s')} && \text{for all } s, s' \in S, \end{aligned} \tag{1-2}$$

where 1 denotes the identity in W.

Most of the time, we just write W for (W, S). In general, we follow the notation of [Björner and Brenti 2005], and we refer to this and to [Humphreys 1990] for details on Coxeter groups and their Hecke algebras. Unless stated otherwise, we always assume that W is finite, and denote its generators by $S = (s_i)_{i\in I}$, where $I = \{1, 2, \dots, n\}$ is the *index set* of W.

The prototypical example is the Coxeter group of type A_{n-1} which is the n-th symmetric group $(W, S) := (\mathfrak{S}_n, \{s_1, \dots, s_{n-1}\})$, where s_i denotes the *elementary transposition* which exchanges i and $i+1$. The relations are given by

$$\begin{aligned} s_i^2 &= 1 && \text{for } 1 \le i \le n-1,\\ s_i s_j &= s_j s_i && \text{for } |i-j| \ge 2,\\ s_i s_{i+1} s_i &= s_{i+1} s_i s_{i+1} && \text{for } 1 \le i \le n-2\,; \end{aligned} \tag{1-3}$$

the last two relations are called the *braid relations*. When writing a permutation $\mu \in \mathfrak{S}_n$ explicitly, we use *one-line notation*, that is the sequence $\mu_1\mu_2\dots\mu_n$, where $\mu_i := \mu(i)$.

A *reduced word* $i_1 \dots i_k$ for an element $w \in W$ corresponds to a decomposition $w = s_{i_1} \cdots s_{i_k}$ of w into a product of generators in S of minimal length $k = \ell(w)$. A *(right) descent* of w is an element $i \in I$ such that $\ell(ws_i) < \ell(w)$. If w is a permutation, this translates into $w_i > w_{i+1}$. *Left descents* are defined analogously. The sets of left and right descents of w are denoted by $\mathrm{D}_L(w)$ and $\mathrm{D}_R(w)$, respectively.

For $J \subseteq I$, we denote by $W_J = \langle s_j \mid j \in J\rangle$ the subgroup of W generated by s_j with $j \in J$. Furthermore, the longest element in W_J and W are denoted by s_J and w_0, respectively. Any finite Coxeter group $W := \langle s_i \mid i \in I\rangle$ can be realized as a finite reflection group; see for example [Humphreys 1990, Chapter 5.6] and [Björner and Brenti 2005, Chapter 4]. The generators s_i of W can be interpreted as reflections on hyperplanes in some $|I|$-dimensional vector space V. The simple roots α_i for $i \in I$ form a basis for V; the set of all roots is given by $\Phi := \{w(\alpha_i) \mid i \in I, w \in W\}$. One can associate reflections s_α to all roots $\alpha \in \Phi$. If $\alpha, \beta \in \Phi$ and $w \in W$, then $w(\alpha) = \beta$ if and only if $w s_\alpha w^{-1} = s_\beta$; see [Humphreys 1990, Chapter 5.7].

1b. *The 0-Hecke monoid.* The 0-*Hecke monoid* $H_0(W) = \langle \pi_i \mid i \in I\rangle$ of a Coxeter group W is generated by the *simple projections* π_i with relations

$$\begin{aligned} \pi_i^2 &= \pi_i && \text{for all } i \in I,\\ \underbrace{\pi_i\pi_j\pi_i\pi_j\cdots}_{m(s_i,s_j)} &= \underbrace{\pi_j\pi_i\pi_j\pi_i\cdots}_{m(s_i,s_j)} && \text{for all } i, j \in I\ . \end{aligned} \tag{1-4}$$

Thanks to these relations, the elements of $H_0(W)$ are canonically indexed by the elements of W by setting $\pi_w := \pi_{i_1} \cdots \pi_{i_k}$ for any reduced word $i_1 \dots i_k$ of w. We further denote by π_J the longest element of the *parabolic submonoid* $H_0(W_J) := \langle \pi_i \mid i \in J\rangle$.

As mentioned before, any finite Coxeter group W can be realized as a finite reflection group, each generator s_i of W acting by reflection along an hyperplane.

The corresponding generator π_i of the 0-Hecke monoid acts as a *folding*, reflecting away from the fundamental chamber on one side of the hyperplane and as the identity on the other side. Both the action of W and of $H_0(W)$ stabilize the set of reflecting hyperplanes and therefore induce an action on chambers.

The right regular representation of $H_0(W)$, or equivalently the action on chambers, induce a concrete realization of $H_0(W)$ as a monoid of operators acting on W, with generators $\pi_1, \dots, \pi_n$ defined by

$$w\,.\pi_i := \begin{cases} w & \text{if } i \in \mathrm{D}_R(w), \\ ws_i & \text{otherwise.} \end{cases} \tag{1-5}$$

In type A, π_i sorts the letters at positions i and $i+1$ decreasingly, and $w\,.\pi_{w_0} = n \cdots 21$ for any permutation w. This justifies naming π_i an *elementary bubble antisorting operator*.

Another concrete realization of $H_0(W)$ can be obtained by considering instead the *elementary bubble sorting operators* $\overline{\pi}_1, \dots, \overline{\pi}_n$, whose action on W are defined by

$$w\,.\overline{\pi}_i := \begin{cases} ws_i & \text{if } i \in \mathrm{D}_R(w), \\ w & \text{otherwise.} \end{cases} \tag{1-6}$$

In geometric terms, this is folding toward the fundamental chamber. In type A, and for any permutation w, one has $w\,.\overline{\pi}_{w_0} = 12 \cdots n$.

Remark 1.1. For a given $w \in W$, define v by $wv = w_0$, where w_0 is the longest element of W. Then

$$i \in \mathrm{D}_R(w) \iff i \notin \mathrm{D}_L(v) \iff i \notin \mathrm{D}_R(v^{-1}) = \mathrm{D}_R(w_0 w).$$

Hence, the action of $\overline{\pi}_i$ on W can be expressed from the action of π_i on W using w_0:

$$w\,.\overline{\pi}_i = w_0[(w_0 w)\,.\pi_i].$$

1c. *The biHecke monoid $M(W)$.* We now introduce our main object of study.

Definition 1.2. Let W be a finite Coxeter group. The *biHecke monoid* is the submonoid of functions from W to W generated simultaneously by the elementary bubble sorting and antisorting operators of (1-5) and (1-6):

$$M := M(W) := \langle \pi_1, \pi_2, \dots, \pi_n, \overline{\pi}_1, \overline{\pi}_2, \dots, \overline{\pi}_n \rangle.$$

As mentioned in [Hivert and Thiéry 2009; Hivert et al. 2009] this monoid admits several natural variants, depending on the choice of the generators:

$$\langle \pi_1, \pi_2, \dots, \pi_n, s_1, s_2, \dots, s_n \rangle,$$
$$\langle \pi_0, \pi_1, \pi_2, \dots, \pi_n \rangle,$$

where π_0 is defined when W is a Weyl group and hence can be affinized. Unlike the algebras they generate, which all coincide with the biHecke algebra (in particular due to the linear relation $1+s_i=\pi_i+\overline{\pi}_i$ which expresses how to recover a reflection by gluing together the two corresponding foldings), these monoids are all distinct as soon as W is large enough. Another close variant is the monoid of all strictly order-preserving functions on the Boolean lattice [Gaucher 2010]. All of these monoids, and their representation theory, remain to be studied.

1d. ***Outline.*** The remainder of this paper consists of two parts: We first introduce and study the new cutting poset structure on finite Coxeter groups, and then proceed to the biHecke monoid and its representation theory.

In Section 2, we recall some needed basic facts, definitions, and properties about posets, Coxeter groups, monoids, and representation theory.

In Section 3, we generalize the notion of blocks of permutation matrices to any Coxeter group, and use it to define a new poset structure on W, which we call the *cutting poset*; we prove that it is (almost) a lattice, and derive that its Möbius function is essentially that of the hypercube.

In Section 4, we study the combinatorial properties of $M(W)$. In particular, we prove that it preserves left and Bruhat order, derive consequences on the fibers and image sets of its elements, prove that it is aperiodic, and study Green's relations and idempotents.

In Section 5, our strategy is to consider a "Borel" triangular submonoid of $M(W)$ whose representation theory is simpler, but with the same number of simple modules, to later induce back information about the representation theory of $M(W)$. Namely, we study the submonoid $M_1(W)$ of the elements fixing 1 in $M(W)$. This monoid not only preserves Bruhat order, but furthermore is regressive. It follows that it is $\mathcal{J}$-trivial (in fact $\mathcal{B}$-trivial) which is the desired triangularity property. It is for example easily derived that $M_1(W)$ has $|W|$ simple modules, all of dimension 1. In fact most of our results about M_1 generalize to any $\mathcal{J}$-trivial monoid, which is the topic of a separate paper on the representation theory of $\mathcal{J}$-trivial monoids [Denton et al. 2010/11]. We also provide properties of the Cartan matrix and a combinatorial description of the quiver of M_1.

In Section 6, we construct, for each $w \in W$, the *translation module* T_w by induction of the corresponding simple $\mathbb{K}M_1(W)$-module. It is a quotient of the indecomposable projective module P_w of $\mathbb{K}M(W)$, and therefore admits the simple module S_w of $\mathbb{K}M(W)$ as top. It further admits a simple combinatorial model using the right classes with the interval $[1, w]_R$ as support, and which passes down to S_w. We derive a formula for the dimension of S_w, using an inclusion-exclusion on the sizes of intervals in $(W, \leq_R)$ along the cutting poset. On the way, we study the algebra $\mathcal{H}W^{(w)}$ induced by the action of $M(W)$ on T_w. It turns out to be a natural

w-analogue of the Hecke group algebra, acting not anymore on the full Coxeter group, but on the interval $[1, w]_R$ in right order. All the properties of the Hecke group algebra pass through this generalization, with the combinatorics of descents being replaced by that of blocks and of the cutting poset. In particular, $\mathscr{H}W^{(w)}$ is Morita equivalent to the incidence algebra of the sublattice induced by the cutting poset on the interval $[1, w]_{\sqsubseteq}$.

In Section 7, we apply the findings of Sections 4, 5, and 6 to derive results on the representation theory of $M(W)$. We conclude in Section 8 with discussions on further research in progress.

There are two appendices. Appendix A summarizes some results on colored graphs which are used in Section 4 to prove properties of the fibers and image sets of elements in the biHecke monoid. Appendix B we present tables of q-Cartan invariant and decomposition matrices for $M(\mathfrak{S}_n)$ for $n = 2, 3, 4$.

2. Background

We review some basic facts about partial orders and finite posets in Section 2a, finite lattices and Birkhoff's theorem in Section 2b, order-preserving functions in Section 2c, the usual partial orders on Coxeter groups (left and right weak order, Bruhat order) in Section 2d, and the notion of $\mathscr{J}$-order (and related orders) and aperiodic monoids in Section 2e. We also prove a result in Proposition 2.4 about the image sets of order-preserving and regressive idempotents on a poset that will be used later in the study of idempotents of the biHecke monoid. Sections 2f and 2g contain reviews of some representation theory of algebras and monoids that will be relevant in our study of translation modules.

2a. *Finite posets.* For a general introduction to posets and lattices, we refer the reader to for example [Pouzet 2013; Stanley 1997] or [Wikipedia 2010, Poset, Lattice]. Throughout this paper, all posets are finite.

A partially ordered set (or *poset* for short) $(P, \preceq)$ is a set P with a binary relation $\preceq$ such that for all $x, y, z \in P$:

(i) $x \preceq x$ (reflexivity);

(ii) if $x \preceq y$ and $y \preceq x$, then $x = y$ (antisymmetry);

(iii) if $x \preceq y$ and $y \preceq z$, then $x \preceq z$ (transitivity).

When we exclude the possibility that $x = y$, we write $x \prec y$.

If $x \preceq y$ in P, we define the *interval*

$$[x, y]_P := \{z \in P \mid x \preceq z \preceq y\}.$$

A pair (x, y) such that $x \prec y$ and there is no $z \in P$ such that $x \prec z \prec y$ is called a *covering*. We denote coverings by $x \to y$. The *Hasse diagram* of $(P, \preceq)$ is the diagram where the vertices are the elements $x \in P$, and there is an upward-directed edge between x and y if $x \to y$.

Definition 2.1. Let $(P, \preceq)$ be a poset and $X \subseteq P$.

(i) X is *convex* if for any $x, y \in X$ with $x \preceq y$ we have $[x, y] \subseteq X$.

(ii) X is *connected* if for any $x, y \in X$ with $x \prec y$ there is a path in the Hasse diagram $x = x_0 \to x_1 \to \cdots \to x_k = y$ such that $x_i \in X$ for $0 \le i \le k$.

The Möbius inversion formula [Stanley 1997, Proposition 3.7.1] generalizes the inclusion-exclusion principle to any poset. Namely, there exists a unique function μ, called the *Möbius function* of P, which assigns an integer to each ordered pair $x \preceq y$ and enjoys the following property: For any two functions $f, g : P \to G$ taking values in an additive group G,

$$g(x) = \sum_{y \preceq x} f(y) \quad \text{if and only if} \quad f(y) = \sum_{x \preceq y} \mu(x, y)\, g(x). \tag{2-1}$$

The Möbius function can be computed thanks to the following recursion:

$$\mu(x, y) = \begin{cases} 1 & \text{if } x = y, \\ -\sum_{x \preceq z \prec y} \mu(x, z) & \text{for } x \prec y. \end{cases}$$

2b. ***Finite lattices and Birkhoff's theorem.*** Let $(P, \preceq)$ be a poset. The *meet* $z = \bigwedge A$ of a subset $A \subseteq P$ is an element such that, first, $z \preceq x$ for all $x \in A$ and, second, $u \preceq x$ for all $x \in A$ implies that $u \preceq z$. When the meet exists, it is unique and is denoted by $\bigwedge A$. The meet of the empty set $A = \{\}$ is the largest element of the poset, if it exists. The meet of two elements $x, y \in P$ is denoted by $x \wedge y$. A poset $(P, \preceq)$ for which every pair of elements has a meet is called a *meet-semilattice*. In that case, P endowed with the meet operation is a commutative $\mathcal{J}$-trivial semigroup, and in fact a monoid with unit the maximal element of P, if the latter exists.

Reversing all comparisons, one can similarly define the *join* $\bigvee A$ of a subset $A \subseteq P$ or $x \vee y$ of two elements $x, y \in P$, and *join-semilattices*. A *lattice* is a poset for which both meets and joins exist for pair of elements. Recall that we only consider finite posets, so we do not have to worry about the distinction between lattices and complete lattices.

A lattice $(L, \vee, \wedge)$ is *distributive* if the following additional identity holds for all $x, y, z \in L$:

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

This condition is equivalent to its dual,

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

Birkhoff's representation theorem (see [Wikipedia 2010, Birkhoff's representation theorem], or [Stanley 1997, Theorem 3.4.1]) states that any finite distributive lattice can be represented as a sublattice of a Boolean lattice, that is, a collection of sets stable under union and intersection. Furthermore, there is a canonical such representation, which we construct now.

An element z in a lattice L is called *join-irreducible* if z is not the smallest element in L and $z = x \vee y$ implies $z = x$ or $z = y$ for any $x, y \in L$ (and similarly for meet-irreducible). Equivalently, since L is finite, z is join-irreducible if and only if it covers exactly one element in L. We denote by $I(L)$ the *poset of join-irreducible elements* of L, that is the restriction of L to its join-irreducible elements. Note that this definition still makes sense for nonlattices. From a monoid point of view, $I(L)$ is the minimal generating set of L.

A *lower set* of a poset P is a subset Y of P such that, for any pair $x \leq y$ of comparable elements of P, x is in Y whenever y is. *Upper sets* are defined dually. The family of lower sets of P ordered by inclusion is a distributive lattice, the *lower sets lattice* $O(P)$. Birkhoff's representation theorem [Birkhoff 1937] states that any finite distributive lattice L is isomorphic to the lattice $O(I(L))$ of lower sets of the poset $I(L)$ of its join-irreducible elements, via the reciprocal isomorphisms:

$$\begin{cases} L & \to O(I(L)), \\ x & \mapsto \{y \in I(L) \mid y \leq x\} \end{cases} \quad \text{and} \quad \bigvee : \begin{cases} O(I(L)) & \to L, \\ I & \mapsto \bigvee I \,. \end{cases}$$

Following Edelman [1986], a meet-semilattice L is *meet-distributive* if for every $y \in L$, if $x \in L$ is the meet of elements covered by y then $[x, y]$ is a Boolean algebra. A stronger condition is that any interval of L is a distributive lattice. A straightforward application of Birkhoff's representation theorem yields that L is then isomorphic to a lower set of $O(I(L))$.

2c. *Order-preserving functions.*

Definition 2.2. Let $(P, \preceq)$ be a poset and $f : P \to P$ a function.

(i) f is called *order-preserving* if $x \preceq y$ implies $f(x) \preceq f(y)$. We also say f preserves the order $\preceq$.

(ii) f is called *regressive* if $f(x) \preceq x$ for all $x \in P$.

(iii) f is called *extensive* if $x \preceq f(x)$ for all $x \in P$.

Lemma 2.3. *Let $(P, \preceq)$ be a poset and $f : P \to P$ an order-preserving map. Then, the preimage $f^{-1}(C)$ of a convex subset $C \subseteq P$ is convex. In particular, the preimage of a point is convex.*

Proof. Let $x, y \in f^{-1}(C)$ with $x \preceq y$. Since f is order-preserving, for any $z \in [x, y]$, we have $f(x) \preceq f(z) \preceq f(y)$, and therefore $f(z) \in C$. □

Proposition 2.4. *Let $(P, \preceq)$ be a poset and $f : P \to P$ be an order-preserving and regressive idempotent. Then, f is determined by its image set. Namely, for $u \in P$ we have*

$$f(u) = \sup_{\preceq}(\downarrow u \cap \operatorname{im}(f)),$$

the supremum being always well-defined. Here $\downarrow u = \{x \in P \mid x \preceq u\}$.

An equivalent statement is that, for $v \in \operatorname{im}(f)$,

$$f^{-1}(v) = \uparrow v \setminus \bigcup_{\substack{v' \in \operatorname{im}(f) \\ v' \succ v}} \uparrow v', \quad \textit{where } \uparrow v = \{x \in P \mid x \succeq v\}.$$

Proof. We first prove that $\downarrow u \cap \operatorname{im}(f) = f(\downarrow u)$. The inclusion $\supseteq$ follows from the fact that f is regressive: Taking $v \in \downarrow u$, we have $f(v) \preceq v \preceq u$ and therefore $f(v) \in \downarrow u \cap \operatorname{im}(f)$. The inclusion $\subseteq$ follows from the assumption that f is an idempotent: For $v \in \operatorname{im}(f)$ with $v \preceq u$, one has $v = f(v)$, so $v \in f(\downarrow u)$.

Since f is order-preserving, $f(\downarrow u)$ has a unique maximal element, namely $f(u)$. The first statement of the proposition follows. The second statement is a straightforward reformulation of the first one. □

An *interior operator* (sometimes also called a kernel operator) is a function $L \to L$ on a lattice L that is order-preserving, regressive and idempotent; see for example [Wikipedia 2010, Moore Family]. A subset $A \subseteq L$ is a *dual Moore family* if it contains the smallest element $\bot_L$ of L and is stable under joins. The image set of an interior operator is a *dual Moore family*. Reciprocally, any dual Moore family A defines an interior operator by

$$L \to L, \quad x \mapsto \operatorname{red}(x) := \bigvee_{a \in A, a \preceq x} a, \tag{2-2}$$

where $\bigvee_{\{\}} = \bot_L$ by convention.

A (dual) Moore family is itself a lattice with the order and join inherited from L. The meet operation usually differs from that of L and is given by $x \wedge_A y = \operatorname{red}(x \wedge_L y)$.

2d. ***Classical partial orders on Coxeter groups.*** A Coxeter group $W = \langle s_i \mid i \in I \rangle$ comes endowed with several natural partial orders: left (weak) order, right (weak) order, left-right (weak) order, and Bruhat order. All of these play an important role for the representation theory of the biHecke monoid $M(W)$.

Fix $u, w \in W$. Then, in *right* (*weak*) *order*,

$$u \leq_R w \quad \text{if } w = u s_{i_1} \cdots s_{i_k} \text{ for some } i_j \in I \text{ and } \ell(w) = \ell(u) + k.$$

Similarly, in *left* (*weak*) *order*,

$$u \leq_L w \quad \text{if } w = s_{i_1} \cdots s_{i_k} u \text{ for some } i_j \in I \text{ and } \ell(w) = \ell(u) + k,$$

and in *left-right* (*weak*) *order*,

$$u \leq_{LR} w \quad \text{if } w = s_{i_1} \cdots s_{i_k} u s_{i'_1} \cdots s_{i'_\ell} \text{ for some } i_j, i'_j \in I \text{ and } \ell(w) = \ell(u) + k + \ell.$$

Note that left-right order is the transitive closure of the union of left and right order. Thanks to associativity, this is equivalent to the existence of a $v \in W$ such that $u \leq_L v$ and $v \leq_R w$.

Let $w = s_{i_1} s_{i_2} \cdots s_{i_\ell}$ be a reduced expression for w. Then, in *Bruhat order*,

$$u \leq_B w \quad \begin{array}{l}\text{if there exists a reduced expression } u = s_{j_1} \cdots s_{j_k} \\ \text{where } j_1 \dots j_k \text{ is a subword of } i_1 \dots i_\ell.\end{array}$$

For any finite Coxeter group W, the posets $(W, \leq_R)$ and $(W, \leq_L)$ are graded lattices [Björner and Brenti 2005, Section 3.2]. The following proposition states that any interval is isomorphic to some interval starting at 1:

Proposition 2.5 [Björner and Brenti 2005, Proposition 3.1.6]. *Let* $\mathbb{O} \in \{L, R\}$ *and* $u \leq_{\mathbb{O}} w \in W$. *Then* $[u, w]_{\mathbb{O}} \cong [1, t]_{\mathbb{O}}$ *where* $t = wu^{-1}$.

Definition 2.6. The *type* of an interval in left and right order are defined to be $\operatorname{type}([u, w]_L) := wu^{-1}$ and $\operatorname{type}([u, w]_R) := u^{-1}w$, respectively.

It is easily shown that, if $\mathbb{O}$ is considered as a colored poset, then the converse of Proposition 2.5 holds as well:

Remark 2.7. Fix a type t. Then, the collection of all intervals in left weak order of type t is in bijection with $[1, t^{-1}w_0]_R$, and the operators π_i and $\overline{\pi}_i$ act transitively on the right on this collection. More precisely: π_a induces an isomorphism from $[1, ba^{-1}]_L$ to $[a, b]_L$, and $\overline{\pi}_{a^{-1}}$ induces an isomorphism from $[a, b]_L$ to $[1, ba^{-1}]_L$.

Proof. Take $u \in [a, b]_L$, and let $s_{i_1} \cdots s_{i_k}$ be a reduced decomposition of a. Let $s_{j_1} \cdots s_{j_\ell}$ be a reduced decomposition of $ua^{-1} = us_{i_k} \cdots s_{i_1}$. Then

$$u = (s_{j_1} \cdots s_{j_\ell})(s_{i_1} \cdots s_{i_k})$$

is a reduced decomposition of u and $u.\overline{\pi}_{a^{-1}} = s_{j_1} \cdots s_{j_\ell} = ua^{-1}$. Reciprocially, applying π_a to an element $u \in [1, ba^{-1}]_L$ progressively builds up a reduced word for a. The result follows. □

2e. ***Preorders on monoids.*** J. A. Green [1951] introduced several preorders on monoids, which are essential for the study of their structures; see for example

[Pin 2012, Chapter V]. *Throughout this paper, we only consider finite monoids.* Define $\leq_{\mathcal{R}}, \leq_{\mathcal{L}}, \leq_{\mathcal{J}}, \leq_{\mathcal{H}}$ for $x, y \in M$ as follows:

$$\begin{aligned}
x \leq_{\mathcal{R}} y &\quad \text{if and only if } x = yu \text{ for some } u \in M,\\
x \leq_{\mathcal{L}} y &\quad \text{if and only if } x = uy \text{ for some } u \in M,\\
x \leq_{\mathcal{J}} y &\quad \text{if and only if } x = uyv \text{ for some } u, v \in M,\\
x \leq_{\mathcal{H}} y &\quad \text{if and only if } x \leq_{\mathcal{R}} y \text{ and } x \leq_{\mathcal{L}} y.
\end{aligned}$$

These preorders give rise to equivalence relations:

$$\begin{aligned}
x \,\mathcal{R}\, y &\quad \text{if and only if } xM = yM,\\
x \,\mathcal{L}\, y &\quad \text{if and only if } Mx = My,\\
x \,\mathcal{J}\, y &\quad \text{if and only if } MxM = MyM,\\
x \,\mathcal{H}\, y &\quad \text{if and only if } x\mathcal{R}y \text{ and } x\mathcal{L}y.
\end{aligned}$$

Strict comparisons are defined by $x <_{\mathcal{R}} y$ if $x \leq_{\mathcal{R}} y$ but $x \notin \mathcal{R}(y)$, or equivalently $\mathcal{R}(x) \subset \mathcal{R}(y)$, and similarly for $<_{\mathcal{L}}, <_{\mathcal{J}}, <_{\mathcal{H}}$.

We further add the relation $\leq_{\mathcal{B}}$ (and its associated equivalence relation $\mathcal{B}$) defined as the finest preorder such that $x \leq_{\mathcal{B}} 1$, and

$$x \leq_{\mathcal{B}} y \text{ implies that } uxv \leq_{\mathcal{B}} uyv \text{ for all } x, y, u, v \in M.$$

(One can view $\leq_{\mathcal{B}}$ as the intersection of all preorders with the property above. There exists at least one such preorder, namely $x \leq y$ for all $x, y \in M$). In the semigroup community, this order is sometimes colloquially referred to as the *multiplicative $\mathcal{J}$-order*.

Beware that 1 is the largest element of those (pre)-orders. This is the usual convention in the semigroup community, but is the converse convention from the closely related notions of left/right/left-right/Bruhat order in Coxeter groups as introduced in Section 2d.

Example 2.8. For the 0-Hecke monoid of Section 1b, $\mathcal{K}$-order for $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}, \mathcal{B}\}$ corresponds to the reverse of right, left, left-right and Bruhat order of Section 2d. More precisely for $x, y \in H_0(W)$, $x \leq_{\mathcal{K}} y$ if and only if $x \geq_K y$ for $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}, \mathcal{B}\}$ and $K \in \{R, L, LR, B\}$ the corresponding letter.

Definition 2.9. Elements of a monoid M in the same $\mathcal{K}$-equivalence class are called $\mathcal{K}$-classes, where $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}, \mathcal{H}, \mathcal{B}\}$. The $\mathcal{K}$-class of $x \in M$ is denoted by $\mathcal{K}(x)$.

A monoid M is called *$\mathcal{K}$-trivial* if all $\mathcal{K}$-classes are of cardinality one.

An element $x \in M$ is called *regular* if it is $\mathcal{J}$-equivalent to an idempotent.

An equivalent formulation of $\mathcal{K}$-triviality is given in terms of *ordered* monoids. A monoid M is called

right-ordered	if $xy \le x$ for all $x, y \in M$,
left-ordered	if $xy \le y$ for all $x, y \in M$,
left-right-ordered	if $xy \le x$ and $xy \le y$ for all $x, y \in M$,
two-sided-ordered	if $xy = yz \le y$ for all $x, y, z \in M$ with $xy = yz$,
ordered with 1 *on top*	if $x \le 1$, and $x \le y$ implies $uxv \le uyv$ for all $x, y, u, v \in M$

for some partial order $\le$ on M.

Proposition 2.10. *M is right-ordered (respectively left-ordered, left-right-ordered, two-sided-ordered, ordered with* 1 *on top) if and only if M is $\mathcal{R}$-trivial (respectively $\mathcal{L}$-trivial, $\mathcal{J}$-trivial, $\mathcal{H}$-trivial, $\mathcal{B}$-trivial).*

When M is $\mathcal{K}$-trivial for $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}, \mathcal{H}, \mathcal{B}\}$, the partial order $\le$ is finer than $\le_{\mathcal{K}}$; that is, for any $x, y \in M$, $x \le_{\mathcal{K}} y$ implies $x \le y$.

Proof. We give the proof for right-order as the other cases can be proved in a similar fashion.

Suppose M is right-ordered and that $x, y \in M$ are in the same $\mathcal{R}$-class. Then $x = ya$ and $y = xb$ for some $a, b \in M$. This implies that $x \le y$ and $y \le x$, so that $x = y$. Conversely, suppose that all $\mathcal{R}$-classes are singletons. Then $x \le_{\mathcal{R}} y$ and $y \le_{\mathcal{R}} x$ imply that $x = y$, so that the $\mathcal{R}$-preorder turns into a partial order. Hence M is right-ordered using $xy \le_{\mathcal{R}} x$. □

Definition 2.11. A monoid M is *aperiodic* if there is an integer $N > 0$ such that $x^N = x^{N+1}$ for each $x \in M$.

Since we are only dealing with finite monoids, it is enough to find such an $N = N_x$ depending on the element x. Indeed, taking $N := \max\{N_x\}$ gives a uniform bound. From this definition it is clear that, for an aperiodic monoid M, the sequence $(x^n)_{n \in \mathbb{N}}$ eventually stabilizes for every $x \in M$. We write x^ω for the stable element, which is idempotent, and $E(M) := \{x^\omega \mid x \in M\}$ for the set of idempotents.

Equivalent characterizations of (finite) aperiodic monoids M are that they are $\mathcal{H}$-trivial, or that the sub-semigroup S of M (the identity of S is not necessarily the one of M), which are also groups, are trivial; see for example [Pin 2012, VII, 4.2, Aperiodic monoids]. In this sense, the notion of aperiodic monoids is orthogonal to that of groups as they contain no group-like structure. By the same token, their representation theory is orthogonal to that of groups.

As we will see in Section 4d, the biHecke monoid $M(W)$ of Definition 1.2 is aperiodic. Its Borel submonoid $M_1(W)$ of functions fixing the identity is $\mathcal{J}$-trivial (see Section 5).

2f. ***Representation theory of algebras.*** We refer to [Curtis and Reiner 1962] for an introduction to representation theory, and to [Benson 1991] for more advanced notions such as Cartan matrices and quivers. Here we mostly review composition series and characters.

Let A be a finite-dimensional algebra. Given an A-module X, any strictly increasing sequence $(X_i)_{i\le k}$ of submodules

$$\{0\} = X_0 \subset X_1 \subset X_2 \subset \cdots \subset X_k = X$$

is called a *filtration* of X. A filtration $(Y_j)_{i\le \ell}$ such that, for any i, $Y_i = X_j$ for some j is called a *refinement of* $(X_i)_{i\le k}$. A filtration $(X_i)_{i\le k}$ without a nontrivial refinement is called a *composition series*. For a composition series, each quotient module X_j/X_{j-1} is simple and is called a *composition factor*. The multiplicity of a simple module S in the composition series is the number of indices j such that X_j/X_{j-1} is isomorphic to S. The Jordan–Hölder theorem states that this multiplicity does not depend on the choice of the composition series. Hence, we may define the *generalized character* (or *character* for short) of a module X as the formal sum

$$[X] := \sum_{i\in I} c_i[S_i],$$

where I indexes the simple modules of A and c_i is the multiplicity of the simple module S_i in any composition series for X.

The additive group of formal sums $\sum_{i\in I} m_i[S_i]$, with $m_i \in \mathbb{Z}$, is called the *Grothendieck group of the category of A-modules* and is denoted by $G_0(A)$. By definition, the character satisfies that, for any exact sequence

$$0 \to X \to Y \to Z \to 0,$$

the equality

$$[Y] = [X] + [Z]$$

holds in the Grothendieck group. See [Serre 1977] for more information about Grothendieck groups.

Suppose that B is a subalgebra of A. Any A-module X naturally inherits an action from B. The constructed B-module thereby is called the *restriction of X to B* and its B-character $[X]_B$ depends only on its A-character $[X]_A$. Indeed, any A-composition series can be refined to a B-composition series and the resulting multiplicities depend only on those in the A-composition series and in the composition series of the simple modules of A restricted to B. This defines a $\mathbb{Z}$-linear map $[X]_A \mapsto [X]_B$, called the *decomposition map*. Let $(S_i^A)_{i\in I}$ and $(S_j^B)_{j\in J}$ be complete families of simple module representatives for A and B, respectively. The matrix of the decomposition map is called the *decomposition matrix* of A over B;

its coefficient (i, j) is the multiplicity of S_j^B as a composition factor of S_i^A, viewed as a B-module.

The adjoint construction of restriction is called *induction*: For any right B-module X the space

$$X{\uparrow}_B^A := X \otimes_B A$$

is naturally endowed with a right A-module structure by right multiplication by elements of A, and is called the *module induced by X from B to A*.

The next subsection, and in particular the statement of Theorem 2.13, requires a slightly more general setting, where the identity e of B does not coincide with that of A. More precisely, let B be a subalgebra of eAe for some idempotent e of A. Then, for any A-module Y, the *restriction* of Y to B is defined as Ye, whereas, for any B-module X, the *induction* of X to A is defined as $X{\uparrow}_B^A := X \otimes_B eA$.

2g. ***Representation theory of monoids.*** Although representation theory started at the beginning of the 20th century with groups before being extended to more general algebraic structures such as algebras, one has to wait until [Clifford 1942] for the first results on the representation theory of semigroups and monoids. Renewed interest in this subject was sparked more recently by the emergence of connections with probability theory and combinatorics; see for example [Brown 2000; Saliola 2007]. Compared to groups, only a few general results are known, the most important one being the construction of the simple modules. It is originally due to Clifford, Munn, and Ponizovskiĭ, and we recall here the construction of [Ganyushkin et al. 2009] (see also the historical references therein) from the regular $\mathcal{J}$-classes and corresponding right class modules.

In principle, one should be specific about the ground field $\mathbb{K}$; in other words, one should consider the representation theory of the *monoid algebra* $\mathbb{K}M$ of a monoid M, and not of the monoid itself. However, the monoids under study in this paper are aperiodic, and their representation theory only depends on the characteristic. We focus on the case where $\mathbb{K}$ is of characteristic 0. Note that the general statements mentioned in this section may further require $\mathbb{K}$ to be large enough (e.g., $\mathbb{K} = \mathbb{C}$) for nonaperiodic monoids.

Let M be a finite monoid. Fix a *regular* $\mathcal{J}$-class J, that is, a $\mathcal{J}$-class containing an idempotent. Consider the sets

$$M_{\geq J} := \bigcup_{K \in \mathcal{J}(M),\ K \geq_{\mathcal{J}} J} K \quad \text{and} \quad I_J := M - M_{\geq J}.$$

Then, I_J is an ideal of M, so that the vector space $\mathbb{K}M_{\geq J}$ can be endowed with an algebra structure by identifying it with the quotient $\mathbb{K}M/\mathbb{K}I_J$. Note that any $\mathbb{K}M_{\geq J}$-module is then a $\mathbb{K}M$-module.

Definition 2.12. Let $f \in M$. Set $\mathbb{K}\mathcal{R}_<(f) := \mathbb{K}\{b \in fM \mid b <_{\mathcal{R}} f\}$. The *right class module* of f (also known as *right Schützenberger representation*) is the $\mathbb{K}M$-module

$$\mathbb{K}\mathcal{R}(f) := \mathbb{K}fM/\mathbb{K}\mathcal{R}_<(f).$$

$\mathbb{K}\mathcal{R}(f)$ is clearly a right module since $\mathbb{K}\mathcal{R}_<(f)$ is a submodule of $\mathbb{K}fM$. Also, as suggested by the notation, $\mathcal{R}(f)$ forms a basis of $\mathbb{K}\mathcal{R}(f)$. Moreover, for a fixed $\mathcal{J}$-class J and thanks to associativity and finiteness, the right class module $\mathbb{K}\mathcal{R}(f)$ does not depend on the choice of $f \in J$ (up to isomorphism). Our main tool for studying the representation theory of the biHecke monoid will be a combinatorial model for its right class modules, which we will call *translation modules* (see Section 6a).

We now choose a $\mathcal{J}$-class J, fix an idempotent e_J in J, and set $\mathbb{K}\mathcal{R}_J := \mathbb{K}\mathcal{R}(e_J)$. Recall that

$$\mathcal{R}(e_J) = e_J M \cap J = e_J M_{\geq J} \cap J.$$

Define similarly

$$G_J := G_{e_J} := e_J M e_J \cap J = e_J M_{\geq J} e_J \cap J.$$

Then, G_J is a group that does not depend on the choice of e_J. More precisely, if e and f are two idempotents in J, the ideals MeM and MfM are equal and the groups G_e and G_f are conjugate and isomorphic. Note that when working with the quotient algebra $\mathbb{K}M_{\geq J}$, the equations above simplify to

$$\mathbb{K}\mathcal{R}_J = e_J \mathbb{K}M_{\geq J} \quad \text{and} \quad \mathbb{K}G_J = e_J \mathbb{K}M_{\geq J} e_J.$$

With these notations, the simple $\mathbb{K}M$-modules can be constructed as follows:

Theorem 2.13 (Clifford, Munn, and Ponizovskiĭ; see [Ganyushkin et al. 2009, Theorem 7]). *Let M be a monoid, and $\mathfrak{U}(M)$ be the set of its regular $\mathcal{J}$-classes. For any $J \in \mathfrak{U}(M)$, define the right class module $\mathbb{K}\mathcal{R}_J$ and groups G_J as above, let $S_1^J, \dots, S_{n_J}^J$ be a complete family of simple $\mathbb{K}G_J$-modules, and set*

$$X_i^J := \operatorname{top}(S_i^J \uparrow_{\mathbb{K}G_J}^{\mathbb{K}M_{\geq J}}) = \operatorname{top}(S_i^J \otimes_{\mathbb{K}G_J} e_J \mathbb{K}M_{\geq J}) = \operatorname{top}(S_i^J \otimes_{\mathbb{K}G_J} \mathbb{K}\mathcal{R}_J), \quad (2\text{-}3)$$

where $\operatorname{top}(X) := X/\operatorname{rad} X$ is the semisimple quotient of the module X. Then, (X_i^J for $J \in \mathfrak{U}(M)$ and $i = 1, \dots, n_J$) is a complete family of simple $\mathbb{K}M$-modules.

In the present paper we only need the very particular case of aperiodic monoids. The key point is that a monoid is aperiodic if and only if all the groups G_J are trivial [Pin 2012, Proposition 4.9]: $G_J = \{e_J\}$. As a consequence, the only $\mathbb{K}G_J$-module is the trivial one, 1, so that the previous construction boils down to the following theorem:

Theorem 2.14. *Let M be an aperiodic monoid. Choose an idempotent transversal $E = \{e_J \mid J \in \mathscr{U}(M)\}$ of the regular $\mathscr{J}$-classes. Further set*

$$X^J := \operatorname{top}(1\uparrow_{\mathbb{K}e_J}^{\mathbb{K}M_{\geq J}}) = \operatorname{top}(e_J \mathbb{K}M_{\geq J}) = \operatorname{top}(\mathbb{K}\mathscr{R}_J). \tag{2-4}$$

Then, the family $(X^J)_{J\in\mathscr{U}(M)}$ is a complete family of representatives of simple $\mathbb{K}M$-modules. In particular, there are as many isomorphic types of simple modules as regular $\mathscr{J}$-classes.

Since the top of $\mathbb{K}\mathscr{R}_J$ is simple, one obtains immediately the following corollary; see [Curtis and Reiner 1962, Corollary 54.14].

Corollary 2.15. *Each regular right class module $\mathbb{K}\mathscr{R}_J$ is indecomposable and a quotient of the projective module P_J corresponding to S_J.*

For a nonaperiodic finite monoid, each right class module remains indecomposable even if its top is not necessarily simple; see [Zalcstein 1971, Corollary 1.10].

The top of a right class module $\mathbb{K}\mathscr{R}_J$ is easy to compute; indeed, the radical of this module is nothing but the annihilator of J acting on it. This in turn boils down to the calculation of the kernel of a matrix as we see below.

Rees matrix monoids [Rees 1940] play an important role in the representation theory of monoids, because any $\mathscr{J}$-class J of any monoid M is, roughly speaking, isomorphic to such a monoid. We give here the definition of aperiodic Rees matrix monoids, which we use in a couple of examples (see Examples 7.8 and 7.9).

Definition 2.16 (aperiodic Rees matrix monoid). Let $P = (p_{ij})$ be an $n \times m$ 0-1-matrix. The *aperiodic Rees matrix monoid $M(P)$* is obtained by endowing the disjoint union

$$\{1\} \cup \{1, \dots, m\} \times \{1, \dots, n\} \cup \{0\}$$

with the product

$$(i, j)(i', j') := \begin{cases} (i, j') & \text{if } p_{ji'} = 1, \\ 0 & \text{otherwise,} \end{cases}$$

1 being neutral and 0 being the zero element.

Note that (i, j) is an idempotent if and only if $p_{j,i} = 1$; hence $M(P)$ can be alternatively described by specifying which elements (i, j) are idempotent.

Without entering into the details, we note that the radical of the unique (up to isomorphism) nontrivial right class modules of $\mathbb{K}M(P)$ is given by the kernel of the matrix P, and thus the dimension of the nontrivial simple module of $\mathbb{K}M(P)$ is given by the rank of P [Clifford and Preston 1961; Lallement and Petrich 1969; Rhodes and Zalcstein 1991; Margolis and Steinberg 2011].

3. Blocks of Coxeter group elements and the cutting poset

In this section, we develop the combinatorics underlying the representation theory of the translation modules studied in Section 6. The key question is, Given $w \in W$, for which subsets $J \subseteq I$ does the canonical bijection between a Coxeter group W and the Cartesian product $W_J \times {}^J W$ of a parabolic subgroup W_J by its set of coset representatives ${}^J W$ in W restrict properly to an interval $[1, w]_R$ in right order (see Figure 1)? In type A, the answer is given by the so-called blocks in the permutation matrix of w, and we generalize this notion to any Coxeter group.

We start with some results on parabolic subgroups and quotients in Section 3a, which are used to define *blocks* and *cutting points* of Coxeter group elements in Section 3b. Then, we illustrate the notion of blocks in type A in Section 3c, recovering the usual blocks in permutation matrices. In Section 3d it is shown that $(W, \sqsubseteq)$ with the cutting order $\sqsubseteq$ is a poset (see Theorem 3.19). In Section 3e we show that blocks are closed under unions and intersections, and relate these to meets and joins in left and right order, thereby endowing the set of cutting points of a Coxeter group element with the structure of a distributive lattice (see Theorem 3.26). In Section 3f, we discuss various indexing sets for cutting points, which leads to the notion of w-analogues of descent sets in Section 3g. Properties of the *cutting poset* are studied in Section 3h (see Theorem 3.41, which also recapitulates the previous theorems).

Throughout this section $W := \langle s_i \mid i \in I \rangle$ denotes a finite Coxeter group.

3a. *Parabolic subgroups and cosets representatives.* For a subset $J \subseteq I$, the *parabolic subgroup* W_J of W is the Coxeter subgroup of W generated by s_j for $j \in J$. A complete system of minimal length representatives of the right cosets $W_J w$ and of the left cosets wW_J are given respectively by

$$
\begin{aligned}
{}^J W &:= \{x \in W \mid \mathrm{D}_L(x) \cap J = \varnothing\},\\
W^J &:= \{x \in W \mid \mathrm{D}_R(x) \cap J = \varnothing\}.
\end{aligned}
$$

Every $w \in W$ has a unique decomposition $w = w_J\, {}^J w$ with $w_J \in W_J$ and ${}^J w \in {}^J W$. Similarly, there is a unique decomposition $w = w^K\, {}_K w$ with ${}_K w \in {}_K W = W_K$ and $w^K \in W^K$.

Lemma 3.1. *Take $w \in W$.*

(i) *For $J \subseteq I$ consider the unique decomposition $w = uv$, where $u = w_J$ and $v = {}^J w$. Then, the unique decomposition of ws_k is $ws_k = (us_j)v$ if $vs_k v^{-1}$ is a simple reflection s_j with $j \in J$ and $ws_k = u(vs_k)$ otherwise.*

(ii) *For $K \subseteq I$ consider the unique decomposition $w = vu$, where $u = {}_K w$ and $v = w^K$. Then, the unique decomposition of $s_j w$ is $s_j w = v(s_k u)$ if $v^{-1} s_j v$ is a simple reflection s_k with $k \in K$ and $s_j w = (s_j v)u$ otherwise.*

Proof. This follows directly from [Björner and Brenti 2005, Lemma 2.4.3 and Proposition 2.4.4]. □

Note in particular that, if we are in case (i) of Lemma 3.1, we have the following:

- If k is a right descent of w, then $(ws_k)_J \in [1, w_J]_R$ and ${}^J(ws_k) \in [1, {}^Jws_k]_R$.
- If k is not a right descent of w, then either s_k skew commutes with ${}^Jw$ (that is, there exists an i such that $s_i{}^Jw = {}^Jws_k$), or ${}^J(ws_k) = {}^Jws_k$. In particular, ${}^J(ws_k) \leq_R {}^Jws_k$.

Definition 3.2. A subset $J \subseteq I$ is *left reduced* with respect to w if $J' \subset J$ implies ${}^Jw <_L {}^{J'}w$ (or equivalently, if for any $j \in J$, s_j appears in some and hence all reduced words for w_J).

We say $K \subseteq I$ is *right reduced* with respect to w if $K' \subset K$ implies $w^K <_R w^{K'}$.

Lemma 3.3. *Let $w \in W$ and $J \subseteq I$ be left reduced with respect to w. Then*

(i) *$v = {}^Jw \leq_R w$ if and only if there exists $K \subseteq I$ and a bijection $\phi_R : J \to K$ such that $s_jv = vs_{\phi_R(j)}$ for all $j \in J$.*

For $K \subseteq I$ right reduced with respect to w, we have

(i) *$v = w^K \leq_L w$ if and only if there exists $J \subseteq I$ and a bijection $\phi_L : K \to J$ such that $vs_k = s_{\phi_L(k)}v$ for all $k \in K$.*

Proof. Assume first that the bijection ϕ_R exists, and write $w = s_{j_1} \cdots s_{j_\ell}v$, where the product is reduced and $j_i \in J$. Then,

$$w = s_{j_1} \cdots s_{j_\ell}v = s_{j_1} \cdots s_{j_{\ell-1}}vs_{\phi_R(j_\ell)} = vs_{\phi_R(j_1)} \cdots s_{\phi_R(j_\ell)},$$

where the last product is reduced. Therefore $v \leq_R w$.

Assume conversely that $v = {}^Jw \leq_R w$, write the reduced expression $w = vs_{k_1} \cdots s_{k_\ell} \geq_R v$, and set $K = \{k_1, \dots, k_\ell\}$. By Lemma 3.1, the sequence

$$v = {}^Jv,\ {}^J(vs_{k_1}), \dots, {}^J(vs_{k_1} \cdots s_{k_\ell}) = {}^Jw = v$$

preserves right order, and therefore is constant. Hence, at each step i

$${}^J(vs_{k_1} \cdots s_{k_i}) = {}^J({}^J(vs_{k_1} \cdots s_{k_{i-1}})s_{k_i}) = {}^J(vs_{k_i}) = v.$$

Applying Lemma 3.1 again, it follows that there is a subset $J' \subseteq J$, and a bijective map $\phi_R : J' \to K$ such that $s_jv = vs_{\phi_R(j)}$ for all $j \in J'$. Then, $w = s_{\phi_R^{-1}(k_1)} \cdots s_{\phi_R^{-1}(k_\ell)}v$, and, since J is left reduced, $J = J'$.

The second part is the symmetric statement. □

By Lemma 3.1, for any $w \in W$ and $J \subseteq I$ we have $[1, w]_R \subseteq [1, w_J]_R[1, {}^Jw]_R$ and similarly for any $K \subseteq I$ we have $[1, w]_L \subseteq [1, w^K]_L[1, {}_Kw]_L$.

Lemma 3.4. *Take $w \in W$, $K \subseteq I$, and assume that $s_i w = w s_k$ for $i \in I$ and $k \in K$, where the products are reduced. Then, there exists $k' \in K$ such that $s_i w^K = w^K s_{k'}$, where the products are again reduced.*

Proof. We have $w^K = (w s_k)^K = (s_i w)^K = (s_i w^K)^K$. Hence, by Lemma 3.1(ii) there exists $k' \in K$ such that $w^K s_{k'} = s_i w^K$, as desired. $\square$

3b. ***Definition and characterizations of blocks and cutting points.*** We now come to the definition of *blocks* of Coxeter group elements and associated *cutting points*. They will lead to a new poset on the Coxeter group W, which we coin the *cutting poset* in Section 3d.

Definition 3.5 (blocks and cutting points). Let $w \in W$. We call $K \subseteq I$ a *right block* (or $J \subseteq I$ a *left block*) of w, if there exists $J \subseteq I$ (respectively $K \subseteq I$) such that

$$W_J w = w W_K.$$

In that case, $v := w^K$ is called a *cutting point* of w, which we denote by $v \sqsubseteq w$. Furthermore, K is *proper* if $K \neq \varnothing$ and $K \neq I$; it is *nontrivial* if $w^K \neq w$ (or equivalently ${}_K w \neq 1$); analogous definitions are made for left blocks.

We denote by $\mathcal{B}_{\mathcal{R}}(w)$ the set of all right blocks for w, and by $\mathcal{R}\mathcal{B}_{\mathcal{R}}(w)$ the set of all (right) reduced (see Definition 3.2) right blocks for w. The sets $\mathcal{B}_{\mathcal{L}}(w)$ and $\mathcal{R}\mathcal{B}_{\mathcal{L}}(w)$ are similarly defined on the left.

Here is an equivalent characterization of blocks, which also shows that cutting points can be equivalently defined using ${}^J w$ instead of w^K.

Proposition 3.6. *Let $w \in W$ and $J, K \subseteq I$. Then, the following are equivalent*:

(i) $W_J w = w W_K$.

(ii) *There exists a bijection $\phi : K \to J$ such that $w^K s_k = s_{\phi(k)} w^K$ (or equivalently $w^K(\alpha_k) = \alpha_{\phi(k)}$) for all $k \in K$.*

Furthermore, when any, and therefore all, of the above hold then,

(iii) $w^K = {}^J w$.

Proof. Suppose (i) holds. Then $W_J {}^J w = w^K W_K$. Since ${}^J w$ has no left descents in J and w^K has no right descents in K, we know that on both sides ${}^J w$ and w^K are the shortest elements and hence have to be equal: ${}^J w = w^K$; this proves (iii). Furthermore, every reduced expression $w^K s_k$ with $k \in K$ must correspond to some reduced expression $s_j {}^J w$ for some $j \in J$, and vice versa. Hence there exists a bijection $\phi : K \to J$ such that $w^K s_k = s_{\phi(k)} {}^J w = s_{\phi(k)} w^K$. Therefore point (ii) holds.

Suppose now that point (ii) holds. Then, for any expression $s_{k_1} \cdots s_{k_\ell} \in W_K$, we have

$$w^K s_{k_1} \cdots s_{k_\ell} = s_{\phi(k_1)} w^K s_{k_2} \cdots s_{k_\ell} = \cdots = s_{\phi(k_1)} \cdots s_{\phi(k_\ell)} w^K.$$

It follows that
$$w^K W_K = W_J w^K.$$
In particular $w \in W_J w^K$ and therefore
$$W_J w = W_J w^K = w^K W_K = w W_K. \qquad \square$$

In general, condition (iii) of Proposition 3.6 is only a necessary, but not sufficient condition for K to be a block. See Example 3.12.

Proposition 3.7. *If K is a right block of w (or more generally if $w^K = w^{K'}$ with K' a right block), then the bijection*
$$W^K \times {}_K W \to W, \quad (v, u) \mapsto vu$$
restricts to a bijection $[1, w^K]_L \times [1, {}_K w]_L \to [1, w]_L$.

Similarly, if J is a left block (or more generally if ${}^J w = {}^{J'} w$ with J' a left block), then the bijection
$$W_J \times {}^J W \to W, \quad (u, v) \mapsto uv$$
restricts to a bijection $[1, w_J]_R \times [1, {}^J w]_R \to [1, w]_R$ (see Figure 1).

Proof. By Proposition 3.6 we know that, if K is a right block, then there exists a bijection $\phi : K \to J$ such that $w^K s_k = s_{\phi(k)} w^K$. Hence the map $y \mapsto w^K y$ induces a skew-isomorphism between $[1, {}_K w]_L$ and $[w^K, w]_L$, where an edge k is mapped to edge $\phi(k)$. It follows in particular that $uv \leq_L w^K\ v \leq_L w^K\ {}_K w = w$ for any $u \in [1, w^K]_L$ and $v \in [1, {}_K w]_L$, as desired.

Assume now that K is not a block, but $w^K = w^{K'}$ with K' a block. Then, $[1, w^K]_L = [1, w^{K'}]_L$ and $[1, {}_K w]_L = [1, {}_{K'} w]_L$ and we are reduced to the previous case.

The second statement can be proved in the same fashion. $\square$

Due to Proposition 3.7, we also say that $[1, v]_R$ *tiles* $[1, w]_R$ if $v = {}^J w$ for some left block J (or equivalently $v = w^K$ for some right block K).

Proposition 3.8. *Let $w \in W$ and K be right reduced with respect to w. Then, the following are equivalent:*

(i) *K is a reduced right block of w.*

(ii) *$w^K \leq_L w$.*

The analogous statement can be made for left blocks.

See also Proposition 6.7 for yet another equivalent condition of reduced blocks.

Proof of Proposition 3.8. If K is a right block, then by Proposition 3.6 we have $w^K = {}^J w$, where J is the associated left block. In particular, $w^K = {}^J w \leq_L w$.

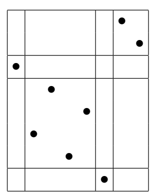
The converse statement follows from Lemma 3.3 and Proposition 3.6. $\square$

Example 3.9. For $w = w_0$, any $K \subseteq I$ is a reduced right block; of course $w_0^K \leq_L w_0$ and ${}_K w_0$ is the maximal element of the parabolic subgroup $W_K = {}_K W$. The cutting point $w^K \sqsubseteq w$ is the maximal element of the right descent class for the complement of K.

The associated left block is given by $J = \phi(K)$, where ϕ is the automorphism of the Dynkin diagram induced by conjugation by w_0 on the simple reflections. The tiling corresponds to the usual decomposition of W into right W_K cosets, or of W into left W_J cosets.

3c. ***Blocks of permutations.*** In this section we illustrate the notion of blocks and cutting points introduced in the previous section for type A. We show that, for a permutation $w \in \mathfrak{S}_n$, the blocks of Definition 3.5 correspond to the usual notion of blocks of the permutation matrix of w (or unions thereof), and the cutting points w^K for right blocks K correspond to putting the identity in those blocks.

A *matrix-block* of a permutation w is an interval $[k', k'+1, \dots, k]$ that is mapped to another interval. Pictorially, this corresponds to a square submatrix of the matrix of w that is again a permutation matrix (that of the *associated permutation*). For example, the interval $[2, 3, 4, 5]$ is mapped to the interval $[4, 5, 6, 7]$ by the permutation $w = 36475812 \in \mathfrak{S}_8$, and is therefore a matrix-block of w with associated permutation 3142. Similarly, $[7, 8]$ is a matrix-block with associated permutation 12:



For any permutation w, the singletons $[i]$ and the full set $[1, 2, \dots, n]$ are always matrix-blocks; the other matrix-blocks of w are called *proper*. A permutation with no proper matrix-block, such as 58317462, is called *simple*. See [Nozaki et al. 1995; Albert et al. 2003; Albert and Atkinson 2005] for a review of simple permutations. Simple permutations are also strongly related to dimension 2 posets.

A permutation $w \in \mathfrak{S}_n$ is *connected* if it does not stabilize any subinterval $[1, \dots, k]$ with $1 \leq k < n$, that is, if w is not in any proper parabolic subgroup $\mathfrak{S}_k \times \mathfrak{S}_{n-k}$. Pictorially, this means that there are no diagonal matrix-blocks. A matrix-block is *connected* if the corresponding induced permutation is connected. In the example above, the matrix-block $[2, 3, 4]$ is connected, but the matrix-block $[7, 8]$ is not.

Proposition 3.10. *Let $w \in \mathfrak{S}_n$. The right blocks of w are in bijection with disjoint unions of* (*nonsingleton*) *matrix-blocks for w*; *each matrix-block with column set*

$[i, i+1, \dots, k]$ contributes $\{i, i+1, \dots, k-1\}$ to the right block; each matrix-block with row set $[i, i+1, \dots, k]$ contributes $\{i, i+1, \dots, k-1\}$ to the left block.

In addition, trivial right blocks correspond to unions of identity matrix-blocks. Also, reduced right blocks correspond to unions of connected matrix-blocks.

Proof. Suppose $w \in \mathfrak{S}_n$ with a disjoint union of matrix-blocks with consecutive column sets $[i_1, \dots, k_1]$ up to $[i_\ell, \dots, k_\ell]$. Set $K_j = \{i_j, \dots, k_j - 1\}$ for $1 \le j \le \ell$ and $K = K_1 \cup \cdots \cup K_\ell$. Define similarly J according to the rows of the blocks.

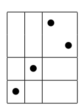
Then multiplying w on the right by some element of W_K permutes some columns of w, while stabilizing each block. Therefore, the same transformation can be achieved by some permutation of the rows stabilizing each block, that is, by multiplication of w on the left by some element of W_J. Hence, using symmetry, $W_J w = w W_K$, that is, J and K are corresponding left and right blocks for w.

Conversely, if K is a right block of w, then w^K maps each α_k with $k \in K$ to another simple root by Proposition 3.6. But then, splitting $K = K_1 \cup \cdots \cup K_\ell$ into consecutive subsets with $K_j = \{i_j, \dots, k_j - 1\}$, the permutation w^K must contain the identity permutation in each matrix-block with column indices $[i_j, \dots, k_j]$. This implies that w itself has matrix-blocks with column indices $[i_j, \dots, k_j]$ for $1 \le j \le \ell$.

Note that, in the described correspondence, $w^K = w$ if and only if all matrix-blocks contain the identity. This proves the statement about trivial right blocks.

A reduced right block K has the property that $w^{K'} \neq w^K$ for every $K' \subset K$. This implies that no matrix-block is in a proper parabolic subgroup, and hence they are all connected. $\square$

Example 3.11. As in Figure 1, consider the permutation 4312, whose permutation matrix is



The reduced (right)-blocks are $K = \{\}$, $\{1\}$, $\{2, 3\}$, and $\{1, 2, 3\}$. The cutting points are 4312, 3412, 4123, and 1234, respectively. The corresponding left blocks are $J = \{\}$, $\{3\}$, $\{1, 2\}$ and $\{1, 2, 3\}$, respectively. The nonreduced (right) blocks are $\{3\}$ and $\{1, 3\}$, as they are respectively equivalent to the blocks $\{\}$ and $\{1\}$. The trivial blocks are $\{\}$ and $\{3\}$.

Example 3.12. In general, condition (iii) of Proposition 3.6 is only a necessary, but not sufficient condition for K to be a block. For example, for $w = 43125$ (similar to 4312 of Example 3.11, but embedded in $\mathfrak{S}_5$), $J = \{3, 4\}$, and $K = \{1, 4\}$, one has ${}^J w = w^K$ yet neither J nor K are blocks. On the other hand (iii) of Proposition 3.6 becomes both necessary and sufficient for reduced blocks.

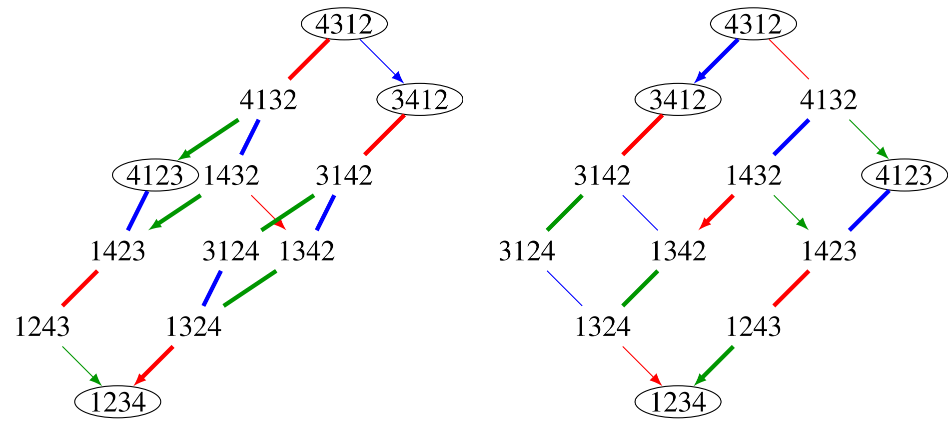


Figure 1. Two pictures of the interval $[1234, 4312]_R$ in right order in $\mathfrak{S}_4$ illustrating its proper tilings, for $J := \{3\}$ and $J := \{1, 2\}$, respectively. The thick edges highlight the tiling. The circled permutations are the cutting points, which are at the top of the tiling intervals. Blue, red, green lines correspond to s_1, s_2, s_3, respectively. See Section 6d for the definition of the orientation of the edges (this is $G^{(4312)}$); edges with no arrow tips point in both directions.

Remark 3.13. It is obvious that the union and intersection of overlapping (possibly with a trivial overlap) matrix-blocks in $\mathfrak{S}_n$ are again matrix-blocks; we will see in Proposition 3.22 that this property generalizes to all types.

Problem 3.14. Fix $J \subseteq \{1, 2, \ldots, n-1\}$ and enumerate the permutations $w \in \mathfrak{S}_n$ for which J is a left block.

3d. ***The cutting poset.*** In this section, we show that $(W, \sqsubseteq)$ indeed forms a poset. We start by showing that for a fixed $u \in W$, the set of elements w such that $u \sqsubseteq w$ admits a simple description. Recall that for $J \subseteq I$, we denote by s_J the longest element of W_J. Proposition 3.6 suggests the following definition.

Definition 3.15. Let $u \in W$. We call $k \in I$ a *short right nondescent* (or $j \in I$ a *short left nondescent*) of u if there exists $j \in I$ (respectively $k \in I$) such that

$$s_j u = u s_k,$$

where the product is reduced (that is, j and k are nondescents). An equivalent condition is that u maps the simple root α_k to a simple root (respectively the preimage of α_j is a simple root).

Set further

$$U_u := u W_K = [u, u s_K]_R = W_J u = [u, s_J u]_L,$$

where $K := K(u)$ and $J := J(u)$ are the sets of short right and left, respectively, nondescents of u.

Pictorially, one takes left and right order on W and associates to each vertex u the translate U_u above u of the parabolic subgroup generated by the short nondescents of u, which correspond to the simultaneous covers of u in both left and right order.

Example 3.16. In type A, i is short for $u \in \mathfrak{S}_n$ if $u(i+1) = u(i)+1$, that is, there is a 2×2 identity block in columns $(i, i+1)$ of the permutation matrix of u. Furthermore U_u is obtained by looking at all identity blocks in u and replacing each by any permutation matrix.

The permutation 4312 of Example 3.11 has a single nondescent 3 that is short, and $U_{4312} = \{4312, 4321\}$.

Proposition 3.17. *U_u is the set of all w such that $u \sqsubseteq w$.*

In particular, it follows that

- *if $u \leq_R v \leq_R w$ and $u \sqsubseteq w$, then $u \sqsubseteq v$; and*
- *if $u \sqsubseteq w$ and $u \sqsubseteq w'$, then $u \sqsubseteq w \vee_R w'$.*

Proof. Note that w is in U_u if and only if there exists K such that $K \subseteq K(u)$ and $w^K = u$. By Proposition 3.6, this is equivalent to the existence of a block K such that $w^K = u$, that is, $u \sqsubseteq w$. $\square$

The following related lemma is used to prove that $(W, \sqsubseteq)$ is a poset.

Lemma 3.18. *If $u \sqsubseteq w$, then the set of short nondescents of w is a subset of the short nondescents of u, namely $K(w) \subseteq K(u)$.*

Proof. Let $k \in K(w)$, so that $ws_k = s_j w$ for some $j \in I$ and both sides are reduced. It follows from Lemma 3.4 that there exists $k' \in K(w)$ such that $s_j u = us_{k'}$ and both sides are reduced. Hence $k' \in K(u)$. Since the map $k \mapsto k'$ is injective it follows that $K(w) \subseteq K(u)$. $\square$

Theorem 3.19. *$(W, \sqsubseteq)$ is a subposet of both left and right order.*

Proof. The relation $\sqsubseteq$ is reflexive since v is a cutting point of v with right block $\varnothing$; hence $v \sqsubseteq v$. Applying Proposition 3.6, it is a subrelation of left and right order: If $v \sqsubseteq w$ then $v = w^K \leq_R w$ for some K and $v = {}^J w \leq_L w$ for some J. Antisymmetry follows from the antisymmetry of left (or right) order.

For transitivity, let $v \sqsubseteq w$ and $w \sqsubseteq z$. Then $v = w^K$ and $w = z^{K'}$ for some right block K of w and K' of z. We claim that $v = z^{K \cup K'}$ with $K \cup K'$ a right block of z. Certainly $k \notin \mathrm{D}_R(v)$ for $k \in K$ since $v = w^K$. Since $w = z^{K'}$ with K' a block of z, all $k' \in K'$ are short nondescents of w and hence by Lemma 3.18 also short nondescents of v. This proves the claim. Therefore $v \sqsubseteq z$. $\square$

Example 3.20. The cutting poset for $\mathfrak{S}_3$ and $\mathfrak{S}_4$ is given in Figure 2. As we can see on those figures, the cutting poset is not the intersection of the right and left order since w_0 is maximal for left and right order but not for cutting poset.

3e. ***Lattice properties of intervals.*** In this section we show that the set of blocks and the set of cutting points $\{u \mid u \sqsubseteq w\}$ of a fixed $w \in W$ are endowed with the structure of distributive lattices (see Theorem 3.26).

We begin with a lemma that gives some properties of blocks that are contained in each other.

Lemma 3.21. *Fix $w \in W$. Let $K \subseteq K'$ be two right blocks of w and $J \subseteq J'$ be the corresponding left blocks, so that*

$$W_J w = w W_K, \quad W_{J'} w = w W_{K'}, \quad {}^J w = w^K \sqsubseteq w, \quad \text{and} \quad {}^{J'} w = w^{K'} \sqsubseteq w.$$

Then,

(i) $w^{K'} \leq_R w^K$ *and* $w^{K'} \leq_L w^K$,

(ii) K' *is a right block of* w^K *and* $w^{K'} \sqsubseteq w^K$,

(iii) K *is a right block of* ${}_{K'}w$ *and* ${}_{K'}w^K \sqsubseteq {}_{K'}w$.
Furthermore K is reduced for ${}_{K'}w$ if and only if it is reduced for w.

The same statements hold for left blocks.

Proof. Part (i) holds because $w^{K'} = (w^K)^{K'} \leq_R w^K \leq_R w$, and similarly on the left.

Part (ii) is a trivial consequence of (i) and Proposition 3.17.

For (iii), first note that $({}_{K'}w)^K = {}_{K'}(w^K)$, so that the notation ${}_{K'}w^K$ is unambiguous. Consider the bijection ϕ from K' to J' of Proposition 3.6, and note that $W_J w^{K'} = w^{K'} W_{\phi^{-1}(J)}$. Therefore,

$$w^{K'}\, {}_{K'}w W_K = w W_K = W_J\, w = W_J\, w^{K'}\, {}_{K'}w = w^{K'} W_{\phi^{-1}(J)}\, {}_{K'}w.$$

Simplifying by $w^{K'}$ on the left, one obtains that

$${}_{K'}w\, W_K = W_{\phi^{-1}(J)}\, {}_{K'}w,$$

proving that K is also a block of ${}_{K'}w$. The reduction statement is trivial. □

We saw in Remark 3.13 that the set of blocks is closed under unions and intersections in type A. This holds for general type.

Proposition 3.22. *The set $\mathcal{B}_{\mathcal{R}}(w)$ (or $\mathcal{B}_{\mathcal{L}}(w)$) of right (respectively left) blocks is stable under union and intersection. Hence, it forms a distributive sublattice of the Boolean lattice $\mathcal{P}(I)$.*

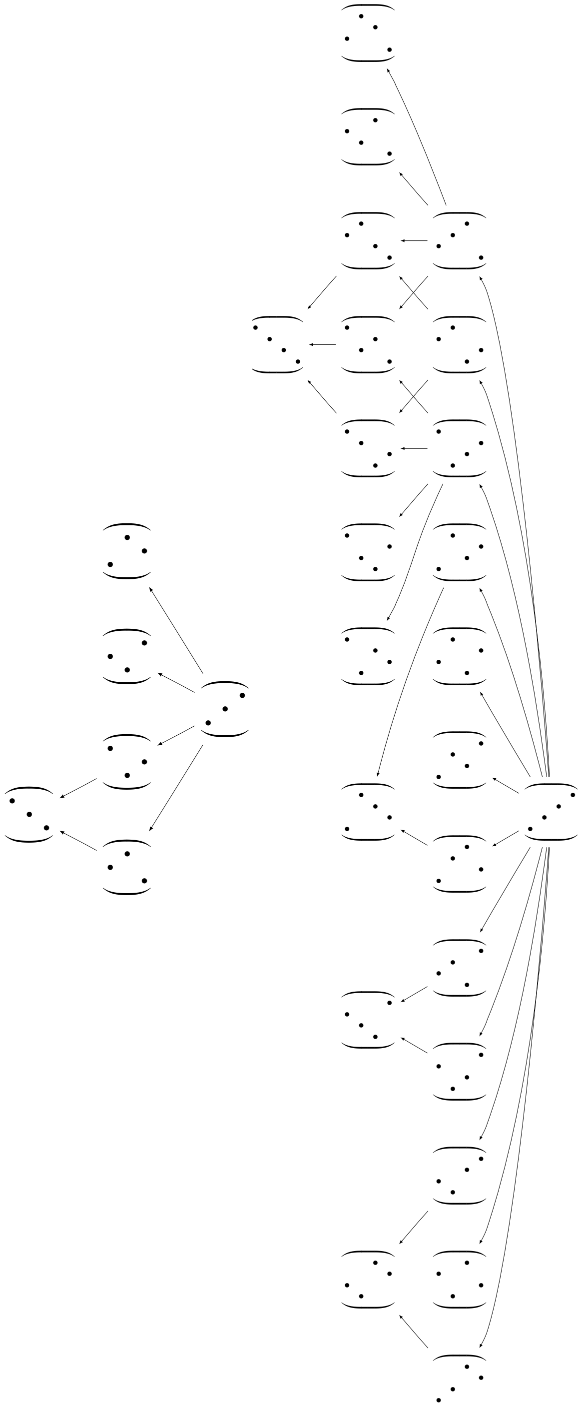


Figure 2. The cutting posets for the symmetric groups $\mathfrak{S}_3$ and $\mathfrak{S}_4$. Each permutation is represented by its permutation matrix, with the bullets marking the positions of the ones. Notice the Boolean sublattice appearing as the interval between the identity permutation at the bottom and the maximal permutation at the top; its elements are the minimal elements of the descent classes.

Proof. Let K and K' be right blocks for $w \in W$, and J and J' be the corresponding left blocks, so that

$$wW_K = W_J w \quad \text{and} \quad wW_{K'} = W_{J'} w.$$

Take $u \in W_{K \cap K'} = W_K \cap W_{K'}$. Then, wuw^{-1} is both in W_J and $W_{J'}$ and therefore in $W_J \cap W_{J'} = W_{J \cap J'}$. This implies $wW_{K\cap K'}w^{-1} \subseteq W_{J\cap J'}$. By symmetry, the inclusion $w^{-1}W_{J\cap J'}w \subseteq W_{K\cap K'}$ holds as well, and therefore $W_{J\cap J'}w = wW_{K\cap K'}$. In conclusion, $K \cap K'$ is a right block, with $J \cap J'$ as corresponding left block.

Now take $u \in W_{K\cup K'} = \langle W_K, W_{K'} \rangle$, and write u as a product $u_1u_1'u_2u_2' \cdots u_\ell u_\ell'$, where $u_i \in W_K$ and $u_i' \in W_{K'}$ for all $1 \le i \le \ell$. Then, for each i, $wu_iw^{-1} \in W_J$ and $wu_i'w^{-1} \in W_{J'}$. By composition, $wuw^{-1} \in W_J W_{J'} W_J W_{J'} \cdots W_J W_{J'} \subseteq W_{J\cup J'}$. Using symmetry as above, we conclude that $wW_{K\cup K'} = W_{J\cup J'}w$. In summary, $K \cup K'$ is a right block, with $J \cup J'$ as corresponding left block.

Finally, since blocks are stable under union and intersection, they form a sublattice of the Boolean lattice. Any sublattice of a distributive lattice is distributive. □

Next we relate the union and intersection operation on blocks with the meet and join operations in right and left order. We start with the following general statement which must be classical, though we have not found it in the literature.

Lemma 3.23. *Take $w \in W$ and $J, J', K, K' \subseteq I$. Then*

$$w^{K\cap K'} = w^K \vee_R w^{K'} \quad \textit{and} \quad {}^{J\cap J'}w = {}^J w \vee_L {}^{J'}w.$$

Proof. We include a proof for the sake of completeness. By Lemma 3.21(i), $w^K, w^{K'} \le_R w^{K\cap K'}$, and therefore $v \le_R w^{K\cap K'}$, where $v = w^K \vee_R w^{K'}$. Suppose that v has a right descent $k \in K \cap K'$. Then vs_k is still bigger than w^K and $w^{K'}$ in right order, a contradiction to the definition of v. Hence $w^{K\cap K'} = w^K \vee_R w^{K'}$, as desired. The statement on the left follows by symmetry. □

Corollary 3.24. *Take $w \in W$. Let $K, K' \subseteq I$ be two right blocks of w and $J, J' \subseteq I$ the corresponding left blocks. Then, for the right block $K \cap K'$ and left block $J \cap J'$,*

$$w^{K\cap K'} = {}^{J\cap J'}w = w^K \vee_R w^{K'} = {}^J w \vee_L {}^{J'}w.$$

The analogous statement of Lemma 3.23 for unions fails in general: Take for example $w = 4231$ and $K = \{3\}$ and $K' = \{1,2\}$, so that $w^K = 4213$ and $w^{K'} = 2341$; then $w^{K\cup K'} = 1234$, but $w^K \wedge_R w^{K'} = 2134$. However, it holds for blocks:

Lemma 3.25. *Take $w \in W$. Let $K, K' \subseteq I$ be two right blocks of w and $J, J' \subseteq I$ the corresponding left blocks. Then, for the right block $K \cup K'$ and left block $J \cup J'$,*

$$w^{K\cup K'} = {}^{J\cup J'}w = w^K \wedge_R w^{K'} = {}^J w \wedge_L {}^{J'}w.$$

Furthermore, $K \cup K'$ is reduced whenever K and K' are reduced, and similarly for the left blocks.

Proof. By symmetry, it is enough to prove the statements for right blocks.

By Lemma 3.21(i), $w^{K\cup K'} \leq_R w^K, w^{K'}$, and therefore $w^{K\cup K'} \leq_R w^K \wedge_R w^{K'}$.

Note that the interval $[w^{K\cup K'}, w]_R$ contains all the relevant points: w^K, $w^{K'}$, and $w^K \wedge_R w^{K'}$. Consider the translate of this interval obtained by dividing on the left by $w^{K\cup K'}$, or equivalently by using the map $u \mapsto {}_{K\cup K'}u$. By Lemma 3.21(iii), K and K' are still blocks of ${}_{K\cup K'}w$. From now on, we may therefore assume without loss of generality that $w^{K\cup K'} = 1$. It follows at once that $[1, w]_R$ lies in the parabolic subgroup $W_{K\cup K'}$ and that $J \cup J' = K \cup K'$.

If $w^K \wedge_R w^{K'} = 1 = w^{K\cup K'}$, then we are done. Otherwise, let $i \in K\cup K' = J\cup J'$ be the first letter of some reduced word for $w^K \wedge_R w^{K'}$. Since $w^K \wedge_R w^{K'}$ is in the interval $[1, w^K]_R$, i cannot be in J; by symmetry i cannot be in J' either, a contradiction.

Assume further that K and K' are reduced. Then, any $k \in K$ appears in any reduced word for ${}_K w$, and therefore in any reduced word for ${}_{K\cup K'}w$ since ${}_K w \leq_L {}_{K\cup K'}w$. By symmetry, the same holds for $k' \in K'$. Hence $K \cup K'$ is reduced. □

Theorem 3.26. *The map $K \mapsto w^K$ (or $J \mapsto {}^J w$) defines a lattice antimorphism from the lattice $\mathcal{B}_{\mathcal{R}}(w)$ (respectively $\mathcal{B}_{\mathcal{L}}(w)$) of right (respectively left) blocks of w to both right and left order on W.*

The set of cutting points for w, which is the image set

$$\{w^K \mid K \in \mathcal{B}_{\mathcal{R}}(w)\} = \{{}^J w \mid J \in \mathcal{B}_{\mathcal{L}}(w)\}$$

of the previous map, is a distributive sublattice of right (respectively left) order.

Proof. The first statement is the combination of Lemmas 3.23 and 3.25. The second statement follows from Proposition 3.22, since the quotient of a distributive sublattice by a lattice morphism is a distributive lattice. □

Corollary 3.27. *Every interval of $(W, \sqsubseteq)$ is a distributive sublattice and an induced subposet of both left and right order.*

Proof. Take an interval in $(W, \sqsubseteq)$; without loss of generality, we may assume that it is of the form $[1, w]_\sqsubseteq = \{w^K \mid K \in \mathcal{R}\mathcal{B}_{\mathcal{R}}(w)\}$. The interval $[1, w]_\sqsubseteq$ is not only a subposet of left (respectively right) order, but actually the induced subposet; indeed for K and K' right reduced blocks, and J and J' the corresponding left blocks,

$$w^K \leq_L w^{K'} \iff w^K \leq_R w^{K'} \iff J' \subseteq J \iff K' \subseteq K \iff w^K \leq_\sqsubseteq w^{K'}.$$

Therefore, using Theorem 3.26, it is a distributive sublattice of left (respectively right) order. □

Let us now consider the lower covers in the cutting poset for a fixed $w \in W$. They correspond to nontrivial blocks J that are minimal for inclusion, and in particular reduced.

Lemma 3.28. *Each minimal nontrivial (left) block J for $w \in W$ contains at least one element which is in no other minimal nontrivial block for w.*

Proof. Assume otherwise. Then, J is the union of its intersections with the other nontrivial blocks. Each such intersection is necessarily a trivial block, and a union of trivial blocks is a trivial block. Therefore, J is a trivial block, a contradiction. □

Corollary 3.29. *The semilattice of unions of minimal nontrivial blocks for a fixed $w \in W$ is free.*

Proof. This is a straightforward consequence of Lemma 3.28. Alternatively, this property is also a direct consequence of Corollary 3.27, since it holds in general for any distributive lattice. □

3f. ***Index sets for cutting points.*** Recall that by Theorem 3.26 the cutting points of w form a distributive lattice. Hence, by Birkhoff's representation theorem, they can be indexed by some collection of subsets closed under unions and intersections. We therefore now aim at finding a suitable choice of indexing scheme for the cutting points of w. More precisely, for each w, we are looking for a pair $(\mathcal{K}^{(w)}, \phi^{(w)})$, where $\mathcal{K}^{(w)}$ is a subset of some Boolean lattice (typically $\mathcal{P}(I)$) such that $\mathcal{K}^{(w)}$ ordered by inclusion is a lattice, and

$$\phi^{(w)} : \mathcal{K}^{(w)} \to [1, w]_{\sqsubseteq}$$

is an isomorphism (or antimorphism) of lattices.

Here are some of the desirable properties of this indexing:

(1) The indexing gives a Birkhoff representation of the lattice of cutting points of w. Namely, $\mathcal{K}^{(w)}$ is a sublattice of the chosen Boolean lattice, and unions and intersections of indices correspond to joins and meets of cutting points.

(2) The isomorphism $\phi^{(w)}$ is given by the map $J \mapsto {}^{J}w$. In that case the choice amounts to defining a section of those maps.

(3) The indexing generalizes the usual combinatorics of descents.

(4) The indices are blocks: $\mathcal{K}^{(w)} \subseteq \mathcal{B}_{\mathcal{L}}(w)$.

(5) We may actually want to have two indexing sets $\mathcal{K}^{(w)}$ and $\mathcal{K}^{(w)}$, one on the left and one on the right, with a natural isomorphism between them.

(6) The index of u in $\mathcal{K}^{(w)}$ does not depend on w (as long as u is a cutting point of w). One may further ask for this index to not depend on W, so that the indexing does not change through embedding of parabolic subgroups.

Unfortunately, there does not seem to be an ideal choice satisfying all of these properties at once, and we therefore propose several imperfect alternatives.

3f1. *Indexing by reduced blocks.* The first natural choice is to take reduced blocks as indices; then, $\mathcal{K}^{(w)} = \mathcal{RB}_{\mathcal{R}}(w)$ (and similarly $\mathcal{J}^{(w)} = \mathcal{RB}_{\mathcal{L}}(w)$ on the left). This indexing scheme satisfies most of the desired properties, except that it does not provide a Birkhoff representation, and depends on w.

Remark 3.30. By Lemma 3.25, if $K, K' \subseteq I$ are reduced right blocks for w, then $K \cup K'$ is also reduced. However, this is not necessarily the case for $K \cap K'$: consider for example the permutation $w = 4231$, $K = \{1, 2\}$ and $K' = \{2, 3\}$; then $K \cap K' = \{2\}$ is a block which is equivalent to the reduced block $\{\}$: $4231^{\{2\}} = 4231 = 4231^{\{\}}$.

The union $K \cup K'$ of two blocks may be reduced even when the blocks are not both reduced. Consider for example the permutation $w = 4312$ as in Figure 1. Then $K = \{1, 3\}$ and $K' = \{2, 3\}$ are blocks and their union $K \cup K' = \{1, 2, 3\}$ is reduced, yet K is not reduced.

Proposition 3.31. *The poset $(\mathcal{RB}_{\mathcal{R}}(w), \subseteq)$ of reduced right blocks is a distributive lattice, with the meet and join operation given respectively by*

$$K \vee K' = K \cup K' \quad \text{and} \quad K \wedge K' = \mathrm{red}(K \cap K'),$$

where, for a block K, $\mathrm{red}(K)$ is the unique largest reduced block contained in K.

The map $\phi^{(w)} : K \mapsto w^K$ restricts to a lattice antiisomorphism from the lattice $\mathcal{B}_{\mathcal{R}}(w)$ of reduced right blocks of w to $[1, w]_{\sqsubseteq}$.

The same statements hold on the left.

Proof. By Proposition 3.22 and Lemma 3.25, $\mathcal{RB}_{\mathcal{R}}(w)$ is a dual Moore family of the Boolean lattice of I, or even of $\mathcal{B}_{\mathcal{R}}(w)$. Therefore, using Section 2a, it is a lattice, with the given join and meet operations.

The lattice antiisomorphism of property follows from Lemma 3.25 and the coincidence of right order and $\sqsubseteq$ on $[1, w]_{\sqsubseteq}$ (Theorem 3.26). □

3f2. *Indexing by largest blocks.* The indexing by reduced blocks corresponds to the section of the lattice morphism $K \mapsto w^K$ by choosing the smallest block K in the fiber of a cutting point u. Instead, one could choose the largest block in the fiber of u, which is given by the set of short nondescents of u. This indexing scheme is independent of w. Also, by the same reasoning as above, the indexing sets $\mathcal{J}^{(w)}$ come endowed with a natural lattice structure. However, it does not give a Birkhoff representation: The meet is given by intersection, but the join is not given by union (take $w = 2143$; its cutting points are 1234, 1243, 2134, and 2143, indexed respectively by $\{1, 2, 3\}$, $\{1\}$, $\{3\}$, and $\{\}$).

3f3. *Birkhoff's representation using nonblocks.* We now relax the condition for the indices to be blocks. That is, we consider $K \mapsto w^K$ as a function from the full Boolean lattice $\mathcal{P}(I)$ to the minimal coset representatives of w. Beware that this map is no longer a lattice antimorphism; yet, the fiber of any u still admits a largest set $K = \overline{\mathrm{D}}_R(u) \subseteq I$, which is the complement of the right descent set of u. One can define a similar indexing on the left by $J = \overline{\mathrm{D}}_L(u)$. These indexings are independent of w and provide a Birkhoff representation for the lattice of cutting points (see Proposition 3.34). Define

$$\mathcal{DB}_{\mathcal{L}}(w) = \{\overline{\mathrm{D}}_L(u) \mid u \sqsubseteq w\} \quad \text{and} \quad \mathcal{DB}_{\mathcal{R}}(w) = \{\overline{\mathrm{D}}_R(u) \mid u \sqsubseteq w\}\,. \tag{3-1}$$

Remark 3.32. Since $\overline{\mathrm{D}}_L(u)$ and $\overline{\mathrm{D}}_R(u)$ are not necessarily blocks anymore, the bijection between $\overline{\mathrm{D}}_L(u)$ and $\overline{\mathrm{D}}_R(u)$ is no longer induced by a bijection at the level of descents: For example, for $u = 3142$, one has $\overline{\mathrm{D}}_L(u) = \{1, 3\}$ and $\overline{\mathrm{D}}_R(u) = \{2\}$.

Remark 3.33. Using $\mathrm{D}_R(u)$ instead of $\overline{\mathrm{D}}_R(u)$ would give an isomorphism instead of an antiisomorphism, and make the indexing further independent of W, at the price of slightly cluttering the notation w^K for cutting points.

Proposition 3.34 (Birkhoff representation for the lattice of cutting points). *The set $\mathcal{DB}_{\mathcal{R}}(w)$ of Equation (3-1) is a sublattice of the Boolean lattice, and the maps $K \mapsto w^K$ and $u \mapsto \overline{\mathrm{D}}_R(u)$ form a pair of reciprocal lattice antiisomorphisms with the lattice of cutting points of w. The same statement holds on the left.*

The proof of this proposition uses the following property of left and right order (recall that $[1, w]_{\sqsubseteq}$ is a sublattice thereof).

Lemma 3.35 [Le Conte de Poly-Barbut 1994, Lemme 5]. *The maps*

$$(W, \leq_L) \to \mathcal{P}(I), \quad w \mapsto \mathrm{D}_R(w), \quad \textit{and} \quad (W, \leq_R) \to \mathcal{P}(I), \quad w \mapsto \mathrm{D}_L(w)$$

are surjective lattice morphisms.

Proof of Proposition 3.34. By construction, $\overline{\mathrm{D}}_L$ is a section of $K \mapsto w^K$, and these maps form a pair of reciprocal bijections between $\mathcal{DB}_{\mathcal{L}}(w)$ and the cutting points of w. Using Lemma 3.35, the map $\overline{\mathrm{D}}_L$ is a lattice antimorphism. Therefore its image set $\mathcal{DB}_{\mathcal{R}}(w)$ is a sublattice of the Boolean lattice. The argument on the left is the same. □

3g. ***A w-analogue of descent sets.*** For each $w \in W$, we now provide a definition of a w-analogue on the interval $[1, w]_R$ of the usual combinatorics of (non)descents on W. From now on, we assume that we have chosen an indexation scheme so that the cutting points of w are given by $(w^K)_{K \in \mathcal{K}(w)}$ or equivalently by $({}^J w)_{J \in \mathcal{J}(w)}$.

Lemma 3.36. *Take a cutting point of w, and write it as $w^K = {}^J w$ for some $J, K \subseteq I$, which are not necessarily blocks. Then*

(i) *for $u \in [1, w]_R$, $u \in [1, {}^J w]_R$ if and only if $\mathrm{D}_L(u) \cap J = \varnothing$;*

(ii) *for $u \in [1, w]_L$, $u \in [1, w^K]_L$ if and only if $\mathrm{D}_R(u) \cap K = \varnothing$.*

Proof. This is a straightforward corollary of Proposition 3.7: Any element u of $[1, w]_R$ can be written uniquely as a product $u'v$ with $u' \in W_J$ and $v \in [1, {}^J w]_R$. So u is in $[1, {}^J w]_R$ if and only if $u' = 1$, which in turn is equivalent to v having no descents in J. This proves (i). The argument for (ii) is analogous. □

Example 3.37. For $w = w_0$, ${}^J w$ is the maximal element of a left descent class, and $[1, {}^J w]_R$ gives all elements of W whose left descent set is a subset of the left descent set of w.

Definition 3.38 (w-nondescent sets). For $u \in [1, w]_R$, define $J^{(w)}(u)$ to be the index $J \in \mathcal{J}^{(w)}$ of the lowest cutting point ${}^J w$ such that $u \in [1, {}^J w]_R$ (or the equivalent condition of Lemma 3.36). Define similarly $K^{(w)}(u)$ as the index in $\mathcal{K}^{(w)}$ of this cutting point.

Example 3.39. When $w = w_0$, $J^{(w_0)}(u)$ and $K^{(w_0)}(u)$ are respectively the sets $\overline{\mathrm{D}}_L(u)$ and $\overline{\mathrm{D}}_R(u)$ of left and right nondescents of u.

Problem 3.40. Given J, describe all the elements $w \in W$ such that J is a left block. This essentially only depends on ${}^J w$.

3h. ***Properties of the cutting poset.*** In this section we study the properties of the cutting poset $(W, \sqsubseteq)$ of Theorem 3.19 for the cutting relation $\sqsubseteq$ introduced in Definition 3.5 (see also Figure 2). The following theorem summarizes the results.

Theorem 3.41. $(W, \sqsubseteq)$ *is a meet-distributive meet-semilattice with* 1 *as minimal element, and a subposet of both left and right order.*

Every interval of $(W, \sqsubseteq)$ *is a distributive sublattice and a sublattice of both left and right order.*

Let $w \in W$ *and denote by* $\mathrm{Pred}(w)$ *the set of its* $\sqsubseteq$*-lower covers. Thanks to meet-distributivity, the meet-semilattice* L_w *generated by* $\mathrm{Pred}(w)$ *using* $\wedge_\sqsubseteq$ *(or equivalently* $\wedge_L$, $\wedge_R$ *if viewed as a sublattice of left or right order) is free, that is, isomorphic to a Boolean lattice.*

In particular, the Möbius function of $(W, \sqsubseteq)$ *is given by* $\mu(u, w) = (-1)^{r(u,w)}$ *if* $u \in L_w$ *and* 0 *otherwise, where* $r(u, w) := |\{v \in \mathrm{Pred}(w) \mid u \sqsubseteq v\}|$.

This Möbius function is used in Section 6d to compute the size of the simple modules of $\mathbb{K}M$.

Since $(W, \sqsubseteq)$ is almost a distributive lattice, Birkhoff's representation theorem suggests that we embed it in the distributive lattice $O(I((W, \sqsubseteq)))$ of the lower sets of its join-irreducible elements (note that a block is join-irreducible if there is only one minimal nontrivial block below it).

Problem 3.42. Describe the set $I(W, \sqsubseteq)$ of join-irreducible elements of $(W, \sqsubseteq)$.

Problem 3.43. Determine the distributive lattice associated with the cutting poset from the join-irreducibles, via Birkhoff's theory.

The join-irreducible elements of $(\mathfrak{S}_n, \sqsubseteq)$, for n small, are counted by the sequence 0, 1, 4, 16, 78, 462, 3224. Figure 2 seems to suggest that they form a tree, but this already fails for $n = 5$. We now briefly comment on the simplest join-irreducible elements, namely the immediate successors w of 1 in the cutting poset. Equivalent statements are that w admits exactly two reduced blocks $\{\}$ and B, possibly with $B = I$, or that the simple module S_w is of dimension $|[1, w]_R| - 1$. For a Coxeter group W, we denote by $S(W)$ the set of elements $w \neq 1$ having no proper reduced blocks, and $T(W)$ those having exactly two reduced blocks. Note that $T(W)$ is the disjoint union of the $S(W_J)$ for $J \subseteq I$.

Example 3.44. In type A, a permutation $w \in S(\mathfrak{S}_n)$ is uniquely obtained by taking a simple permutation, and inflating each 1 of its permutation matrix by an identity matrix. An element of $T(\mathfrak{S}_n)$ has a block diagonal matrix with one block in $S(\mathfrak{S}_m)$ for $m \leq n$, and $n-m$ 1×1 blocks. This gives an easy way to construct the generating series for $S(\mathfrak{S}_n)_{n\in\mathbb{N}}$ and for $T(\mathfrak{S}_n)_{n\in\mathbb{N}}$ from that of the simple permutations given in [Albert and Atkinson 2005].

We now turn to the proof of Theorem 3.41.

Lemma 3.45. $(W, \sqsubseteq)$ *is a partial join-semilattice. That is, when the join exists, it is unique and given by the join in left and in right order*:

$$v \vee_{\sqsubseteq} v' = v \vee_L v' = v \vee_R v'.$$

Proof. Take v and v' with at least one common successor. Applying Corollary 3.27 to the interval $[1, w]_{\sqsubseteq}$ for any such common successor w, one obtains $v, v' \sqsubseteq v \vee_R v' = v \vee_L v' \sqsubseteq w$. Therefore, $v \vee_R v' = v \vee_L v'$ is the join of v and v' in the cutting order. □

Lemma 3.46. $(W, \sqsubseteq)$ *is a meet-semilattice. That is, for $v, v' \in W$*

$$v \wedge_{\sqsubseteq} v' = \bigvee_{u \sqsubseteq v, v'} u,$$

where $\bigvee$ is the join for the cutting order (or equivalently for left or right order). If further v and v' have a common successor, then

$$v \wedge_{\sqsubseteq} v' = v \wedge_R v' = v \wedge_L v'.$$

Proof. The first part follows from a general result. Namely, for any poset, the following statements are equivalent (see for example [Pouzet 2013, Proposition 7.3]):

(i) Any bounded nonempty part has an upper bound.

(ii) Any bounded nonempty part has a lower bound.

Here we prove again this fact for the sake of self-containment. Take u and u' two common cutting points for v and v'. Then, using Lemma 3.45, their join exists and $u \vee_{\sqsubseteq} u' = u \vee_R u' = u \vee_L u'$ is also a cutting point for v and v'. The first statement follows by repeated iteration over all common cutting points.

Now assume that v and v' have a common successor w. Then by applying Corollary 3.27 to the interval $[1, w]_{\sqsubseteq}$, we find that $v \wedge_R v' = v \wedge_L v'$ is the meet of v and v' in the cutting order. □

Proof of Theorem 3.41. $(W, \sqsubseteq)$ is a meet-semilattice by Lemma 3.46. Meet-distributivity follows from Corollary 3.29. The argument is in fact general: Any poset with a minimal element 1 such that all intervals $[1, x]$ are distributive lattices and such that any two elements admit either a join or no common successor is a meet-distributive meet-semilattice (see [Edelman 1986] for literature on such). The end of the first statement is Theorem 3.19.

The statement about intervals is Corollary 3.27.

The $\sqsubseteq$-lower covers of an element w correspond to the nontrivial blocks of w that are minimal for inclusion. The top part L_w of an interval $[1, w]_{\sqsubseteq}$ is further described in Corollary 3.29, through the bijection $\phi^{(w)}$ between blocks of w and the interval $[1, w]_{\sqsubseteq}$ of Proposition 3.31. The value of $\mu(u, w)$ depends only on this interval. The remaining statements follow using Rota's crosscut theorem [1964] on Möbius functions for lattices; see also [Blass and Sagan 1997, Theorem 1.3]. □

4. Combinatorics of $M(W)$

In this section we study the combinatorics of the biHecke monoid $M(W)$ of a finite Coxeter group W. In particular, we prove in Sections 4a and 4b that its elements preserve left order and Bruhat order, and derive in Section 4c properties of their image sets and fibers. In Sections 4d and 4e, we prove the key combinatorial ingredients for the enumeration of the simple modules of $\mathbb{K}M(W)$ in Section 7: $M(W)$ is aperiodic and its $\mathcal{J}$-classes of idempotents are indexed by W. Finally, in Section 4f we study Green's relations as introduced in Section 2e and involutions on $M(W)$ in Section 4g.

4a. ***Preservation of left order.*** Recall that $M(W)$ is defined by its right action on elements in W by (1-5) and (1-6). The following key proposition, illustrated in Figure 3, states that it therefore preserves properties on the left.

Proposition 4.1. *Take $f \in M(W)$, $w \in W$, and $j \in I$. Then, $(s_j w).f$ is either $w.f$ or $s_j(w.f)$.*

The proof of Proposition 4.1 is a consequence of the associativity of the 0-Hecke monoid and relies on the following lemma, which is a nice algebraic (partial) formulation of the exchange property [Björner and Brenti 2005, Section 1.5].

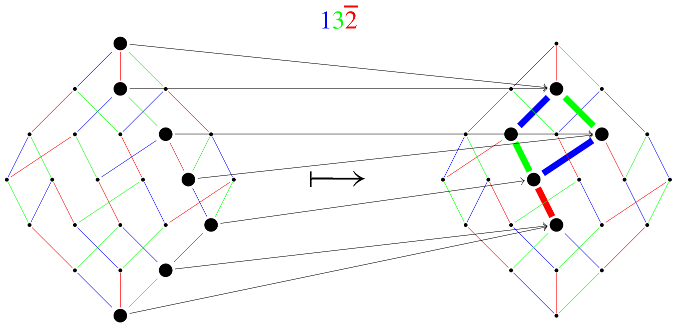


Figure 3. A partial picture of the graph of the element $f := \pi_1\pi_3\overline{\pi}_2$ of the monoid $M(\mathfrak{S}_4)$. On both sides, the underlying poset is left order of $\mathfrak{S}_4$ (with 1 at the bottom, and the same color code as in Figure 1); on the right, the bold dots depict the image set of f. The arrows from the left to the right describe the image of each point along some chain from 1 to w_0.

Lemma 4.2. *Let $w \in W$ and $i, j \in I$ such that $j \notin \mathrm{D}_L(w)$. Then*

$$(s_j w).\pi_i = \begin{cases} w.\pi_i & \text{if } j \in \mathrm{D}_L(w.\pi_i), \\ s_j(w.\pi_i) & \text{otherwise.} \end{cases}$$

The same result holds with π_i replaced by $\overline{\pi}_i$.

Proof. Recall that $w.\pi_v = 1.(\pi_w\pi_v)$ for any $w, v \in W$. Set $w' = w.\pi_i$. Then

$$\begin{aligned}(s_j w).\pi_i &= 1.(\pi_{s_j w}\pi_i) = 1.((\pi_j\pi_w)\pi_i) = 1.(\pi_j(\pi_w\pi_i)) = 1.(\pi_j\pi_{w'}) \\ &= \begin{cases} 1.\pi_{w'} = w' & \text{if } j \in \mathrm{D}_L(w'), \\ 1.\pi_{s_j w'} = s_j w' & \text{otherwise.} \end{cases}\end{aligned}$$

The result for $\overline{\pi}_i$ follows from Remark 1.1 and the fact that $w_0 s_j = s_{j'} w_0$ for some $j' \in I$ by Example 3.9 and Lemma 3.3 with $w = w_0$ and $K = \{j\}$. ☐

Proof of Proposition 4.1. Any element $f \in M(W)$ can be written as a product of π_i and $\overline{\pi}_i$. Lemma 4.2 describes the action of π_i and $\overline{\pi}_i$ on the Hasse diagram of left order. By induction, each π_i and $\overline{\pi}_i$ in the expansion of f satisfies all desired properties, and hence so does f (the statement holds trivially for the identity). ☐

Proposition 4.3. *For $f \in M(W)$, the following holds.*

(i) *f preserves left order*:

$$w \leq_L w' \quad \textit{implies} \quad w.f \leq_L w'.f \quad \textit{for } w, w' \in W.$$

(ii) *Take $w \leq_L w'$ in W, and consider a maximal chain*

$$w.f = v_1 \xrightarrow{i_1} v_2 \xrightarrow{i_2} \cdots \xrightarrow{i_{k-1}} v_k = w'.f.$$

Then, there is a maximal chain

$$w = u_{1,1} \to \cdots \to u_{1,\ell_1} \xrightarrow{i_1} u_{2,1} \to \cdots \to u_{2,\ell_2} \xrightarrow{i_2} \cdots$$
$$\cdots \xrightarrow{i_{k-1}} u_{k,1} \to \cdots \to u_{k,\ell_k} = w', \quad (4\text{-}1)$$

such that $u_{j,l} \,.\, f = v_j$ *for all* $1 \le j \le k$ *and* $1 \le l \le \ell_j$.

(iii) f *is length contracting; that is, for* $w \le_L w'$

$$\ell(w' \,.\, f) - \ell(w \,.\, f) \le \ell(w') - \ell(w).$$

Furthermore, when equality holds, $(w' \,.\, f)(w \,.\, f)^{-1} = w'w^{-1}$.

(iv) *Let* $J = [a, b]_L$ *be an interval in left order. Then the image of* J *under* f *denoted by* $J \,.\, f$ *has* $a \,.\, f$ *and* $b \,.\, f$ *as minimal and maximal element, respectively. Furthermore,* $J \,.\, f$ *is connected. If* $\ell(b \,.\, f) - \ell(a \,.\, f) = \ell(b) - \ell(a)$, *then* $J \,.\, f$ *is isomorphic to* J, *that is,* $x \,.\, f = (xa^{-1})(a \,.\, f)$ *for* $x \in J$.

Proof. Parts (i) and (ii) are direct consequences of Proposition 4.1, using induction.

Part (iii) follows from (ii).

Part (iv) follows from (i), (ii), and (iii) applied to $a \le_L x$ for all $x \in [a, b]_L$. □

4b. *Preservation of Bruhat order.* Recall the following well-known property of Bruhat order of Coxeter groups.

Proposition 4.4 (lifting property [Björner and Brenti 2005, p. 35]). *Suppose* $u <_B v$ *and* $i \in \mathrm{D}_R(v)$ *but* $i \notin \mathrm{D}_R(u)$. *Then,* $u \le_B vs_i$ *and* $us_i \le_B v$.

The next proposition is a consequence of the lifting property.

Proposition 4.5. *The elements* f *of* $M(W)$ *preserve Bruhat order. That is, for* $u, v \in W$

$$u \le_B v \quad \textit{implies} \quad u \,.\, f \le_B v \,.\, f.$$

Proof. It suffices to show the property for π_i and $\overline{\pi}_i$ since they generate $M(W)$. For these, the claim of the proposition is trivial if i is a right descent of u, or i is not a right descent of v. Otherwise, we can apply the lifting property:

$$u \,.\, \pi_i = us_i \le_B v = v \,.\, \pi_i,$$
$$u \,.\, \overline{\pi}_i = u \le_B vs_i = v \,.\, \overline{\pi}_i.$$

□

Remark 4.6. By Lemma 2.3, the preimage of a point is a convex set, but need not be an interval. For example, the preimage of $s_1s_3 \in \mathfrak{S}_4$ (or 2143 in one-line notation) of $f = \overline{\pi}_1\pi_2\pi_1\pi_3\overline{\pi}_2\overline{\pi}_3\overline{\pi}_1\overline{\pi}_2$ is

$$\{2413, 2341, 4213, 3412, 3241, 2431, 4312, 4231, 3421, 4321\},$$

which in Bruhat order has two maximal elements 2413 and 2341 and hence is not an interval.

Corollary 4.7 (of Proposition 4.3). *Let $f \in M(W)$.*

(i) *If $1\,.\,f = 1$, then f is regressive for Bruhat order: $w\,.\,f \leq_B w$ for all $w \in W$.*

(ii) *If $w_0\,.\,f = w_0$, then f is extensive for Bruhat order: $w\,.\,f \geq_B w$ for all $w \in W$.*

Proof. First suppose that $1\,.\,f = 1$. Let $w\,.\,f = s_{i_k} \cdots s_{i_1}$ be a reduced decomposition of $w\,.\,f$. This defines a maximal chain

$$1\,.\,f = 1 = v_0 \xrightarrow{i_1} \cdots \xrightarrow{i_{k-2}} v_{k-2} \xrightarrow{i_{k-1}} v_{k-1} \xrightarrow{i_k} v_k = w\,.\,f$$

in left order. By Proposition 4.3(ii) there is a larger chain from 1 to w so that there is a reduced word for w which contains $s_{i_k} \cdots s_{i_1}$ as a subword. Hence by the subword property of Bruhat order $w\,.\,f \leq_B w$. This proves (i).

Now let $w_0\,.\,f = w_0$. By arguments similar to the above, constructing a maximal chain from $w\,.\,f$ to $w_0\,.\,f$ in left order, one finds that $w_0(w\,.\,f)^{-1} \leq_B w_0 w^{-1}$. By [Björner and Brenti 2005, Proposition 2.3.4], the map $v \mapsto w_0 v$ is a Bruhat antiautomorphism and by the subword property $v \mapsto v^{-1}$ is a Bruhat automorphism. This implies $w \leq_B w\,.\,f$ as desired for (ii). □

4c. *Fibers and image sets.* Viewing elements of the biHecke monoid $M(W)$ as functions on W, we now study properties of their fibers and image sets.

Proposition 4.8. (i) *The image set* $\mathrm{im}(f)$ *for any $f \in M(W)$ is connected (see Definition 2.1) with a unique minimal and maximal element in left order.*

(ii) *The image set of an idempotent in $M(W)$ is an interval in left order.*

Proof. Part (i) follows immediately from Proposition 4.3(iv) with $J = [1, w_0]_L$.

For part (ii), we let $e \in M(W)$ be an idempotent with image set $\mathrm{im}(e)$. By Proposition 4.3(iv) with $J = [1, w_0]_L$, we have that $1\,.\,e$ and $w_0\,.\,e$ are the minimal and maximal, respectively, elements of $\mathrm{im}(e)$. Then by Proposition 4.3(ii), for every maximal chain in left order between $1\,.\,e$ and $w_0\,.\,e$, there is a maximal chain in left order of preimage points. Since e is an idempotent, there must be such a chain that contains the original chain. Hence all chains in left order between $1\,.\,e$ and $w_0\,.\,e$ are in $\mathrm{im}(e)$, proving that $\mathrm{im}(e)$ is an interval. □

Note that the proof above, in particular Proposition 4.3(ii), heavily uses the fact that the edges in left order are colored.

Definition 4.9. For any $f \in M(W)$, we call the set of fibers of f, denoted by $\mathrm{fibers}(f)$, the (unordered) set-partition of W associated by the equivalence relation $w \equiv w'$ if $w\,.\,f = w'\,.\,f$.

Proposition 4.10. *Take $f \in M(W)$, and consider the Hasse diagram of left order contracted with respect to the fibers of f. Then, this graph is isomorphic to left order restricted on the image set.*

Proof. See Appendix A on colored graphs. □

Proposition 4.11. *Any $f \in M(W)$ is characterized by its set of fibers and $1 . f$.*

Proof. Fix a choice of fibers. Contract the left order with respect to the fibers. By Proposition 4.10 this graph has to be isomorphic to the left order on the image set.

Once the lowest element in the image set $1 . f$ is fixed, this isomorphism is forced, since by Proposition 4.8(i) the graphs are (weakly) connected, have a unique minimal element, and there is at most one arrow of a given color leaving each node. □

Proposition 4.11 makes it possible to visualize nontrivial elements of the monoid (see Figure 4).

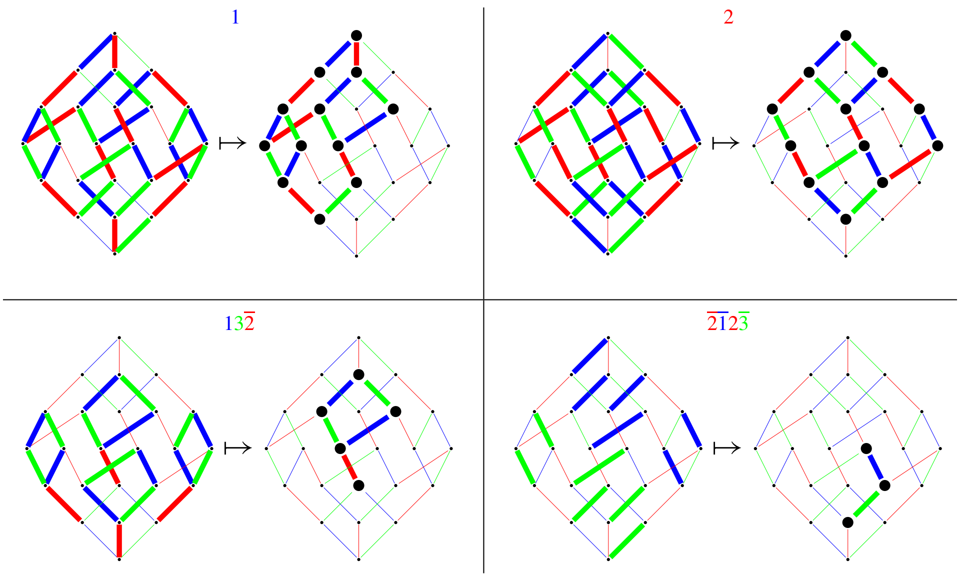


Figure 4. The elements $f = \pi_1$, π_2, $\pi_1\pi_3\overline{\pi}_2$ and $\overline{\pi}_2\overline{\pi}_1\pi_2\overline{\pi}_3$ of $M(\mathfrak{S}_4)$. As in Figure 3, the underlying poset on both sides is left order on $\mathfrak{S}_4$, and the bold dots on the right sides depict the image set of f. On the left side, an edge between two elements of W is thick if they are not in the same fiber. This information completely describes f; indeed $u = 1$ on the left is mapped to the lowest element of the image set on the right; each time one moves u up along a thick edge on the left, its image $u . f$ is moved up along the edge of the same color on the right.

Recall that a set-partition $\Lambda = \{\Lambda_i\}$ is said to be finer than the set-partition $\Lambda' = \{\Lambda'_i\}$ if for all i there exists a j such that $\Lambda_i \subseteq \Lambda'_j$. This is denoted by $\Lambda \preceq \Lambda'$. The refinement relation is a partial order.

For $f \in M(W)$, define the *type* of f by

$$\operatorname{type}(f) := \operatorname{type}([1 . f, w_0 . f]_L) = (w_0 . f)(1 . f)^{-1}. \tag{4-2}$$

The *rank* of $f \in M(W)$ is the cardinality of the image set $\operatorname{im}(f)$.

Lemma 4.12. *Fix $f \in M(W)$. For $h = fg \in fM(W)$,*

(1) $\operatorname{fibers}(f) \preceq \operatorname{fibers}(h)$,

(2) $\operatorname{type}(h) \leq_B \operatorname{type}(f)$,

(3) $\operatorname{rank}(h) \leq \operatorname{rank}(f)$.

Furthermore, the following are equivalent:

(i) $\operatorname{fibers}(h) = \operatorname{fibers}(f)$,

(ii) $\operatorname{rank}(h) = \operatorname{rank}(f)$,

(iii) $\operatorname{type}(h) = \operatorname{type}(f)$,

(iv) $\ell(w_0 . h) - \ell(1 . h) = \ell(w_0 . f) - \ell(1 . f)$.

If any, and therefore all, of the above hold, then h is completely determined (within $fM(W)$) by $1 . h$.

Proof. For $f, g \in M(W)$, the statement $\operatorname{fibers}(f) \preceq \operatorname{fibers}(fg)$ is obvious.

By Proposition 4.3(iii) and (iv), we know for $f, g \in M(W)$ that either $\operatorname{type}(fg) = \operatorname{type}(f)$ or $\ell(w_0.(fg)) - \ell(1.(fg)) < \ell(w_0 . f) - \ell(1 . f)$. In the latter case by Proposition 4.5, $\operatorname{type}(fg) <_B \operatorname{type}(f)$. The second case occurs precisely when $\operatorname{fibers}(f)$ is strictly finer than $\operatorname{fibers}(fg)$, or equivalently $\operatorname{rank}(fg) < \operatorname{rank}(f)$.

The last statement, that if $\operatorname{fibers}(h) = \operatorname{fibers}(f)$ then h is determined by $1 . h$, follows from Proposition 4.11. □

4d. *Aperiodicity*. Recall from Section 2e that a monoid M is called *aperiodic* if for any $f \in M$, there exists $k > 0$ such that $f^{k+1} = f^k$. Note that, in this case, $f^\omega := f^k = f^{k+1} = \cdots$ is an idempotent.

Proposition 4.13. *The biHecke monoid $M(W)$ is aperiodic.*

Proof. From Proposition 4.3(iv), we know that $\operatorname{im}(f^k)$ has a minimal element $a_k = 1 . f^k$ and a maximal element $b_k = w_0 . f^k$ in left order. Since $\operatorname{im}(f^{k+1}) \subseteq \operatorname{im}(f^k)$, we have $a_{k+1} \geq_L a_k$ and $b_{k+1} \leq_L b_k$. Therefore, both sequences a_k and b_k must ultimately be constant.

This implies that, for N big enough, a_N and b_N are fixed points. Applying Proposition 4.3(iii) yields that all elements in $[a_N, b_N]_L$ are fixed points under f.

It follows successively that $\mathrm{im}(f^N) = [a_N, b_N]_L$, $f^N = f^{N+1} = \cdots$, and $\mathrm{fix}(f) = [a_N, b_N]_L$. □

Corollary 4.14. *The set of fixed points of an element $f \in M(W)$ is an interval in left order.*

Proof. The set of fixed point of f is the image set of f^ω, which is an interval in left order by Proposition 4.8(ii). □

4e. *Idempotents.* We now study the properties of idempotents in $M(W)$.

Proposition 4.15. (i) *For $w \in W$*

$$e_w := \pi_{w^{-1}w_0} \overline{\pi}_{w_0 w}$$

is the unique idempotent such that $1 . e_w = 1$ and $w_0 . e_w = w$. Its image set is $[1, w]_L$, and it satisfies

$$u . e_w = \max_{\leq_B}\bigl([1, u]_B \cap [1, w]_L\bigr).$$

(ii) *Similarly, for $w \in W$,*

$$\tilde{e}_w := \overline{\pi}_{w^{-1}} \pi_w$$

is the unique idempotent with image set $[w, w_0]_L$, and it satisfies a dual formula.

(iii) *Furthermore,*

$$e_{a,b} := \overline{\pi}_{a^{-1}} e_{ba^{-1}} \pi_a$$

is an idempotent with image set $[a, b]_L$.

Proof. (i) Clearly, the image of e_w is a subset of $[1, w]_L$. Applying Remark 2.7 shows that $[1, w]_L$ is successively mapped bijectively to $[w^{-1}w_0, w_0]_L$ and back to $[1, w]_L$. So e_w is an idempotent with image set $[1, w]_L$. Reciprocally, let f be an idempotent such that $1 . f = 1$ and $w_0 . f = w$. Then, by Proposition 4.5, f preserves Bruhat order and by Corollary 4.7(i), $u . f \leq_B u$ for all $u \in W$. Furthermore, by Proposition 4.8, the image set of f is the interval $[1, w]_L$. Using Proposition 2.4, uniqueness and the given formula follow.

Statement (ii) is dual to (i) and is proved similarly.

(iii) The image set of $e_{ba^{-1}}$ is $[1, ba^{-1}]_L$; hence the image set of $e_{a,b}$ is a subset of $[a, b]_L$. We conclude by checking that $[a, b]_L$ is mapped bijectively at each step $\overline{\pi}_{a^{-1}}$, $e_{ba^{-1}}$ and π_a (see also Remark 2.7), and therefore consists of fixed points. □

Remark 4.16. For $f \in M(W)$, $f e_v = f e_{u.e_v}$, where $u = w_0 . f$.

Proof. Use the formula of Proposition 4.15(i). □

Corollary 4.17. *For $u, w \in W$, the intersection $[1, u]_B \cap [1, w]_L$ is a $\leq_L$-lower set with a unique maximal element v in Bruhat order. The maximum is given by $v = u\,.\,e_w$.*

4f. *Green's relations.* We have now gathered enough information about the combinatorics of $M(W)$ to give a partial description of its Green's relations, which will be used in the study of the representation theory of $M(W)$.

As an example, Figure 5 completely describes Green's relations $\mathscr{L}$, $\mathscr{R}$, and $\mathscr{J}$ for $M(\mathfrak{S}_3)$. The vertices are the 23 elements of $M(\mathfrak{S}_3)$, each drawn as in Figure 4. The edges give both the left and right Cayley graph of $M(\mathfrak{S}_3)$; for example, there are arrows

$$f \xrightarrow{\times\pi_1} g \quad \text{if } g = f\pi_1 \quad \text{and} \quad f \xrightarrow{\pi_1\times\pi_1} g \quad \text{if } g = f\pi_1 = \pi_1 f.$$

The picture also highlights the $\mathscr{J}$-classes of $M(\mathfrak{S}_3)$, and the corresponding eggbox pictures (that is, the decomposition of the $\mathscr{J}$-classes into $\mathscr{L}$ and $\mathscr{R}$-classes); namely, from top to bottom, there is one $\mathscr{J}$-class of size $1 = 1 \times 1$, two $\mathscr{J}$-classes of size $2 = 1 \times 2$, two $\mathscr{J}$-classes of size $6 = 2 \times 3$, and one $\mathscr{J}$-class of size $6 = 1 \times 6$, where $n \times m$ gives the dimension of the eggbox picture. In other words the $\mathscr{J}$-class splits into n $\mathscr{R}$-classes of size m and also into m $\mathscr{L}$-classes of size n. This example is specific in that all $\mathscr{J}$-classes are regular.

In the sequel, we describe $\mathscr{R}$-classes for general elements, as well as $\mathscr{J}$-order on regular elements. In particular, we obtain that the $\mathscr{J}$-classes of idempotents are indexed by the elements of W, and that $\mathscr{J}$-order on regular classes is given by left-right order $<_{LR}$ on W. Note that the latter is *not* a lattice, unlike for the variety $\mathscr{DA}$ (which consists of all aperiodic monoids all of whose simple modules are dimension 1; see for example [Ganyushkin et al. 2009]).

Proposition 4.18. *Two elements $f, g \in M(W)$ are in the same $\mathscr{R}$-class if and only if they have the same fibers. In particular, the $\mathscr{R}$-class of f is given by*

$$\mathscr{R}(f) = \{h \in f\,M(W) \mid \operatorname{rank}(h) = \operatorname{rank}(f)\} = \{f_u \mid u \in [1, \operatorname{type}(f)^{-1}w_0]_R\}, \quad (4\text{-}3)$$

where f_u is the unique element of $M(W)$ such that $\operatorname{fibers}(f_u) = \operatorname{fibers}(f)$ *and* $1\,.\,f_u = u$.

Proof. It is a general easy fact about monoids of functions that elements in the same $\mathscr{R}$-class have the same fibers (see also Lemma 4.12). Reciprocally, if g has the same fibers as f, then one can use Remark 2.7 to define $g' = g\overline{\pi}_{(1\,.\,g)^{-1}}\pi_{1\,.\,f}$ such that $\operatorname{fibers}(g') = \operatorname{fibers}(f)$ and $1\,.\,g' = 1\,.\,f$. Also by Proposition 4.11, $f = g' \in gM(W)$, and similarly, $g \in f\,M(W)$.

Equation (4-3) follows using Lemma 4.12 and Remark 2.7. $\square$

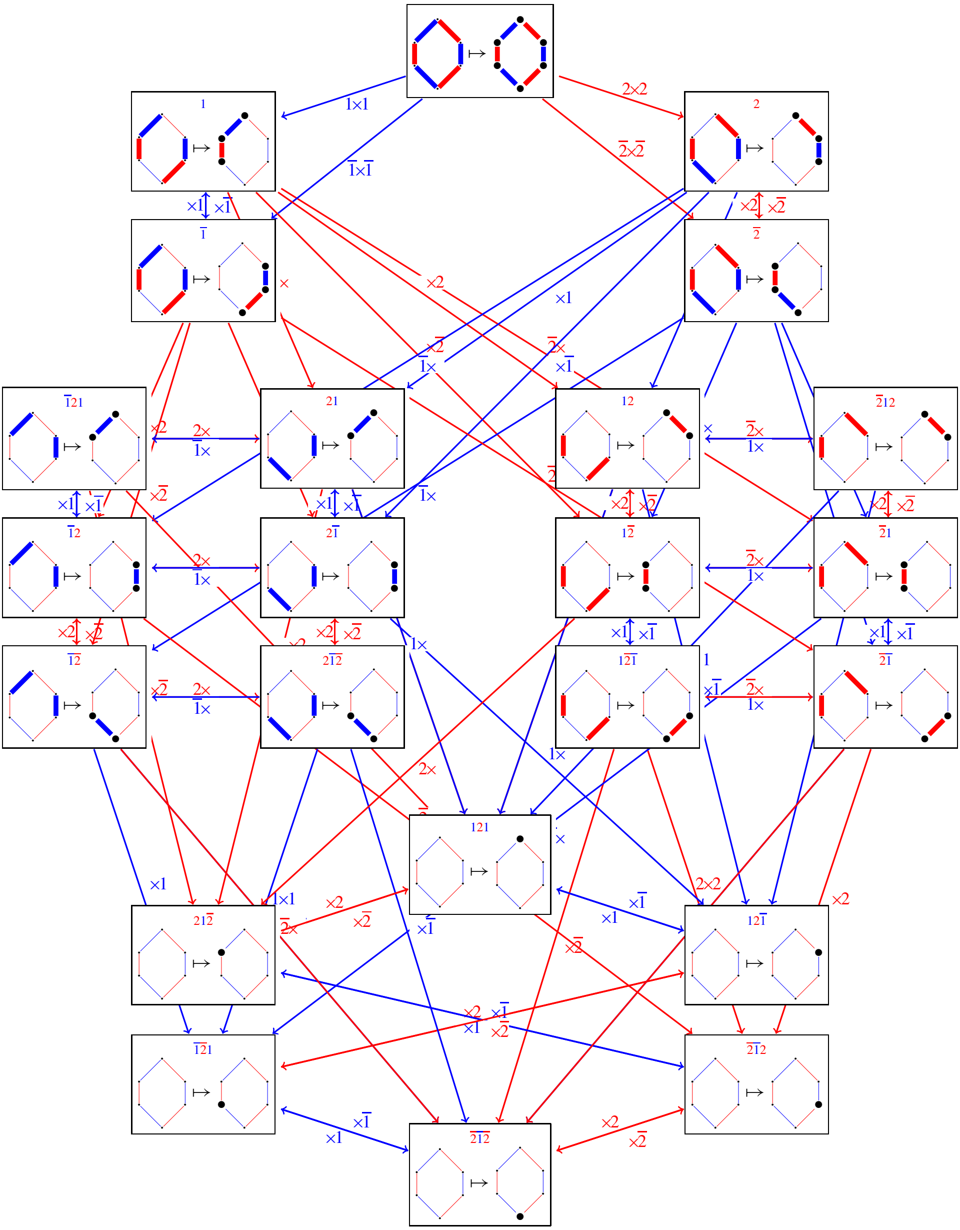


Figure 5. The graph of $\mathscr{J}$-order for $M(\mathfrak{S}_3)$, as described on page 636.

Lemma 4.19. *Let e and f be idempotents of $M(W)$ with respective image sets $[a,b]_L$ and $[c,d]_L$. Then, $f \leq_{\mathscr{J}} e$ if and only if $dc^{-1} \leq_{LR} ba^{-1}$.*

In particular, two idempotents e and f are $\mathscr{J}$-equivalent if and only if the intervals $[a,b]_L$ and $[c,d]_L$ are of the same type: $dc^{-1} = ba^{-1}$.

The properties above extend to any two regular elements (*elements whose* $\mathcal{J}$-*class contains an idempotent*).

Proof. First note that an interval $[c, d]_L$ is isomorphic to a subinterval of $[a, b]_L$ if and only $dc^{-1} \leq_{LR} ba^{-1}$. This follows from Proposition 2.5 and the fact that $[c, d]_L$ is a subinterval of $[a, b]_L$ if and only if $[ca^{-1}, da^{-1}]_L$ is a subinterval of $[1, ba^{-1}]_L$. But then dc^{-1} is a subfactor of ba^{-1}.

Assume first that $dc^{-1} \leq_{LR} ba^{-1}$, and let $[c', d']_L$ be a subinterval of $[a, b]_L$ isomorphic to $[c, d]_L$. Using Proposition 2.5, take $u, v \in M(W)$ that induce reciprocal bijections between $[c, d]_L$ and $[c', d']_L$. Then, $f = fuev$, so that f is $\mathcal{J}$-equivalent to e.

Reciprocally, assume that $f = uev$ with $u, v \in M(W)$. Without loss of generality, we may assume that $u = ue$ so that $\mathrm{im}(u) \subseteq [a, b]_L$. Set $c' = c.u$ and $d' = d.u$. Since $f = ff = fuv$, and using Proposition 4.3, the functions u and v must induce reciprocal isomorphisms between $[c, d]_L$ and $[c', d']_L$, the latter being a subinterval of $[a, b]_L$. Therefore, $dc^{-1} \leq_{LR} ba^{-1}$.

To conclude, note that a regular element has the same type as any idempotent in its $\mathcal{J}$-class. □

Corollary 4.20. *The idempotents* $(e_w)_{w \in W}$ *form a complete set of representatives of regular* $\mathcal{J}$-*classes in* $M(W)$.

Example 4.21. For $w \in W$, the idempotents e_w and $\tilde{e}_{w^{-1}w_0}$ are in the same $\mathcal{J}$-class. This follows immediately from Lemma 4.19, or by direct computation using the explicit expressions for e_w and $\tilde{e}_{w^{-1}w_0}$ in Proposition 4.15:

$$e_w = e_w^2 = \pi_{w^{-1}w_0} \overline{\pi}_{w_0 w} \pi_{w^{-1}w_0} \overline{\pi}_{w_0 w} = \pi_{w^{-1}w_0} \tilde{e}_{w^{-1}w_0} \overline{\pi}_{w_0 w},$$
$$\tilde{e}_{w^{-1}w_0} = \tilde{e}^2_{w^{-1}w_0} = \overline{\pi}_{w_0 w} \pi_{w^{-1}w_0} \overline{\pi}_{w_0 w} \pi_{w^{-1}w_0} = \overline{\pi}_{w_0 w} e_w \pi_{w^{-1}w_0}.$$

Corollary 4.22. *The image of a regular element is an interval in left order.*

Proof. A regular element has the same type, and same size of image set as any idempotent in its $\mathcal{J}$-class. □

Remark 4.23. The reciprocal is false: In type B_3, the element $\overline{\pi}_1 \overline{\pi}_3 \overline{\pi}_2 \pi_1 \overline{\pi}_3 \overline{\pi}_2 \overline{\pi}_1$ has the interval $[1, s_2 s_3 s_2]_L$ as image set, but it is not regular. The same holds in type A_4 with the element $\pi_2 \pi_1 \overline{\pi}_4 \pi_3 \overline{\pi}_2 \overline{\pi}_1 \overline{\pi}_3 \pi_4 \overline{\pi}_2 \overline{\pi}_3 \overline{\pi}_4$.

Problem 4.24. Describe $\mathcal{L}$-classes in general, and $\mathcal{L}$-order, $\mathcal{R}$-order, as well as $\mathcal{J}$-order on nonregular elements.

4g. ***Involutions and consequences.*** Define an involution $*$ on W by

$$w \mapsto w^* := w_0 w,$$

where w_0 is the maximal element of W. Moreover, define the *bar* map $M(W) \to M(W)$ as the conjugacy by $*$: For a given $f \in M(W)$

$$w\,.\,\bar{f} := (w^*\,.\,f)^* \quad \text{for all } w \in W.$$

Proposition 4.25. *The bar involution is a monoid endomorphism of $M(W)$ that exchanges π_i and $\overline{\pi}_i$.*

Proof. This is a consequence of the general fact that for any permutation ϕ of W, conjugation by ϕ is an automorphism of the monoid of maps from W to itself. Moreover, it is easy to see that bar exchanges π_i and $\overline{\pi}_i$, so that it fixes $M(W)$. □

The previous proposition has some interesting consequences when applied to idempotents: For any $w \in W$, the bar involution is a bijection from $e_w M(W)$ to $\overline{e}_w M(W)$. But $\overline{e}_w$ fixes w_0 and sends $1 = w_0^*$ to w^*, so that $\overline{e}_w = e_{w^*,w_0} = \tilde{e}_{w_0 w}$. The latter is in turn $\mathcal{J}$-equivalent to $e_{w_0 w^{-1} w_0}$ by Example 4.21. This implies the following result.

Corollary 4.26. *The ideals $e_w M(W)$ and $e_{w_0 w^{-1} w_0} M(W)$ are in bijection.*

5. The Borel submonoid $M_1(W)$ and its representation theory

In the previous section, we outlined the importance of the idempotents $(e_w)_{w \in W}$. A crucial feature is that they live in a “Borel” submonoid $M_1(W) \subseteq M(W)$ of elements of the biHecke monoid $M(W)$ that fix the identity:

$$M_1(W) := \{f \in M(W) \mid 1\,.\,f = 1\}.$$

In this section we study this monoid and its representation theory, as an intermediate step toward the representation theory of $M(W)$ (see Section 6). For the representation theory of $M(W)$, it is actually more convenient to work with the submonoid fixing w_0 instead of 1:

$$M_{w_0}(W) := \{f \in M(W) \mid w_0\,.\,f = w_0\}.$$

However, since both monoids $M_1(W)$ and $M_{w_0}(W)$ are isomorphic under the involution of Section 4g and since the interaction of $M_1(W)$ with Bruhat order is notationally simpler, we focus on $M_1(W)$ in this section.

Note. In the remainder of this paper, unless explicitly stated, we fix a Coxeter group W and use the shorthand notation $M := M(W)$, $M_1 := M_1(W)$ and $M_{w_0} := M_{w_0}(W)$.

From the definition it is clear that M_1 is indeed a submonoid that contains the idempotents $(e_w)_{w \in W}$. Furthermore, by Proposition 4.5 and Corollary 4.7 its elements are both order-preserving and regressive for Bruhat order. In fact, a bit more can be said.

Remark 5.1. For $w \in W$, $w . M_1$ is the interval $[1, w]_B$ in Bruhat order.

Proof. By Corollary 4.7, for $f \in M_1$, we have $w . f \leq_B w$. Take reciprocally $v \in [1, w]_B$. Then, using Proposition 4.15, $w . e_v = v$. □

As a consequence of the preservation and regressiveness on Bruhat order, M_1 is *an ordered monoid with* 1 *on top*. Namely, for $f, g \in M_1$, define the relation $f \leq g$ if $w . f \leq_B w . g$ for all $w \in W$. Then, $\leq$ defines a partial order on M_1 such that $f \leq 1$, $fg \leq f$ and $fg \leq g$ for all $f, g \in M_1$. In other words, M_1 is $\mathcal{R}$-trivial (see [Denton et al. 2010/11, Proposition 2.2], as well as Section 2.5 there) and in particular $\mathcal{J}$-trivial.

In the next two subsections, we study the combinatorics of M_1 and then apply the general results on the representation theory of $\mathcal{J}$-trivial monoids of [Denton et al. 2010/11] to M_1.

5a. *$\mathcal{J}$-order on idempotents and minimal generating set.* Recall from Section 2e that $\mathcal{J}$-order is the partial order $\leq_{\mathcal{J}}$ defined by $f \leq_{\mathcal{J}} g$ if there exists $x, y \in M_1$ such that $f = xgy$. The restriction of $\mathcal{J}$-order to idempotents has a very simple description:

Proposition 5.2. *For $u, v \in W$, the following are equivalent*:

$$e_u e_v = e_u, \qquad u \leq_L v,$$
$$e_v e_u = e_u, \qquad e_u \leq_{\mathcal{J}} e_v.$$

Moreover, $(e_u e_v)^\omega = e_{u \wedge_L v}$, *where $u \wedge_L v$ is the meet* (*or greatest lower bound*) *of u and v in left order.*

Proof. This follows from [Denton et al. 2010/11, Theorem 3.4, Lemma 3.6] and Proposition 4.15. □

As a consequence the following definition, which plays a central role in the representation theory of $\mathcal{J}$-trivial monoids [Denton et al. 2010/11], makes sense.

Definition 5.3. For any element $x \in M_1$, define

$$\operatorname{lfix}(x) := \min_{\leq_L}\{u \in W \mid e_u x = x\} \quad \text{and} \quad \operatorname{rfix}(x) := \min_{\leq_L}\{u \in W \mid x e_u = x\} = w_0 . x,$$

the min being taken for the left order.

Interestingly, M_1 can be defined as the submonoid of M generated by the idempotents $(e_w)_{w \in W}$, and in fact the subset of these idempotents indexed by Grassmannian elements (an element $w \in W$ is *Grassmannian* if it has at most one descent).

Theorem 5.4. *M_1 has a unique minimal generating set that consists of the idempotents e_w where $w^{-1} w_0$ is right Grassmannian.*

In type A_{n-1} this minimal generating set is of size $2^n - n$ (which is the number of Grassmannian elements in this case [Manivel 2001]).

Proof. Define the length $\ell(f)$ of an element $f \in M$ as the length of a minimal expression of f as a product of the generators π_i and $\overline{\pi}_i$. We now prove by induction on the length that M_1 is generated by $\{e_w \mid w \in W\}$.

Take an element $f \in M_1$ of length l. If $l = 0$ we are done. Otherwise, since $1 . f = 1$, an expression of f as a product of the π_i and $\overline{\pi}_i$ contains at least one $\overline{\pi}_i$. Write $f = gh$ where $g = \pi_w \overline{\pi}_i$ for some $w \in W$ and $h \in M$ so that $\ell(w)+1+\ell(h) = l$.

Claim. *$f = e_{w_0(ws_i)^{-1}} \pi_w h$ and $\pi_w h \in M_1$.*

It follows from the claim that $\ell(\pi_w h) < l$, and hence since $\pi_w h \in M_1$ we can apply induction to conclude that M_1 is generated by $\{e_w \mid w \in W\}$.

Let us prove the claim. By minimality of l, i is not a descent of w (otherwise, we would obtain a shorter expression for f: $f = \pi_w \overline{\pi}_i h = \pi_{w'} \pi_i \overline{\pi}_i h = \pi_{w'} \overline{\pi}_i h$, where $\ell(w') < \ell(w)$). Therefore, $1 . g = 1 . (\pi_w \overline{\pi}_i) = w$. Since $f \in M_1$ it follows that $w . h = 1$ and therefore $\pi_w h \in M_1$. It further follows that $\overline{\pi}_{w^{-1}} \pi_w$ acts trivially on the image set $[w, w_0]_L$ of g, and therefore $f = g \overline{\pi}_{w^{-1}} \pi_w h$. Note that $g \overline{\pi}_{w^{-1}} = \pi_w \overline{\pi}_i \overline{\pi}_{w^{-1}} = \pi_w \pi_i \overline{\pi}_i \overline{\pi}_{w^{-1}} = e_{w_0(ws_i)^{-1}}$.

By Proposition 5.2, the idempotents of M_1 are generated by the meet-irreducible idempotents e_w in $\mathcal{J}$ order. Here x is meet-irreducible if and only if $x = a$ or $x = b$ whenever $x = a \wedge b$ for some $a, b \in M_1$. These meet-irreducible elements are indexed by the elements w of W that are meet-irreducible in left order (or equivalently that have at most one left nondescent, that is, $w_0 w^{-1}$ is right Grassmannian).

The uniqueness of the minimal generating set holds for any $\mathcal{J}$-trivial monoid with a minimal generating set [Doyen 1984, Theorem 2; Doyen 1991, Theorem 1]. □

Actually one can be much more precise:

Proposition 5.5. *Any element $f \in M_1$ can be written as a product $e_{w_1} \cdots e_{w_k}$, where*

- *$w_1 >_B \cdots >_B w_k$ is a chain in Bruhat order such that any two consecutive terms w_i and w_{i+1} are incomparable in left order*;
- *$w_i = \mathrm{rfix}(e_{w_1} \cdots e_{w_i}) = \mathrm{lfix}(e_{w_i} \cdots e_{w_k})$.*

Proof. Start from any expression $e_{w_1} \cdots e_{w_k}$ for f. We show that if any of the conditions of the proposition is not satisfied, the expression can be reduced to a strictly smaller (in length, or in Bruhat, term by term) expression, so that induction can be applied.

- If $u \not>_B v$, then by Remark 4.16 $e_u e_v = e_u e_{u.e_v}$ with $u . e_v <_B v$.
- If $u <_L v$, then $e_u e_v = e_u$, and similarly on the right.

- If the left symbol e_u for $e_{w_i} \cdots e_{w_k}$ is not e_{w_i}, then $u <_L w_i$ and

$$e_{w_i} \cdots e_{w_k} = e_u e_{w_i} \cdots e_{w_k} = e_u e_{w_{i+1}} \cdots e_{w_k}.$$

Similarly on the right. □

Corollary 5.6. *For $f \in M_1$, $\operatorname{lfix}(f) \geq_B \operatorname{rfix}(f)$, with equality if and only if f is an idempotent.*

Lemma 5.7. *If $v \leq_B u$ in Bruhat order and $u' = \operatorname{lfix}(e_u e_v)$, then*

$$v \leq_B u' \quad \text{and} \quad u' \leq_L u.$$

Proof. By Definition 5.3, $u' \leq_L u$ since $e_u(e_u e_v) = e_u e_v$ and for M_1 the minimum is measured in left order. Also by Proposition 4.15

$$v = w_0 . e_u e_v = w_0 . e_{u'} e_u e_v \leq_B u'.$$ □

Lemma 5.8. *If u covers v in Bruhat order and $u' = \operatorname{lfix}(e_u e_v)$, then either $u' = u$, or $u' = v$ and $e_u e_v = e_v e_u$.*

Proof. By Lemma 5.7, we have that either $u' = u$ or $u' = v$, since u covers v in Bruhat order. When $u' = v$, we have again by Lemma 5.7 that $v \leq_L u$. Hence $e_u e_v = e_v = e_v e_u$. □

5b. *Representation theory.* In this subsection, we specialize general results about the representation theory of finite $\mathcal{J}$-trivial monoids to describe some of the representation theory of the Borel submonoid M_1, such as its simple modules, radical, Cartan invariant matrix and quiver. The description also applies to M_{w_0}, mutatis mutandis. We follow the presentation of [Denton et al. 2010/11] (also see this paper for the proofs), though many of the general results have been previously known; see for example [Almeida et al. 2009; Clifford and Preston 1961; Ganyushkin et al. 2009; Lallement and Petrich 1969; Rhodes and Zalcstein 1991] and references therein.

5b1. *Simple modules and radical.* For each $w \in W$ define S_w^1 (written $S_w^{w_0}$ for M_{w_0}) to be the one-dimensional vector space with basis $\{\epsilon_w\}$ together with the right operation of any $f \in M_1$ given by

$$\epsilon_w . f := \begin{cases} \epsilon_w & \text{if } w . f = w, \\ 0 & \text{otherwise.} \end{cases}$$

The basic features of the representation theory of M_1 can be stated as follows:

Theorem 5.9. *The radical of $\mathbb{K}M_1$ is the ideal with basis $f^\omega - f$ for $f \in M_1$ nonidempotent. The quotient of $\mathbb{K}M_1$ by its radical is commutative. Therefore, all simple $\mathbb{K}M_1$-modules are one-dimensional. In fact, the family $\{S_w^1\}_{w \in W}$ forms a complete system of representatives of the simple $\mathbb{K}M_1$-modules.*

5b2. *Cartan matrix and projective modules.* The projective modules and Cartan invariants can be described as follows:

Theorem 5.10. *There exists an explicit basis $(b_x)_{x\in M_1}$ of $\mathbb{K}M_1$ such that, for all $w \in W$,*

- *the family $\{b_x \mid x \in M_1$ with $\operatorname{lfix}(x) = w\}$ is a basis for the right indecomposable projective module P^1_w associated to S^1_w;*
- *the family $\{b_x \mid \operatorname{rfix}(x) = w\}_{x\in M_1}$ is a basis for the left indecomposable projective module associated to S^1_w.*

Moreover, the Cartan invariant of $\mathbb{K}M_1$ defined by $c_{u,v} := \dim(e_u \mathbb{K}M_1 e_v)$ for $u, v \in W$ is given by $c_{u,v} = |C_{u,v}|$, where

$$C_{u,v} := \{f \in M_1 \mid \operatorname{lfix}(f) = u \text{ and } \operatorname{rfix}(f) = v\}.$$

In particular, the Cartan matrix of $\mathbb{K}M_1$ is upper-unitriangular with respect to Bruhat order.

Proof. Apply [Denton et al. 2010/11, Section 3.4] and conclude with Corollary 5.6. □

Remark 5.11. In terms of characters, the previous theorem can be restated as

$$[P^1_u] = \sum_{f\in M_1, \operatorname{lfix}(f)=u} [S^1_{w_0\cdot f}], \tag{5-1}$$

which gives the following character for the right regular representation:

$$[\mathbb{K}M_1] = \sum_{f\in M_1} [S^1_{w_0\cdot f}]. \tag{5-2}$$

Problem 5.12. Describe the Cartan matrix and projective modules of $\mathbb{K}M_1$ more explicitly, if at all possible in terms of the combinatorics of the Coxeter group W.

5b3. *Quiver.* We now turn to a description of the quiver of $\mathbb{K}M_1$ in terms of the combinatorics of left and Bruhat order. Recall that M_1 is a submonoid of the monoid of regressive and order preserving functions. As such, it is not only $\mathcal{J}$-trivial but also ordered with 1 on top, that is $\mathcal{B}$-trivial; see [Denton et al. 2010/11, Section 2.5 and Proposition 2.2]. By [Denton et al. 2010/11, Theorem 3.35 and Corollary 3.44] we know that the vertices of the quiver of a $\mathcal{J}$-trivial monoid generated by idempotents are labeled by its idempotents $(e_x)_x$ and there is an edge from vertex e_x to vertex e_z, if $q := e_x e_z$ is not idempotent, has $\operatorname{lfix}(q) = x$ and $\operatorname{rfix}(q) = z$, and does not admit any factorization $q = uv$ that is nontrivial: $eu \neq e$ and $vf \neq f$. By [Denton et al. 2010/11, Proposition 3.31] the condition that q has a nontrivial factorization is equivalent to q having a compatible factorization $q = uv$, meaning that u, v are nonidempotents and $\mathrm{lfix}(q) = \mathrm{lfix}(u)$, $\mathrm{rfix}(u) = \mathrm{lfix}(v)$ and $\mathrm{rfix}(v) = \mathrm{rfix}(q)$.

Let $e_x, e_y, e_z \in M_1$ be idempotents. Call e_y an *intermediate factor* for $q := e_x e_z$ if $e_x e_y e_z = e_x e_z$. Call further e_y a *nontrivial intermediate factor* if $e_x e_y \neq e_x$, and $e_y e_z \neq e_z$.

Lemma 5.13. *The quiver of $\mathbb{K}M_1$ is the graph with W as vertex set and edges (x,z) for all $x \neq z$ such that $q := e_x e_z$ satisfies* $\mathrm{lfix}(q) = x$ *and* $\mathrm{rfix}(q) = z$ *and admits no nontrivial intermediate factor e_y with $y \in W$.*

Proof. Take $q := e_x e_z$ admitting a nontrivial intermediate factor e_y. Then q admits a nontrivial factorization $q = (e_x e_y)(e_y e_z)$ in the sense of [Denton et al. 2010/11, Definition 3.25], and is therefore not in the quiver.

Reciprocally, assume that q admits a compatible factorization, that is $q = uv$ with $\mathrm{lfix}(u) = x$, $\mathrm{rfix}(u) = \mathrm{lfix}(v)$ and $\mathrm{rfix}(v) = z$. By [Denton et al. 2010/11, Lemma 3.29], this factorization is nontrivial: $e_x u \neq e_x$ and $v e_z \neq e_z$. Using Proposition 5.5, write u and v as $u = e_x e_{y_1} \cdots e_{y_k}$ and $v = e_{y_k} \cdots e_{y_\ell} e_z$, with $x >_B y_1 >_B \cdots >_B y_\ell >_B z$. Then, $e_x e_{y_i} e_z = e_x e_z$ for any i; indeed, since M_1 is $\mathcal{B}$-trivial,

$$e_x e_z = e_x e_{y_1} \cdots e_{y_\ell} e_z \leq_{\mathcal{B}} e_x e_{y_i} e_z \leq_{\mathcal{B}} e_x e_z.$$

If any e_{y_i} is a nontrivial intermediate factor for q, we are done by setting $y = y_i$. Otherwise, $e_{y_i} e_z = e_z$ for any i ($e_x e_{y_i} = e_x$ is impossible since $x >_B y_i$). But then, $v = e_{y_k} \cdots e_{y_\ell} e_z = e_z$, a contradiction. □

Problem 5.14. Can Lemma 5.13 be generalized to any $\mathcal{B}$-trivial monoid? Its statement has been tested successfully on the 0-Hecke monoid in type $A_1 - A_6$, $B_3 - B_4$, $D_4 - D_5$, $H_3 - H_4$, G_2, I_{135}, F_4.

Lemma 5.13 admits a combinatorial reformulation in terms of the combinatorics of W. For $x, y, z \in W$ such that $x >_B z$, call $y \in W$ an *intermediate factor* for x, z if $[1, y]_L$ intersects all intervals $[c, a]_B$ with $a \in [1, x]_L$ and $c \in [1, z]_L$ nontrivially. Call further y a *nontrivial intermediate factor* if $x >_B y >_B z$ and $y \not>_L z$.

Theorem 5.15. *The quiver of $\mathbb{K}M_1$ is the graph with W as vertex set, and edges (x,z) for all $x >_B z$ and $x \not>_L z$ admitting no nontrivial intermediate factor y. Each such edge can be associated with the element $q := e_x e_z$ of the monoid.*

In particular, the quiver of $\mathbb{K}M_1$ is acyclic and every cover $x \succ_B z$ in Bruhat order that is not a cover in left order contributes one edge to the quiver.

Proof. Consider a nonidempotent product $e_x e_z$. Using Proposition 5.5, we may assume without loss of generality that $x >_B z$ and $x \not>_L z$, and furthermore that $\mathrm{lfix}(e_x e_z) = x$ and $\mathrm{rfix}(e_x e_z) = z$.

We now show that the combinatorial definition of intermediate factor on an element of $y \in W$ is a reformulation of the monoidal one on the idempotent e_y of M_1.

Assume that e_y is an intermediate factor for $e_x e_z$, that is, $e_x e_y e_z = e_x e_z$. Take $a \in [1, x]_L$ and $c \in [1, z]_L$ with $a \geq_B c$, and write $b = a.e_y \in [1, y]_L$. Using Proposition 4.15, $a \geq_B b$ and $a.e_z \geq_B c$. Furthermore, since a is in the image set of e_x, one has $b.e_z = a.e_y.e_z = a.e_z \geq_B c$. Therefore, $[1, y]_L$ intersects $[c, a]_B$ at least in b. Hence, y is an intermediate factor for x, z.

For the reciprocal, take any $a \in [1, x]_L$. Since M_1 preserves Bruhat order and is regressive, $a.e_y.e_z \leq_B a.e_z$. Set $c = a.e_z$, and take b in $[c, a]_B \cap [1, y]_L$. Using Proposition 4.15,

$$a.e_y.e_z \geq_B b.e_z \geq_B c = a.e_z,$$

and equality holds. Hence, e_y is an intermediate factor for e_x, e_y: $e_x e_y e_z = e_x e_z$.

The combinatorial reformulation of nontriviality for intermediate factors is then straightforward using Proposition 5.5. $\square$

Problem 5.16. Exploit the interrelations between left order and Bruhat order to find a more satisfactory combinatorial description of the quiver of $\mathbb{K}M_1$.

5b4. *Connection with the representation theory of the* 0*-Hecke monoid.* Recall that the 0-Hecke monoid $H_0(W)$ is a submonoid of $M_{w_0}(W)$. As a consequence any $\mathbb{K}M_{w_0}(W)$-module is a $H_0(W)$-module and one can consider the decomposition map $G_0(M_{w_0}(W)) \to G_0(H_0(W))$. It is given by the following formula:

Proposition 5.17. *For $w \in W$, let $S_w^{w_0}$ be the simple $\mathbb{K}M_{w_0}(W)$-module defined by*

$$\epsilon_w . f := \begin{cases} \epsilon_w & \text{if } w.f = w, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, for $J \subseteq I$, let $S_J^{H_0}$ be the simple $H_0(W)$-module defined by

$$\mu_J . \pi_i := \begin{cases} \mu_I & \text{if } i \in J, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the restriction of $S_w^{w_0}$ to $H_0(W)$ is isomorphic to $S_{\mathrm{D}_R(w)}^{H_0}$. The decomposition map is therefore given by

$$G_0(M_{w_0}(W)) \to G_0(H_0(W)), \quad [S_w^{w_0}] \mapsto [S_{\mathrm{D}_R(w)}^{H_0}].$$

Proof. By definition of the action, $w.\pi_i = w$ if and only if $i \in \mathrm{D}_R(W)$. $\square$

5b5. *The tower of $M_1(\mathfrak{S}_n)$ monoids (type A).*

Problem 5.18. The monoids $M_1(\mathfrak{S}_n)$, for $n \in \mathbb{N}$, form a tower of monoids with the natural embeddings $M_1(\mathfrak{S}_n) \times M_1(\mathfrak{S}_m) \hookrightarrow M_1(\mathfrak{S}_{m+n})$. Due to the involution of Section 4g, one has also embeddings $M_{w_0}(\mathfrak{S}_n) \times M_{w_0}(\mathfrak{S}_m) \hookrightarrow M_{w_0}(\mathfrak{S}_{m+n})$. As outlined in the introduction, it would hence be interesting to understand the induction and restriction functors in this setting, and in particular to describe the bialgebra obtained from the associated Grothendieck groups. This would give a representation theoretic interpretation of some bases of FQSym.

In this context, Proposition 5.17 provides an interpretation of the surjective coalgebra morphism FQSym $\twoheadrightarrow$ QSym, through the restriction along the following commutative diagram of monoid inclusions (see [Duchamp et al. 2002] for more details):

$$\begin{tikzcd} H_0(\mathfrak{S}_n) \times H_0(\mathfrak{S}_m) \arrow[r, hook] \arrow[d, hook] & M_{w_0}(\mathfrak{S}_n) \times M_{w_0}(\mathfrak{S}_m) \arrow[d, hook] \\ H_0(\mathfrak{S}_{n+m}) \arrow[r, hook] & M_{w_0}(\mathfrak{S}_{n+m}). \end{tikzcd}$$

6. Translation modules and w-biHecke algebras

The main purpose of this section is to pave the ground for the construction of the simple modules S_w of the biHecke monoid $M := M(W)$ in Section 7a.

As for any aperiodic monoid, each such simple module is associated with some regular $\mathcal{J}$-class D of the monoid, and can be constructed as a quotient of the span $\mathbb{K}\mathcal{R}(f)$ of the $\mathcal{R}$-class of any idempotent f in D, endowed with its natural right $\mathbb{K}M$-module structure (see Section 2g).

In Section 6a, we endow the interval $[1, w]_R$ with a natural structure of a combinatorial $\mathbb{K}M$-module T_w, called *translation module*, and show that, for any $f \in M$, regular or not, the right $\mathbb{K}M$-module $\mathbb{K}\mathcal{R}(f)$ is always isomorphic to some T_w.

The translation modules will play an ubiquitous role for the representation theory of $\mathbb{K}M$ in Section 7: indeed T_w can be obtained by induction from the simple modules S_w of $\mathbb{K}M$, and the right regular representation of $\mathbb{K}M$ admits a filtration in terms of the T_w that mimics the composition series of the right regular representation of $\mathbb{K}M_{w_0}$ in terms of its simple modules S_w. Reciprocally T_w, and therefore the right regular representation of $\mathbb{K}M$, restricts naturally to M_{w_0}. Finally, T_w is closely related to the projective module P_w of $\mathbb{K}M$ (Corollary 7.4).

By taking the quotient of $\mathbb{K}M$ through its representation on T_w, we obtain a w-analogue $\mathcal{H}W^{(w)}$ of the biHecke algebra $\mathcal{H}W$. This algebra turns out to be interesting in its own right, and we proceed by generalizing most of the results of [Hivert and Thiéry 2009] on the representation theory of $\mathcal{H}W$.

As a first step, we introduce in Section 6b a collection of submodules $P_J^{(w)}$ of T_w, which are analogues of the projective modules of $\mathcal{H}W$. Unlike for $\mathcal{H}W$, not any subset J of I yields such a submodule, and this is where the combinatorics of the blocks of w as introduced in Section 3 enters the game. In a second step, we derive in Section 6c a lower bound on the dimension of $\mathscr{H}W^{(w)}$; this requires a (fairly involved) combinatorial construction of a family of functions on $[1, w]_R$ that is triangular with respect to Bruhat order. In Section 6d we combine these results to derive the dimension and representation theory of $\mathscr{H}W^{(w)}$: projective and simple modules, Cartan matrix, quiver, etc. (see Theorem 6.17).

6a. ***Translation modules and w-biHecke algebras.*** In this section we study the combinatorics of the right class modules for the biHecke monoid, in particular a combinatorial model for them. Indeed, we show that the right class modules correspond to uniform translations of image sets, hence the name "translation modules".

Fix $f \in M$. Recall from Definition 2.12 that the right class module associated to f is defined as the quotient

$$\mathbb{K}\mathscr{R}(f) := \mathbb{K} f M / \mathbb{K}\mathscr{R}_<(f).$$

The basis of $\mathbb{K}\mathscr{R}(f)$ is the right class $\mathscr{R}(f)$ described in Proposition 4.18. Recall from there that f_u denote the unique element of $M(W)$ such that $\mathrm{fibers}(f_u) = \mathrm{fibers}(f)$ and $1\,.\,f_u = u$.

Proposition 6.1. *Set $w = \mathrm{type}(f)^{-1} w_0$. Then $(f_u)_{u\in[1,w]_R}$ forms a basis of $\mathbb{K}\mathscr{R}(f)$ such that*

$$\begin{aligned} f_u\,.\,\pi_i &= \begin{cases} f_u & \text{if } i \in \mathrm{D}_R(u),\\ f_{us_i} & \text{if } i \notin \mathrm{D}_R(u) \text{ and } us_i \in [1, w]_R,\\ 0 & \text{otherwise};\end{cases}\\ f_u\,.\,\overline{\pi}_i &= \begin{cases} f_{us_i} & \text{if } i \in \mathrm{D}_R(u),\\ f_u & \text{if } i \notin \mathrm{D}_R(u) \text{ and } us_i \in [1, w]_R,\\ 0 & \text{otherwise.}\end{cases} \end{aligned} \tag{6-1}$$

In particular, the action of any $g \in M$ on a basis element f_u of the right class module either annihilates f_u or agrees with the usual action on W: $f_u\,.\,g = f_{u\,.\,g}$.

Proof. By Definition 2.12 and Proposition 4.18, $(f_u)_{u\in[1,w]_R}$ forms a basis of $\mathbb{K}\mathscr{R}(f)$.

The action of π_i agrees with right multiplication, except when the index v of the new f_v is no longer in $[1, w]_R$, in which case the element is annihilated. The action of $\overline{\pi}_i$ also agrees with right multiplication. However, due to the relations $\pi_i\overline{\pi}_i = \overline{\pi}_i$ and $\overline{\pi}_i\pi_i = \pi_i$, we need that $\overline{\pi}_i$ annihilates f_u if $i \notin \mathrm{D}_R(u)$ and $us_i \notin [1, w]_R$.

The last statement follows by induction writing $f \in M$ in terms of the generators π_i and $\overline{\pi}_i$ and using (6-1). □

Proposition 6.1 gives a combinatorial model for right class modules. It is clear that two functions with the same type yield isomorphic right class modules. The converse also holds:

Proposition 6.2. *For any $f, f' \in M$, the right class modules $\mathbb{K}\mathcal{R}(f)$ and $\mathbb{K}\mathcal{R}(f')$ are isomorphic if and only if* $\operatorname{type}(f) = \operatorname{type}(f')$.

Proof. By Proposition 6.1, it is clear that if $\operatorname{type}(f) = \operatorname{type}(f')$, then $\mathbb{K}\mathcal{R}(f) \cong \mathbb{K}\mathcal{R}(f')$.

Conversely, suppose $\operatorname{type}(f) \neq \operatorname{type}(f')$. Then we also have $w \neq w'$, where $w = \operatorname{type}(f)^{-1} w_0$ and $w' = \operatorname{type}(f')^{-1} w_0$. Without loss of generality, we may assume that $\ell(w) \geq \ell(w')$. Using the combinatorial model of Proposition 6.1, we then have

$$f_1 \,.\, \pi_w = f_w \neq 0 \quad \text{and} \quad f'_1 \,.\, \pi_w = 0\,,$$

so that $\mathbb{K}\mathcal{R}(f) \ncong \mathbb{K}\mathcal{R}(f')$. □

It is not obvious from the combinatorial action of π_i and $\overline{\pi}_i$ of Proposition 6.1 that the result indeed gives a module. However, since it agrees with the right action on the quotient space as in Definition 2.12, this is true. By Proposition 6.2, we may choose a canonical representative for right class modules.

Definition 6.3. The module $T_w := \mathbb{K}\mathcal{R}(e_{w,w_0})$ for all $w \in W$ is called the *translation module associated to w*. We identify its basis with $[1, w]_R$ via $u \mapsto f_u$, where $f = e_{w,w_0}$.

For the remainder of this section for $f \in M$ and $u \in [1, w]_R$, unless otherwise specified, $u \,.\, f$ means the action of f on u in the translation module T_w.

Definition 6.4. The w-biHecke algebra $\mathcal{H}W^{(w)}$ is the natural quotient of $\mathbb{K}M$ through its representation on T_w. In other words, it is the subalgebra of $\operatorname{End}(T_w)$ generated by the operators π_i and $\overline{\pi}_i$ of Proposition 6.1.

6b. ***Left antisymmetric submodules.*** By analogy with the simple reflections in the Hecke group algebra, we define for each $i \in I$ the operator $s_i := \pi_i + \overline{\pi}_i - 1$. For $u \in [1, w]_R$, the action on the translation module T_w is given by

$$u \,.\, s_i = \begin{cases} us_i & \text{if } us_i \in [1, w]_R, \\ -u & \text{otherwise.} \end{cases} \tag{6-2}$$

These operators are still involutions, but do not always satisfy the braid relations.

Example 6.5. Take W of type A_2 and $w = s_1$. The translation module T_w has two basis elements $B = (1, s_1)$ and the matrices for s_1 and s_2 on this basis are given by

$$s_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad s_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is not hard to check that then $s_1 s_2 s_1 \neq s_2 s_1 s_2$.

Similarly, one can define operators $\overleftarrow{s_i}$ acting on the left on the translation module T_w:

$$\overleftarrow{s_i}\,.u = \begin{cases} s_i u & \text{if } s_i u \in [1, w]_R, \\ -u & \text{otherwise.} \end{cases} \tag{6-3}$$

Definition 6.6. For $J \subseteq I$, set $P_J^{(w)} := \{v \in T_w \mid \overleftarrow{s_i}\,.v = -v \text{ for all } i \in J\}$.

For $w = w_0$, these are the projective modules P_J of the biHecke algebra [Hivert and Thiéry 2009].

Proposition 6.7. *Take $w \in W$ and $J \subseteq I$. Then, the following are equivalent*:

(i) *${}^J w$ is a cutting point of w*;

(ii) *$P_J^{(w)}$ is an $\mathbb{K}M$-submodule of T_w.*

Furthermore, when any, and therefore all, of the above hold, $P_J^{(w)}$ is isomorphic to $T_{{}^J w}$, and its basis is indexed by $[1, {}^J w]_R$, that is, assuming $J \in \mathcal{J}^{(w)}$, we have $\{v \in [1, w]_R, J \subset J^{(w)}(v)\}$.

Proof. (ii) $\Rightarrow$ (i): Set

$$v_J^w := \sum_{u \in [1, w_J]_R} (-1)^{\ell(u) - \ell(w_J)} u.$$

Up to a scalar factor, this is the unique vector in $P_J^{(w)}$ with support contained in $[1, w_J]_R$. Then,

$$v_J^w\,.\pi_{{}^J w} = \sum_{\substack{u \in [1, w_J]_R \\ \text{s.t. } u\,{}^J w \in [1, w]_R}} (-1)^{\ell(uv) - \ell(w_J v)} u\,{}^J w,$$

$$v_J^w\,.\pi_v \overline{\pi}_{v^{-1}} = \sum_{\substack{u \in [1, w_J]_R \\ \text{s.t. } u\,{}^J w \in [1, w]_R}} (-1)^{\ell(u) - \ell(w_J)} u.$$

Therefore, if ${}^J w \not\leq_R w$, then $v_J^w\,.\pi_{{}^J w} \overline{\pi}_{{}^J w^{-1}}$ is a nonzero vector with support strictly included in $[1, w_J]_R$ and therefore not in $P_J^{(w)}$. By Proposition 3.8 this proves that (ii) implies (i).

(i) $\Rightarrow$ (ii): If (i) holds, then the action of π_i (resp. $\overline{\pi}_i$) on $v_J^w\,.\pi_v$ either leaves it unmodified, kills it (if $vs_i = s_j v$ for some j) or maps it to $v_J^w\,.\pi_{vs_i}$. The vectors $(v_J^w\,.\pi_v)_{v \in [1, {}^J w]_R}$ form a basis of $P_J^{(w)}$ that is stable by M.

The last statement follows straightforwardly. ☐

It is clear from the definition that $P_{J_1 \cup J_2}^{(w)} = P_{J_1}^{(w)} \cap P_{J_2}^{(w)}$ for $J_1, J_2 \subseteq I$. Since the set $\mathcal{RB}_{\mathcal{L}}(w)$ of left blocks of w is stable under union, the set of $\mathbb{K}M$-modules $(P_J^{(w)})_{J \in \mathcal{RB}_{\mathcal{L}}(w)}$ is stable under intersection. On the other hand, unless J_1 and J_2 are comparable, $P_{J_1 \cup J_2}^{(w)}$ is a strict subspace of $P_{J_1}^{(w)} + P_{J_2}^{(w)}$. This motivates the following definition.

Definition 6.8. For $J \in \mathscr{J}^{(w)}$, we define the module

$$S_J^{(w)} := P_J^{(w)} \Big/ \sum_{J' \supsetneq J,\, J' \in \mathscr{RB}_{\mathscr{L}}(w)} P_{J'}^{(w)}. \tag{6-4}$$

Remark 6.9. By the last statement of Proposition 6.7, and the triangularity of the natural basis of the modules $P_{J'}^{(w)}$, the basis of $S_J^{(w)}$ is given by

$$[1, {}^J w]_R \setminus \bigcup_{v \sqsubset {}^J w} [1, v]_R = \{v \in [1, w]_R,\, J \subset J^{(w)}(v)\}. \tag{6-5}$$

6c. *A (maximal) Bruhat-triangular family of the w-biHecke algebra.* Consider the submonoid F in $\mathscr{H}W^{(w)}$ generated by the operators π_i, $\bar{\pi}_i$, and s_i, for $i \in I$. For $f \in F$ and $u \in [1, w]_R$, we have $u\,.\,f = \pm v$ for some $v \in [1, w]_R$. For our purposes, the signs can be ignored and f be considered as a function from $[1, w]_R$ to $[1, w]_R$.

Definition 6.10. For $u, v \in [1, w]_R$, a function $f \in F$ is called (u, v)-*triangular* (for Bruhat order) if v is the unique minimal element of $\operatorname{im}(f)$ and u is the unique maximal element of $f^{-1}(v)$ (all minimal and maximal elements in this context are with respect to Bruhat order).

Recall the notion of maximal reduced right block $K^{(w)}(u)$ of Definition 3.38.

Proposition 6.11. *Take $u, v \in [1, w]_R$ such $K^{(w)}(u) \subseteq K^{(w)}(v)$. Then, there exists a (u, v)-triangular function $f_{u,v}$ in F.*

For example, for $w = 4312$ in $\mathfrak{S}_4$, the condition on u and v is equivalent to the existence of a path from u to v in the digraph $G^{(4312)}$ (see Figure 1 and Section 6d).

The proof of Proposition 6.11 relies on several remarks and lemmas that are given in the sequel of this section. The construction of $f_{u,v}$ is explicit, and the triangularity derives from $f_{u,v}$ being either in M, or close enough to be bounded below by an element of M. It follows from the upcoming Theorem 6.17 that the condition on u and v is not only sufficient but also necessary.

Remark 6.12. If f is (u, v)-triangular and g is (v, v')-triangular, then fg is (u, v')-triangular.

Remark 6.13. Take $x \in [1, w]_R$ and let $i \in I$. Then, $x\,.\,\bar{\pi}_i \leq_R x\,.\,s_i$.

By repeated application, for $S \subseteq I$, and $i_1, \ldots, i_k \in S$, $x\,.\,\bar{\pi}_S \leq_R x\,.\,s_{i_1} \cdots s_{i_k}$, where recall that $\bar{\pi}_S$ is the longest element in the generators $\{\bar{\pi}_j \mid j \in S\}$.

Lemma 6.14. *Take $u \in [1, w]_R$, and define $f_{u,u} := e_{u,w_0} = \bar{\pi}_{u^{-1}} \pi_u$. Then*

(i) *$f_{u,u}$ is (u, u)-triangular*;

(ii) *for $v \in [1, w]_R$, either $v\,.\,f_{u,u} = 0$ or $v\,.\,f_{u,u} \geq_B v$*;

(iii) $\operatorname{im}(f_{u,u}) = [u, w_0]_L \cap [1, w]_R$.

Proof. First consider the case $w = w_0$. Then, (ii) and (iii) hold by Proposition 4.15.

Now take any $w \in W$. By Proposition 6.1 the action of $f \in M$ on the translation module T_w either agrees with the action on W or yields 0. Hence in particular Proposition 4.5 still applies, which yields (ii). This also implies the inclusion $\operatorname{im}(f_{u,u}) \setminus \{0\} \subset [u, w_0]_L \cap [1, w]_R$. The reverse inclusion is straightforward: If $u' = xu$, then $u' . f_{u,u} = xu . \overline{\pi}_{u^{-1}} \pi_u = x\pi_u = xu = u'$. Therefore (iii) holds as well.

Finally, (iii) implies that u is the unique minimal element of $\operatorname{im}(f_{u,u})$, and (ii) implies that u is the unique maximal element in $f_{u,u}^{-1}(u)$; therefore (i) holds. □

Lemma 6.15. *If $u >_R v$, then $f_{u,v} := f_{u,u}\overline{\pi}_{u^{-1}v}$ is (u, v)-triangular.*

Proof. By Lemma 6.14(iii), the image set of $f_{u,u}$ is a subset of $[u, w_0]_L$. Therefore, by Remark 2.7, $\overline{\pi}_{u^{-1}v}$ translates it isomorphically to the interval $[v, w_0u^{-1}v]_L$. In particular, the fibers are preserved: $f_{u,v}^{-1}(v) = f_{u,u}^{-1}(u)$, and the triangularity of $f_{u,v}$ follows. □

Lemma 6.16. *Take $u \in [1, w]_R$. Then, either u is a cutting point of w, or there exists a (u, v)-triangular function $f_{u,v}$ in F with $u <_R v \leq_R w$.*

Proof. Let J be the set of short nondescents i of u, and set $V := U_u \cap [1, w]_R$ (recall from Definition 3.15 that $U_u := uW_J$). By Proposition 3.17, V is the set of $w' \in [1, w]_R$ such that $u \sqsubseteq w'$. Furthermore, V is a lattice (it is the intersection of the two lattices $(uW_J, <_R)$ and $[1, w]_R$) with u as unique minimal element; in particular, $V \subset [u, w]_R$.

If $w \in V$ (which includes the case $u = w$ and $J = \{\}$), then u is a cutting point for w and we are done.

Otherwise, consider a shortest sequence $i_1, \dots, i_k$ such that $\{i_1, \dots, i_k\}$ does not intersect $\mathrm{D}_R(u)$, and $v' = us_{i_1} \cdots s_{i_k} \notin V$. Such a sequence must exist since $w \notin V$. Set $S := \{i_1, \dots, i_k\}$. Note that $i_1, \dots, i_{k-1}$ are in J but i_k is not. Furthermore, $u \not\sqsubseteq v'$ while $u = v'^S$ because $v' \in uW_S$ and $S \cap \mathrm{D}_R(u) = \varnothing$.

Case 1: $v' \in \operatorname{im}(f_{u,u})$. Then, $u <_L v'$. Combining this with $u = v'^S$ yields that $u \sqsubseteq v'$, a contradiction.

Case 2: $v' \notin \operatorname{im}(f_{u,u})$. Set $v := us_{i_1}$, and define $f_{u,v} := f_{u,u}\sigma\pi_{i_1}$, where

$$\sigma := s_{i_2} \cdots s_{i_{k-1}} s_{i_k} s_{i_{k-1}} \cdots s_{i_2}. \tag{6-6}$$

Note that for $k = 1$, we have $\sigma = 1$. We now prove that $f_{u,v}$ is (u, v)-triangular.

First, we consider the fiber $f_{u,v}^{-1}(v)$. By minimality of k, and up to sign, s_{i_k} fixes all the elements of V at distance at most $k-2$ of u. Hence, $\sigma^{-1}(u) = u$. Simultaneously,

$$v . \sigma^{-1} = v . s_{i_2} \cdots s_{i_{k-1}} s_{i_k} s_{i_{k-1}} \cdots s_{i_2} = v' . s_{i_{k-1}} \cdots s_{i_2} \in v' W_J. \tag{6-7}$$

Hence, $v.\sigma^{-1} \notin \mathrm{im}(f_{u,u})$ because $v' \notin \mathrm{im}(f_{u,u})$ and $\mathrm{im}(f_{u,u})$ is stable under right multiplication by s_j for $j \in J$. Putting everything together, we have

$$f_{u,v}^{-1}(v) = f_{u,u}^{-1}(\sigma^{-1}(\pi_{i_1}^{-1}(v))) = f_{u,u}^{-1}(\sigma^{-1}(\{u, v\})) = f_{u,u}^{-1}(\{u\}) = [1, u]_B \cap [1, w]_R.$$

Therefore, u is the unique length maximal element of $f_{u,v}^{-1}(v)$, as desired.

We take now $x \in \mathrm{im}(f_{u,u})$, and apply Proposition 4.5 repeatedly. To start with,

$$u = 1\,.\,f_{u,u} \leq_B x\,.\,f_{u,u}. \tag{6-8}$$

Using Remark 6.13, we have

$$u = u\,.\bar{\pi}_S \leq_B (x\,.\,f_{u,u})\,.\bar{\pi}_S \leq_B (x\,.\,f_{u,u})\,.\sigma = x\,.\,f_{u,u}\,.\sigma. \tag{6-9}$$

It follows that

$$v = u\,.\pi_{i_1} \leq_B (x\,.\,f_{u,u}\,.\sigma)\,.\pi_{i_1} = x\,.\,f_{u,v}. \tag{6-10}$$

In particular, v is the unique Bruhat minimal element of $\mathrm{im}(f_{u,v})$, as desired. □

Proof of Proposition 6.11. Since W is finite, repeated application of Lemma 6.16 yields a finite sequence of triangular functions

$$f_{u,u_1}, \ldots, f_{u_{k-1},u_k}, \quad \text{where} \quad u <_R u_1 <_R \cdots <_R u_k$$

and u_k is a cutting point w^J of w. Since $u <_R w^J$, one has $J \subset K^{(w)}(u) \subset K^{(w)}(v)$, and therefore $u_k = w^J >_R v$. Then, applying Lemma 6.15 one can construct a (u_k, v)-triangular function $f_{u_k,v}$. Finally, by Remark 6.12, composing all these triangular functions gives a (u, v)-triangular function $f_{u,u_1} \cdots f_{u_{k-1},u_k} f_{u_k,v}$. □

6d. ***Representation theory of the w-biHecke algebra.*** Consider the digraph $G^{(w)}$ on $[1, w]_R$ with an edge $u \mapsto v$ if $u = vs_i$ for some i and $J^{(w)}(u) \subseteq J^{(w)}(v)$. Up to orientation, this is the Hasse diagram of right order (see for example Figure 1). The following theorem is a generalization of [Hivert and Thiéry 2009, Section 3.3].

Theorem 6.17. *$\mathcal{H}W^{(w)}$ is the maximal algebra stabilizing all modules $P_J^{(w)}$ for $J \in \mathcal{R}\mathcal{B}_{\mathcal{L}}(w)$*

$$\mathcal{H}W^{(w)} = \{f \in \mathrm{End}(T_w) \mid f(P_J^{(w)}) \subseteq P_J^{(w)}\}$$

The elements $f_{u,v}$ of Proposition 6.11 form a basis $\mathcal{H}W^{(w)}$; in particular,

$$\dim \mathcal{H}W^{(w)} = \left|\{(u, v) \mid J^{(w)}(u) \subseteq J^{(w)}(v)\}\right|. \tag{6-11}$$

$\mathcal{H}W^{(w)}$ is the digraph algebra of the graph $G^{(w)}$.

The family $(P_J^{(w)})_{J \in \mathcal{R}\mathcal{B}_{\mathcal{L}}(w)}$ forms a complete system of representatives of the indecomposable projective modules of $\mathcal{H}W^{(w)}$.

The family $(S_J^{(w)})_{J \in \mathcal{R}\mathcal{B}_{\mathcal{L}}(w)}$ forms a complete system of representatives of the simple modules of $\mathcal{H}W^{(w)}$. The dimension of $S_J^{(w)}$ is the size of the corresponding w-nondescent class.

$\mathcal{H}W^{(w)}$ is Morita equivalent to the poset algebra of the lattice $[1, w]_{\sqsubseteq}$. In particular, its Cartan matrix is the incidence matrix and its quiver the Hasse diagram of this lattice.

Proof. From Proposition 6.11, one derives by triangularity that $\dim \mathcal{H}W^{(w)} \geq |\{(u, v) \mid K^{(w)}(u) \subseteq K^{(w)}(v)\}|$. The stability of all the subspaces $P_J^{(w)}$ imposes the converse equality. Hence, $\mathcal{H}W^{(w)}$ is exactly the subalgebra of $\operatorname{End}(T_w)$ stabilizing each $P_J^{(w)}$. The remaining statements follow straightforwardly, as in [Hivert and Thiéry 2009, Section 3.3]. See also for example [Denton et al. 2010/11, Section 3.7.4] for the Cartan matrix and quiver of a poset algebra. □

7. Representation theory of $M(W)$

In this section, we gather all results of the preceding sections in order to describe the representation theory of $M := M(W)$. The main result is Theorem 7.1, which gives the simple modules of $\mathbb{K}M$. We further relate the representation theory of $\mathbb{K}M$ to the representation theory of $\mathbb{K}M_{w_0}$. In particular, we prove that the translation modules are exactly the modules induced by the simple modules of $\mathbb{K}M_{w_0}$. We then conclude by computing some characters and the decomposition map from $\mathbb{K}M$ to $\mathbb{K}M_{w_0}$.

7a. *Simple modules.* We now study the simple modules of the biHecke monoid $\mathbb{K}M$ and also show that the translation modules are indecomposable.

Theorem 7.1. (i) *The biHecke monoid M admits $|W|$ nonisomorphic simple modules $(S_w)_{w\in W}$ (resp. projective indecomposable modules $(P_w)_{w\in W}$).*

(ii) *The simple module S_w is isomorphic to the top simple module*

$$S_{\{\}}^{(w)} = T_w \Big/ \sum_{v \sqsubset w} T_v$$

of the translation module T_w. Its dimension is given by

$$\dim S_w = \Big| [1, w]_R \setminus \bigcup_{v \sqsubset w} [1, v]_R \Big|.$$

In general, the simple quotient module $S_J^{(w)}$ of T_w is isomorphic to S_{J_w} of M.

Proof. Since M is aperiodic (Proposition 4.13), we may apply the special form of Clifford, Munn, and Ponizovskiĭ's construction of the simple modules (see Theorem 2.14). Namely, the simple modules are indexed by the regular $\mathcal{J}$- classes of M; by Corollary 4.20, there are $|W|$ of them. This yields (i), since for any finite-dimensional algebra, the simple and indecomposable projective modules share the same indexing set (see [Curtis and Reiner 1962, Corollary 54.14]).

Clifford, Munn, and Ponizovskiĭ further construct S_w as the top of the right class modules, that is, in our case, of the translation module T_w. Our explicit description of the radical of T_w as $\sum_{v \sqsubset w} T_v$ in (ii) is a straightforward application of Theorem 6.17. The dimension formula follows using Remark 6.9. □

For a direct proof that $\operatorname{rad} T_w = \sum_{v \sqsubset w} T_v$, without using Theorem 6.17, one would want to show that $\sum_{v \sqsubset w} T_v$ is exactly the annihilator of $\mathscr{J}(e_{w,w_0})$. One inclusion is easy, thanks to the following remark.

Remark 7.2. The submodule T_v is annihilated by $\mathscr{J}(e_{w,w_0}) = \mathscr{J}(\overline{\pi}_w)$.

Proof. Fix w and take v such that $v \sqsubset w$. Then $\overline{\pi}_w$ annihilates $T_v \subset T_w$. Indeed, combining $\overline{\pi}_w(w) = 1$ with Propositions 6.1 and 4.5, one obtains that $\overline{\pi}_w$ either annihilates f_u or maps it to f_1. Take now $x \in T_v$, and write $x\,.\overline{\pi}_w = \lambda f_1$. Since T_v is a submodule, λf_1 lies in T_v; however the basis elements of T_v have disjoint support and since $v \sqsubset w$ none of them are collinear to f_1. Therefore $x\,.\overline{\pi}_w = 0$. □

Type	$\lvert W\rvert$	$\lvert M_{w_0}\rvert$	$\lvert M\rvert$	$(\dim S_w)_w$	$\sum \dim S_w$
A_0	1	1	1	1	1
A_1	2	2	3	1^2	2
$A_2 = I_2(3)$	6	8	23	$1^4 2^2$	8
A_3	24	71	477	$1^8 2^4 3^4 4^6 5^2$	62
A_4	120	1646	31103	$1^{16} 2^{10} 3^8 4^{16} 5^{16} 6^6 \cdots 20^6$	770
A_5	720	118929	7505009	$1^{32} 2^{24} 3^{20} 4^{42} 5^{38} 6^{40}$ $\cdots 120^2$	13080
$B_2 = I_2(4)$	8	14	49	$1^4 2^2 3^2$	14
B_3	48	498	5455	$1^8 2^4 3^4 4^6 5^7 6^4 7^4 8^4 9^1$ $\cdot 10^2 11^2 12^2$	246
B_4	384	149622	6664553	$1^{16} 2^{10} 3^{10} 4^{14} 5^{17} 6^{16} \cdots 80^2$	6984
$G_2 = I_2(6)$	12	32	153	$1^4 2^2 3^2 4^2 5^2$	32
H_3	120	87	1039	$1^8 2^4 3^4 4^8 5^6 6^7 \cdots 36^2$	1404
$A_1 \times A_1$	4	4	9	$1^2 1^2$	4
$I_2(p)$	$2p$	p^2-p+2	$\frac{2}{3}p^3+\frac{4}{3}p+1$	$1^4 2^2 \cdots (p-1)^2$	p^2-p+2

Table 1. Statistics on the biHecke monoids $M := M(W)$ for the small Coxeter groups. In column four, $1^8 2^4 \cdots 5^2$ means that there are 8 simple modules of dimension 1, 4 of dimension 2, and so on. The sum in the last column is over w. The sequence p^2-p+2 is #A014206 in [OEIS Foundation 2012].

Example 7.3. The simple module S_{4312} is of dimension 3, with basis indexed by $\{4312, 4132, 1432\}$ (see Figure 1). The other simple modules S_{3412}, S_{4123}, and S_{1234} are of dimension 5, 3, and 1, respectively. See also Table 1.

In general, the two extreme cases are, on the one hand, when w is the maximal element of a parabolic subgroup, in which case the simple module is of dimension 1 and, on the other hand, when w is an immediate successor of 1 in the cutting poset (see Example 3.44), in which case the simple module is of dimension $|T_w| - 1$. In the other cases, one can use Theorem 3.41 to calculate the dimension of S_w by inclusion-exclusion from the sizes of the intervals $[1, {}^J w]_R$, where ${}^J w$ runs through the free sublattice at the top of the interval $[1, w]_\sqsubseteq$ of the cutting poset. Note that the sizes of the intervals in W can also be computed by a similar inclusion-exclusion (the Möbius function for right order is given by $\mu(u, w) = (-1)^k$ if the interval $[u, w]_R$ is isomorphic to some W_J with $|J| = k$, and 0 otherwise). This may open the door for some generating series manipulations to derive statistics like the sum of the dimension of the simple modules.

Corollary 7.4. *The translation module T_w is an indecomposable $\mathbb{K}M$-module, quotient of the projective module P_w of $\mathbb{K}M$.*

Proof. Direct application of Corollary 2.15 □

7b. *From $M_{w_0}(W)$ to $M(W)$.* In this section, we use our knowledge of M_{w_0} to learn more about M.

Proposition 7.5. *The translation module T_w is isomorphic to the induction to $\mathbb{K}M$ of the simple module $S_w^{w_0}$ of $\mathbb{K}M_{w_0}$.*

The proof of this proposition follows from the upcoming lemmas giving some simple conditions on a general inclusion of monoids $B \subseteq A$ under which the (regular) right class modules of $\mathbb{K}A$ are induced from those of $\mathbb{K}B$.

Lemma 7.6. *Let $B \subseteq A$ be two finite monoids and $f \in B$. If*

$$\mathbb{K}\mathcal{R}^A_<(f) = \mathbb{K}\mathcal{R}^B_<(f)A,$$

then the right class module $\mathbb{K}\mathcal{R}^A(f)$ is isomorphic to the induction from $\mathbb{K}B$ to $\mathbb{K}A$ of the right class module $\mathbb{K}\mathcal{R}^B(f)$:

$$\mathbb{K}\mathcal{R}^A(f) \cong \mathbb{K}\mathcal{R}^B(f)\uparrow_{\mathbb{K}B}^{\mathbb{K}A}.$$

Proof. Recall that for a $\mathbb{K}B$-module Y, the module $Y\uparrow_{\mathbb{K}B}^{\mathbb{K}A}$ induced by Y from $\mathbb{K}B$ to $\mathbb{K}A$ is given by $Y\uparrow_{\mathbb{K}B}^{\mathbb{K}A} := Y \otimes_{\mathbb{K}B} \mathbb{K}A$.

By construction of the right class modules (see Definition 2.12), we have the following exact sequences:

$$0 \to \mathbb{K}\mathcal{R}^B_<(f) \to \mathbb{K}fB \to \mathbb{K}\mathcal{R}^B(f) \to 0, \tag{7-1}$$

$$0 \to \mathbb{K}\mathcal{R}^A_<(f) \to \mathbb{K}fA \to \mathbb{K}\mathcal{R}^A(f) \to 0. \tag{7-2}$$

Consider now the sequence obtained by tensoring (7-1) by $\mathbb{K}A$:

$$0 \to \mathbb{K}\mathcal{R}^B_<(f) \otimes_{\mathbb{K}B} \mathbb{K}A \to \mathbb{K}fB \otimes_{\mathbb{K}B} \mathbb{K}A \to \mathbb{K}\mathcal{R}^B(f) \otimes_{\mathbb{K}B} \mathbb{K}A \to 0\,. \tag{7-3}$$

We want to prove that it is exact and isomorphic to (7-2).

First note that, since $\mathbb{K}B$ is a subalgebra of $\mathbb{K}A$, we have $b \otimes a = 1 \otimes ba$ for $b \in B$ and $a \in A$. Therefore the product map

$$\mu : \mathbb{K}fB \otimes_{\mathbb{K}B} \mathbb{K}A \to \mathbb{K}fA, \qquad fb \otimes a \to fba$$

is an isomorphism of $\mathbb{K}A$-modules.

Consider the restriction of μ to $\mathbb{K}\mathcal{R}^B_<(f) \otimes_{\mathbb{K}B} \mathbb{K}A$. Its image set is $\mathbb{K}\mathcal{R}^B_<(f)A$, which is equal to $\mathbb{K}\mathcal{R}^A_<(f)$ by hypothesis. Therefore, μ restricts to an A-module isomorphism from $\mathbb{K}\mathcal{R}^B_<(f) \otimes_{\mathbb{K}B} \mathbb{K}A$ to $\mathbb{K}\mathcal{R}^A_<(f)$. As a consequence, the following diagram is commutative, all vertical arrows being isomorphisms (for short we write here $\otimes$ for $\otimes_{\mathbb{K}B}$):

$$\begin{tikzcd}
0 \arrow[r] & \mathbb{K}\mathcal{R}^B_<(f) \otimes \mathbb{K}A \arrow[r] \arrow[d, "\mu"'] & \mathbb{K}fB \otimes \mathbb{K}A \arrow[r] \arrow[d, "\mu"'] & \mathbb{K}\mathcal{R}^B(f) \otimes \mathbb{K}A \arrow[r] \arrow[d, "\mathrm{id}"'] & 0 \\
0 \arrow[r] & \mathbb{K}\mathcal{R}^A_<(f) \arrow[r] & \mathbb{K}fA \arrow[r] & \mathbb{K}\mathcal{R}^B(f) \otimes \mathbb{K}A \arrow[r] & 0
\end{tikzcd}$$

It is a well-known fact that the functor $\cdot \otimes_{\mathbb{K}B} \mathbb{K}A$ is right exact, so that the middle and right part of the top sequence is exact. The left part of the bottom sequence is clearly exact. Therefore they are both exact sequences.

Comparing with (7-2), we obtain that

$$\mathbb{K}\mathcal{R}^A(f) \cong \mathbb{K}\mathcal{R}^B(f) \otimes_{\mathbb{K}B} \mathbb{K}A,$$

where the latter is isomorphic to $\mathbb{K}\mathcal{R}^B(f)\uparrow^{\mathbb{K}A}_{\mathbb{K}B}$ by definition. □

In the next lemma we denote by $<_{\mathcal{R}^A}$ the strict right preorder on a monoid A; that is, $x <_{\mathcal{R}^A} y$ if $x \leq_{\mathcal{R}^A} y$ but $x \notin \mathcal{R}^A(y)$.

Lemma 7.7. *Let $B \subseteq A$ be two finite monoids and assume that*:

(i) *$\mathcal{R}$-order on B is induced by $\mathcal{R}$-order on A*; *that is, for all $x, y \in B$,*

$$x <_{\mathcal{R}^A} y \quad\Longleftrightarrow\quad x <_{\mathcal{R}^B} y.$$

(ii) *Any $\mathcal{R}$-class of A intersects B.*

Then, for any $f \in B$, the equality $\mathbb{K}\mathcal{R}^B_<(f)A = \mathbb{K}\mathcal{R}^A_<(f)$ holds. In particular,

$$\mathbb{K}\mathcal{R}^A(f) \cong \mathbb{K}\mathcal{R}^B(f)\uparrow^{\mathbb{K}A}_{\mathbb{K}B}.$$

Moreover, condition (i) may be replaced by the stronger condition

(i′) $$x \leq_{\mathcal{R}^A} y \iff x \leq_{\mathcal{R}^B} y.$$

Proof. Inclusion $\subseteq$: Take $b \in B$ with $b <_{\mathcal{R}^B} f$ and $a \in A$. Then, using (i), we have $ba \in \mathbb{K}\mathcal{R}^A_<(f)$, since

$$ba \leq_{\mathcal{R}^A} b <_{\mathcal{R}^A} f.$$

Inclusion $\supseteq$: Take $a \in A$ with $a <_{\mathcal{R}^A} f$. Using (ii) choose an element $b \in B$ such that $b \,\mathcal{R}^A\, a$. Then $b \leq_{\mathcal{R}^A} a <_{\mathcal{R}^A} f$ and therefore, by (i), $b \in \mathbb{K}\mathcal{R}^B_<(f)$. It follows that $a \in \mathbb{K}\mathcal{R}^B_<(f)A$.

The statement $\mathbb{K}\mathcal{R}^A(f) \cong \mathbb{K}\mathcal{R}^B(f)\uparrow^{\mathbb{K}A}_{\mathbb{K}B}$ follows from Lemma 7.6. □

Here is an example of what can go wrong when Condition (i) fails.

Example 7.8. Let A be the (multiplicative) submonoid of $M_2(\mathbb{Z})$ with elements given by the matrices

$$1 := \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right),\ b_{11} := \left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right),\ b_{12} := \left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right),\ a_{21} := \left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right),\ b_{22} := \left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right),\ 0 := \left(\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right).$$

Alternatively, A is the aperiodic Rees matrix monoids (see Definition 2.16) whose nontrivial $\mathcal{J}$-class is described by

$$\begin{pmatrix} b_{11}^* & b_{12} \\ a_{21} & b_{22}^* \end{pmatrix},$$

where the * marks the elements that are idempotent. In other words, $A = M(P)$, where $P := \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, and for convenience the matrix above specifies names for the elements of the nontrivial $\mathcal{J}$-class. Recall that the nontrivial left and right classes of A are given respectively by the columns and rows of this matrix.

Let B be the submonoid $\{1, b_{11}, b_{12}, b_{22}, 0\}$. Then B satisfies condition (ii) but not condition (i): indeed $b_{11} \,\mathcal{R}^A\, b_{12}$ whereas $b_{11} <_{\mathcal{R}^B} b_{12}$. Then, taking $f = b_{11}$, one obtains $\mathcal{R}^B_<(b_{11}) = \{0, b_{12}\}$ so that $\mathcal{R}^B_<(b_{11})A = \{0, b_{11}, b_{12}\}$, and therefore

$$\mathbb{K}\{0\} = \mathbb{K}\mathcal{R}^A_<(b_{11}) \subset \mathbb{K}\mathcal{R}^B_<(b_{11})A = \mathbb{K}\{0, b_{11}, b_{12}\}.$$

Now $\mathbb{K}\mathcal{R}^B(b_{11}) = \mathbb{K}\{0, b_{11}, b_{12}\}/\mathbb{K}\{0, b_{12}\}$, so that $\mathbb{K}\mathcal{R}^B(b_{11})$ is one-dimensional, spanned by $x := b_{11} \mod (\mathbb{K}\{0, b_{12}\})$. The action of B is given by $x\,.\,1 = x\,.\,b_{11} = x$ and $x\,.\,m = 0$ for any $m \in B \setminus \{1, b_{11}\}$.

We claim that

$$\mathbb{K}\mathcal{R}^B(b_{11})\uparrow^{\mathbb{K}A}_{\mathbb{K}B} = \mathbb{K}\mathcal{R}^B(b_{11}) \otimes_{\mathbb{K}B} \mathbb{K}A = 0.$$

Indeed, $x \otimes 1 = x\,.b_{11} \otimes 1 = x \otimes b_{11} = x \otimes b_{12}a_{21} = x\,.b_{12} \otimes a_{21} = 0$. Thus

$$\mathbb{K}\mathcal{R}^A(b_{11}) \ncong \mathbb{K}\mathcal{R}^B(b_{11})\uparrow_{\mathbb{K}B}^{\mathbb{K}A}.$$

As shown in the following example, Condition (i′) may be strictly stronger than Condition (i) because $<_{\mathcal{R}}$ is only a preorder.

Example 7.9. Let A be the aperiodic Rees matrix monoid with nontrivial $\mathcal{J}$-class given by

$$\begin{pmatrix} a_{11}^* & b_{12} & b_{13} \\ a_{21}^* & b_{22}^* & a_{23} \\ a_{31}^* & a_{32} & b_{33}^* \end{pmatrix},$$

Let B be the submonoid $\{1, b_{12}, b_{13}, b_{22}, b_{33}, 0\}$. Then B satisfies conditions (i) and (ii), but not condition (i′): b_{12} and b_{13} are incomparable for $\leq_{\mathcal{R}^B}$ whereas they are in the same right class for A.

We now turn to the proof of Proposition 7.5 by showing that $M_{w_0}(W) \subseteq M(W)$ satisfy the conditions of Lemma 7.7. We use the stronger condition (i′).

Lemma 7.10. *The biHecke monoid and its Borel submonoid $M_{w_0}(W) \subseteq M(W)$ satisfy conditions* (i′) and (ii) *of Lemma 7.7.*

Proof. By Proposition 4.18, for any $f \in M$ there exists a unique $f_1 \in \mathcal{R}(f) \cap M_1$. Using the bar involution of Section 4g, one finds similarly a unique $\overline{f_1} \in \mathcal{R}(f) \cap M_{w_0}$. This proves condition (ii).

We now prove the nontrivial implication in condition (i′). Take $f, g \in M_{w_0}$ with $f \leq_{\mathcal{R}^M} g$. Then, $f = gx$ for some $x \in M$. Note that $w_0\,.\,f = w_0\,.\,g = w_0$, which implies that $w_0\,.x = w_0$ as well. Hence x is in fact in M_{w_0} and $f \leq_{\mathcal{R}^{M_{w_0}}} g$. □

Proof of Proposition 7.5. Let $g_w := e_{w,w_0}$. By definition, the translation module is the quotient $T_w = \mathbb{K}g_w M/\mathbb{K}\mathcal{R}_<(g_w)$, whereas $S_w^{w_0} = \mathbb{K}g_w M_{w_0}/\mathbb{K}\mathcal{R}_<^{w_0}(g_w)$. By Lemma 7.10, $M_{w_0} \subseteq M$ satisfy the two conditions of Lemma 7.7; Proposition 7.5 follows. □

Theorem 7.11. *The right regular representation of $\mathbb{K}M$ admits a filtration with factors all isomorphic to translation modules, and its character is given by*

$$[\mathbb{K}M] = \sum_{f \in M_{w_0}} [T_{1\,.\,f}]. \tag{7-4}$$

Proof. As any monoid algebra, $\mathbb{K}M$ admits a filtration where each composition factor is given by (the linear span of) an $\mathcal{R}$-class of M. By Proposition 6.2, each such composition factor is isomorphic to the translation module $T_{1\,.\,f}$, where f is the unique element of the $\mathcal{R}$-class that lies in M_{w_0}. The character formula follows.

Alternatively, it can be obtained using Proposition 7.5 and the character formula for the right regular representation of M_{w_0} (see Remark 5.11):

$$[\mathbb{K}M_{w_0}]_{M_{w_0}} = \sum_{f \in M_{w_0}} [S^{w_0}_{1.f}]_{M_{w_0}}, \tag{7-5}$$

which completes the proof. $\square$

Proposition 7.12. *For any $w \in W$, the translation module T_w is multiplicity-free as an $\mathbb{K}M_{w_0}$-module and its character is given by*

$$[T_w]_{M_{w_0}} = \sum_{u \in [1,w]_R} [S^{w_0}_u]_{M_{w_0}}. \tag{7-6}$$

Proof. Let f be an element in M that yields the translation module T_w, and define f_u as in Proposition 4.18.

Take some sequence $u_1, \dots, u_m$ (for $m = |[1, w]_R|$) of the elements of $[1, w]_R$ that is length increasing, and define the corresponding sequence of subspaces by $X_i := \mathbb{K}\{u_1, \dots, u_i\}$. Using Lemma 6.14, each such subspace is stable by M_{w_0}, and $X_0 \subset \cdots \subset X_m$ forms an M_{w_0}-composition series of T_w since X_i/X_{i-1} is of dimension 1.

Consider now a composition factor X_i/X_{i-1}. Again, by Lemma 6.14, e_{v,w_0} fixes u_i if and only if $v \leq_L u_i$ (that is, if the image set $[u_i, w^{-1}w_0u_i]_L$ of f_{u_i} is contained in the image set $[v, w_0]_L$ of e_{v,w_0}), and kills it otherwise. Hence, X_i/X_{i-1} is isomorphic to $S^{w_0}_{u_i}$. $\square$

Theorem 7.13. *The decomposition map of $\mathbb{K}M$ over $\mathbb{K}M_{w_0}$ is lower uni-triangular for right order, with 0, 1 entries. More explicitly,*

$$[S_w]_{M_{w_0}} = \sum_{u \in [1,w]_R \setminus \bigcup_{v \sqsubset w} [1,v]_R} [S^{w_0}_u]_{M_{w_0}}. \tag{7-7}$$

Proof. Since S_w is a quotient of T_w, its composition factors form a subset of the composition factors for T_w. Hence, using Proposition 7.12, the decomposition matrix of M over M_{w_0} is lower triangular for right order, with 0, 1 entries. Furthermore, by construction (see Remark 6.9 and Theorem 7.1(ii)), $S_w = T_w / \sum_{v \sqsubset w} T_v$; using Proposition 7.12 the sum on the right hand side contains at least one composition factor isomorphic to $S^{w_0}_u$ for each u in $[1, v]_R$ with $v \sqsubset w$; therefore S_w has no such composition factor. We conclude using the dimension formula of Theorem 7.1(ii). $\square$

Example 7.14. Following up on Example 7.3, the decomposition of the $\mathbb{K}M$-simple module S_{4312} over $\mathbb{K}M_{w_0}$ is given by $[S_{4312}]_{M_{w_0}} = [S^{w_0}_{4312}] + [S^{w_0}_{4132}] + [S^{w_0}_{1432}]$. See also Figure 1 and the decomposition matrices given in Appendix A.

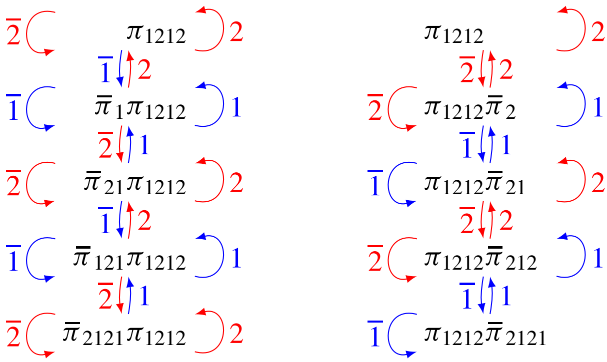


Figure 6. The left and right class modules indexed by $w := s_1s_2s_1s_2$ for the biHecke monoid $M(I_p)$ with $p \geq 5$. The left picture also describes the left simple module S_w of $M(I_p)$, and the projective module $P_w^{w_0}$ of the Borel submonoid $M_{w_0}(I_p)$.

7c. ***Example: the rank 2 Coxeter groups.*** We now give a complete description of the representation theory of the biHecke monoid for each rank 2 Coxeter group I_p. The proofs are left as exercises for the reader.

Example 7.15. Let M be the biHecke monoid for the dihedral group $W := I_p$ of order $2p$. Then, M is a regular monoid.

The right class module $\mathbb{K}\mathcal{R}_w := \mathbb{K}\mathcal{R}(e_{w,w_0})$ is the translation module spanned by $[1, w]_R$. It is of dimension $2p$ for $w = w_0$, and $\ell(w)$ otherwise. The left class modules $\mathbb{K}\mathcal{L}_1$ and $\mathbb{K}\mathcal{L}_{w_0}$ are respectively the trivial module spanned by 1 and the zero module spanned by w_0. For $w \neq 1, w_0$, the left class module $\mathbb{K}\mathcal{L}_w$ is of dimension $\ell(w) - 1$, and its structure is as in Figure 6. In particular,

$$|M| = 2p + 1 + 2\sum_{k=1}^{p-1} k(k+1) = \tfrac{2}{3}p^3 + \tfrac{4}{3}p + 1.$$

The simple right module S_w can be constructed from the cutting poset. Namely, S_1 is the trivial module spanned by 1, while S_{w_0} is the zero module spanned by w_0 and, for $w \neq 1, w_0$, S_w is the quotient of the right class module by the line spanned by alternating sum of $[1, w]_R$. The simple left module S_w is directly given by the left class module L_w.

The quiver of M is given by the cutting poset (see Figure 7). The q-Cartan matrix is given by the path algebra of this quiver; namely, there is an extra arrow from 1 to w_0 with weight q^2. In particular, it is upper unitriangular and of determinant 1.

Example 7.16. Let M_{w_0} be the Borel submonoid of the biHecke monoid for the dihedral group $W := I_p$ of order $2p$.

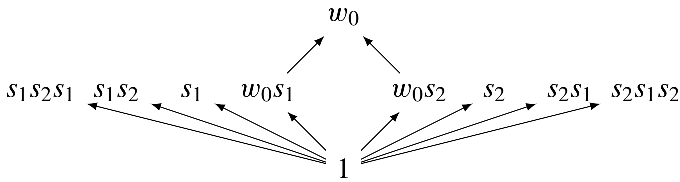


Figure 7. The Hasse diagram of the cutting poset for the dihedral group $W := I_5$. This is also the quiver of the biHecke monoid for that group.

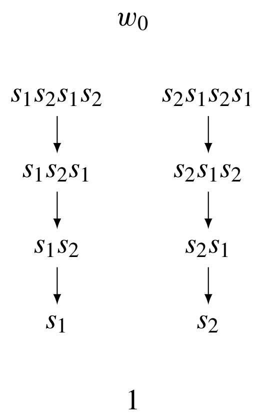


Figure 8. The quiver of the Borel submonoid $M_{w_0}(I_5)$ of the biHecke monoid for the dihedral group I_5.

The projective module P_w of M_{w_0} is given by the left simple modules S_w, or equivalently the left-class-module L_w of M. In particular,

$$|M_{w_0}| = 1 + 1 + 2\sum_{k=1}^{p-1} k = p^2 - p + 2.$$

The quiver of M_{w_0} is given by the cover relations in Bruhat order (or equivalently right order) that are not covers in left order (see Figure 8); this gives two chains of length $p-1$. The monoid algebra is isomorphic to the path algebra of this quiver, which gives right away its radical filtration. Combinatorially speaking, every non-idempotent element f of the monoid admits a unique minimal factorization $e_w e_u$, with $\ell(u) < \ell(w)$ and $u \not\leq_L w$; namely, $u := f(1)$ and w is the smallest element such that $f(w) = 1$.

8. Research in progress

Our guiding problem is the search for a formula for the cardinality of the biHecke monoid. Using a standard result of the representation theory of finite-dimensional algebras together with the results of this paper, we can now write

$$|M(W)| = \sum_{w \in W} \dim S_w \dim P_w,$$

where $\dim S_w$ is given by an inclusion-exclusion formula. It remains to determine the dimensions of the projective modules P_w.

While studying the representation theory of the Borel submonoid M_1 as an intermediate step, the authors realized that many of the combinatorial ingredients that arose were well-known in the semigroup community (for example the Green's relations and related classes, automorphism groups, etc.), and hence the representation theory of M_1 is naturally expressed in the context of $\mathscr{J}$-trivial monoids; see [Denton et al. 2010/11]. This sparked their interest in the representation theory of more general classes of monoids, in particular aperiodic monoids.

At the current stage, it appears that the Cartan matrix of an aperiodic monoid (and therefore the composition series of its projective modules, and by consequence their dimensions) is completely determined by the knowledge of the composition series for both left and right class modules. In other words, the study in this paper of right class modules (that is, translation modules), whose original purpose was to construct the simple modules using [Ganyushkin et al. 2009, Theorem 7], turns out to complete half of this program. The remaining half, in progress, is the decomposition of left class modules.

At the combinatorial level, this requires one to control $\mathscr{L}$-order. Loosely speaking, $\mathscr{L}$-order is essentially given by left and right order in W; however, within $\mathscr{L}$-classes the structure seems more elusive, in particular because fibers are more difficult to describe than image sets. Another difficulty is that, unlike for $\mathscr{R}$-class modules, $\mathscr{L}$-class modules are not all isomorphic to regular ones (that is, classes containing idempotents).

Yet, the general theory gives that the decomposition matrix should be upper triangular for left-right order for regular classes, and upper triangular for Bruhat order for nonregular ones, with no left-right "arrow" for left-right order. Pushing this further gives that the Cartan matrix has determinant 1.

We conclude by illustrating the above for $W = \mathfrak{S}_4$ in Figure 9. The blue arrows are the covering relations of the cutting poset, which encode the composition series of the translation modules (that is, right class modules). Namely, the character of T_w is given by the sum of $q^k[S_u]$ for u below w in the cutting poset, with k the

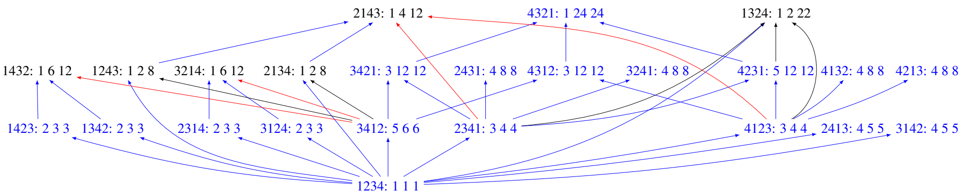


Figure 9. Graph encoding the characters of left and right class modules, and therefore the Cartan invariant matrix for $M(\mathfrak{S}_4)$. See the text for details.

distance from u to w in that poset. For example,

$$\begin{aligned}
[T_{2143}] &= [S_{2143}] + q[S_{1243}] + q[S_{2134}] + q^2[S_{1234}],\\
[T_{2341}] &= [S_{2341}] + q[S_{1234}],\\
[T_{4123}] &= [S_{4123}] + q[S_{4123}].
\end{aligned}$$

Similarly the black and red arrows encode the composition series of regular and, respectively, nonregular left classes. In this simple example, the q-character of a right projective module P_w is then given by

$$[P_w] = [T_w] + \sum_u q[T_u],$$

where (u, w) is a black or red arrow in the graph. For example,

$$\begin{aligned}
[P_{2143}] &= [T_{2143}] + q[T_{2341}] + q[T_{4123}]\\
&= [S_{2143}] + q[S_{1243}] + q[S_{2134}] + q[S_{2341}] + q[S_{4123}] + 3q^2[S_{1234}].
\end{aligned}$$

Acknowledgments

A short version with the announcements of (some of) the results of this paper was published in [Hivert et al. 2010]. The discovery and analysis of the cutting poset is due to Schilling and Thiéry. The study of the biHecke monoid and of its representation theory is joint work of all three authors, under an initial suggestion of Alain Lascoux.

We would like to thank Jason Bandlow, Mahir Can, Brant Jones, Jean-Christophe Novelli, Jean-Éric Pin, Nathan Reading, Vic Reiner, Franco Saliola, Benjamin Steinberg, and Jean-Yves Thibon for enlightening discussions. In particular, Saliola pointed us to the large body of literature on semigroups, and the recent progress in their representation theory. Special thanks to the anonymous referee for carefully reading the paper, suggesting many improvements, and many insightful comments. We are grateful to Aladin Virmaux for pointing out several misprints and helping to work out the details of Examples 7.15 and 7.16.

This research was driven by computer exploration, using the open-source mathematical software `Sage` [Stein et al. 2009] and its algebraic combinatorics features developed by the `Sage-Combinat` community [Sage-Combinat 2008].

Schilling would like to thank the Université Paris Sud, Orsay and the Kastler foundation for support during her visit in November 2008. Thiéry would like to thank the Department of Mathematics at UC Davis for hospitality during his visit in 2007–2008 and spring 2009.

Appendix A. Monoid of edge surjective morphism of a colored graph

Let C be a set whose elements are called colors. We consider colored simple digraphs without loops. More precisely, a C-colored graph is a triple $G=(V, E, c)$, where V is the set of vertices of G, $E\subset V\times V/\{(x, x)\mid x\in V\}$ is the set of (oriented) edges of G, and $c: E\to C$ is the coloring map.

Definition A.1. Let $G=(V, E, c)$ and $G'=(V', E', c')$ be two colored graphs. An *edge surjective* morphism (or ES-morphism) from G to G' is a map $f: V\to V'$ such that

- For any edge $(a, b)\in E$, either $f(a)=f(b)$, or $(f(a), f(b))\in E'$ and $c(a, b)=c'(f(a), f(b))$.
- For any edge $(a', b')\in E'$ with a' and b' in the image set of f there exists an edge $(a, b)\in E$ such that $f(a)=a'$ and $f(b)=b'$.

Note that by analogy to categories, instead of ES-morphism, we can speak about full morphisms.

The following proposition shows that colored graphs together with edge surjective morphisms form a category.

Proposition A.2. *For any colored graphs G, G_1, G_2, G_3,*

- *the identity* $\mathrm{id}: G\to G$ *is an ES-morphism*;
- *for any ES-morphism $f: G_1\to G_2$ and $g: G_2\to G_3$ the composed function $g\circ f: G_1\to G_3$ is an ES-morphism.*

Corollary A.3. *For any colored graph G, the set of ES-morphisms from G to G is a submonoid of the monoid of the functions from G to G.*

Here are some general properties of ES-morphisms:

Proposition A.4. *Let G_1 and G_2 be two colored graphs and f an ES-morphism from G_1 to G_2. Then the image of any path in G_1 is a path in G_2.*

In our particular case, we have some more properties:

(i) The graph is acyclic, with unique source and sink. In particular, it is (weakly) connected.

	12	21	Proj.
12	1	.	1
21	.	1	1
Proj.	1	1	

Table 2. q-Cartan invariant matrix of $M_{w_0}(\mathfrak{S}_2)$ (type A_1).

	123	132	213	231	312	321	Proj.
123	1	.	.	.	.	.	1
132	.	1	.	.	q	.	2
213	.	.	1	q	.	.	2
231	.	.	.	1	.	.	1
312	.	.	.	.	1	.	1
321	.	.	.	.	.	1	1
Proj.	1	1	1	2	2	1	

Table 3. q-Cartan invariant matrix of $M_{w_0}(\mathfrak{S}_3)$ (type A_2).

(ii) The graph is ranked by the integers, and edges occur only between two consecutive ranks.

(iii) The graph is C-regular, which means that for any vertex v

Remarks A.5. Proposition 4.1 gives that our monoid is a submonoid of the $M(G)$ monoid for left order.

Propositions 4.3 and 4.11 are generic, and would apply to any $M(G)$. For the latter, we just need that G is C-regular.

A natural source of colored graphs are crystal graphs. A question that arises is what the G-monoid of a crystal looks like.

Appendix B. Tables

B1. ***q*-*Cartan invariant matrices.*** In Tables 2–7, we give the Cartan invariant matrix for $\mathbb{K}M_{w_0}$ and $\mathbb{K}M$ in types A_1, A_2 and A_3. The q-parameter records the layer in the radical filtration. The extra rows and columns entitled “Simp.” and “Proj.” give the dimension of the simple and projective modules, on the right for right modules and below for left modules. When all simple modules are one-dimensional, the column is omitted.

Using [Thiéry 2012], it is possible to go further, and compute for example the Cartan invariant matrix for M in type A_4 in about one hour (though at $q = 1$ only).

	1234	1243	1324	1342	1423	1432	2134	2143	2314	2341	2413	2431	3124	3142	3214	3241	3412	3421	4123	4132	4213	4231	4312	4321	Proj.
1234	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
1243	.	1	.	.	q	.	.	.	.	.	q	q^2	.	.	.	.	q	.	q^2	.	.	.	.	.	6
1324	.	.	1	q	.	.	.	.	.	q	.	.	q	q^2	.	q^2	q^3	.	q	q^2	.	q	.	.	10
1342	.	.	.	1	.	.	.	.	.	.	.	.	.	q	.	.	q^2	.	.	q	.	.	.	.	4
1423	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	q	.	.	.	.	.	2
1432	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	q	.	.	q	.	.	q^2	.	4
2134	.	.	.	.	.	.	1	.	q	q^2	q	.	.	.	.	.	q	.	.	.	q^2	.	.	.	6
2143	.	.	.	.	.	.	.	1	.	q	q	q^2	.	.	.	.	.	.	q	.	q^2	q^3	.	.	7
2314	.	.	.	.	.	.	.	.	1	q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
2341	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
2413	.	.	.	.	.	.	.	.	.	.	1	q	.	.	.	.	.	.	.	.	q	q^2	.	.	4
2431	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	q	.	.	2
3124	.	.	.	.	.	.	.	.	.	.	.	.	1	q	.	q	q^2	.	.	.	.	.	.	.	4
3142	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	q	.	.	.	.	.	.	.	2
3214	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	q	q	q^2	.	.	.	.	.	.	4
3241	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	q	.	.	.	.	.	.	2
3412	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1
3421	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1
4123	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1
4132	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	q	.	2
4213	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	q	.	.	2
4231	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1
4312	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1
4321	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
Proj.	1	1	1	2	2	1	1	1	2	5	4	4	2	4	1	4	9	3	5	4	4	6	3	1	

Table 4. q-Cartan invariant matrix of $M_{w_0}(\mathfrak{S}_4)$ (type A_3).

	12	21	Simp.	Proj.
12	1	.	1	1
21	q	1	1	2
Simp.	1	1		
Proj.	2	1		

Table 5. q-Cartan invariant matrix of $M(\mathfrak{S}_2)$ (type A_1).

B2. ***Decomposition matrices.*** Since M_{w_0} is a submonoid of M, any simple M-module is also a simple M_{w_0}-module. The matrices of Tables 8–10 give the (generalized) M_{w_0} character of the simple M-module. The table reads as follows: for any two permutations σ, τ, the coefficient $m_{\sigma,\tau}$ gives the Jordan–Hölder multiplicity of the M_{w_0}-module $S_\tau^{w_0}$ in the M-module S_σ. In particular, since the simple M_{w_0}-modules are of dimension 1, summing each line one recovers the dimension of the simple M-modules.

	123	132	213	231	312	321	Simp.	Proj.
123	1	.	.	.	.	.	1	1
132	q	1	.	.	.	.	1	2
213	q	.	1	.	.	.	1	2
231	q	.	.	1	.	.	2	3
312	q	.	.	.	1	.	2	3
321	q^2	.	.	q	q	1	1	6
Simp.	1	1	1	2	2	1		
Proj.	8	1	1	3	3	1		

Table 6. q-Cartan invariant matrix of $M(\mathfrak{S}_3)$ (type A_2).

	1234	1243	1324	1342	1423	1432	2134	2143	2314	2341	2413	2431	3124	3142	3214	3241	3412	3421	4123	4132	4213	4231	4312	4321	Simp.	Proj.
1234	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
1243	q^2+q	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	q	.	.	.	.	.	.	.	1	8
1324	q^3+2q^2+q	.	1	.	.	.	.	.	.	q^2+q	.	.	.	.	.	.	.	.	q^2+q	.	.	q	.	.	1	22
1342	q	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	3
1423	q	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	3
1432	$2q^2$	.	.	q	q	1	.	.	.	.	.	.	.	.	.	.	q	.	.	.	.	.	.	.	1	12
2134	q^2+q	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	q	.	.	.	.	.	.	.	1	8
2143	$3q^2$	q	.	.	.	.	q	1	.	q	.	.	.	.	.	.	.	.	q	.	.	.	.	.	1	12
2314	q	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	3
2341	q	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	4
2413	q	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	4	5
2431	q^2	.	.	.	.	.	.	.	.	q	.	1	.	.	.	.	.	.	.	.	.	.	.	.	4	8
3124	q	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	2	3
3142	q	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	4	5
3214	$2q^2$	.	.	.	.	.	.	.	q	.	.	.	q	.	1	.	q	.	.	.	.	.	.	.	1	12
3241	q^2	.	.	.	.	.	.	.	.	q	.	.	.	.	.	1	.	.	.	.	.	.	.	.	4	8
3412	q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	5	6
3421	q^2	.	.	.	.	.	.	.	.	q	.	.	.	.	.	.	q	1	.	.	.	.	.	.	3	12
4123	q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	3	4
4132	q^2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	q	1	.	.	.	.	4	8
4213	q^2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	q	.	1	.	.	.	4	8
4231	q^2	.	.	.	.	.	.	.	.	q	.	.	.	.	.	.	.	.	q	.	.	1	.	.	5	12
4312	q^2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	q	.	q	.	.	.	1	.	3	12
4321	q^3	.	.	.	.	.	.	.	.	q^2	.	.	.	.	.	.	q^2	q	q^2	.	.	q	q	1	1	24
Simp.	1	1	1	2	2	1	1	1	2	3	4	4	2	4	1	4	5	3	3	4	4	5	3	1		
Proj.	71	2	1	3	3	1	2	1	3	23	4	4	3	4	1	4	16	4	23	4	4	7	4	1		

Table 7. q-Cartan invariant matrix of $M(\mathfrak{S}_4)$ (type A_3).

References

[Albert and Atkinson 2005] M. H. Albert and M. D. Atkinson, "Simple permutations and pattern restricted permutations", *Discrete Math.* **300**:1-3 (2005), 1–15. MR 2006d:05007 Zbl 1073.05002

	12	21	Simp.
12	1	.	1
21	.	1	1

Table 8. Decomposition matrix of $M(\mathfrak{S}_2)$ on $M_{w_0}(\mathfrak{S}_2)$ (type A_1).

	123	132	213	231	312	321	Simp.
123	1	.	.	.	.	.	1
132	.	1	.	.	.	.	1
213	.	.	1	.	.	.	1
231	.	.	1	1	.	.	2
312	.	1	.	.	1	.	2
321	.	.	.	.	.	1	1

Table 9. Decomposition matrix of $M(\mathfrak{S}_3)$ on $M_{w_0}(\mathfrak{S}_3)$ (type A_2).

	1234	1243	1324	1342	1423	1432	2134	2143	2314	2341	2413	2431	3124	3142	3214	3241	3412	3421	4123	4132	4213	4231	4312	4321	Simp.
1234	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
1243	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
1324	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
1342	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
1423	.	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
1432	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
2134	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
2143	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
2314	.	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
2341	.	.	.	.	.	.	1	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3
2413	.	1	.	.	.	.	1	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	4
2431	.	1	.	.	.	.	.	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	4
3124	.	.	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	2
3142	.	.	1	1	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	4
3214	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1
3241	.	.	1	.	.	.	.	.	.	.	.	.	1	.	1	1	.	.	.	.	.	.	.	.	4
3412	.	.	1	1	.	.	.	.	.	.	.	.	1	1	.	.	1	.	.	.	.	.	.	.	5
3421	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	1	.	.	.	.	.	.	3
4123	.	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	3
4132	.	.	1	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	4
4213	.	.	.	.	.	.	1	1	.	.	1	.	.	.	.	.	.	.	.	.	1	.	.	.	4
4231	.	.	.	.	.	.	.	1	.	.	1	1	.	.	.	.	.	.	.	.	1	1	.	.	5
4312	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	3
4321	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1

Table 10. Decomposition matrix of $M(\mathfrak{S}_4)$ on $M_{w_0}(\mathfrak{S}_4)$ (type A_3).

[Albert et al. 2003] M. H. Albert, M. D. Atkinson, and M. Klazar, "The enumeration of simple permutations", *J. Integer Seq.* **6**:4 (2003), article 03.4.4. MR 2051958 Zbl 1065.05001

[Almeida et al. 2009] J. Almeida, S. Margolis, B. Steinberg, and M. Volkov, "Representation theory of finite semigroups, semigroup radicals and formal language theory", *Trans. Amer. Math. Soc.* **361**:3 (2009), 1429–1461. MR 2010b:20116 Zbl 1185.20058

[Benson 1991] D. J. Benson, *Representations and cohomology, I: Basic representation theory of finite groups and associative algebras*, Cambridge Studies in Advanced Mathematics **30**, Cambridge University Press, 1991. MR 92m:20005 Zbl 0718.20001

[Birkhoff 1937] G. Birkhoff, "Rings of sets", *Duke Math. J.* **3**:3 (1937), 443–454. MR 1546000 Zbl 0017.19403

[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

[Blass and Sagan 1997] A. Blass and B. E. Sagan, "Möbius functions of lattices", *Adv. Math.* **127**:1 (1997), 94–123. MR 98c:06001 Zbl 0872.06004

[Brown 2000] K. S. Brown, "Semigroups, rings, and Markov chains", *J. Theoret. Probab.* **13**:3 (2000), 871–938. MR 2001e:60141 Zbl 0980.60014

[Carter 1986] R. W. Carter, "Representation theory of the 0-Hecke algebra", *J. Algebra* **104**:1 (1986), 89–103. MR 88a:20014 Zbl 0624.20007

[Clifford 1942] A. H. Clifford, "Matrix representations of completely simple semigroups", *Amer. J. Math.* **64** (1942), 327–342. MR 4,4a Zbl 0061.02404

[Clifford and Preston 1961] A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, vol. 1, Mathematical Surveys **7**, American Mathematical Society, Providence, RI, 1961. MR 24 #A2627 Zbl 0111.03403

[Curtis and Reiner 1962] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Pure and Applied Mathematics **11**, Interscience Publishers, New York, 1962. MR 26 #2519 Zbl 0131.25601

[Denton 2010] T. Denton, "A combinatorial formula for orthogonal idempotents in the 0-Hecke algebra of S_N", pp. 701–711 in *22nd international conference on formal power series and algebraic combinatorics* (San Francisco, CA, 2010), Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010. MR 2012m:05418

[Denton 2011] T. Denton, "A combinatorial formula for orthogonal idempotents in the 0-Hecke algebra of the symmetric group", *Electron. J. Combin.* **18**:1 (2011), paper 28. MR 2012d:20009 Zbl 1214.20003

[Denton et al. 2010/11] T. Denton, F. Hivert, A. Schilling, and N. M. Thiéry, "On the representation theory of finite $\mathcal{J}$-trivial monoids", *Sém. Lothar. Combin.* **64** (2010/11), art. B64d. MR 2012f:20184 Zbl 06004424

[Doyen 1984] J. Doyen, "Équipotence et unicité de systèmes générateurs minimaux dans certains monoïdes", *Semigroup Forum* **28**:1-3 (1984), 341–346. MR 85g:20081 Zbl 0524.20033

[Doyen 1991] J. Doyen, "Quelques propriétés des systèmes générateurs minimaux des monoïdes", *Semigroup Forum* **42**:3 (1991), 333–343. MR 92b:20073 Zbl 0727.20038

[Duchamp et al. 2002] G. Duchamp, F. Hivert, and J.-Y. Thibon, "Noncommutative symmetric functions, VI: Free quasi-symmetric functions and related algebras", *Internat. J. Algebra Comput.* **12**:5 (2002), 671–717. MR 2003j:05126 Zbl 1027.05107

[Edelman 1986] P. H. Edelman, "Abstract convexity and meet-distributive lattices", pp. 127–150 in *Combinatorics and ordered sets* (Arcata, CA, 1985), edited by I. Rival, Contemp. Math. **57**, American Mathematical Society, Providence, RI, 1986. MR 87m:52003 Zbl 0596.52003

[Ganyushkin et al. 2009] O. Ganyushkin, V. Mazorchuk, and B. Steinberg, "On the irreducible representations of a finite semigroup", *Proc. Amer. Math. Soc.* **137**:11 (2009), 3585–3592. MR 2010h:20150 Zbl 1184.20054

[Gaucher 2010] P. Gaucher, "Combinatorics of labelling in higher-dimensional automata", *Theoret. Comput. Sci.* **411**:11-13 (2010), 1452–1483. MR 2011c:18007 Zbl 1191.68384

[Gessel 1984] I. M. Gessel, "Multipartite P-partitions and inner products of skew Schur functions", pp. 289–317 in *Combinatorics and algebra* (Boulder, CO, 1983), edited by C. Greene, Contemp. Math. **34**, American Mathematical Society, Providence, RI, 1984. MR 86k:05007 Zbl 0562.05007

[Green 1951] J. A. Green, "On the structure of semigroups", *Ann. of Math.* (2) **54** (1951), 163–172. MR 13,100d Zbl 0043.25601

[Hivert and Thiéry 2009] F. Hivert and N. M. Thiéry, "The Hecke group algebra of a Coxeter group and its representation theory", *J. Algebra* **321**:8 (2009), 2230–2258. MR 2010a:20010 Zbl 1211.20006

[Hivert et al. 2009] F. Hivert, A. Schilling, and N. M. Thiéry, "Hecke group algebras as quotients of affine Hecke algebras at level 0", *J. Combin. Theory Ser. A* **116**:4 (2009), 844–863. MR 2010a:20011 Zbl 1185.20004

[Hivert et al. 2010] F. Hivert, A. Schilling, and N. M. Thiéry, "The biHecke monoid of a finite Coxeter group", pp. 307–318 in *22nd international conference on formal power series and algebraic combinatorics* (San Francisco, CA, 2010), Discrete Math. Theor. Comput. Sci. Proc., Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010. MR 2012k:05452

[Humphreys 1990] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge Studies in Advanced Mathematics **29**, Cambridge University Press, 1990. MR 92h:20002 Zbl 0725.20028

[Krob and Thibon 1997] D. Krob and J.-Y. Thibon, "Noncommutative symmetric functions, IV: Quantum linear groups and Hecke algebras at $q = 0$", *J. Algebraic Combin.* **6**:4 (1997), 339–376. MR 99c:05196 Zbl 0881.05120

[Lallement and Petrich 1969] G. Lallement and M. Petrich, "Irreducible matrix representations of finite semigroups", *Trans. Amer. Math. Soc.* **139** (1969), 393–412. MR 39 #4300 Zbl 0205.02504

[Macdonald 1995] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford University Press, New York, 1995. MR 96h:05207 Zbl 0824.05059

[Manivel 2001] L. Manivel, *Symmetric functions, Schubert polynomials and degeneracy loci*, SMF/AMS Texts and Monographs **6**, American Mathematical Society, Providence, RI, 2001. MR 2002h:05161 Zbl 0998.14023

[Margolis and Steinberg 2011] S. Margolis and B. Steinberg, "The quiver of an algebra associated to the Mantaci-Reutenauer descent algebra and the homology of regular semigroups", *Algebr. Represent. Theory* **14**:1 (2011), 131–159. MR 2012b:20156 Zbl 1227.20058

[Norton 1979] P. N. Norton, "0-Hecke algebras", *J. Austral. Math. Soc. Ser. A* **27**:3 (1979), 337–357. MR 80e:16015 Zbl 0407.16019

[Nozaki et al. 1995] A. Nozaki, M. Miyakawa, G. Pogosyan, and I. G. Rosenberg, "The number of orthogonal permutations", *European J. Combin.* **16**:1 (1995), 71–85. MR 95k:05010 Zbl 0828.05004

[OEIS Foundation 2012] OEIS Foundation, *The on-line encyclopedia of integer sequences*, 2012, available at http://oeis.org.

[Pin 2012] J.-E. Pin, "Mathematical foundations of automata theory", course notes, Laboratoire d'Informatique Algorithmique: Fondements et Applications, 2012, available at http://www.liafa.jussieu.fr/~jep/PDF/MPRI/MPRI.pdf.

[Le Conte de Poly-Barbut 1994] C. Le Conte de Poly-Barbut, "Sur les treillis de Coxeter finis", *Math. Inform. Sci. Humaines* **125** (1994), 41–57. MR 95f:20064 Zbl 0802.06016

[Pouzet 2013] M. Pouzet, "Théorie de l'ordre: une introduction", book, to appear on arXiv, 2013.

[Rees 1940] D. Rees, "On semi-groups", *Proc. Cambridge Philos. Soc.* **36** (1940), 387–400. MR 2, 127g Zbl 0028.00401

[Rhodes and Zalcstein 1991] J. Rhodes and Y. Zalcstein, "Elementary representation and character theory of finite semigroups and its application", pp. 334–367 in *Monoids and semigroups with applications* (Berkeley, CA, 1989), edited by J. Rhodes, World Sci. Publ., River Edge, NJ, 1991. MR 92k:20129 Zbl 0799.20062

[Rota 1964] G.-C. Rota, "On the foundations of combinatorial theory, I: Theory of Möbius functions", *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** (1964), 340–368. MR 30 #4688 Zbl 0121.02406

[Sage-Combinat 2008] Sage-Combinat community, "Sage-Combinat", software project, 2008, available at http://combinat.sagemath.org.

[Saliola 2007] F. V. Saliola, "The quiver of the semigroup algebra of a left regular band", *Internat. J. Algebra Comput.* **17**:8 (2007), 1593–1610. MR 2009a:20113 Zbl 1148.16024

[Serre 1977] J.-P. Serre, *Linear representations of finite groups*, Graduate Texts in Mathematics **42**, Springer, New York, 1977. MR 56 #8675 Zbl 0355.20006

[Stanley 1997] R. P. Stanley, *Enumerative combinatorics*, vol. 1, Cambridge Studies in Advanced Mathematics **49**, Cambridge University Press, 1997. MR 98a:05001

[Stein et al. 2009] W. Stein et al., "Sage mathematics software (version 3.3)", 2009, available at http://www.sagemath.org.

[Thiéry 2012] N. Thiéry, "Cartan invariant matrices for finite monoids", pp. 887–898 in *24th International Conference on Formal Power Series and Algebraic Combinatorics* (Nagoya, 2012), edited by N. Broutin and L. Devroye, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2012.

[Wikipedia 2010] "Wikipedia, the free encyclopedia", web page, Wikipedia, 2010, available at http://en.wikipedia.org.

[Zalcstein 1971] Y. Zalcstein, "Studies in the representation theory of finite semigroups", *Trans. Amer. Math. Soc.* **161** (1971), 71–87. MR 44 #337 Zbl 0228.20063

[Zelevinsky 1981] A. V. Zelevinsky, *Representations of finite classical groups: A Hopf algebra approach*, Lecture Notes in Mathematics **869**, Springer, Berlin, 1981. MR 83k:20017 Zbl 0465.20009

Communicated by Victor Reiner
Received 2011-06-08 Revised 2012-02-20 Accepted 2012-04-04

florent.hivert@lri.fr — *Laboratoire de Recherche en Informatique (UMR CNRS 8623), Université Paris-Sud 11, 91405 Orsay cedex, France*

anne@math.ucdavis.edu — *Department of Mathematics, University of California, One Shields Avenue, Davis, CA 95616-8633, United States*

nicolas.thiery@u-psud.fr — *Laboratoire de Mathématiques d'Orsay, Université Paris-Sud 11, 91405 Orsay cedex, France*

Current address: — *Laboratoire de Recherche en Informatique (UMR CNRS 8623), Université Paris-Sud 11, 91405 Orsay cedex, France*

ALGEBRA AND NUMBER THEORY 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.673

Shuffle algebras, homology, and consecutive pattern avoidance

Vladimir Dotsenko and Anton Khoroshkin

Shuffle algebras are monoids for an unconventional monoidal category structure on graded vector spaces. We present two homological results on shuffle algebras with monomial relations, and use them to prove exact and asymptotic results on consecutive pattern avoidance in permutations.

1. Introduction

The goal of this paper is twofold. First of all, it is intended to develop some homological algebra tools for shuffle algebras defined by Maria Ronco [2011] (called also *permutads* in a recent paper [Loday and Ronco 2011]). Namely, our main result can be viewed as the computation of appropriate Tor groups for shuffle algebras with monomial relations (the case of nonmonomial relations may be handled in a usual way by Gröbner bases and homological perturbation [Dotsenko and Khoroshkin 2010]). This generalizes for the case of shuffle algebras a celebrated construction of Anick [1986].

On the other hand, our result has a transparent combinatorial meaning. Shuffle algebras with monomial relations have bases that can be naturally described via (generalized colored) permutations avoiding given consecutive patterns. A permutation τ is said to occur in a permutation σ as a consecutive pattern if there exists a subword of σ which is order-isomorphic to τ. Free resolutions that we construct allow us to give combinatorial formulae for inverses of the corresponding exponential generating functions. A simple example one can have in mind is as follows. Permutations avoiding the consecutive pattern 12 are precisely the *decreasing* permutations; there is exactly one such permutation of each length n. "On the dual level", the space of generators of the corresponding free resolution is spanned by permutations where all the subwords of length two are order-isomorphic

Khoroshkin's research was supported by grants NSh3349.2012.2, RFBR-10-01-00836, RFBR-CNRS-10-01-93111, RFBR-CNRS-10-01-93113, and a grant of Ministry of Education and Science of the Russian Federation under the contract 14.740.11.081.

MSC2010: primary 05E15; secondary 18G10, 16E05, 05A16, 05A15, 05A05.

Keywords: shuffle algebra, consecutive pattern avoidance, free resolution.

to 12, that is *increasing* permutations. There is also exactly one such permutation of each length n. This leads to the inversion formula

$$\sum_{n\geq 0}\frac{t^n}{n!}=\frac{1}{1-t+\sum_{q\geq 2}\frac{(-1)^q}{q!}t^q},$$

where one recognizes an elementary formula

$$\exp(t)=\frac{1}{\exp(-t)}.$$

One particular application of our approach for longer patterns is a proof of a conjecture of Elizalde [2004] on patterns without self-overlaps. When we prepared the first draft of this paper, we learned that this conjecture was independently proved by Adrian Duane and Jeffrey Remmel [2011] based on methods developed in [Mendes and Remmel 2006].

The example above, as well as many similar ones, fits into a very simple combinatorial proof using the inclusion-exclusion principle. The combinatorial formalism for that is called the cluster method of Goulden and Jackson [1979]; see also [Noonan and Zeilberger 1999]. However, the formulas provided in that way have many terms canceling for somewhat trivial reasons. In contrast, our approach gives formulas free from those trivial cancellations. Further progress in algorithmic and computational approaches to consecutive pattern avoidance is presented in recent preprints [Baxter et al. 2011; Nakamura 2011]. We also wish to mention a follow-up [Khoroshkin and Shapiro 2011] to an earlier version of this paper showing the relevance of homological methods for studying consecutive patterns.

The paper is organized as follows. In Section 2 we give the definition of a shuffle algebra, and explain how shuffle algebras can be used to study consecutive pattern avoidance. Then, before constructing our free resolutions in full detail, we begin with exploring the low homological degrees in Section 3. It turns out that they can be used to obtain various asymptotic results on consecutive pattern avoidance, in the spirit of the approach of Golod and Shafarevich [1964]. We re-prove several results in that direction previous obtained by Elizalde [2006], and derive various new ones. Finally in Section 4, we construct a free resolution of the trivial module over a shuffle algebra with monomial relations, and discuss applications of this resolution. A reader primarily interested in applications to combinatorics should refer to Sections 3.2 and 4.2; though these sections contain references to results proved in more algebraic parts of the paper, they are close to being self-contained in all other respects.

All vector spaces throughout this work are defined over an arbitrary field $\Bbbk$ of zero characteristic. We adopt the usual notation $[n]$ for the set $\{1,2,\dots,n\}$. The group of permutations of a finite set I is denoted by $\mathrm{Sym}(I)$. In case $I = [n]$, we use a more concise notation S_n for the permutation group.

2. Shuffle algebras

2.1. ***Nonsymmetric collections and shuffle products.*** In this section, we shall recall the definition of a shuffle algebra, as defined by Ronco [2011]; see also [Loday and Ronco 2011]. Our definitions and methods, though equivalent to the original definition of Ronco (and the subsequent definition of Loday and Ronco), are different, and rather follow the approach of [Dotsenko and Khoroshkin 2010].

We denote by Ord_+ the category whose objects are finite ordered sets (with order-preserving bijections as morphisms). Also, we denote by Vect the category of vector spaces (with linear operators as morphisms).

Definition 1. (1) A (*nonsymmetric*) *collection* is a contravariant functor from the category Ord_+ to the category Vect.

(2) Let $\mathcal{P}$ and $\mathcal{Q}$ be two nonsymmetric collections. Define their *shuffle tensor product* $\mathcal{P} \boxtimes \mathcal{Q}$ by the formula

$$(\mathcal{P} \boxtimes \mathcal{Q})(I) := \bigoplus_{J \sqcup K = I} \mathcal{P}(J) \otimes \mathcal{Q}(K),$$

where the sum is taken over all partitions of I into two disjoint subsets J and K.

Remark 2. (1) Nonsymmetric collections are in one-to-one correspondence with (nonnegatively) graded vector spaces (for a functor $\mathcal{F}$, the graded component F_n of the corresponding graded vector space F is $\mathcal{F}([n])$). However, the functorial definition makes the monoidal structure much easier to handle, with one exception: To *define* a nonsymmetric collection, it is sufficient to define the spaces $\mathcal{F}([n])$, with all other spaces defined automatically because of functoriality. We shall use this observation many times throughout the paper.

(2) If we define the tensor product of two nonsymmetric collections by a similar-looking formula

$$(\mathcal{P} \otimes \mathcal{Q})(I) := \bigoplus_{J + K = I} \mathcal{P}(J) \otimes \mathcal{Q}(K),$$

where the sum is taken over all partitions of I into two consecutive intervals J and K, this would indeed give the standard tensor product of graded vector spaces.

The following proposition is straightforward; we omit the proof.

Proposition 3. *The shuffle tensor product endows the category of nonsymmetric collections with a structure of a monoidal category. The unit object in each case is the functor $\mathcal{I}$ that vanishes on all nonempty sets and is one-dimensional for the empty set.*

The following proposition shows that the shuffle tensor product provides a "categorification" of the product of exponential generating functions in the same way as the usual tensor product provides a categorification of the product of "normal" generating functions.

Proposition 4. *For a nonsymmetric collection $\mathcal{P}$, let us define its exponential generating series $f_{\mathcal{P}}(t)$ as the power series*

$$\sum_{n\geq 0}\frac{\dim \mathcal{P}([n])}{n!}t^n.$$

Then we have

$$f_{\mathcal{P}\boxtimes\mathcal{Q}}(t)=f_{\mathcal{P}}(t)\cdot f_{\mathcal{Q}}(t). \tag{1}$$

Proof. Indeed, the number of ways to split $[n]$ into a disjoint union $[n]=J\sqcup K$ with $|J|=j$ and $|K|=k$ is equal to

$$\binom{n}{j}=\frac{n!}{j!(n-j)!}=\frac{n!}{j!\,k!},$$

so

$$\dim((\mathcal{P}\boxtimes\mathcal{Q})([n]))=\sum_{0\leq j\leq n}\frac{n!}{j!\,k!}\dim(\mathcal{P}([j]))\dim(\mathcal{Q}([k])),$$

and the result follows. □

2.2. *Shuffle algebras.*

Definition 5. A *shuffle* (*associative*) *algebra* is a monoid in the category of nonsymmetric collections with the monoidal structure given by the shuffle tensor product.

In other words, to define a shuffle algebra structure on a nonsymmetric collection $\mathcal{A}$, one has to define the *structure maps*

$$\mu_{J,K}:\mathcal{A}(J)\otimes\mathcal{A}(K)\to\mathcal{A}(J\sqcup K)$$

satisfying the obvious associativity conditions.

Remark 6. Shuffle algebras are closely related to twisted associative algebras (see for instance [Stover 1993]), namely, they are in the same relationship with them as shuffle operads are with symmetric operads. Also, the category of shuffle algebras admits an embedding into the category of shuffle operads, and this embedding is behind some of the constructions of this paper. We shall not discuss these topics in detail here.

Example 7. Every graded associative algebra V gives rise to a shuffle algebra $\widetilde{V}$ with $\widetilde{V}(I) = V_{|I|}$, where for every partition $I = J \sqcup K$ the corresponding product map

$$\mu_{J,K} : \widetilde{V}(J) \otimes \widetilde{V}(K) = V_{|J|} \otimes V_{|K|} \to V_{|J|+|K|} = \widetilde{V}(I)$$

is given by the product in V.

Example 8. Consider the shuffle algebra $\mathscr{A}_{MR}$ with $\mathscr{A}_{MR}(I) = \Bbbk\,\mathrm{Sym}(I)$, where for every partition $I = J \sqcup K$ the corresponding product map

$$\mu_{J,K} : \mathscr{A}_{MR}(J) \otimes \mathscr{A}_{MR}(K) = \Bbbk\,\mathrm{Sym}(J) \otimes \Bbbk\,\mathrm{Sym}(K) \to \Bbbk\,\mathrm{Sym}(I) = \mathscr{A}_{MR}(I)$$

is somewhat tautological: The product of two permutations is the permutation of $I = J \sqcup K$ obtained from the respective permutations of J and K by concatenation.

As shown in [Ronco 2011], the algebra from the previous example is isomorphic to the free shuffle algebra with one generator of degree 1. This shuffle algebra gives a refinement of (the underlying graded algebra of) the Malvenuto–Reutenauer Hopf algebra of permutations [Malvenuto and Reutenauer 1995]. Many other Hopf algebras of combinatorial nature, for example, the Hopf algebra of quasisymmetric functions, the Hopf algebra of parking functions, the Hopf algebra of set partitions, etc. (for definitions, see [Loday and Ronco 2010] and references therein), are shuffle algebras as well, with the associative product being the sum over all possible shuffle products.

Let us give the combinatorial construction of a free algebra generated by a given nonsymmetric collection. Let $\mathscr{M}$ be a nonsymmetric collection with $\mathscr{M}(\varnothing) = \{0\}$, and let B be a nonsymmetric collection of finite ordered sets (that is, a functor from the category Ord_+ to itself) such that for every ordered set I the set $\mathsf{B}(I)$ is a basis of $\mathscr{M}(I)$. We shall describe a nonsymmetric collection of finite ordered sets that will form a bases in components of the free shuffle algebra. By definition, elements of $\mathbb{B}(I)$ correspond to the following combinatorial data:

(1) an ordered partition of I into subsets, $I = \bigsqcup_{j=1}^{m} I_j$,

(2) a "monomial" $c_1 c_2 \cdots c_m$ with $c_j \in \mathsf{B}(I_j)$ for every $j = 1, \ldots, m$.

The shuffle product $\mu_{J,K}$ concatenates both the ordered partitions and the monomials.

Note that if we assume that $\mathscr{M}(I) = \{0\}$ for $|I| \neq 1$ and $\dim \mathscr{M}(I) = 1$ for $|I| = 1$, we see that every subset I_j has to consist of one element, and therefore any ordered partition that contributes is just a permutation (and the monomials do not carry additional information, capturing the lengths of the permutations). Therefore, we recover the free algebra with one generator of degree 1 from Example 8 above.

The following proposition is straightforward.

Proposition 9. *The collection $F\langle\mathcal{M}\rangle$ with $F\langle\mathcal{M}\rangle(I) = \operatorname{span}\mathbb{B}(I)$ is (isomorphic to) the free shuffle algebra generated by $\mathcal{M}$.*

2.2.1. *Shuffle ideals and modules.* Since shuffle algebras are monoids in a monoidal category, the usual definitions of ideals, quotients, modules, etc. can be immediately given in this context. To make the article self-contained, we present them here. All shuffle algebras in this paper are assumed to be *connected*, that is having $\Bbbk$ as the empty set component.

Definition 10. Let $\mathcal{A}$ be a shuffle algebra with the product $\mu\colon \mathcal{A}\boxtimes\mathcal{A}\to\mathcal{A}$.

- A right module over $\mathcal{A}$ is a nonsymmetric collection $\mathcal{M}$ together with a structure map $\gamma\colon \mathcal{M}\boxtimes\mathcal{A}\to\mathcal{M}$ satisfying the associativity condition
$$\gamma(\gamma\boxtimes\mathrm{id}_{\mathcal{A}}) = \gamma(\mathrm{id}_{\mathcal{M}}\boxtimes\mu).$$
- The trivial right module over $\mathcal{A}$ is the collection $\mathcal{I}$ that has $\Bbbk$ as the empty set component and zero for all other components, where the only nonzero part of the structure map is
$$\mathcal{I}(\varnothing)\otimes\mathcal{A}(\varnothing) = \Bbbk\otimes\Bbbk \simeq \Bbbk = \mathcal{I}(\varnothing).$$
- The regular right module over $\mathcal{A}$ is the collection $\mathcal{A}$ itself, with the structure map $\gamma = \mu$.
- A right ideal of $\mathcal{A}$ is a subcollection of the regular right module that is closed under the structure map.
- For a subcollection $\mathcal{R}$ of $\mathcal{A}$, the right ideal $(\mathcal{R})$ generated by $\mathcal{R}$ is the minimal right ideal of $\mathcal{A}$ that contains $\mathcal{R}$.
- The free right module over $\mathcal{A}$ generated by the nonsymmetric collection $\mathcal{V}$ is the collection $\mathcal{V}\boxtimes\mathcal{A}$ with the structure map $\gamma = \mathrm{id}_{\mathcal{V}}\boxtimes\mu$. A free module is said to be finitely generated if all components $\mathcal{V}(I)$ are finite-dimensional, and moreover they vanish for $|I|$ sufficiently large.

The respective definitions of left modules, left ideals, bimodules, and two-sided ideals are completely analogous.

The following is an example of how graded associative algebras can be presented as shuffle algebras with generators and relations, that is, as quotients of free shuffle algebras.

Example 11. Let us take the algebra $\mathcal{A}_{MR}$ discussed in Example 8, and compute its quotient modulo the two-sided ideal generated by the difference $12-21\in\Bbbk S_2$. This quotient is isomorphic to the algebra $\widetilde{V}$ from Example 7 with $V=\Bbbk[x]$.

2.2.2. *Consecutive patterns.* In this section, we shall explain how our definitions are related to the combinatorial concept of consecutive pattern avoidance.

Let us recall some definitions and notation. To every sequence s of length k consisting of k distinct numbers, we assign a permutation $\mathrm{st}(s)$ of length k called the standardization of s; it is uniquely determined by the condition that $s_i < s_j$ if and only if $\mathrm{st}(s)_i < \mathrm{st}(s)_j$. For example, $\mathrm{st}(153) = 132$. In other words, $\mathrm{st}(s)$ is a permutation whose relative order of entries is the same as that of s. We say that a permutation σ of length n avoids the given permutation τ of length j as a consecutive pattern if for each $j < n - i + 1$ we have $\mathrm{st}(\sigma_i \sigma_{i+1} \cdots \sigma_{i+j-1}) \neq \tau$; otherwise we say that σ contains τ as a consecutive pattern. Throughout this paper, we only deal with consecutive patterns, so the word "consecutive" will be omitted. For historical information on pattern avoidance in general and the state of the art for consecutive patterns, we refer the reader to [Kitaev and Mansour 2003; Steingrímsson 2010].

The central question arising in the theory of pattern avoidance is that of enumeration of permutations of given length that avoid the given set of forbidden patterns P or, more generally, contain exactly l occurrences of patterns from P. This question naturally leads to the following equivalence relations. Two sets of patterns P and P' are said to be Wilf equivalent (notation: $P \simeq_W P'$) if for every n, the number of P-avoiding permutations of length n is equal to the number of P'-avoiding permutations of length n. This notion (in the case of one pattern) is due to Wilf [2002]. More generally, P and P' are said to be equivalent (notation: $P \simeq P'$) if for every n and every $k \geq 0$, the number of permutations of length n with k occurrences of patterns from P is equal to the number of permutations of length n with k occurrences of patterns from P'.

While studying the equivalence classes of patterns, sometimes it is possible to replace the set of forbidden patterns by a Wilf equivalent one with fewer patterns in it. Namely, we have a partial ordering on the set of all permutations (of all possible lengths): $\tau < \sigma$ if σ contains τ as a consecutive pattern. Given a set P of "forbidden" patterns, to enumerate the permutations avoiding all patterns from P, we may assume that P is an antichain with respect to this partial ordering. Indeed, ignoring all patterns from P that contain a smaller forbidden subpattern does not change the set of P-avoiding permutations. Therefore, further on we shall assume that forbidden patterns do indeed form an antichain.

2.2.3. *Shuffle algebras and consecutive patterns.* The following result, however simple, provides a bridge between algebra and combinatorics, defining for each forbidden set P of patterns a shuffle algebra whose exponential generating series is precisely the exponential generating function for the numbers of permutations avoiding P. Let us denote by a_n^P the number of permutations of length n that avoid all patterns from P, and by $g_P(t)$ the corresponding exponential generating function,

$$g_P(t) := 1 + \sum_{n\geq 1} \frac{a_n^P}{n!} t^n.$$

Theorem 12. *For every set P of forbidden patterns, let us define the shuffle algebra $\mathcal{A}_{MR}^P$ as the quotient of the algebra $\mathcal{A}_{MR}$ modulo the two-sided ideal generated by all patterns from P. Then the (classes of) permutations avoiding all patterns from P form a basis of the quotient. Consequently,*

$$f_{\mathcal{A}_{MR}^P}(t) = g_P(t).$$

Proof. Since the products in $\mathcal{A}_{MR}$ are defined via concatenations, it is clear that the ideal generated by P consists precisely of permutations containing patterns from P. This means that we may identify classes in the quotient $\mathcal{A}_{MR}/(P)$ with permutations avoiding patterns from P. We shall use this identification throughout the paper. □

A similar result for the free shuffle algebra with more than one generator provides technical tools to deal with pattern avoidance in colored permutations [Mansour 2001/02], and more general consecutive pattern avoidance where, for instance, each occurrence of a rise of length 2 may or may not be colored. We shall not discuss the corresponding applications in this paper, but want to draw the reader's attention that all our methods generalize immediately to those settings.

2.2.4. *Modules over the associative operad.* This short section is intended for readers whose intuition, as ours does, comes from operad theory. Essentially, it retells the shuffle algebra approach in a slightly different way, explaining also the place for classical pattern avoidance in the story (recall that classical patterns are those occurring as subsequences rather than as factors in permutations).

Studying varieties of algebras, that is, algebras satisfying certain identities, goes back to work of Specht [1950]. The notions of T-ideals and T-spaces formalize the ways to derive identities from one another. One natural way to study identities is to define an analogue of a Gröbner basis for an ideal of identities. This approach is taken in works of Latyshev [2005; 2008], who suggested a combinatorial approach to studying associative algebras with additional identities via standard bases of the corresponding T-spaces. His approach can be described as follows. For each "T-space" (in other words, right ideal in the associative operad), he defines a version of a Gröbner basis; such a basis would allow to study arbitrary relations via monomials avoiding certain patterns. Here, for once, by a pattern we mean a classical pattern (its occurrence does not have to be as a consecutive subword, but rather a subsequence). This approach has a slight disadvantage. Namely, even though the actual Gröbner bases of relations are expected to be finite (at least, the famous result of Kemer [1988] states that in principle there exists a finite set of generating identities), they are difficult to compute, as there is no algorithm comparable to the one due to Buchberger in the associative algebra case [Ufnarovskij 1995]. Remarkably, this trouble disappears if we study left ideals in the associative operad. In terms of combinatorics, studying left ideals also has a very clear meaning: The corresponding notion of divisibility corresponds to *consecutive* pattern avoidance! For consecutive patterns, the intuition of [Dotsenko and Khoroshkin 2010; 2012] for Gröbner bases and resolutions applies directly, and it turns out to be possible to describe the relevant resolutions explicitly, in fact the level of complexity here being closer to the case of associative algebras than to the case of operads.

2.3. ***Shuffle homological algebra.*** One of the central concepts of homological algebra is that of a derived functor. Computing derived functors relies on being able to construct "nice" (free, projective, injective etc.) resolutions of objects to that we want to apply our derived functors. The category of objects of primary interest to us is the category of left modules over the given shuffle algebra $\mathscr{A}$, and a typical functor we want to derive is "shuffle torsion groups", that is, the derived functor of the shuffle tensor product over A with a given module, for instance with the trivial right module. This paper is focused on combinatorial applications of shuffle algebras, so in the view of Theorem 12 the shuffle algebras of main interest for us are quotients of the algebra $\mathscr{A}_{MR}$ modulo the ideal generated by several patterns. In the following sections, we shall present two results of homological algebra for such shuffle algebras, and derive from these results various statements on enumerative combinatorics of consecutive patterns. One technical result that we shall be using to translate between the two languages is the following standard statement on Euler characteristics, applied to nonsymmetric collections.

Proposition 13. *Let $\cdots \to \mathscr{C}_n \to \cdots \to \mathscr{C}_2 \to \mathscr{C}_1 \to \mathscr{C}_0$ be a chain complex of nonsymmetric collections with homology groups $\mathscr{H}_0, \mathscr{H}_1, \dots, \mathscr{H}_n, \dots$. Then we have*

$$f_{\mathscr{C}_0}(t) - f_{\mathscr{C}_1}(t) + \cdots + (-1)^n f_{\mathscr{C}_n}(t) + \cdots \\ = f_{\mathscr{H}_0}(t) - f_{\mathscr{H}_1}(t) + \cdots + (-1)^n f_{\mathscr{H}_n}(t) + \cdots,$$

provided that the sums on the left and on the right make sense (for every integer l only finitely many summands have nonzero coefficients of t^l).

3. Golod–Shafarevich-type complex and its applications

3.1. ***Golod–Shafarevich-type inequality.*** In this section, we shall exhibit a very simple application of homological algebra philosophy to combinatorics, mimicking the idea used by Golod and Shafarevich [1964] in their study of the class field tower, which has been used a lot in algebra and combinatorics since then; see for instance [Piotkovskii 1993] and the later papers [Bell and Small 2002; Bell and Goh 2007; Etingof and Ginzburg 2007; Rampersad 2011]. Namely, we shall construct the low homological degree part of the minimal resolution of the trivial right module over the shuffle algebra $\mathcal{A}_{MR}^{P}$ by free modules. More precisely, we shall prove the following theorem.

Theorem 14. *Let $\mathcal{V}$ be the one-dimensional space generating the free algebra $\mathcal{A}_{MR}$, and $\mathcal{P}$ be the subcollection of $\mathcal{A}_{MR}$ spanned by forbidden patterns. There exists a chain complex*

$$\mathcal{P} \boxtimes \mathcal{A}_{MR}^{P} \to \mathcal{V} \boxtimes \mathcal{A}_{MR}^{P} \to \mathcal{A}_{MR}^{P} \to \mathcal{I} \to 0, \tag{2}$$

which is exact everywhere except for the leftmost term.

Proof. The boundary maps are as follows:

(1) $\mathcal{A}_{MR}^{P} \to \mathcal{I}$ is the augmentation, mapping all permutations of positive length to zero,

(2) $\mathcal{V} \boxtimes \mathcal{A}_{MR}^{P} \to \mathcal{A}_{MR}^{P}$ is the product in the algebra $\mathcal{A}_{MR}^{P}$ (generated by $\mathcal{V}$),

(3) $\mathcal{P} \boxtimes \mathcal{A}_{MR}^{P} \to \mathcal{V} \boxtimes \mathcal{A}_{MR}^{P}$ is the composition of the inclusion

$$\mathcal{P} \boxtimes \mathcal{A}_{MR}^{P} \to \mathcal{V} \boxtimes \mathcal{A}_{MR} \boxtimes \mathcal{A}_{MR}^{P}$$

(which exists because $\mathcal{V}$ generates the algebra $\mathcal{A}_{MR}$), the projection

$$\mathcal{V} \boxtimes \mathcal{A}_{MR} \boxtimes \mathcal{A}_{MR}^{P} \to \mathcal{V} \boxtimes \mathcal{A}_{MR}^{P} \boxtimes \mathcal{A}_{MR}^{P},$$

and the product in the algebra $\mathcal{A}_{MR}^{P}$.

The exactness of this complex in the terms $\mathcal{I}$ and $\mathcal{A}_{MR}^{P}$ is obvious. Let us show the exactness in the term $\mathcal{V} \boxtimes \mathcal{A}_{MR}^{P}$. Since the relations of the algebra $\mathcal{A}_{MR}^{P}$ are monomial, the kernel of the boundary map is spanned by "monomials" $j \boxtimes \rho$ with j being an element of degree 1 and ρ being a permutation avoiding patterns from P. Such an element belongs to the kernel of the boundary map if $j\rho = 0$; therefore, $j\rho$ contains a pattern from P. Such a pattern has to be an initial segment of $j\rho$, otherwise ρ would contain a pattern from P itself. This instantly implies that our element is in the image of the boundary map from $\mathcal{P} \boxtimes \mathcal{A}_{MR}^{P}$. □

Corollary 15. *Let $P = \bigsqcup_{n\geq 2} P_n$ be a collection of forbidden consecutive patterns in permutations. Then the following coefficient-wise inequality holds*

$$\Big(1 - t + \sum_{k\geq 2} \frac{|P_k|}{k!} t^k\Big) g_P(t) \geq 1. \tag{3}$$

Proof. Let us denote by $\mathcal{H}_3$ the only nontrivial piece of homology of the chain complex (2). Computing Euler characteristics according to Proposition 13, we see that

$$f_{\mathcal{I}}(t) - f_{\mathcal{A}^P_{MR}}(t) + f_{\mathcal{V}}(t) f_{\mathcal{A}^P_{MR}}(t) - f_{\mathcal{P}}(t) f_{\mathcal{A}^P_{MR}}(t) = -f_{\mathcal{H}_3}(t),$$

or

$$1 - g_P(t) + t g_P(t) - \Big(\sum_{k\geq 2} \frac{|P_k|}{k!} t^k\Big) g_P(t) = -f_{\mathcal{H}_3}(t),$$

which implies

$$\Big(1 - t + \sum_{k\geq 2} \frac{|P_k|}{k!} t^k\Big) g_P(t) = 1 + f_{\mathcal{H}_3}(t) \geq 1. \qquad \square$$

3.2. *Applications to consecutive pattern avoidance.* The following contains on key application of Corollary 15.

Corollary 16. *Assume that the power series*

$$f(t) = 1 - t + \sum_{k\geq 2} \frac{|P_k|}{k!} t^k$$

has a root $\alpha > 0$. *Then* $a_n^P \geq \alpha^{-n} n!$.

Proof. Let

$$\sum_{l\geq 0} b_l t^l := \frac{1}{1 - t + \sum_{k\geq 2} \frac{|P_k|}{k!} t^k},$$

so that $b_0 = 1$ and

$$b_n - b_{n-1} + \sum_{k=2}^{n} \frac{|P_k|}{k!} b_{n-k} = 0.$$

Let us prove by induction that $b_n \geq \alpha^{-1} b_{n-1}$. Indeed, for $n = 1$ this statement is obvious ($\alpha \geq 1$ because otherwise $f(\alpha)$ is evidently positive), and for $n > 1$ we note that by the induction hypothesis $b_{n-1} \geq \alpha^{1-k} b_{n-k}$, so

$$\begin{aligned} b_n &= b_{n-1} - \sum_{k=2}^{n} \frac{|P_k|}{k!} b_{n-k} \geq b_{n-1} - \sum_{k=2}^{n} \frac{|P_k|}{k!} \alpha^{k-1} b_{n-1} \\ &\geq b_{n-1} - \sum_{k\geq 2} \frac{|P_k|}{k!} \alpha^{k-1} b_{n-1} = \alpha^{-1} b_{n-1} \Big(\alpha - \sum_{k\geq 2} \frac{|P_k|}{k!} \alpha^k\Big) = \alpha^{-1} b_{n-1}, \end{aligned}$$

which proves the step of induction. Therefore $b_n \geq \alpha^{-n}$, and the series

$$\frac{1}{1 - t + \sum_{k\geq 2} \frac{|P_k|}{k!} t^k}$$

has positive coefficients. Hence multiplying the inequality (3) by that series preserves the inequality, and we obtain

$$g_P(t) \geq \frac{1}{1-t+\sum_{k\geq 2}\frac{|P_k|}{k!}t^k}, \quad \text{so } \frac{a_n^P}{n!} \geq \alpha^{-n}. \qquad \square$$

Using the corollary above, one can obtain good asymptotic results on enumeration of permutations avoiding the given set of consecutive patterns, thus rediscovering a result of Elizalde [2006] in the case of one pattern, but also recovering some much stronger results. Let use give several examples.

Corollary 17. *The number of permutations of length n avoiding the given single pattern τ of length k is at least $\alpha_k^{-n}n!$, where α_k is the smallest positive root of the equation*

$$1-t+\frac{t^k}{k!}=0.$$

(*For example*, $\alpha_4 \approx 1.050800769$, $\alpha_5 \approx 1.008702295$, $\alpha_6 \approx 1.001400601$.)

Corollary 18. *Let the set of forbidden patterns P contain one pattern of each length $l \geq 4$. Then the number of permutations of length n avoiding P is at least $\alpha^{-n}n!$, where $\alpha \approx 1.068290263$ is the root of the equation*

$$e^t-2t-\tfrac{1}{2}t^2-\tfrac{1}{6}t^3=0.$$

In particular, there are infinitely many permutations avoiding P regardless of the actual choice of patterns in P.

Proof. In this case, $|P_n|=1$ for all $n\geq 4$, so

$$1-t+\sum_{k\geq 2}\frac{|P_k|}{k!}t^k = 1-t+e^t-1-t-\tfrac{1}{2}t^2-\tfrac{1}{6}t^3 = e^t-2t-\tfrac{1}{2}t^2-\tfrac{1}{6}t^3. \qquad \square$$

4. Anick-type resolution and its applications

4.1. ***Anick-type resolution.*** In this section, we shall explain how to extend the complex we constructed above to a resolution of the trivial right module by free $\mathscr{A}_{MR}^P$-modules. The generators of those modules are defined combinatorially. Once the set P of forbidden patterns is fixed, we define, for each nonnegative integer q, the notion of a q-chain and the tail of a given q-chain associated to P inductively as follows:

- The empty permutation is a 0-chain on the empty set; it coincides with its tail.
- The only permutation of a one-element set I is a 1-chain on I; it also coincides with its tail.

- Each q-chain is a permutation σ represented as a concatenation $\sigma'\tau$, where τ is the tail of σ, and σ' is a $(q-1)$-chain on its underlying set.
- If we denote by τ' the tail of σ' in the representation above, then $\tau'\tau$ contains exactly one occurrence of a pattern from P, and this occurrence is a terminal segment of $\tau'\tau$.

The way we define the chains here is slightly different from the original approach of Anick [1986]; the reader familiar with the excellent textbook of Ufnarovskii [1995] will rather notice similarities with the approach to Anick resolution adopted there.

Informally, a q-chain is a "minimal" way to form a permutation by linking together $(q-1)$ prohibited patterns. The word "minimal" is justified by the following:

Lemma 19. *No proper beginning of a q-chain is a q-chain.*

Proof. We shall prove this by induction on q, the basis of induction ($q = 0, 1, 2$) being obvious.

Assume there is a q-chain $\sigma = \sigma'\tau$ that has a proper beginning that is a q-chain as well, so $\tau = \mu\nu$, and $\sigma'\mu$ is a q-chain. By the induction hypothesis, no proper beginning of a $(q-1)$-chain is a $(q-1)$-chain, which implies that μ is the tail of the q-chain $\sigma'\mu$. However, this immediately shows that (in the notation of the definition of chains and tails above) $\tau'\tau$ contains at least two different occurrences of patterns from P, which is a contradiction. □

One more fact about chains that makes the definition above more transparent is that, even though we defined a chain as a permutation together with a factorization, in fact the factorization carries no additional information:

Lemma 20. *If σ is a q-chain, the way to link $(q-1)$ patterns from P to one another to form σ is unique.*

Proof. Assume that there are two ways to link q patterns to form σ. Obviously, for each $m < q$, the endpoints of the m-th (from left to right) patterns in these two linkages should coincide, otherwise we shall find an m-chain whose proper beginning is an m-chain as well, which is not the case by the previous lemma. Once we know that the endpoints of the m-th patterns are the same, the beginnings have to be the same because P is assumed to be an antichain (and so patterns from P cannot be contained in one another). □

Let us give some examples clarifying the notion of a chain. For example, if $P = \{12\}$, the only q-chain for each q is $12\cdots q$, while if $P = \{123\}$, we can easily see that 123 is the only 1-chain, and 1234 is the only 2-chain, but 12345 is not a 2-chain because it starts from a 2-chain 1234, and it is not a 3-chain because in the only way to cover this permutation by three copies of our pattern, the first and the third occurrences overlap:

$$\underbrace{1\quad 2\quad \rlap{$\overbrace{}$}3}\quad 4\quad 5\,.$$

Theorem 21. *Denote by $\mathscr{C}_q$ the subcollection of the free algebra $\mathscr{A}_{MR}$ spanned by all q-chains. There exists a chain complex*

$$\cdots\mathscr{C}_q\boxtimes\mathscr{A}_{MR}^P\to\mathscr{C}_{q-1}\boxtimes\mathscr{A}_{MR}^P\to\cdots\to\mathscr{C}_1\boxtimes\mathscr{A}_{MR}^P\to\mathscr{A}_{MR}^P\to\mathscr{I}\to 0, \tag{4}$$

which is exact in every term.

This result is a direct generalization of the one in Theorem 14 since $\mathscr{C}_2=\mathscr{P}$ and $\mathscr{C}_1=\mathscr{V}$.

Proof. The boundary map $\mathscr{C}_q\boxtimes\mathscr{A}_{MR}^P\to\mathscr{C}_{q-1}\boxtimes\mathscr{A}_{MR}^P$ is defined as a composition of the inclusion

$$\mathscr{C}_q\boxtimes\mathscr{A}_{MR}^P\to\mathscr{C}_{q-1}\boxtimes\mathscr{A}_{MR}\boxtimes\mathscr{A}_{MR}^P$$

(which exists because we can factorize a q-chain as a product of a $(q-1)$-chain and a tail), the projection

$$\mathscr{C}_{q-1}\boxtimes\mathscr{A}_{MR}\boxtimes\mathscr{A}_{MR}^P\to\mathscr{C}_{q-1}\boxtimes\mathscr{A}_{MR}^P\boxtimes\mathscr{A}_{MR}^P,$$

and the product in the algebra $\mathscr{A}_{MR}^P$.

Let us prove the exactness of this complex in the term $\mathscr{C}_q\boxtimes\mathscr{A}_{MR}^P$. Since the relations of the algebra $\mathscr{A}_{MR}^P$ are monomial, the kernel of the boundary map is spanned by "monomials" $\sigma\otimes\rho$, where σ is a q-chain and ρ is a permutation avoiding patterns from P. Such an element belongs to the kernel of the boundary map if $\sigma'\otimes\tau\rho=0$, where τ is the tail of σ and $\sigma=\sigma'\tau$. Therefore, $\tau\rho$ contains a pattern from P. Since ρ avoids patterns from P, this means that there exists a decomposition $\rho=\rho'\rho''$ such that $\tau\rho'$ contains a pattern from P as its terminal segment, and this is the only occurrence of a pattern from P in $\tau\rho'$ (take for ρ' the smallest initial segment of ρ with this property). This immediately implies that $\sigma\rho'$ is a $(q+1)$-chain with the tail ρ', so our element is the image under the boundary map of the element $\sigma\rho'\otimes\rho''$. □

Let us denote by $c_{n,q}$ the number of q-chains that are permutations of length n.

Corollary 22. *We have*

$$g_P(t)=\frac{1}{1-t+\sum\limits_{q\geq 2,n\geq 1}\frac{(-1)^q c_{n,q}}{n!}t^n}. \tag{5}$$

Proof. Computing Euler characteristics according to Proposition 13, we see that

$$f_{\mathscr{I}}(t)-f_{\mathscr{A}_{MR}^P}(t)+f_{\mathscr{C}_1}(t)f_{\mathscr{A}_{MR}^P}(t)-f_{\mathscr{C}_2}(t)f_{\mathscr{A}_{MR}^P}(t)+f_{\mathscr{C}_3}(t)f_{\mathscr{A}_{MR}^P}(t)-\cdots=0$$

or

$$1-g_P(t)+tg_P(t)-\Big(\sum_{k\geq 2}\frac{(-1)^q c_{n,q}}{n!}t^n\Big)g_P(t)=0,$$

which implies

$$g_P(t)=\frac{1}{1-t+\sum\limits_{q\geq 2,n\geq 1}\frac{(-1)^q c_{n,q}}{n!}t^n}. \qquad \square$$

Let us give a simple example in which both the left hand side and the right hand side of the (5) can be easily computed (we already mentioned it in the introduction). Let P consist of a single pattern 12. Then, for each q we have one q-chain $12\cdots q$ of length q. Also, for every m the only permutation of length m avoiding 12 is $m(m-1)\cdots 21$. Therefore, the inversion formula above becomes

$$\sum_{n\geq 0}\frac{t^n}{n!}=\frac{1}{1-t+\sum\limits_{q\geq 2}\frac{(-1)^q}{q!}t^q},$$

and we recognize the well-known formula $\exp(t)\exp(-t)=1$.

Equation (5) bears a striking resemblance to a celebrated result of Goulden and Jackson [1979] expressing the inverses of generating functions for consecutive pattern avoidance in terms of *clusters*:

$$g_P(t)=\frac{1}{1-t+\sum\limits_{q\geq 2,n\geq 1}\frac{(-1)^q cl_{n,q}}{n!}t^n}, \tag{6}$$

where $cl_{n,q}$ is the number of q-clusters of length n. A q-cluster is, roughly speaking, an indecomposable covering of a permutation by patterns from the forbidden set P, but, unlike chains, without any minimality condition. As a consequence, the number of chains is potentially much smaller than the number of clusters, and our result is a strengthening of the result of Goulden and Jackson. A good way to think of it is to say that many "obvious" cancellations happen in the cluster formula (6), and our approach takes care of these "obvious" cancellations.[1] For example, we already saw that for $P=\{123\}$ the permutation 12345 is not a chain. However, it can be covered by two copies of 123 as well as by three copies of 123, and these coverings give it a structure of a 2-cluster and a 3-cluster respectively. The contributions of these two clusters in (6) occur with opposite signs, and the total contribution of this permutation is equal to zero, exactly as (5) suggests. Among the applications below, for some of the examples it does not really matter if we are dealing with chains or clusters, whereas for other ones chains give more compact formulas.

[1] We want to note, however, that our approach can be used to prove the cluster inversion formula too, if one adapts the method of [Dotsenko and Khoroshkin 2012] for constructing free resolutions.

4.2. ***Applications to consecutive pattern avoidance.*** Before moving on to particular results, let us state a general remark. Our results suggest that the class of power series that contains all inverses of pattern avoidance enumerators is related to some nice combinatorics. Results of Elizalde and Noy [2003] that we re-prove below describe some of these series as solutions to particular differential equations. Our formulas for other cases we considered can be rewritten as more complicated functional equations. What can be said about other series of that sort? So far we have not able to describe a reasonable class of series that cover all of these. A wild guess is that all these series satisfy algebraic differential equations, that is, if $f(x)$ is such a series, then $P(x, f(x), f'(x), \dots, f^{(d)}(x)) = 0$ for some nonzero polynomial $P(x, t_0, t_1, \dots, t_d)$.

4.2.1. *Patterns without self-overlaps, linking schemes, and posets.* In this section, we shall enumerate chains in one particular case, namely, the case of an arbitrary pattern without self-overlaps, which will allow us to prove a conjecture of Elizalde [2004]. In fact, in this case chains coincide with clusters, so one could refer to results of Goulden and Jackson instead of Theorem 21.

Definition 23. A pattern τ is said to have no self-overlaps if every permutation of length at most $2m-2$ has at most one occurrence of τ. (Clearly, there always exist permutations of length $2m-1$ with two occurrences of τ.)

For example, the pattern 132 is of that form: Clearly, we can only link it with itself using the last entry. A more general example studied in [Elizalde and Noy 2003] is $12\cdots a\ \tau\ (a+1) \in \Sigma_n$, where $a+1 < n$, and τ is an arbitrary permutation of the numbers $a+2, \dots, n$.

For a pattern τ without self-overlaps, there exists a simple way to reformulate the enumeration problem for chains in terms of total orderings on posets. The first author used this method in [Dotsenko and Vejdemo Johansson 2012] in a similar setting, dealing with tree monomials in the free shuffle operad. To a q-chain σ obtained by linking $q-1$ copies of τ, let us assign a "linking scheme" of the shape that we expect, replacing each entry in σ by the symbol $\bullet$ (a bullet), and marking the segments of consecutive bullets that are "traces" of (occurrences of) τ. For example, for the pattern 1243 and 4-chains we get

$$\bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\,.$$

For such a linking scheme, let us define a partial ordering on bullets as follows: For each j, we equip the j-th trace of τ with a total ordering identical to the ordering of the corresponding entries of τ. Let us denote by $\Pi_{q,\tau}$ the thus-defined poset.

Example 24. Let us take the linking scheme above, and replace bullets by letters, to make it easier to distinguish between different bullets:

$$\underbrace{a\ b\ c\ d}\ e\ f\ \underbrace{g\ h\ i\ j}\,.$$

Then the orderings inherited from 1243 are $a < b < d < c$, $d < e < g < f$, and $g < h < i < j$, so we obtain the poset $\Pi_{4,1243}$:

$$\begin{tikzcd}[column sep=small, row sep=small]
 & & & & j \\
 & & & & i \arrow[u, dash] \\
 & & f & & h \arrow[u, dash] \\
 & & & g \arrow[ul, dash] \arrow[ur, dash] & \\
c & & e \arrow[ur, dash] & & \\
 & d \arrow[ul, dash] \arrow[ur, dash] & & & \\
 & b \arrow[u, dash] & & & \\
 & a \arrow[u, dash] & & &
\end{tikzcd}$$

(the covering relation of the poset is, as usual, represented by edges; v is covered by w if w is the top vertex of the corresponding edge).

The following proposition is obvious.

Proposition 25. *The set of q-chains for $P = \{\tau\}$, where τ has no self-overlaps, is in one-to-one correspondence with the set of all total orderings on posets $\Pi_{q,\tau}$.*

Now we shall see how this approach can be applied in some cases.

4.2.2. *Case of the pattern* $12\cdots a\ \tau\ (a+1)$. Let $a < m$, and let $12\cdots a\ \tau\ (a+1)$ be a permutation of length $m+1$ that starts with the increasing run $1, 2, \ldots, a$, followed by some permutation τ of $(a+2), \ldots, m+1$, followed by the number $(a+1)$. Clearly, this pattern has no self-overlaps, so to enumerate chains we may count total orderings of posets. Note that every $(q+1)$-chain for $q \geq 0$ is of length $q(m+1) - (q-1) = qm + 1$.

Proposition 26. *For $P = \{12\cdots a\ \tau\ (a+1)\}$, the number of $(q+1)$-chains is equal to*

$$\prod_{j=1}^{q} \binom{jm-a}{m-a}.$$

Proof. This proof serves as a starting example of how to use posets to study chains. The poset $\Pi_{q,\tau}$ in this case looks like a tree of height $m+1$ with the only branch growing on the height $a+1$, this branch being of length $m+1$ and having a smaller branch growing at the distance $a+1$ from the starting point, etc. (An example of such a poset for the case of the permutation 1243 with $a=2$, $m=3$ is given above.) To extend such a partial ordering to a total ordering, we should make the lowest $a+1$ elements for such a tree the smallest elements $1, 2, \ldots, a+1$ of the resulting ordering. Then, there are $\binom{qm-a}{m-a}$ ways to choose $(m+1)-(a+1)=m-a$ remaining elements forming the stem of our tree, and we are left with the same question for a smaller tree, where we may proceed by induction. $\square$

Corollary 27 (see [Elizalde and Noy 2003; Kitaev 2005] for $t=0$). *For $a<m$, the multiplicative inverse of the generating function $g_P(t)$ of permutations avoiding $12\cdots a\ \tau\ (a+1) \in S_{m+1}$ is given by the formula*

$$1-t-\sum_{q\geq 1}\frac{(-1)^{q+1}t^{qm+1}}{(qm+1)!}\prod_{j=1}^{q}\binom{jm-a}{m-a}. \tag{7}$$

In particular, all these patterns, for different τ, are Wilf equivalent to each other.

Except for the case of the pattern $123 \simeq_W 321$, this covers all patterns of length 3, because $132 \simeq_W 312 \simeq_W 231 \simeq_W 213$ (the equivalence provided by either reversing the order of entries in the pattern from the left to the right, or reversing the relative order of entries in the pattern). We shall deal with the pattern 123 and, more generally, $12\cdots a$, in further sections.

4.2.3. *Case of one arbitrary pattern without self-overlaps.* Generalizing the previous result, let us consider an arbitrary pattern τ of length $m+1$ without self-overlaps. For such a pattern, every $(q+1)$-chain for $q\geq 0$ is still of length $qm+1$. The following result was conjectured in [Elizalde 2004], where it was proved in some particular cases. Another proof in the general case was, as we discovered after the first version of this paper got in circulation, obtained by Adrian Duane and Jeffrey Remmel [2011]; it is based on entirely different techniques developed in [Mendes and Remmel 2006].

Theorem 28. *For a pattern τ of length $m+1$ without self-overlaps, the number of permutations of length n with k occurrences of τ depends only on n, k, m, $\tau(1)$, and $\tau(m+1)$. In other words, two non-self-overlapping permutations of length $m+1$ are equivalent if their first and last entries are the same.*

Proof. Since for patterns without self-overlaps clusters coincide with chains, and cluster inversion can be used to count permutations with a given number of occurrences of forbidden patterns [Goulden and Jackson 1979], it is enough to show that the number of $(q+1)$-chains depends only on the first and the last entry of τ. This result is also very easy to derive using posets. To make formulas compact, let us put $a=\tau(1)-1$ and $b=\tau(m+1)-1$. The poset $\Pi_{q,\tau}$ whose total orderings enumerate q-chains is obtained from q totally ordered sets of cardinality $m+1$ as follows: The element $a+1$ of the second set is identified with the element $b+1$ of the first set, the element $a+1$ of the third set is identified with the element $b+1$ of the second set, etc. Clearly, this poset depends only on m, a, and b.

The actual number of q-chains in this case can be computed as follows. Let us denote by $f_k(p)$ the number of q-chains σ whose first element is $p+1$. Then it is easy to see that the following recurrence relation holds (here we assume, without the loss of generality, that $a<b$):

$$f_k(p)=\sum_q \binom{p}{a}\binom{km-q}{m-b}\binom{q-p-1}{b-a-1} f_{k-1}(q-b). \tag{8}$$

Writing $q+1=\sigma(m+1)$, there are $\binom{p}{a}$ ways to choose elements less than $p+1$ in the first pattern in the chain, $\binom{km-q}{m-b}$ ways to choose elements greater than $\sigma(m+1)$ there, $\binom{q-p-1}{b-a-1}$ to fill the space between these elements, and $f_{k-1}(q-b)$ ways to choose the remaining $(k-1)$-chain. □

Example 29. Theorem 28 shows that the two patterns 23154 and 21534 are equivalent to each other. Computing the first ten cluster numbers and inverting the corresponding series, we get the first ten entries 1, 1, 2, 6, 24, 119, 708, 4914, 38976, 347776 of the sequence counting permutations that avoid either of them.

4.2.4. *Case of one pattern of length* 4. Let us now consider the case of a single pattern of length 4. The equivalence classes of these are as follows [Elizalde 2004]:

I. $1234 \simeq 4321$.

II. $2413 \simeq 3142$.

III. $2143 \simeq 3412$.

IV. $1324 \simeq 4231$.

V. $1423 \simeq 3241 \simeq 4132 \simeq 2314$.

VI. $1342 \simeq 2431 \simeq 4213 \simeq 3124 \simeq 1432 \simeq 2341 \simeq 4123 \simeq 3214$.

VII. $1243 \simeq 3421 \simeq 4321 \simeq 2134$.

The case I will be considered later. In each of the cases VI and VII, the pattern has no self-overlaps, so Corollary 27 applies.

A very special feature of all patterns of length 4 (except for the case I) is that they have self-overlaps of length at most 2, so however we try to link several patterns together, it will be automatically true that only neighbors overlap. Moreover, even if we are dealing with a pattern τ with self-overlaps, every labeling of a linking scheme that is compatible with ordering of each of the patterns gives a genuine chain. Assume that γ is a linking scheme for q copies of τ. By induction, we may assume that the linking scheme provided by the first $q-1$ traces of τ only gives chains, and we only need to check the chain condition for the terminal segment, for which the statement follows from the fact that if two patterns of length 4 overlap by a segment of length 1 or 2, then every pattern of length 4 overlapping with the both of them overlaps with at least one of them by a segment of length 3. Guided by this observation, we compute all the exponential generating functions of consecutive pattern avoidance. Since in this case chains coincide with clusters, our results can be easily adapted for enumeration of permutations with a given number of occurrences of a given pattern.

Theorem 30. *The numbers $c_{n,l}$ for the pattern* 1324 *satisfy the recurrence relations*

$$c_{n,l} = \sum_{4\le 2k+2\le n} \frac{1}{k+1}\binom{2k}{k} c_{n-2k-1,l-k} \tag{9}$$

with initial conditions $c_{1,l} = \delta_{0,l}$ (the Kronecker delta symbol), $c_{2,l} = 0,\ \ c_{3,l} = 0$. Consequently, the generating function for avoidance of 1324 *is*

$$\Big(1 - t - \sum_{n\ge 2, l\ge 1} \frac{c_{n,l}t^n(-1)^l}{n!}\Big)^{-1}.$$

Proof. As we discussed above, counting chains is reduced to counting total orderings of the corresponding posets. Let us assume that the first $k+1$ patterns have two-element overlaps, and the following overlap involves just one element. For a chain $\sigma = a_1a_2\cdots a_{2k+1}a_{2k+2}a_{2k+3}\cdots$, this means that

$$a_1 < a_3 < a_2 < a_4, \quad a_3 < a_5 < a_4 < a_6,\ \ldots,\ a_{2k-1} < a_{2k+1} < a_{2k} < a_{2k+2}, \tag{10}$$

that $\{a_1, \ldots, a_{2k+2}\} = \{1, \ldots, 2k+2\}$, and that $\mathrm{st}(a_{2k+2}a_{2k+3}\cdots)$ is an $(l-k)$-chain. To prove (9), we notice that the number of permutations $a_1a_2\cdots a_{2k+2}$ of $\{1, \ldots, 2k+2\}$ for which the conditions (10) are satisfied is given by the number of standard Young tableaux of size $2\times k$: Clearly, $a_1 = 1,\ \ a_{2k+2} = 2k+2$, and

$$a_2, a_3, a_4, \ldots, a_{2k+1} \quad \longleftrightarrow \quad \begin{array}{|c|c|c|c|c|}\hline a_3 & a_5 & a_7 & \cdots & a_{2k+1} \\ \hline a_2 & a_4 & a_6 & \cdots & a_{2k} \\ \hline\end{array}$$

gives a bijection with standard Young tableaux. The number of such tableaux is equal to the Catalan number $\frac{1}{k+1}\binom{2k}{k}$ (see, for example [Stanley 1999]), and the recurrence relation (9) follows. □

Example 31. Computing the first ten of those numbers and inverting the corresponding series, we get the first ten entries 1, 1, 2, 6, 23, 110, 632, 4229, 32337, 278204 of the sequence, which is indeed counting permutations that avoid 1324 (A113228 in [Sloane 2010]).

Theorem 32. *The numbers $c_{n,l}$ for the pattern* 1423 *satisfy the recurrence relations*

$$c_{n,l} = \sum_{4 \leq 2k+2 \leq n} \binom{n-k-2}{k} c_{n-2k-1,l-k} \tag{11}$$

with initial conditions $c_{1,l} = \delta_{0,l},\ c_{2,l} = 0,\ c_{3,l} = 0$. Consequently, the generating function for avoidance of 1423 *is*

$$\Big(1 - t - \sum_{n \geq 2, l \geq 1} \frac{c_{n,l}(-1)^l}{n!}\Big)^{-1}.$$

Proof. Similarly to the proof of Theorem 30, counting chains is reduced to counting total orderings of the corresponding posets. Let us assume that the first $k+1$ patterns have two-element overlaps, and the following overlap involves just one element. For a chain $\sigma = a_1 a_2 \cdots a_{2k+1} a_{2k+2} a_{2k+3} \cdots$, this means that

$$a_1 < a_3 < a_4 < a_2,\ a_3 < a_5 < a_6 < a_4,\ \ldots,\ a_{2k-1} < a_{2k+1} < a_{2k+2} < a_{2k}, \tag{12}$$

so

$$a_1 < a_3 < \cdots < a_{2k-1} < a_{2k+1} < a_{2k+2} < a_{2k} < \cdots < a_4 < a_2, \tag{13}$$

$\{a_1, a_3, \ldots, a_{2k+1}\} = \{1, 2, \ldots, k+1\}$, $a_{2k+2} = k+2$, and $\mathrm{st}(a_{2k+2}a_{2k+3}\cdots)$ is an $(l-k)$-chain. To prove (11), we notice that the number of ways to distribute numbers between the increasing sequence (13) and the $(l-k)$-chain $\mathrm{st}(a_{2k+2}a_{2k+3}\cdots)$ is equal to the number of way to choose the k numbers $a_{2k}, \ldots, a_2$. The latter is clearly the binomial coefficient $\binom{n-k-2}{k}$, and the recurrence relation (11) follows. □

Example 33. Computing the first ten of those numbers and inverting the corresponding series, we get the first ten entries 1, 1, 2, 6, 23, 110, 631, 4218, 32221, 276896 of the sequence counting permutations that avoid 1423.

Theorem 34. *The numbers $c_{n,l}$ for the pattern* 2143 *satisfy the recurrence relations*

$$c_{n,l} = \sum_{2 \leq p < n-2} c_{n,l}(p),$$

where the numbers $c_{n,l}(p)$ satisfy the recurrence relations

$$c_{n,l}(p) = \sum_{4 \leq 2k+2 \leq q \leq n} \binom{q-p-1}{2k-2} (p-1)(n-q) c_{n-2k-1,l-k}(q-2k) \tag{14}$$

with initial conditions $c_{1,l}(p) = \delta_{0,l}\delta_{1,p},\ c_{2,l}(p) = 0,\ c_{3,l}(p) = 0$. Consequently, the generating function for avoidance of 2143 *is*

$$\Big(1 - t - \sum_{n \geq 2, l \geq 1} \frac{c_{n,l}t^n(-1)^l}{n!}\Big)^{-1}.$$

Proof. Similarly to the proof of Theorem 30, counting chains is reduced to counting total orderings of the corresponding posets. Let $c_{n,l}(p)$ be the number of l-chains σ of length n with $\sigma(1) = p$. Let us assume that the first $k+1$ copies of 2143 in σ have two-element overlaps, and the following overlap involves just one element. For a chain $\sigma = a_1 a_2 \cdots a_{2k+1} a_{2k+2} a_{2k+3} \cdots$, this means that

$$a_2 < a_1 < a_4 < a_3,\ a_4 < a_3 < a_6 < a_5,\ \ldots,\ a_{2k} < a_{2k-1} < a_{2k+2} < a_{2k+1}, \tag{15}$$

so

$$a_2 < a_1 < a_4 < a_3 < \cdots < a_{2k} < a_{2k-1} < a_{2k+2} < a_{2k+1}, \tag{16}$$

and $\mathrm{st}(a_{2k+2}a_{2k+3}\cdots)$ is an $(l-k)$-chain. Assume that $a_1 = p$. To prove (14), we notice that if $a_{2k+2} = q$, then there are $\binom{q-p-1}{2k-2}$ ways to pick the numbers $a_3, \ldots, a_{2k}$, $p-1$ ways to pick a_2, $(n-q)$ ways to pick a_{2k+1}, and $c_{n-2k-1,l-k}$ ways to pick the remaining $(l-k)$-chain (where the entry q is the $(q-2k)$-th biggest). This completes the proof. □

Example 35. Computing the first ten of those numbers and inverting the corresponding series, we get the first ten entries 1, 1, 2, 6, 23, 110, 631, 4223, 32301, 277962 of the sequence counting permutations that avoid 2143.

In the last remaining case $2413 \simeq 3142$ (II in the list above), we have no trick like above that would simplify the computations, so we shall use the most general strategy for chain enumeration, which allows us to compute the chain numbers rather fast (polynomially in n) for all sets of forbidden patterns. There is an obvious similarity with the approach in [Kitaev and Mansour 2005].

Theorem 36. *The numbers $c_{n,l}$ for the pattern* 2413 *are given by the formulas*

$$c_{n,l} = \sum_{1<p<q-1<n} c_{n,l}(p,q),$$

where the numbers $c_{n,l}(p,q)$ satisfy the recurrence relations

$$c_{n,l}(p,q) = \sum_{r<p<s<q} c_{n-2,l-1}(r, s-1)$$
$$+ \sum_{p<r<s<q} (p-1)c_{n-3,l-1}(r-1, s-1) + \sum_{p<r<q<s} (p-1)c_{n-3,l-1}(r-1, s-2) \tag{17}$$

with initial conditions $c_{2,l}(p,q) = 0$, $c_{3,l}(p,q) = 0$, $c_{4,l}(p,q) = \delta_{l,1}\delta_{p,2}\delta_{q,4}$. Consequently, the generating function for avoidance of 2143 *is*

$$\left(1 - t - \sum_{n\geq 2, l\geq 1} \frac{c_{n,l} t^n (-1)^l}{n!}\right)^{-1}.$$

Proof. This statement is straightforward. Let us consider an n-chain $\sigma = a_1a_2a_3\cdots$. The first pattern in that chain intersects with its neighbor by either two or one elements. In the first case, we have $a_3 < a_2 < a_4 < a_1$, so if we fix a_1 and a_2 and forget about them, we are left with an $(n-1)$-chain, and we should sum over all choices of a_3 and a_4 for its first entries. If, on the contrary, the first overlap uses just one element, then there are $(a_1 - 1)$ choices for a_3, and we should distinguish between the cases $a_5 > a_2$ and $a_5 < a_2$: In the first case a_5 is the (a_5-1)-st biggest in the remaining cluster, while in the second case it is the (a_5-2)-nd biggest. □

Example 37. Computing the first ten of those numbers and inverting the corresponding series, we get the first ten entries 1, 1, 2, 6, 23, 110, 632, 4237, 32465, 279828 of the sequence counting permutations that avoid 2413.

4.2.5. *Case of two patterns* $\{132, 231\}$.

Theorem 38. *The number $c_{n,l}$ for $P = \{132, 231\}$ is not equal to zero only for $n = 2l + 1$, and in this case is equal to E_{2l+1}, the tangent number* [Stanley 1999], *so the generating function for avoidance of* $\{132, 231\}$ *is*

$$(1 - \tanh t)^{-1}. \tag{18}$$

Proof. This pair of patterns has no self-overlaps at all (both for a pattern with itself, and two patterns with each other), so every linking scheme clearly provides only chains. Clearly, chains are nothing but "up-down" permutations, that is, permutations $a_1a_2\cdots a_{2l}a_{2l+1}$ for which

$$a_1 < a_2 > a_3 < a_4 > \cdots < a_{2l} > a_{2l+1}.$$

It is well known that the number of such permutations is equal to the tangent number. □

4.2.6. *Case of the pattern* $12\cdots k$. The case we consider in this section is the case of the single pattern $12\cdots k$, which marks increasing runs of length k in permutations. The enumeration result in this case is well known, however, we want to show that it can also be obtained as a direct application of our results.

Theorem 39 [Elizalde and Noy 2003; Goulden and Jackson 1983; Kitaev 2005]. *The multiplicative inverse of the exponential generating function for patterns avoiding* $12\cdots k$ *is given by the formula*

$$\sum_{q\geq 0}\frac{x^{kq}}{(kq)!} - \sum_{q\geq 0}\frac{x^{kq+1}}{(kq+1)!}. \tag{19}$$

Proof. Indeed, q-chains for $q \geq 2$ are as follows:

- the only 2-chain is $12\cdots k$,
- the only 3-chain is $12\cdots(k+1)$,

- the only 4-chain is $12\cdots(2k)$,
- the only 5-chain is $12\cdots(2k+1)$,
- $\cdots$
- the only $(2l)$-chain is $12\cdots(kl)$,
- the only $(2l+1)$-chain is $12\cdots(kl+1)$,
- $\cdots$. □

4.2.7. *Case of the pattern* $\lambda(\lambda+m)\cdots(\lambda+(k-1)m)$. The result of this section gives one way to somewhat generalize both Theorem 28 and Theorem 39. Let λ be a pattern of length m without self-overlaps. Denote by $\lambda+j$ the permutation of numbers $\{j+1,\ldots,j+m\}$ obtained by adding j to each entry of λ. Let $\tau=\tau_{k,\lambda}=\lambda(\lambda+m+1)\cdots(\lambda+(k-1)m)$ be the "ordered sum" of k copies of λ.

Theorem 40. *The number of permutations of length n avoiding τ depends only on n, m, $\tau(1)$, $\tau(m)$, and k. In other words, for two non-self-overlapping patterns of length m the corresponding k-fold ordered sums are Wilf equivalent if their first and last entries are the same.*

Proof. For the k-fold ordered sum of a pattern without self-overlaps, it is very easy to exhibit the linking schemes that actually give rise to chains. Such a linking scheme is a genuine mixture of linking schemes for patterns without self-overlaps and linking schemes for the pattern $12\cdots k$. Namely, for each $l\geq 2$ there is one basic "building block", a linking scheme modeled on the l-chains

- $\lambda(\lambda+m)\cdots(\lambda+(k-1)m)$ for $l=2$,
- $\lambda(\lambda+m)\cdots(\lambda+(k-1)m)(\lambda+km)$ for $l=3$,
- $\lambda(\lambda+m)\cdots(\lambda+(2k-2)m)(\lambda+(2k-1)m)$ for $l=4$,
- $\lambda(\lambda+m)\cdots(\lambda+(2k-1)m)(\lambda+2km)$ for $l=5$,
- $\cdots$
- $\lambda(\lambda+m)\cdots(\lambda+(pk-2)m)(\lambda+(pk-1)m)$ for $l=2p$,
- $\lambda(\lambda+m)\cdots(\lambda+(pk-1)m)(\lambda+pkm)$ for $l=2p+1$,
- $\cdots$,

and every linking scheme producing a chain is a linkage of several building blocks like that overlapping only by one element. The poset defined by such a linking scheme obviously depends only on the first and the last element of τ but not on the relative order of other elements. The corresponding recurrence relations can easily be derived from this description as well. □

4.2.8. *Case of the pattern* $12\cdots k$ *and a pattern without self-overlaps.* This section gives another way to somewhat generalize both Theorem 28 and Theorem 39. Let λ be a pattern of length m without self-overlaps. We shall study the enumeration problem for avoidance of $P_{\lambda,k} = \{\lambda, 12\cdots k\}$. Let us introduce several parameters important for enumeration. Denote by $l_I(\lambda)$ the length of the maximal initial segment of λ that is an increasing rise, and by $l_T(\lambda)$ the length of the maximal terminal segment of λ that is an increasing rise. Since we always assume patterns of P to not contain one another, and we assume λ to have no self-overlaps, we conclude that $l_I(\lambda), l_T(\lambda) < k,\ l_I(\lambda) + l_T(\lambda) < m$ and $\min(l_I(\lambda), l_T(\lambda)) = 1$.

Theorem 41. *The number of permutations of length n avoiding* $P_{\lambda,k}$ *depends only on* m, $\lambda(1)$, $\lambda(m)$, $l_I(\lambda)$, $l_T(\lambda)$, *and* k. *In particular, if we adjoin to two non-self-overlapping patterns* λ_1 *and* λ_2 *of the same length m an increasing rise of length k, the corresponding two-element sets are Wilf equivalent if the first and last entries, and the lengths of the initial and terminal increasing rises of* λ_1 *and* λ_2 *are the same.*

Proof. Both reversing the direction in which we read permutations (left-to-right becomes right-to-left) and reversing the order of entries (increasing becomes decreasing) in all permutations considered preserve Wilf classes, and doing both these changes keeps the permutation $12\cdots k$ intact, we may assume that $l_T(\lambda) = 1$.

It is easy to exhibit the linking schemes that actually give rise to chains. Basically, there are two basic types of "building blocks" for the linking schemes: A linking scheme modeled on a single copy of λ and linking schemes modeled on chains for a single pattern $12\cdots k$, as in the proof of Theorem 39. There is no freedom in linking copies of λ together: Since λ has no self-overlaps, two copies of λ may only overlap by a single element. Since we assume that $l_T(\lambda) = 1$, we conclude that an occurrence of λ can only overlap with a building block coming from an overlap of several rises by a single element as well. For an overlap of several rises followed by an occurrence of λ the situation is different. Namely, if we are talking about the scheme modeled on the $(2l)$-chain $12\cdots(kl)$, it should overlap with the following copy of λ by the initial increasing rise of that copy, that is, by the first $l_I(\lambda)$ elements (since no proper beginning of a q-chain may be a q-chain). However, for the scheme modeled on the $(2l+1)$-chain $12\cdots(kl+1)$, it should overlap with the following copy of λ by a single element (since only neighboring patterns in a chain may overlap). Similarly to the proof of Theorem 28, the posets defined by such linking schemes are completely determined by the first and the last entry of λ, and the lengths of its initial and terminal increasing rises. □

Acknowledgments

We wish to thank Sergi Elizalde and Sergey Kitaev for their remarks about consecutive patterns, and Jean-Louis Loday and Maria Ronco for useful conversations on shuffle algebras. We are also grateful to Dmitri Piontkovski who drew our attention to the fact that some of our techniques had previously been used in [Piotkovskii 1993] in the case of associative algebras. The work on this paper started when the first author was working at, and the second author was visiting, Dublin Institute for Advanced Studies; they express their deepest gratitude to all the staff there for their hospitality.

References

[Anick 1986] D. J. Anick, "On the homology of associative algebras", *Trans. Amer. Math. Soc.* **296**:2 (1986), 641–659. MR 87i:16046 Zbl 0598.16028

[Baxter et al. 2011] A. Baxter, B. Nakamura, and D. Zeilberger, "Automatic generation of theorems and proofs on enumerating consecutive-Wilf classes", preprint, 2011. arXiv 1101.3949

[Bell and Goh 2007] J. P. Bell and T. L. Goh, "Exponential lower bounds for the number of words of uniform length avoiding a pattern", *Information and Computation* **205**:9 (2007), 1295–1306. MR 2008h:68093 Zbl 1127.68073

[Bell and Small 2002] J. P. Bell and L. W. Small, "A question of Kaplansky", *J. Algebra* **258**:1 (2002), 386–388. MR 2003m:16030 Zbl 1020.16014

[Dotsenko and Khoroshkin 2010] V. Dotsenko and A. Khoroshkin, "Gröbner bases for operads", *Duke Math. J.* **153**:2 (2010), 363–396. MR 2011h:18010 Zbl 1208.18007

[Dotsenko and Khoroshkin 2012] V. Dotsenko and A. Khoroshkin, "Quillen homology via Gröbner bases", preprint, 2012. arXiv 1203.5053

[Dotsenko and Vejdemo Johansson 2012] V. Dotsenko and M. Vejdemo Johansson, "Implementing Gröbner bases for operads", pp. 77–98 in *Operads 2009* (Luminy, 2009), edited by J.-L. Loday and B. Vallette, Séminaires et Congrès **26**, Société Mathématique de France, Paris, 2012.

[Duane and Remmel 2011] A. Duane and J. Remmel, "Minimal overlapping patterns in colored permutations", *Electron. J. Combin.* **18**:2 (2011), paper 25. MR 2861417 Zbl 1229.05017

[Elizalde 2004] S. Elizalde, *Consecutive patterns and statistics on restricted permutations*, thesis, Universitat Politècnica de Catalunya, 2004, available at http://hdl.handle.net/10803/5839.

[Elizalde 2006] S. Elizalde, "Asymptotic enumeration of permutations avoiding generalized patterns", *Adv. in Appl. Math.* **36**:2 (2006), 138–155. MR 2006h:05004 Zbl 1089.05007

[Elizalde and Noy 2003] S. Elizalde and M. Noy, "Consecutive patterns in permutations", *Adv. in Appl. Math.* **30**:1-2 (2003), 110–125. MR 2004d:05004 Zbl 1016.05002

[Etingof and Ginzburg 2007] P. Etingof and V. Ginzburg, "Noncommutative complete intersections and matrix integrals", *Pure Appl. Math. Q.* **3**:1 (2007), 107–151. MR 2008b:16044 Zbl 1151.14006

[Golod and Shafarevich 1964] E. S. Golod and I. R. Shafarevich, "On the class field tower", *Izv. Akad. Nauk SSSR Ser. Mat.* **28** (1964), 261–272. In Russian. MR 28 #5056 Zbl 0136.02602

[Goulden and Jackson 1979] I. P. Goulden and D. M. Jackson, "An inversion theorem for cluster decompositions of sequences with distinguished subsequences", *J. London Math. Soc.* (2) **20**:3 (1979), 567–576. MR 83b:05005 Zbl 0467.05008

[Goulden and Jackson 1983] I. P. Goulden and D. M. Jackson, *Combinatorial enumeration*, John Wiley, New York, 1983. Zbl 0519.05001

[Kemer 1988] A. R. Kemer, "Solution of the problem as to whether associative algebras have a finite basis of identities", *Sov. Math., Dokl.* **37**:1 (1988), 60–64. MR 89d:16023 Zbl 0739.16016

[Khoroshkin and Shapiro 2011] A. Khoroshkin and B. Shapiro, "Using homological duality in consecutive pattern avoidance", *Electron. J. Combin.* **18**:2 (2011), paper 9. MR 2012e:05031 Zbl 1217.05025

[Kitaev 2005] S. Kitaev, "Partially ordered generalized patterns", *Discrete Math.* **298**:1-3 (2005), 212–229. MR 2006f:05007 Zbl 1070.05002

[Kitaev and Mansour 2003] S. Kitaev and T. Mansour, "A survey of certain pattern problems", preprint, 2003, available at https://personal.cis.strath.ac.uk/sergey.kitaev/index_files/Papers/survey.ps.

[Kitaev and Mansour 2005] S. Kitaev and T. Mansour, "On multi-avoidance of generalized patterns", *Ars Combin.* **76** (2005), 321–350. MR 2006a:05011 Zbl 1164.05300

[Latyshev 2005] V. N. Latyshev, "A general version of standard basis and its application to T-ideals", *Acta Appl. Math.* **85**:1-3 (2005), 219–223. MR 2006a:16031 Zbl 1085.16015

[Latyshev 2008] V. N. Latyshev, "Combinatorial generators of multilinear polynomial identities", *J. Math. Sci. (N. Y.)* **149**:2 (2008), 1107–1112. MR 2007e:16030 Zbl 1163.16013

[Loday and Ronco 2010] J.-L. Loday and M. Ronco, "Combinatorial Hopf algebras", pp. 347–383 in *Quanta of maths* (Paris, 2007), edited by E. Blanchard et al., Clay Math. Proc. **11**, American Mathematical Society, Providence, RI, 2010. MR 2011k:16068 Zbl 1217.16033

[Loday and Ronco 2011] J.-L. Loday and M. Ronco, "Permutads", preprint, 2011. arXiv 1105.5271

[Malvenuto and Reutenauer 1995] C. Malvenuto and C. Reutenauer, "Duality between quasi-symmetric functions and the Solomon descent algebra", *J. Algebra* **177**:3 (1995), 967–982. MR 97d:05277 Zbl 0838.05100

[Mansour 2001/02] T. Mansour, "Pattern avoidance in coloured permutations", *Sém. Lothar. Combin.* **46** (2001/02), B46g. MR 2002j:05004 Zbl 0986.05004

[Mendes and Remmel 2006] A. Mendes and J. Remmel, "Permutations and words counted by consecutive patterns", *Adv. in Appl. Math.* **37**:4 (2006), 443–480. MR 2008c:05003 Zbl 1110.05005

[Nakamura 2011] B. Nakamura, "Computational approaches to consecutive pattern avoidance in permutations", preprint, 2011. arXiv 1102.2480

[Noonan and Zeilberger 1999] J. Noonan and D. Zeilberger, "The Goulden–Jackson cluster method: extensions, applications and implementations", *J. Differ. Equations Appl.* **5**:4-5 (1999), 355–377. MR 2000f:05005 Zbl 0935.05003

[Piotkovskii 1993] D. I. Piotkovskii, "On the growth of graded algebras with a small number of defining relations", *Russian Math. Surveys* **48**:3 (1993), 211–212. MR 94g:16031 Zbl 0810.16022

[Rampersad 2011] N. Rampersad, "Further applications of a power series method for pattern avoidance", *Electron. J. Combin.* **18**:1 (2011), paper 134. MR 2012e:68300 Zbl 1219.68128

[Ronco 2011] M. Ronco, "Shuffle bialgebras", *Ann. Inst. Fourier* (*Grenoble*) **61**:3 (2011), 799–850. MR 2918719 Zbl 1239.16032

[Sloane 2010] N. J. A. Sloane, *On-line encyclopedia of integer sequences*, 2010, available at http://oeis.org/.

[Specht 1950] W. Specht, "Gesetze in Ringen, I", *Math. Z.* **52** (1950), 557–589. MR 11,711i Zbl 0032.38901

[Stanley 1999] R. P. Stanley, *Enumerative combinatorics*, vol. 2, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, 1999. MR 2000k:05026 Zbl 0928.05001

[Steingrímsson 2010] E. Steingrímsson, "Generalized permutation patterns—a short survey", pp. 137–152 in *Permutation patterns* (St. Andrews, 2007), edited by S. Linton et al., London Math. Soc. Lecture Note Ser. **376**, Cambridge Univ. Press, 2010. MR 2011m:05009 Zbl 1217.05017

[Stover 1993] C. R. Stover, "The equivalence of certain categories of twisted Lie and Hopf algebras over a commutative ring", *J. Pure Appl. Algebra* **86**:3 (1993), 289–326. MR 94e:16031 Zbl 0793.16016

[Ufnarovskij 1995] V. A. Ufnarovskij, "Combinatorial and asymptotic methods in algebra", pp. 1–196 in *Algebra, VI*, vol. 4, edited by I. R. Shafarevich, Encyclopaedia Math. Sci. **57**, Springer, Berlin, 1995. MR 96f:00042 Zbl 0826.16001

[Wilf 2002] H. S. Wilf, "The patterns of permutations", *Discrete Math.* **257**:2-3 (2002), 575–583. MR 2003h:05005 Zbl 1028.05002

Communicated by Victor Reiner
Received 2011-09-13 Accepted 2012-04-08

vdots@maths.tcd.ie *School of Mathematics, Trinity College, Dublin 2, Ireland*

khorosh@itep.ru *Simons Center for Geometry and Physics, Stony Brook University, Stony Brook, NY 11794-3636, United States*

ITEP, Bolshaya Cheremushkinskaya 25, 117259, Moscow, Russia

ALGEBRA AND NUMBER THEORY 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.701

Preperiodic points for families of polynomials

Dragos Ghioca, Liang-Chung Hsia and Thomas J. Tucker

Let $a(\lambda), b(\lambda) \in \mathbb{C}[\lambda]$, and let $f_\lambda(x) \in \mathbb{C}[x]$ be a one-parameter family of polynomials indexed by all $\lambda \in \mathbb{C}$. We study whether there exist infinitely many $\lambda \in \mathbb{C}$ such that both $a(\lambda)$ and $b(\lambda)$ are preperiodic for f_λ.

1. Introduction

The classical Manin–Mumford conjecture for abelian varieties (now a theorem due to Raynaud [1983a; 1983b]) predicts that the set of torsion points of an abelian variety A defined over $\mathbb{C}$ is not Zariski dense in a subvariety V of A unless V is a translate of an algebraic subgroup of A by a torsion point. Pink and others have suggested extending the Manin–Mumford conjecture to a more general question regarding unlikely intersections between a subvariety V of a semiabelian scheme A and algebraic subgroups of the fibers of A having codimension greater than the dimension of V [Bombieri et al. 1999; Habegger 2009; Masser and Zannier 2010; Masser and Zannier 2012; Pink 2005]. Here we state a special case of the question when V is a curve:

Question 1.1. Let $\mathcal{S}$ be a semiabelian scheme over a variety $\mathcal{Y}$ defined over $\mathbb{C}$, and let $V \subset \mathcal{S}$ be a curve that is not contained in any proper algebraic subgroup of $\mathcal{S}$. We define

$$\mathcal{S}^{[2]} := \bigcup_{y \in \mathcal{Y}} B_y,$$

where B_y is the union of all algebraic subgroups of the fiber $\mathcal{S}_y$ of codimension at least equal to 2. Must the intersection of V with $\mathcal{S}^{[2]}$ be finite?

Ghioca was partially supported by an NSERC Discovery Grant. Hsia was partially supported by the National Center of Theoretical Sciences of Taiwan and NSC grant 99-2115-M-008-007-MY3. Tucker was partially supported by NSF grants 0801072 and 0854839.
MSC2010: primary 37P05; secondary 37P10.
Keywords: preperiodic points, heights.

Bertrand [2011] recently showed that the answer to Question 1.1 is sometimes "no". The question may, however, have a positive answer in many instances. For example, Masser and Zannier [2010; 2012] study Question 1.1 when $\mathcal{S}$ is the square of the Legendre family of elliptic curves E_λ (over the base $\mathbb{A}^1 \setminus \{0, 1\}$) given by the equation $y^2 = x(x-1)(x-\lambda)$. They show that for any two independent points P and Q on the generic fiber, there are at most finitely many $\lambda \in \mathbb{C}$ such that the specializations P_λ and Q_λ are both torsion points for E_λ. Their work thus gives a positive answer to Question 1.1 in this special case.

The result of Masser and Zannier has a distinct dynamical flavor. Indeed, one may consider the following more general problem. Let $\{X_\lambda\}$ be an algebraic family of quasiprojective varieties defined over $\mathbb{C}$, let $\Phi_\lambda : X_\lambda \to X_\lambda$ be an algebraic family of endomorphisms, and let $P_\lambda \in X_\lambda$ and $Q_\lambda \in X_\lambda$ be two algebraic families of points. Under what conditions do there exist infinitely many λ such that both P_λ and Q_λ are preperiodic for Φ_λ? Indeed, the problem from [Masser and Zannier 2010; 2012] fits into this general dynamical framework by letting $X_\lambda = E_\lambda$ be the Legendre family of elliptic curves and letting Φ_λ be the multiplication-by-2 map on each elliptic curve in this family.

Baker and DeMarco [2011] study an interesting special case of the general dynamical question above first suggested by Zannier at an American Institute of Mathematics workshop in 2008. Given complex numbers a and b and an integer $d \geq 2$, when do there exist infinitely many $\lambda \in \mathbb{C}$ such that both a and b are preperiodic for the action of $f_\lambda(x) := x^d + \lambda$ on $\mathbb{C}$? They show that this happens if and only if $a^d = b^d$. We prove this generalization of the main result of [Baker and DeMarco 2011]:

Theorem 1.2. *Let $f \in \mathbb{C}[x]$ be any polynomial of degree $d \geq 2$, and let $a, b \in \mathbb{C}$. Then there exist infinitely many $\lambda \in \mathbb{C}$ such that both a and b are preperiodic for $f(x) + \lambda$ if and only if $f(a) = f(b)$.*

We will derive Theorem 1.2 from a more technical result, Theorem 2.3, which also treats the case of "nonconstant starting points" $\boldsymbol{a}$ and $\boldsymbol{b}$, a topic that was raised in [Baker and DeMarco 2011].

One might hope to formulate a general dynamical version of Question 1.1 for polarizable endomorphisms of projective varieties more general than multiplication-by-m maps on abelian varieties (an endomorphism Φ of a projective variety X is polarizable if there exists $d \geq 2$ and a line bundle $\mathcal{L}$ on X such that $\Phi^*(\mathcal{L})$ is linearly equivalent to $\mathcal{L}^{\otimes d}$ in $\operatorname{Pic}(X)$) by using the analogy between abelian subschemes and preperiodic subvarieties. Because of the results of Baker and DeMarco, along with the results of this paper, we believe it is reasonable to ask the following dynamical analog of Question 1.1:

Question 1.3. Let Y be any quasiprojective curve defined over $\mathbb{C}$, and let F be the function field of Y. Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{P}^1(F)$, and let $V \subset \mathscr{X} := \mathbb{P}^1_F \times_F \mathbb{P}^1_F$ be the $\mathbb{C}$-curve $(\boldsymbol{a}, \boldsymbol{b})$. Let $\boldsymbol{f} : \mathbb{P}^1 \to \mathbb{P}^1$ be a rational map of degree $d \geq 2$ defined over F. Then for all but finitely many $\lambda \in Y$, $\boldsymbol{f}$ induces a well-defined rational map $f_\lambda : \mathbb{P}^1 \to \mathbb{P}^1$ defined over $\mathbb{C}$. If there exist infinitely many $\lambda \in Y$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic points of $\mathbb{P}^1(\mathbb{C})$ under the action of f_λ, then must V be contained in a proper preperiodic subvariety of $\mathscr{X}$ under the action of $\Phi := (\boldsymbol{f}, \boldsymbol{f})$?

Theorem 1.2 is a special case of Question 1.3 for $f_\lambda(x) = f(x) + \lambda$ and constant starting points $\boldsymbol{a}(\lambda) = a$ and $\boldsymbol{b}(\lambda) = b$. Theorem 2.3 also allows us to prove some other special cases of Question 1.3 such as the following:

Theorem 1.4. *Let $f \in \mathbb{C}[x]$ be any polynomial of degree $d \geq 2$, let $g \in \mathbb{C}[x]$ be any nonconstant polynomial, and let $c \in \mathbb{C}^*$. Then there exist at most finitely many $\lambda \in \mathbb{C}$ such that either*

(1) *both $g(\lambda)$ and $g(\lambda + c)$ are preperiodic for $f(x) + \lambda$ or*

(2) *both $g(\lambda)$ and $g(\lambda) + c$ are preperiodic for $f(x) + \lambda$.*

The next result is for the case when the family of maps $\boldsymbol{f}$ is constant:

Theorem 1.5. *Let $f \in \mathbb{C}[x]$ be a polynomial of degree $d \geq 2$, and let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}[\lambda]$ be two polynomials of same degree and with the same leading coefficient. If there exist infinitely many $\lambda \in \mathbb{C}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f, then $\boldsymbol{a} = \boldsymbol{b}$.*

A special case of Theorem 1.5 is that for any fixed $c \in \mathbb{C}^*$, there can be only finitely many $\lambda \in \mathbb{C}$ such that both λ and $\lambda + c$ are preperiodic for f. In fact, more generally it provides a positive answer to a special case of Zhang's dynamical Manin–Mumford conjecture, which states that for a polarizable endomorphism $\Phi : X \to X$ on a projective variety, the only subvarieties of X containing a dense set of preperiodic points are those subvarieties that are themselves preperiodic under f; see [Zhang 1995, Conjecture 2.5; 2006, Conjecture 1.2.1, Conjecture 4.1.7] for details. This conjecture turns out to be false in general [Ghioca et al. 2011], but it may be true in many cases. For example, let $X := \mathbb{P}^1 \times \mathbb{P}^1$, $\Phi(x, y) := (f(x), f(y))$ for a polynomial f of degree $d \geq 2$, and Y be the Zariski closure in X of the set $\{(\boldsymbol{a}(z), \boldsymbol{b}(z)) : z \in \mathbb{C}\}$, where $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}[x]$ are polynomials of same degree and with the same leading coefficient; Theorem 1.5 implies that if Y contains infinitely many points preperiodic under Φ, then Y is the diagonal subvariety of X and thus is itself preperiodic under Φ. Theorem 1.5 also has consequences for a case of a revised dynamical Manin–Mumford conjecture [Ghioca et al. 2011, Conjecture 1.4]; see Section 11 for details.

The plan of our paper is as follows. In Section 2, we state our main result, Theorem 2.3, and some of its consequences and then describe the method of our proof. In Section 3, we set up our notation while in Section 4 we give a brief overview of Berkovich spaces. Then in Section 5, we introduce some basic preliminaries regarding the iterates of a generic starting point $\boldsymbol{c}$ under a family of maps $\boldsymbol{f}$. Section 6 contains computations of the capacities of the generalized v-adic Mandelbrot sets associated with a generic point $\boldsymbol{c}$ under the action of $\boldsymbol{f}$. In Section 7, we prove an explicit formula for the Green function for the generalized v-adic Mandelbrot sets when v is an archimedean valuation. We proceed with our proof of the direct implication in Theorem 2.3 in Section 8 (for the case $f_\lambda \in \overline{\mathbb{Q}}[x]$ and $\boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{Q}}[x]$) and in Section 10 (for the general case). In Section 9, we prove the converse implication from Theorem 2.3. Then in Section 11, we conclude our paper by proving Corollary 2.7 and discussing the connections between our Question 1.3 and the dynamical Manin–Mumford conjecture formulated by Ghioca, Tucker, and Zhang [2011].

2. Statement of the main results

A special case of Question 1.3 is when $Y = \mathbb{A}^1$, $\boldsymbol{f} \in R[x]$, where $R = \mathbb{C}[\lambda]$, and $\boldsymbol{a}, \boldsymbol{b} \in R$. In Theorem 2.3, we provide a positive answer to Question 1.3 for any family of polynomials of the form

$$f_\lambda(x) = x^d + \sum_{i=0}^{d-2} c_i(\lambda) x^i, \quad \text{where } c_i(\lambda) \in \mathbb{C}[\lambda] \text{ for } i = 0, \dots, d-2, \tag{2.1}$$

together with some mild restriction on the polynomials $\boldsymbol{a}$ and $\boldsymbol{b}$.

We say that a polynomial $f(x)$ of degree d is in *normal form* if it is monic and its coefficient of x^{d-1} equals 0. (Note that any polynomial of degree $d > 1$ can be put in normal form after a change of coordinates.) As a matter of notation, we rewrite (2.1) as

$$f_\lambda(x) = P(x) + \sum_{i=1}^{r} Q_i(x) \cdot \lambda^{m_i} \tag{2.2}$$

for some polynomial $P \in \mathbb{C}[x]$ in normal form of degree d, some nonnegative integer r, integers $m_0 := 0 < m_1 < \cdots < m_r$, and some polynomials $Q_i \in \mathbb{C}[x]$ of degrees $0 \le e_i \le d-2$. We do not exclude the case $r = 0$, in which case the sum in the sigma notation is empty and $\{f_\lambda\}_\lambda$ is a constant family of polynomials.

Let $\boldsymbol{a}(\lambda), \boldsymbol{b}(\lambda) \in \mathbb{C}[\lambda]$. If $\boldsymbol{a}$ is preperiodic for $\boldsymbol{f}$, that is, $\boldsymbol{f}^k(\boldsymbol{a}) = \boldsymbol{f}^\ell(\boldsymbol{a})$ for some $k \neq \ell$, then for each $\boldsymbol{b}$, one can show that there are infinitely many $\lambda \in \mathbb{C}$ such that $\boldsymbol{b}(\lambda)$ (and thus also $\boldsymbol{a}(\lambda)$) is preperiodic for f_λ (see also Proposition 9.1). Therefore, we may assume that $\boldsymbol{a}$ and $\boldsymbol{b}$ are not preperiodic for $\boldsymbol{f}$. Assuming there exist infinitely many $\lambda \in \mathbb{C}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f_λ, then Question 1.3 predicts that there exist φ_1 and φ_2 commuting with $\boldsymbol{f}$ such that

$\varphi_1(\boldsymbol{a}) = \varphi_2(\boldsymbol{b})$. A natural possibility is for φ_1 and φ_2 to be iterates of $\boldsymbol{f}$; under a mild condition on $\boldsymbol{a}$ and $\boldsymbol{b}$, we prove that this is the *only* possibility.

Theorem 2.3. *Let $\boldsymbol{f} := f_\lambda$ be the family of one-parameter polynomials* (*indexed by all $\lambda \in \mathbb{C}$*) *given by*

$$f_\lambda(x) := x^d + \sum_{i=0}^{d-2} c_i(\lambda)x^i = P(x) + \sum_{j=1}^{r} Q_j(x)\cdot\lambda^{m_j}$$

as above (*see* (2.1) *and* (2.2)). *Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}[\lambda]$, and assume there exist nonnegative integers k and ℓ such that the following conditions hold*:

(i) *$f_\lambda^k(\boldsymbol{a}(\lambda))$ and $f_\lambda^\ell(\boldsymbol{b}(\lambda))$ have the same degree and the same leading coefficient as polynomials in λ, and*

(ii) *if $m = \deg_\lambda(f_\lambda^k(\boldsymbol{a}(\lambda)))\deg_\lambda(f_\lambda^\ell(\boldsymbol{b}(\lambda)))$, then $m \geq m_r$.*

Then there exist infinitely many $\lambda \in \mathbb{C}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic points for f_λ if and only if $f_\lambda^k(\boldsymbol{a}(\lambda)) = f_\lambda^\ell(\boldsymbol{b}(\lambda))$.

Remarks 2.4. (a) The one-dimensional $\mathbb{C}$-scheme $(\boldsymbol{a}, \boldsymbol{b}) \subset \mathcal{X} := \mathbb{P}^1_{\mathbb{C}(\lambda)} \times_{\mathbb{C}(\lambda)} \mathbb{P}^1_{\mathbb{C}(\lambda)}$ in Theorem 2.3 is contained in the two-dimensional $\mathbb{C}$-subscheme $\mathcal{Y}$ of $\mathcal{X}$ given by the equation

$$\boldsymbol{f}^k(x) = \boldsymbol{f}^\ell(y),$$

where (x, y) are the coordinates of $\mathcal{X}$. Such a $\mathcal{Y}$ is fixed by the action of $(\boldsymbol{f}, \boldsymbol{f})$ on $\mathcal{X}$ as predicted by Question 1.3.

(b) It follows from the Lefschetz principle that the same statements in Theorem 2.3 hold if we replace $\mathbb{C}$ by any other algebraically closed complete valued field of characteristic 0.

(c) We note that if $\boldsymbol{c} \in \mathbb{C}[\lambda]$ has the property that there exists $k \in \mathbb{N}$ such that $\deg_\lambda(f_\lambda^k(\boldsymbol{c}(\lambda))) = m$ has the property (ii) from Theorem 2.3, then $\boldsymbol{c}$ is not preperiodic for $\boldsymbol{f}$ (see Lemma 5.2).

(d) If $\boldsymbol{f}$ is not a constant family, then it follows from Benedetto's theorem [2005] that $\boldsymbol{c} \in \mathbb{C}[\lambda]$ is not preperiodic for $\boldsymbol{f}$ if and only if there exists $k \in \mathbb{N}$ such that $\deg_\lambda(f_\lambda^k(\boldsymbol{c}(\lambda))) \geq m_r$. On the other hand, if $\boldsymbol{f}$ is a constant family of polynomials defined over $\mathbb{C}$, that is, $r = 0$ and $m_0 = 0$ in Theorem 2.3, then implicitly $m > 0$. (Otherwise the conclusion holds trivially.)

Theorem 2.3 generalizes known results regarding "unlikely intersections" in the dynamical setting including the dynamical Manin–Mumford questions (see Section 11). First, Theorem 2.3 generalizes the main result of [Baker and DeMarco 2011] in two ways. On one hand, in the case when $\boldsymbol{a}$ and $\boldsymbol{b}$ are both constant, we can prove a generalization of the main result from [ibid.] as follows:

Theorem 2.5. *Let $a, b \in \mathbb{C}$, let $d \geq 2$, and let $c_0, \ldots, c_{d-2} \in \mathbb{C}[\lambda]$ such that $\deg(c_0) > \deg(c_i)$ for each $i = 1, \ldots, d-2$. If there are infinitely many $\lambda \in \mathbb{C}$ such that both a and b are preperiodic for*

$$f_\lambda(x) := x^d + \sum_{i=0}^{d-2} c_i(\lambda)x^i,$$

then $f_\lambda(a) = f_\lambda(b)$.

Proof. We apply Theorem 2.3 for $\boldsymbol{a}(\lambda) := f_\lambda(a)$ and $\boldsymbol{b}(\lambda) := f_\lambda(b)$. $\square$

Consequently, Theorem 2.5 yields the proof of Theorem 1.2.

Proof of Theorem 1.2. Note that in this case, we may drop the hypothesis that $f(x)$ is in normal form since we may conjugate $f(x)$ by some linear polynomial $\delta \in \mathbb{C}[x]$ such that $g := \delta^{-1} \circ f \circ \delta + \delta^{-1}(\lambda)$ is a family of polynomials in normal form. Then apply Theorem 2.5 to the pair of points $\delta^{-1}(a)$ and $\delta^{-1}(b)$. $\square$

On the other hand, using our Theorem 2.3 we are able to treat the case when the pair of points $\boldsymbol{a}$ and $\boldsymbol{b}$ depend algebraically on the parameter. This answers a question raised by Silverman mentioned in [Baker and DeMarco 2011, Section 1.1]. For instance, as an application of Theorem 2.3, by taking $\boldsymbol{f} = f(x) + \lambda$ for *any* nonconstant polynomial $f(x) \in \mathbb{C}[x]$ of degree at least 2, we have the following:

Corollary 2.6. *Let $f \in \mathbb{C}[x]$ be any polynomial of degree $d \geq 2$, and let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}[\lambda]$ be polynomials such that $\boldsymbol{a}$ and $\boldsymbol{b}$ have the same degree and the same leading coefficient. Then there are infinitely many $\lambda \in \mathbb{C}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic under the action of $f(x) + \lambda$ if and only if $\boldsymbol{a}(\lambda) = \boldsymbol{b}(\lambda)$.*

Proof. First, the theorem is vacuously true if $\boldsymbol{a}$ and $\boldsymbol{b}$ are constant polynomials since then they are automatically equal because they have the same leading coefficient. So we may assume that $\deg(\boldsymbol{a}) = \deg(\boldsymbol{b}) \geq 1$.

Second, we conjugate $f(x)$ by some linear polynomial $\delta \in \mathbb{C}[x]$ such that $g := \delta^{-1} \circ f \circ \delta$ is a polynomial in normal form. Then we apply Theorem 2.3 to the family of polynomials $g(x) + \delta^{-1}(\lambda)$ and to the starting points $\delta^{-1}(\boldsymbol{a}(\lambda))$ and $\delta^{-1}(\boldsymbol{b}(\lambda))$. Since $\boldsymbol{a}$ and $\boldsymbol{b}$ are polynomials of the same positive degree and same leading coefficient, it is immediate to check that conditions (i) and (ii) of Theorem 2.3 hold for $k = \ell = 0$. Therefore, $\boldsymbol{a}(\lambda) = \boldsymbol{b}(\lambda)$ as desired. $\square$

An important special case of Corollary 2.6 is Theorem 1.4. Using Theorem 2.3 when $\boldsymbol{f}$ is a constant family of polynomials, we obtain a proof of Theorem 1.5.

Proof of Theorem 1.5. The result is an immediate consequence of Theorem 2.3 once we observe, as before, that we may replace f with a conjugate $\delta^{-1} \circ f \circ \delta$ of itself that is a polynomial in normal form. (Note that in this case, we also replace $\boldsymbol{a}$ and $\boldsymbol{b}$ by $\delta^{-1}(\boldsymbol{a})$ and $\delta^{-1}(\boldsymbol{b})$, respectively, which are also polynomials in λ of the same degree and same leading coefficient.) $\square$

On the other hand, assuming each c_i and also $\boldsymbol{a}$ and $\boldsymbol{b}$ have algebraic coefficients, the exact same proof we have yields stronger statements of Theorems 2.3, 1.5, and Corollary 2.6, allowing us to replace the hypothesis that there are infinitely many $\lambda \in \overline{\mathbb{Q}}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f_λ with the weaker condition that there exists an infinite sequence of $\lambda_n \in \overline{\mathbb{Q}}$ such that

$$\lim_{n\to\infty} \hat{h}_{f_{\lambda_n}}(\boldsymbol{a}(\lambda_n)) + \hat{h}_{f_{\lambda_n}}(\boldsymbol{b}(\lambda_n)) = 0,$$

where for each $\lambda \in \overline{\mathbb{Q}}$, $\hat{h}_{f_\lambda}$ is the canonical height constructed with respect to the polynomial f_λ. (For the precise definition of the canonical height with respect to a polynomial map, see Section 3.) Therefore, we can prove a special case of Zhang's dynamical Bogomolov conjecture [2006].

Corollary 2.7. *Let $Y \subset \mathbb{P}^1 \times \mathbb{P}^1$ be a curve that admits a parametrization given by $(\boldsymbol{a}(z), \boldsymbol{b}(z))$ for $z \in \mathbb{C}$, where $\boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{Q}}[x]$ are polynomials of the same degree and with the same leading coefficient. Let $f \in \overline{\mathbb{Q}}[x]$ be a polynomial of degree at least equal to 2, and let $\Phi(x, y) := (f(x), f(y))$ be the diagonal action of f on $\mathbb{P}^1 \times \mathbb{P}^1$. If there exists an infinite sequence of points $(x_n, y_n) \in Y(\overline{\mathbb{Q}})$ such that*

$$\lim_{n\to\infty} \hat{h}_f(x_n) + \hat{h}_f(y_n) = 0,$$

then $\boldsymbol{a} = \boldsymbol{b}$. In particular, Y is the diagonal subvariety of $\mathbb{P}^1 \times \mathbb{P}^1$ and thus is preperiodic under the action of Φ.

Remark 2.8. In fact, this result holds not only over $\overline{\mathbb{Q}}$ but also over the algebraic closure of any global function field L (whose subfield of constants is K) as long as f is not conjugate to a polynomial with coefficients in $\overline{K}$.

Note that the second author, together with Baker, proved a similar result [Baker and Hsia 2005, Theorem 8.10] in the case when Y is a line; that is, if a line in $\mathbb{P}^1 \times \mathbb{P}^1$ contains an infinite set of points of small canonical height with respect to the coordinatewise action of the polynomial f on $\mathbb{P}^1 \times \mathbb{P}^1$, then the line Y is preperiodic under the action of (f, f) on $\mathbb{P}^1 \times \mathbb{P}^1$.

Laura DeMarco communicated to us that our Theorem 2.3 yields the proof of the first case of a conjecture she made as a dynamical analogue of the André–Oort conjecture. Essentially, the dynamical André–Oort conjecture envisioned by DeMarco aims to characterize subvarieties in the moduli space M_d of complex rational maps $f : \mathbb{P}^1 \to \mathbb{P}^1$ (of degree $d > 1$) that contain a Zariski dense subset of postcritically finite rational maps. A rational map is *postcritically finite* (PCF) if all of its critical points are preperiodic. The PCF rational maps play an important role in complex dynamics; for example, the Lattès maps are PCF.

Our Theorem 2.3 has the following consequence. Let $\boldsymbol{f} = f_\lambda$ be a family of polynomials in normal form of degree d with polynomial coefficients in λ. Furthermore, assume the critical points $\boldsymbol{c}_1(\lambda), \dots, \boldsymbol{c}_{d-1}(\lambda)$ of f_λ are also polynomials in λ. Let I be the collection of indices i such that $\boldsymbol{c}_i$ is not preperiodic for $\boldsymbol{f}$. Suppose for each $i \in I$, there exist iterates $f_\lambda^{m_i}(\boldsymbol{c}_i(\lambda))$ with the same degree and leading coefficients in λ and that this degree is large enough (that is, satisfies the hypothesis from Theorem 2.3). Then there are infinitely many PCF maps in this family if and only if all $\boldsymbol{f}^{m_i}(\boldsymbol{c}_i)$ (for $i \in I$) are equal.

We prove Theorem 2.3 first for the case when both $\boldsymbol{a}$ and $\boldsymbol{b}$ and also each of the c_i have algebraic coefficients, and then we extend our proof to the general case. For the extension to $\mathbb{C}$, we use a result of Benedetto [2005] (see also Baker's extension [2009] to arbitrary rational maps), which states that for a polynomial f of degree at least equal to 2 defined over a function field K of finite transcendence degree over a subfield K_0, if f is not isotrivial (that is, f is not conjugate to a polynomial defined over $\overline{K_0}$), then each $x \in \overline{K}$ is preperiodic if and only if its canonical height $\hat{h}_f(x)$ equals 0. Strictly speaking, Benedetto's result is stated for function fields of transcendence degree 1, but a simple inductive argument on the transcendence degree yields the result for function fields of arbitrary finite transcendence degree. (See also [Baker 2009, Corollary 1.8], where Baker extends Benedetto's result to rational maps defined over function fields of arbitrary finite transcendence degree.)

Our results and proofs are inspired by the results of [Baker and DeMarco 2011] so that the strategy for the proof of Theorem 2.3 essentially follows their ideas. However, there are significantly more technical difficulties in our proofs. The plan of our proof is to use the v-adic generalized Mandelbrot sets introduced therein for the family of polynomials f_λ and then use the equidistribution result of Baker and Rumely [2010]. A key ingredient is Proposition 6.8, which says that the canonical local height of the point in question at the place v is a constant multiple of the Green function associated with the v-adic generalized Mandelbrot set. Then the condition that $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic is translated to the condition that the heights $h_{\mathbb{M}_{\boldsymbol{a}}}(\lambda)$ and $h_{\mathbb{M}_{\boldsymbol{b}}}(\lambda)$, respectively, are zero for the corresponding parameter λ. Therefore, the equidistribution result of Baker–Rumely can be applied to conclude that the v-adic generalized Mandelbrot sets for $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are the same for each place v. Finally, we need to use an explicit formula for the Green function associated with the v-adic generalized Mandelbrot set corresponding to an archimedean valuation v to conclude that the desired equality of $f_\lambda^k(\boldsymbol{a}(\lambda))$ and $f_\lambda^\ell(\boldsymbol{b}(\lambda))$ holds. Extra work is needed for the explicit description of the Green function for a v-adic generalized Mandelbrot set (when v is an archimedean place) due to the fact that in our case, the polynomial f_λ has arbitrary (finitely) many critical points that vary with λ in contrast to the family of polynomials $x^d + \lambda$ from [Baker and DeMarco 2011], which has only one critical point for the entire family.

3. Notation and preliminary

For any quasiprojective variety X endowed with an endomorphism Φ, we call a point $x \in X$ *preperiodic* if there exist two distinct nonnegative integers m and n such that $\Phi^m(x) = \Phi^n(x)$, where by Φ^i we always denote the i-iterate of the endomorphism Φ. If $n = 0$, then, by convention, Φ^0 is the identity map.

Let K be a field of characteristic 0 equipped with a set of inequivalent absolute values (places) Ω_K, normalized so that the product formula holds; more precisely, for each $v \in \Omega_K$ there exists a positive integer N_v such that for all $\alpha \in K^*$ we have $\prod_{v\in\Omega} |\alpha|_v^{N_v} = 1$, where for $v \in \Omega_K$, the corresponding absolute value is denoted by $|\cdot|_v$. Let $\mathbb{C}_v$ be a fixed completion of the algebraic closure of a completion of $(K, |\cdot|_v)$. When v is an archimedean valuation, then $\mathbb{C}_v = \mathbb{C}$. We fix an extension of $|\cdot|_v$ to an absolute value of $(\mathbb{C}_v, |\cdot|_v)$. Examples of product formula fields (or *global fields*) are number fields and function fields of projective varieties that are regular in codimension 1 [Lang 1983, Section 2.3; Bombieri and Gubler 2006, Section 1.4.6].

Let $f \in \mathbb{C}_v[x]$ be any polynomial of degree $d \geq 2$. Following Call and Silverman [1993], for each $x \in \mathbb{C}_v$, we define the *local canonical height* of x by

$$\hat{h}_{f,v}(x) := \lim_{n\to\infty} \frac{\log^+ |f^n(x)|_v}{d^n},$$

where $\log^+ z$ always denotes $\log \max\{z, 1\}$ (for any real number z).

It is immediate that $\hat{h}_{f,v}(f^i(x)) = d^i \hat{h}_{f,v}(x)$, and thus, $\hat{h}_{f,v}(x) = 0$ whenever x is a preperiodic point for f. If v is nonarchimedean and $f(x) = \sum_{i=0}^d a_i x^i$, then $|f(x)|_v = |a_d x^d|_v > |x|_v$ when $|x|_v > r_v$, where

$$r_v := \max\left\{|a_d|_v^{-1/(d-1)},\ \max_{0\leq i<d}\left\{\left|\frac{a_i}{a_d}\right|^{1/(d-i)}\right\}\right\}. \tag{3.1}$$

Moreover, if $|x|_v > r_v$, then $\hat{h}_v(x) = \log|x|_v + \log|a_d|_v/(d-1) > 0$. For more details, see [Ghioca and Tucker 2008; Hsia 2008]. (Although these results are for canonical heights associated with Drinfeld modules, all the proofs go through for any local canonical height associated with any polynomial with respect to any nonarchimedean place.)

Now, if v is archimedean, again it is easy to see that if $|x|_v$ is sufficiently large, then $|f(x)|_v \gg |x|_v^d$, and moreover, $|f^n(x)|_v \to \infty$ as $n \to \infty$.

We fix an algebraic closure $\overline{K}$ of K, and for each $v \in \Omega_K$ we fix an embedding $\overline{K} \hookrightarrow \mathbb{C}_v$. Assume $f \in \overline{K}[x]$. Call and Silverman [1993] also defined the *global canonical height* $\hat{h}(x)$ for each $x \in \overline{K}$ as

$$\hat{h}_f(x) = \lim_{n\to\infty} \frac{h(f^n(x))}{d^n},$$

where h is the usual (logarithmic) Weil height on $\overline{K}$. Call and Silverman show that the global canonical height decomposes into a sum of the corresponding local canonical heights.

For each $\sigma \in \mathrm{Gal}(\overline{K}/K)$, we denote by $\hat{h}_{f^\sigma}$ the global canonical height computed with respect to f^σ, which is the polynomial obtained by applying σ to each coefficient of f. Similarly, for each $v \in \Omega_K$ we denote by $\hat{h}_{f^\sigma,v}$ the corresponding local canonical height constructed with respect to the polynomial f^σ. For $x \in \overline{K}$, we have $\hat{h}_f(x) = 0$ if and only if $\hat{h}_{f^\sigma}(x^\sigma) = 0$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$. More precisely, for $x \in \overline{K}$ we have

$$\hat{h}_f(x) = 0 \iff \hat{h}_{f^\sigma,v}(x^\sigma) = 0 \text{ for all } v \in \Omega_K \text{ and all } \sigma \in \mathrm{Gal}(\overline{K}/K). \quad (3.2)$$

Essentially, (3.2) says that $\hat{h}_f(x) = 0$ if and only if the orbits of x^σ under each polynomial f^σ (for $\sigma \in \mathrm{Gal}(\overline{K}/K)$) are bounded with respect to each absolute value $|\cdot|_v$ for $v \in \Omega_K$.

Benedetto [2005] proved that if a polynomial f defined over a function field K (endowed with a set Ω_K of absolute values) is not *isotrivial* (that is, it cannot be conjugated to a polynomial defined over the constant subfield of K), then each point $c \in \overline{K}$ is preperiodic for f if and only if its global canonical height (computed with respect to f) equals 0. In particular, if $c \in \overline{K}$, then c is preperiodic if and only if

$$\hat{h}_{f^\sigma,v}(c^\sigma) = 0 \quad \text{for all } \sigma \in \mathrm{Gal}(\overline{K}/K) \text{ and for all places } v \in \Omega_K. \quad (3.3)$$

Let

$$\boldsymbol{f} = f_\lambda := x^d + \sum_{i=0}^{d-2} c_i(\lambda) x^i,$$

where $c_i(\lambda) \in \mathbb{C}[\lambda]$ for $i = 0, \ldots, d-2$, and let $\boldsymbol{c}(\lambda) \in \mathbb{C}[\lambda]$. We let K be the field extension of $\mathbb{Q}$ generated by all coefficients of each $c_i(\lambda)$ and of $\boldsymbol{c}(\lambda)$. Assume K is a global field; that is, it has a set Ω_K of inequivalent absolute values with respect to which the nonzero elements of K satisfy a product formula. For each place $v \in \Omega_K$, we define the v-adic Mandelbrot set $M_{\boldsymbol{c},v}$ for $\boldsymbol{c}$ with respect to the family of polynomials $\boldsymbol{f}$ as the set of all $\lambda \in \mathbb{C}_v$ such that $\hat{h}_{f_\lambda,v}(\boldsymbol{c}(\lambda)) = 0$, that is, the set of all $\lambda \in \mathbb{C}_v$ such that the iterates $f_\lambda^n(\boldsymbol{c}(\lambda))$ are bounded with respect to the v-adic absolute value.

4. Berkovich spaces

We now introduce Berkovich spaces and state the equidistribution theorem of Baker and Rumely [2010], which will be key for the proofs of Theorems 2.3 and 2.5.

Let K be a global field of characteristic 0, and let Ω_K be the set of its inequivalent absolute values. For each $v \in \Omega_K$, we let $\mathbb{C}_v$ be the completion of an algebraic closure of the completion of K at v. Let $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ denote the Berkovich affine line over $\mathbb{C}_v$; see [Berkovich 1990; Baker and Rumely 2010, Section 2.1] for details. Then $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ is a locally compact, Hausdorff, path-connected space containing $\mathbb{C}_v$ as a dense subspace (with the topology induced from the v-adic absolute value). As a topological space, $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ is the set consisting of all multiplicative seminorms, denoted by $[\cdot]_x$, on $\mathbb{C}_v[T]$ extending the absolute value $|\cdot|_v$ on $\mathbb{C}_v$ endowed with the weakest topology such that the map $z \mapsto [f]_z$ is continuous for all $f \in \mathbb{C}_v[T]$. It follows from the Gelfand–Mazur theorem that if $\mathbb{C}_v$ is the field of complex numbers $\mathbb{C}$, then $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}}$ is homeomorphic to $\mathbb{C}$. In the following, we will also use $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ to denote the complex line $\mathbb{C}$ whenever $\mathbb{C}_v = \mathbb{C}$. If $(\mathbb{C}_v, |\cdot|_v)$ is nonarchimedean, then the set of seminorms can be described as follows. If $\{D(a_i, r_i)\}_i$ is any decreasing nested sequence of closed disks $D(c_i, r_i)$ centered at points $c_i \in \mathbb{C}_v$ of radius $r_i \geq 0$, then the map $f \mapsto \lim_{i\to\infty} [f]_{D(c_i,r_i)}$ defines a multiplicative seminorm on $\mathbb{C}_v[T]$, where $[f]_{D(c_i,r_i)}$ is the sup norm of f over the closed disk $D(a_i, r_i)$. Berkovich's classification theorem says that there are exactly four types of points: types I, II, III, and IV. The first three can be described in terms of closed disks $\zeta := D(c, r) = \bigcap D(c_i, r_i)$, where $c \in \mathbb{C}_v$ and $r \geq 0$. The corresponding multiplicative seminorm is just $f \mapsto [f]_{D(c,r)}$ for $f \in \mathbb{C}_v[T]$. Then ζ is of type I, II or III if and only if $r = 0$, $r \in |\mathbb{C}_v^*|_v$ or $r \notin |\mathbb{C}_v^*|_v$, respectively. As for type IV points, they correspond to sequences of decreasing nested disks $D(c_i, r_i)$ such that $\bigcap D(c_i, r_i) = \varnothing$ and the multiplicative seminorm is $f \mapsto \lim_{i\to\infty}[f]_{D(c_i,r_i)}$ as described above. For details, see [Berkovich 1990; Baker and Rumely 2010]. For $\zeta \in \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$, we sometimes write $|\zeta|_v$ instead of $[T]_\zeta$.

In order to apply the main equidistribution result from [Baker and Rumely 2010, Theorem 7.52], we recall the potential theory on the affine line over $\mathbb{C}_v$. We will focus on the case when $\mathbb{C}_v$ is a nonarchimedean field; the case $\mathbb{C}_v = \mathbb{C}$ is classical. (We refer the reader to [Ransford 1995].) The right setting for nonarchimedean potential theory is the potential theory on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ developed in [Baker and Rumely 2010]. We quote part of a nice summary of the theory from [Baker and DeMarco 2011, Section 2.2 and Section 2.3] without going into details. We refer the reader to [Baker and DeMarco 2011; Baker and Rumely 2010] for all the details and proofs. Let E be a compact subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. Then analogous to the complex case, the Green function G_E of E relative to ∞ and the logarithmic capacity $\gamma(E) := e^{-V(E)}$ can be defined, where $V(E)$ is the infimum of the *energy integral* with respect to all possible probability measures supported on E. More precisely,

$$V(E) = \inf_{\mu} \iint_{E\times E} -\log \delta(x, y)\, d\mu(x)\, d\mu(y),$$

where the infimum is computed with respect to all probability measures μ supported on E while for $x, y \in \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$, the function $\delta(x, y)$ is the *Hsia kernel* [Baker and

Rumely 2010, Proposition 4.1]:

$$\delta(x,y) := \limsup_{\substack{z,w\in\mathbb{C}_v \\ z\to x,\, w\to y}} |z-w|_v.$$

The following are basic properties of the logarithmic capacity of E.

- If E_1 and E_2 are two compact subsets of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ such that $E_1 \subset E_2$, then $\gamma(E_1) \le \gamma(E_2)$.
- If $E = \{\zeta\}$, where ζ is a type II or III point corresponding to a closed disk $D(c,r)$, then $\gamma(E) = r > 0$ [Baker and Rumely 2010, Example 6.3]. (This can be viewed as an analogue of the fact that a closed disk $D(c,r)$ of positive radius r in $\mathbb{C}_v$ has logarithmic capacity $\gamma(D(c,r)) = r$.)

If $\gamma(E) > 0$, then there exists a unique probability measure μ_E attaining the infimum of the energy integral. Furthermore, the support of μ_E is contained in the boundary of the unbounded component of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} \setminus E$.

The Green function $G_E(z)$ of E relative to infinity is a well-defined nonnegative real-valued subharmonic function on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ that is harmonic on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} \setminus E$ (in the sense of [Baker and Rumely 2010, Chapter 8] for the nonarchimedean setting; see [Ransford 1995] for the archimedean case). If $\gamma(E) = 0$, then there exists no Green function associated with the set E (see [Ransford 1995, Exercise 1, page 115] in the case when $|\cdot|_v$ is archimedean; a similar argument works when $|\cdot|_v$ is nonarchimedean). Indeed, as shown in [Baker and Rumely 2010, Proposition 7.17, page 151], if $\gamma(\partial E) = 0$, then there exists no nonconstant harmonic function on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} \setminus E$ that is bounded below. (This is the strong maximum principle for harmonic functions defined on Berkovich spaces). The following result is [Baker and DeMarco 2011, Lemmas 2.2 and 2.5], and it gives a characterization of the Green function of the set E:

Lemma 4.1. *Let $(\mathbb{C}_v, |\cdot|_v)$ be either an archimedean or a nonarchimedean field. Let E be a compact subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ and U the unbounded component of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} \setminus E$.*

(1) *If $\gamma(E) > 0$ (that is, $V(E) < \infty$), then $G_E(z) = V(E) + \log|z|_v + o(1)$ for all $z \in \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ such that $|z|_v$ is sufficiently large. Furthermore, the $o(1)$ term may be omitted if v is nonarchimedean.*

(2) *If $G_E(z) = 0$ for all $z \in E$, then G_E is continuous on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$, $\mathrm{Supp}(\mu_E) = \partial U$, and $G_E(z) > 0$ if and only if $z \in U$.*

(3) *If $G : \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} \to \mathbb{R}$ is a continuous subharmonic function that is harmonic on U, identically zero on E, and such that $G(z) - \log^+|z|_v$ is bounded, then $G = G_E$. Furthermore, if $G(z) = \log|z|_v + V + o(1)$ (as $|z|_v \to \infty$) for some $V < \infty$, then $V(E) = V$, so $\gamma(E) = e^{-V}$.*

To state the equidistribution result from [Baker and Rumely 2010], we consider the compact *Berkovich adelic sets*, which are of the form

$$\mathbb{E} := \prod_{v \in \Omega} E_v,$$

where E_v is a nonempty compact subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ for each $v \in \Omega$ and where E_v is the closed unit disk $\mathfrak{D}(0, 1)$ in $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ for all but finitely many $v \in \Omega$. The *logarithmic capacity* $\gamma(\mathbb{E})$ of $\mathbb{E}$ is defined by

$$\gamma(\mathbb{E}) := \prod_{v \in \Omega} \gamma(E_v)^{N_v},$$

where the positive integers N_v are the ones associated with the product formula on the global field K. Note that this is a finite product since $\gamma(E_v) = \gamma(\mathfrak{D}(0, 1)) = 1$ for all but finitely many $v \in \Omega$. Let $G_v = G_{E_v}$ be the Green function of E_v relative to ∞ for each $v \in \Omega$. For every $v \in \Omega$, we fix an embedding $\overline{K} \hookrightarrow \mathbb{C}_v$. Let $S \subset \overline{K}$ be any finite subset that is invariant under the action of the Galois group $\mathrm{Gal}(\overline{K}/K)$. We define the height $h_{\mathbb{E}}(S)$ of S relative to $\mathbb{E}$ by

$$h_{\mathbb{E}}(S) := \sum_{v \in \Omega} N_v \Big(\frac{1}{|S|} \sum_{z \in S} G_v(z) \Big). \tag{4.2}$$

Note that this definition is independent of any particular embedding $\overline{K} \hookrightarrow \mathbb{C}_v$ that we choose at $v \in \Omega$. The following is a special case of the equidistribution result [Baker and Rumely 2010, Theorem 7.52].

Theorem 4.3. *Let $\mathbb{E} = \prod_{v \in \Omega} E_v$ be a compact Berkovich adelic set with $\gamma(\mathbb{E}) = 1$. Suppose that S_n is a sequence of $\mathrm{Gal}(\overline{K}/K)$-invariant finite subsets of $\overline{K}$ with $|S_n| \to \infty$ and $h_{\mathbb{E}}(S_n) \to 0$ as $n \to \infty$. For each $v \in \Omega$ and for each n, let δ_n be the discrete probability measure supported equally on the elements of S_n. Then the sequence of measures $\{\delta_n\}$ converges weakly to μ_v, the equilibrium measure on E_v.*

5. General results about the dynamics of polynomials f_λ

In this section, we work with a family of polynomials f_λ as given in Section 2, that is,

$$f_\lambda(x) = x^d + \sum_{i=0}^{d-2} c_i(\lambda) x^i$$

with $c_i(\lambda) \in \mathbb{C}[\lambda]$ for $i = 0, \ldots, d-2$. As before, we may rewrite our family of polynomials as

$$f_\lambda(x) = P(x) + \sum_{j=1}^{r} Q_j(x) \cdot \lambda^{m_j},$$

where $P(x)$ is a polynomial of degree d in normal form, each Q_i has degree at most equal to $d-2$, r is a nonnegative integer, and $m_0 := 0 < m_1 < \cdots < m_r$. Let $\boldsymbol{c}(\lambda) \in \mathbb{C}[\lambda]$ be given, and let K be the field extension of $\mathbb{Q}$ generated by all the coefficients of $c_i(\lambda)$ for $i = 0, \dots, d-2$ and of $\boldsymbol{c}(\lambda)$.

We define $g_{\boldsymbol{c},n}(\lambda) := f_\lambda^n(\boldsymbol{c}(\lambda))$ for each $n \in \mathbb{N}$. Assume $m := \deg(\boldsymbol{c})$ satisfies the property (ii) from Theorem 2.3, that is,

$$m = \deg(\boldsymbol{c}) \geq m_r. \tag{5.1}$$

Furthermore, if $r = 0$, we assume $m \geq 1$ (see also Remarks 2.4(c)). We let q_m be the leading coefficient of $\boldsymbol{c}(\lambda)$. In the next lemma, we compute the degrees of all polynomials $g_{\boldsymbol{c},n}$ for all positive integers n.

Lemma 5.2. *With the hypothesis above, the polynomial $g_{\boldsymbol{c},n}(\lambda)$ has degree $m \cdot d^n$ and leading coefficient $q_m^{d^n}$ for each $n \in \mathbb{N}$.*

Proof. The assertion follows easily by induction on n using (5.1) since the term of highest degree in λ from $g_{\boldsymbol{c},n}(\lambda)$ is $\boldsymbol{c}(\lambda)^{d^n}$. □

We immediately obtain as a corollary of Lemma 5.2 the fact that $\boldsymbol{c}$ is not preperiodic for $\boldsymbol{f}$. The set of all $\lambda \in \mathbb{C}$ such that $\boldsymbol{c}(\lambda)$ is preperiodic for f_λ is denoted by $\operatorname{Prep}(\boldsymbol{c})$. The following result is an immediate consequence of Lemma 5.2:

Corollary 5.3. $\operatorname{Prep}(\boldsymbol{c}) \subset \overline{K}$.

6. Capacities of generalized Mandelbrot sets

We continue with the notation from Sections 4 and 5. Let $\boldsymbol{c} := \boldsymbol{c}(\lambda) \in \mathbb{C}[\lambda]$ be a nonconstant polynomial, and let K be a product formula field containing the coefficients of each $c_i(\lambda)$ for $i = 0, \dots, d-2$ and of $\boldsymbol{c}$. We let Ω_K be the set of inequivalent absolute values of the global field K, and let $v \in \Omega_K$. Assume that $\boldsymbol{c}(\lambda) = q_m\lambda^m + (\text{lower terms})$, where $m = \deg(\boldsymbol{c})$ satisfies the condition (5.1).

Our goal is to compute the logarithmic capacities of the v-adic generalized Mandelbrot sets $M_{\boldsymbol{c},v}$ defined in Section 3. Following [Baker and DeMarco 2011], we extend the definition of our v-adic Mandelbrot set $M_{\boldsymbol{c},v}$ to be a subset of the affine Berkovich line $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ as follows:

$$M_{\boldsymbol{c},v} := \{ \lambda \in \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} : \sup_n [g_{\boldsymbol{c},n}(T)]_\lambda < \infty \}.$$

Note that if $\mathbb{C}_v$ is a nonarchimedean field, then our present definition for $M_{\boldsymbol{c},v}$ yields more points than our definition from Section 3. Let $\lambda \in \mathbb{C}_v$, and recall the local canonical height $\hat{h}_{\lambda,v}(x)$ of $x \in \mathbb{C}_v$ is given by the formula

$$\hat{h}_{\lambda,v}(x) := \hat{h}_{f_\lambda,v}(x) = \lim_{n\to\infty} \frac{\log^+ |f_\lambda^n(x)|_v}{d^n}.$$

Notice that $\hat{h}_{\lambda,v}(x)$ is a continuous function of both λ and x (see [Branner and Hubbard 1988, Proposition 1.2] for polynomials over complex numbers; the proof for the nonarchimedean case is similar). As $\mathbb{C}_v$ is a dense subspace of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$, continuity in λ implies that the canonical local height function $\hat{h}_{\lambda,v}(\boldsymbol{c}(\lambda))$ has a natural extension on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. (Since the topology on $\mathbb{C}_v$ is the restriction of the weak topology on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$, any continuous function on $\mathbb{C}_v$ automatically has a unique extension to $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$.) We will view $\hat{h}_{\lambda,v}(\boldsymbol{c}(\lambda))$ as a continuous function on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ in the following. It follows from the definition of $M_{\boldsymbol{c},v}$ that $\lambda \in M_{\boldsymbol{c},v}$ if and only if $\hat{h}_{\lambda,v}(\boldsymbol{c}(\lambda)) = 0$. Thus, $M_{\boldsymbol{c},v}$ is a closed subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. In fact, the following is true:

Proposition 6.1. *$M_{\boldsymbol{c},v}$ is a compact subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$.*

We already showed that $M_{\boldsymbol{c},v}$ is a closed subset of the locally compact space $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$, and thus, in order to prove Proposition 6.1, we only need to show that $M_{\boldsymbol{c},v}$ is a bounded subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. If $\boldsymbol{f}$ is a constant family of polynomials, then Proposition 6.1 follows from our assumption that $\deg(\boldsymbol{c}) \geq 1$. Indeed, if $|\lambda|_v$ is large, then $|\boldsymbol{c}(\lambda)|_v$ is large, and thus, $|\boldsymbol{f}^n(\boldsymbol{c}(\lambda))|_v \to \infty$ as $n \to \infty$. Furthermore, for nonarchimedean place v, if $|\lambda|_v$ is sufficiently large, then (assuming v is nonarchimedean)

$$|\boldsymbol{f}^n(\boldsymbol{c}(\lambda))|_v = |\boldsymbol{c}(\lambda)|_v^{d^n} = |q_m\lambda^m|_v^{d^n}. \tag{6.2}$$

So now we are left with the case that $\boldsymbol{f}$ is not a constant family, that is, $r \geq 1$.

Lemma 6.3. *Assume $r \geq 1$, that is, $\boldsymbol{f}$ is not a constant family of polynomials. Then $M_{\boldsymbol{c},v}$ is a bounded subset of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$.*

Proof. First we rewrite as before

$$f_\lambda(x) = P(x) + \sum_{j=1}^{r} Q_j(x)\cdot\lambda^{m_j}$$

with $P(x)$ in normal form of degree d and each polynomial Q_j of degree $e_j \leq d-2$; also, $0 < m_1 < \cdots < m_r$. We know $m = \deg(\boldsymbol{c}) \geq m_r$.

Since $q_m\lambda^m$ is the leading monomial in $\boldsymbol{c}$, there exists a positive real number C_1 depending only on v, coefficients of $c_i(\lambda)$ for $i = 0, \ldots, d-2$, and $\boldsymbol{c}$ such that if $|\lambda|_v > C_1$, then $|\boldsymbol{c}(\lambda)|_v > \frac{1}{2}|q_m|_v \cdot |\lambda|_v^m$.

Let $\alpha := \max_{i=1}^{r} m_i/(d-e_i)$; then $\alpha \leq m_r/2$ since $e_i \leq d-2$ for all i. There exist positive real numbers C_2 and C_3 (depending only on v and the coefficients of $c_i(\lambda)$) such that if $|\lambda|_v > C_2$ and $|x|_v > C_3|\lambda|_v^\alpha$, then

$$|f_\lambda(x)|_v > \tfrac{1}{2}|x|_v^d > |x|_v,$$

and thus, $|f_\lambda^n(x)|_v \to \infty$ as $n \to \infty$.

However, since $m \ge m_r \ge 2\alpha > \alpha$, we conclude that if $|\lambda|_v > (2C_3/|q_m|_v)^{1/(m-\alpha)}$, then

$$\tfrac{1}{2}|q_m|_v \cdot |\lambda|_v^m > C_3|\lambda|_v^\alpha.$$

We let $C_4 := \max\{C_1, C_2, (2C_3/|q_m|_v)^{1/(m-\alpha)}, |q_m|_v^{-1/m}\}$. So if $|\lambda|_v > C_4$, then

$$|\boldsymbol{c}(\lambda)|_v > \tfrac{1}{2}|q_m|_v \cdot |\lambda|_v^m > C_3|\lambda|_v^\alpha,$$

and thus, $|f_\lambda^n(\boldsymbol{c}(\lambda))|_v \to \infty$ as $n \to \infty$. We conclude that if $\lambda \in M_{\boldsymbol{c},v}$, then $|\lambda|_v \le C_4$, as desired. □

Remark 6.4. It is possible to make the constants in the proof above explicit. Moreover, for a nonarchimedean place v, the estimate of the absolute values can be precise. For example, if v is nonarchimedean, we can ensure that if $|\lambda|_v > C_4$, then

$$|f_\lambda^n(\boldsymbol{c}(\lambda))|_v = |q_m\lambda^m|_v^{d^n} \text{ for all } n \ge 1. \tag{6.5}$$

Theorem 6.6. *The logarithmic capacity of $M_{\boldsymbol{c},v}$ is $\gamma(M_{\boldsymbol{c},v}) = |q_m|_v^{-1/m}$.*

The strategy for the proof of Theorem 6.6 is to construct a continuous subharmonic function $G_{c,v} : \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v} \to \mathbb{R}$ satisfying Lemma 4.1(3). Analogously to the family $f_\lambda(x) = x^d + \lambda$ treated in [Baker and DeMarco 2011], we let

$$G_{\boldsymbol{c},v}(\lambda) := \lim_{n\to\infty} \frac{1}{\deg(g_{\boldsymbol{c},n})} \log^+[g_{\boldsymbol{c},n}(T)]_\lambda. \tag{6.7}$$

Then by a similar reasoning as in the proof of [ibid., Proposition 3.7], it can be shown that the limit exists for all $\lambda \in \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. In fact, by the definition of canonical local height, for $\lambda \in \mathbb{C}_v$ we have

$$\begin{aligned} G_{\boldsymbol{c},v}(\lambda) &= \lim_{n\to\infty} \frac{1}{md^n} \log^+|f_\lambda^n(\boldsymbol{c}(\lambda))|_v && \text{since } \deg(g_{\boldsymbol{c},n}) = md^n \text{ by Lemma 5.2,} \\ &= \frac{1}{m} \cdot \hat{h}_{f_\lambda,v}(\boldsymbol{c}(\lambda)) && \text{by the definition of canonical local height.} \end{aligned}$$

As a consequence of the computation above, we have the following:

Proposition 6.8 [Silverman 1994a, Theorem II.0.1; Silverman 1994b, Theorem III.0.1 and Corollary III.0.3]. *We have $\hat{h}_{f_\lambda,v}(\boldsymbol{c}(\lambda)) = \deg(\boldsymbol{c})G_{\boldsymbol{c},v}(\lambda)$.*

Remark 6.9. The formula above holds in the more general case of Question 1.3; for example, one may work with a rational function $\boldsymbol{c} \in \mathbb{C}(\lambda)$.

Note that $G_{\boldsymbol{c},v}(\lambda) \ge 0$ for all $\lambda \in \mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. Moreover, we see easily that $\lambda \in M_{\boldsymbol{c},v}$ if and only if $G_{\boldsymbol{c},v}(\lambda) = 0$.

Lemma 6.10. *The function $G_{\boldsymbol{c},v}$ is the Green function for $M_{\boldsymbol{c},v}$ relative to ∞.*

The proof is essentially the same as the proof of [Baker and DeMarco 2011, Proposition 3.7]; we simply give a sketch of the idea.

Proof of Lemma 6.10. We deal with the case that v is nonarchimedean. (The case when v is archimedean follows similarly.) So using the same argument as in the proof of [Branner and Hubbard 1988, Proposition 1.2], we observe that as a function of λ,

$$\frac{\log^+[g_{\boldsymbol{c},n}(T)]_\lambda}{\deg(g_{\boldsymbol{c},n})}$$

converges uniformly on compact subsets of $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. So this function is continuous and subharmonic on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$ and converges to $G_{\boldsymbol{c},v}$ uniformly; hence, it follows from [Baker and Rumely 2010, Proposition 8.26(c)] that $G_{\boldsymbol{c},v}$ is continuous and subharmonic on $\mathbb{A}^1_{\mathrm{Berk},\mathbb{C}_v}$. Furthermore, as remarked above, $G_{\boldsymbol{c},v}$ is 0 on $M_{\boldsymbol{c},v}$.

Arguing as in the proof of Lemma 6.3 (see (6.2) and (6.5)), if $|\lambda|_v > C_4$, then for $n \geq 1$ we have

$$|g_{\boldsymbol{c},n}(\lambda)|_v = |f_\lambda^n(\boldsymbol{c}(\lambda))|_v = |q_m\lambda^m|_v^{d^n}.$$

Hence, for $|\lambda|_v > C_4$ we have

$$G_{\boldsymbol{c},v}(\lambda) = \lim_{n\to\infty} \frac{1}{md^n} \log|g_{\boldsymbol{c},n}(\lambda)|_v = \log|\lambda|_v + \frac{\log|q_m|_v}{m}.$$

It follows from Lemma 4.1(3) that $G_{\boldsymbol{c},v}$ is indeed the Green function of $M_{\boldsymbol{c},v}$. □

Now we are ready to prove Theorem 6.6.

Proof of Theorem 6.6. As in the proof of Lemma 6.10, we have

$$G_{\boldsymbol{c},v}(\lambda) = \log|\lambda|_v + \frac{\log|q_m|_v}{m} + o(1)$$

for $|\lambda|_v$ sufficiently large. By Lemma 4.1(3), we find that $V(M_{\boldsymbol{c},v}) = \log|q_m|_v/m$. Hence, the logarithmic capacity of $M_{\boldsymbol{c},v}$ is

$$\gamma(M_{\boldsymbol{c},v}) = e^{-V(M_{\boldsymbol{c},v})} = 1/|q_m|_v^{1/m}.$$ □

Let $\mathbb{M}_{\boldsymbol{c}} = \prod_{v\in\Omega} M_{\boldsymbol{c},v}$ be the generalized adelic Mandelbrot set associated with $\boldsymbol{c}$. As a corollary to Theorem 6.6, we see $\mathbb{M}_{\boldsymbol{c}}$ satisfies the hypothesis of Theorem 4.3.

Corollary 6.11. *For all but finitely many nonarchimedean places v, we have that $M_{\boldsymbol{c},v}$ is the closed unit disk $\mathfrak{D}(0;1)$ in $\mathbb{C}_v$; furthermore, $\gamma(\mathbb{M}_{\boldsymbol{c}}) = 1$.*

Proof. For each place v where all coefficients of $c_i(\lambda)$ for $i = 0, \ldots, d-2$ and of $\boldsymbol{c}(\lambda)$ are v-adic integral and moreover $|q_m|_v = 1$, we have that $M_{\boldsymbol{c},v} = \mathfrak{D}(0,1)$. Indeed, $\mathfrak{D}(0,1) \subset M_{\boldsymbol{c},v}$ since then $f_\lambda^n(\boldsymbol{c}(\lambda))$ is always a v-adic integer. For the converse implication, we note that each coefficient of $g_{\boldsymbol{c},n}(\lambda)$ is a v-adic integer while the leading coefficient is a v-adic unit for all $n \geq 1$; thus, $|g_{\boldsymbol{c},n}(\lambda)|_v = |\lambda|_v^{md^n} \to \infty$ if $|\lambda|_v > 1$. Note that $q_m \neq 0$, so the second assertion in Corollary 6.11 follows immediately by the product formula in K. □

Using Proposition 6.8 and the decomposition of the global canonical height as a sum of local canonical heights, we obtain the following result:

Corollary 6.12. *Let $\lambda \in \overline{K}$, let S be the set of* $\mathrm{Gal}(\overline{K}/K)$*-conjugates of* λ, *and let* $h_{\mathbb{M}_c}$ *be defined as in (4.2). Then* $\deg(\boldsymbol{c}) \cdot h_{\mathbb{M}_{\boldsymbol{c}}}(\lambda) = \hat{h}_{f_\lambda}(\boldsymbol{c}(\lambda))$.

Remark 6.13. Let $h(\lambda)$ denote a Weil height function corresponding to the divisor ∞ of the parameter space that is the projective line in our case. Then it follows from [Call and Silverman 1993, Theorem 4.1] that

$$\lim_{h(\lambda)\to\infty} \frac{\hat{h}_{f_\lambda}(\boldsymbol{c}(\lambda))}{h(\lambda)} = \hat{h}_{\boldsymbol{f}}(\boldsymbol{c}),$$

where $\hat{h}_{\boldsymbol{f}}(\boldsymbol{c})$ is the canonical height associated with the polynomial map $\boldsymbol{f}$ over the function field $\mathbb{C}(\lambda)$. Corollary 6.12 gives a precise relationship between the canonical height function on the special fiber, the height of the parameter λ, and $\hat{h}_{\boldsymbol{f}}(\boldsymbol{c})$, which is equal to $\deg(\boldsymbol{c})$ in this case.

7. Explicit formula for the Green function

In this section, we work under the assumption that $|\cdot|_v = |\cdot|$ is archimedean, and $\mathbb{C}_v$ simply denotes $\mathbb{C}$ in this case. We show that in this setting we have an alternative way of representing the Green function $G_c := G_{\boldsymbol{c},v}$ for the Mandelbrot set $M_c := M_{\boldsymbol{c},v}$. We continue to work under the same hypothesis on $\boldsymbol{c}(\lambda)$; in particular, we assume that (5.1) holds. Furthermore, if $r = 0$ (that is, $\boldsymbol{f}$ is a constant family of polynomials), then $m = \deg(\boldsymbol{c}) \geq 1$.

Since the degree in x of $f_\lambda(x)$ is d, there exists a unique function ϕ_λ that is an analytic homeomorphism on the set U_{R_λ} for some $R_\lambda \geq 1$ (where for any positive real number R, U_R denotes the open set $\{ z \in \mathbb{C} : |z| > R \}$) satisfying the following conditions:

(1) ϕ_λ has derivative equal to 1 at ∞, or more precisely, the analytic function $\psi_\lambda(z) := 1/\phi_\lambda(1/z)$ has derivative equal to 1 at $z = 0$, and

(2) $\phi_\lambda(f_\lambda(z)) = (\phi_\lambda(z))^d$ for $|z| > R_\lambda$.

We can make (1) above more precise by giving the power series expansion

$$\phi_\lambda(z) = z + \sum_{n=1}^{\infty} \frac{A_{\lambda,n}}{z^n}. \tag{7.1}$$

From (7.1) we immediately conclude that $|\phi_\lambda(z)| = |z| + O_\lambda(1)$, and thus,

$$\log|\phi_\lambda(z)| = \log|z| + O_\lambda(1) \quad \text{for } |z| \text{ large enough.} \tag{7.2}$$

So using that $\phi_\lambda(f_\lambda(z)) = \phi_\lambda(z)^d$, we conclude that if $|z| > R_\lambda$, then

$$\lim_{n\to\infty} \frac{\log^+ |f_\lambda^n(z)|}{d^n} = \lim_{n\to\infty} \frac{\log |\phi_\lambda(f_\lambda^n(z))|}{d^n} = \log |\phi_\lambda(z)|. \tag{7.3}$$

Hence, (7.3) yields that the Green function G^λ for the (filled Julia set of the) polynomial f_λ equals

$$G^\lambda(z) := \lim_{n\to\infty} \frac{\log |f_\lambda^n(z)|}{d^n} = \log |\phi_\lambda(z)| \quad \text{if } |z| > R_\lambda.$$

For details on the Green function associated with any polynomial, see [Carleson and Gamelin 1993], where Chapter III.4 says that the function $\log |\phi_\lambda(z)|$ can be extended to a well-defined harmonic function on the entire basin of attraction A_∞^λ of the point at ∞ for the polynomial map f_λ. The set A_∞^λ is the complement of the filled Julia set of f_λ; more precisely, it is the set of all $z \in \mathbb{C}$ such that the orbit of z under f_λ is unbounded. Thus, on A_∞^λ we have

$$G^\lambda(z) := \log |\phi_\lambda(z)| \tag{7.4}$$

is the Green function for (the filled Julia set of) the polynomial f_λ. Also by [ibid., Chapter III.4], we know that

$$R_\lambda := \max_{f_\lambda'(x)=0} e^{G^\lambda(x)} \geq 1.$$

In Proposition 7.6, we will show that if $|\lambda|$ is sufficiently large, then $\boldsymbol{c}(\lambda)$ is in the domain of analyticity for ϕ_λ. In particular, using (7.2) this would yield

$$G_{\boldsymbol{c}}(\lambda) = \lim_{n\to\infty} \frac{\log^+ |f_\lambda^n(\boldsymbol{c}(\lambda))|}{md^n} = \frac{\log |\phi_\lambda(\boldsymbol{c}(\lambda))|}{m} = \frac{G^\lambda(\boldsymbol{c}(\lambda))}{m} \tag{7.5}$$

for $|\lambda|$ sufficiently large.

Proposition 7.6. *There exists a positive constant C_0 such that if $|\lambda| > C_0$, then $\boldsymbol{c}(\lambda)$ belongs to the analyticity domain of ϕ_λ.*

Proof. The proof is similar to that of [Baker and DeMarco 2011, Lemma 3.2]. If $\boldsymbol{f}$ is a constant family of polynomials, then the conclusion is immediate since R_λ is constant (independent of λ), and thus, for $|\lambda|$ sufficiently large, clearly $|\boldsymbol{c}(\lambda)| > R_\lambda$. So from now on assume $\boldsymbol{f}$ is not a constant family of polynomials, which in particular yields that $r \geq 1$ and $0 < m_1 < \cdots < m_r$.

First we recall that

$$R_\lambda = e^{G^\lambda(x_0)} := \max_{f_\lambda'(x)=0} e^{G^\lambda(x)}.$$

Next we show that $R_\lambda \to \infty$ as $|\lambda| \to \infty$, which will be used later in our proof.

Lemma 7.7. *As $|\lambda| \to \infty$, we have $R_\lambda \to \infty$.*

Proof. We recall that

$$f_\lambda(x) = P(x) + \sum_{i=1}^{r} \lambda^{m_i} \cdot Q_i(x),$$

where $P(x)$ is a polynomial in normal form of degree d and $0 < m_1 < \cdots < m_r$ are positive integers while the Q_i are nonzero polynomials of degrees $e_i \le d-2$. We have two cases.

Case 1. Each $Q_i(x)$ is a constant polynomial. Then the critical points of f_λ are independent of λ, that is, $x_0 = O(1)$. We let $x_1 \in \mathbb{C}$ such that $f_\lambda(x_1) = x_0$. Since each Q_i is a nonzero constant polynomial, we immediately conclude that $|x_1| \gg |\lambda|^{m_r/d}$. On the other hand, since $U_{2R_\lambda} \subset \phi_\lambda^{-1}(U_{R_\lambda})$ [Branner and Hubbard 1988, Corollary 3.3], we conclude that $|x_1| \le 2R_\lambda$, so $R_\lambda \gg |\lambda|^{m_r/d}$. Indeed, if $|x_1| > 2R_\lambda$, then there exists $z_1 \in U_{R_\lambda}$ such that $\phi_\lambda^{-1}(z_1) = x_1$. Using the fact that ϕ_λ is a conjugacy map at ∞ for f_λ, we would obtain that

$$x_0 = f_\lambda(x_1) = f_\lambda(\phi_\lambda^{-1}(z_1)) = \phi_\lambda^{-1}(z_1^d) \in U_{R_\lambda},$$

which contradicts the fact that x_0 is not in the analyticity domain of ϕ_λ.

Case 2. There exists $i = 1, \dots, r$ such that $Q_i(x)$ is not a constant polynomial. Then the critical points of f_λ vary with λ. In particular, there exists a critical point x_λ of maximum absolute value such that $|x_\lambda| \gg |\lambda|^{m_j/(d-e_j)}$ (for some $j = 1, \dots, r$), where for each $i = 1, \dots, r$, we have $e_i = \deg(Q_i) \le d-2$. Now, x_λ is not in the domain of analyticity of ϕ_λ, and thus, $|x_\lambda| \le R_\lambda$, which again shows that $R_\lambda \to \infty$ as $|\lambda| \to \infty$. ☐

Using that $R_\lambda \to \infty$, we will finish our proof. First we note that

$$|\phi_\lambda(f_\lambda(x_0))| = e^{G^\lambda(f_\lambda(x_0))} = e^{dG^\lambda(x_0)} = R_\lambda^d. \tag{7.8}$$

Note that $\phi_\lambda(z)$ is analytic on U_{R_λ} while $\log|\phi_\lambda(z)|$ is continuous for $|z| \ge R_\lambda$. Moreover, whenever it is defined, $G^\lambda(f_\lambda(z)) = dG^\lambda(z)$, so also using (7.4), we obtain (7.8).

Now for $|\lambda|$ sufficiently large, we have that $R_\lambda^d/2 > R_\lambda$ (since $R_\lambda \to \infty$ according to Lemma 7.7). So $U_{R_\lambda^d} \subset \phi_\lambda(U_{R_\lambda^d/2})$ [Branner and Hubbard 1988, Corollary 3.3], and thus,

$$|f_\lambda(x_0)| \ge \tfrac{1}{2} R_\lambda^d. \tag{7.9}$$

Case 1. We have $\deg(Q_i) = 0$ for each i. Then $x_0 = O(1)$ as noticed in Lemma 7.7, and thus, using (7.9) we obtain that $|\lambda|^{m_r} \gg R_\lambda^d$. Since $\deg(\boldsymbol{c}) = m \ge m_r$, we obtain

$$|\boldsymbol{c}(\lambda)| \ge |q_m| \cdot |\lambda|^m - |O(\lambda^{m-1})| \gg R_\lambda^d > R_\lambda$$

if $|\lambda|$ is sufficiently large.

Case 2. If not all of the Q_i are constant polynomials, then we still know that

$$|x_0| \ll |\lambda|^{\max_{i=1}^r m_i/(d-e_i)} \ll |\lambda|^{m_r/2}$$

because $e_i \le d-2$ for each i. Therefore,

$$R_\lambda^d \ll |f_\lambda(x_0)| \ll |\lambda|^{dm_r/2}. \tag{7.10}$$

On the other hand, $|\boldsymbol{c}(\lambda)| \sim |\lambda|^m$ and $m \ge m_r$, which yield that

$$|\boldsymbol{c}(\lambda)| \gg |\lambda|^m \gg R_\lambda^2 \gg R_\lambda$$

by (7.10). This concludes the proof of Proposition 7.6. □

Therefore, for large $|\lambda|$, the point $\boldsymbol{c}(\lambda)$ is in the domain of analyticity for ϕ_λ, which allows us to conclude that (7.5) holds.

We know from [Carleson and Gamelin 1993] that for each $\lambda \in \mathbb{C}$ and for each $z \in \mathbb{C}$ sufficiently large in absolute value, we have

$$\phi_\lambda(z) = z \prod_{n=0}^{\infty} \left(\frac{f_\lambda^{n+1}(z)}{f_\lambda^n(z)^d} \right)^{1/d^{n+1}}, \tag{7.11}$$

and thus,

$$\phi_\lambda(z) = z \prod_{n=0}^{\infty} \left(1 + \frac{Q_0(f_\lambda^n(z)) + \sum_{i=1}^r Q_i(f_\lambda^n(z)) \cdot \lambda^{m_i}}{f_\lambda^n(z)^d} \right)^{1/d^{n+1}}, \tag{7.12}$$

where $Q_0(x) := P(x) - x^d$ is a polynomial of degree at most equal to $d-2$. We showed in Proposition 7.6 that $\phi_\lambda(\boldsymbol{c}(\lambda))$ is well-defined; furthermore, the function $\phi_\lambda(\boldsymbol{c}(\lambda))/\boldsymbol{c}(\lambda)$ can be expressed near ∞ as the infinite product above. Indeed, for each $n \in \mathbb{N}$, the order of magnitude of the numerator in the n-th fraction from the product appearing in (7.12) when we substitute $z = \boldsymbol{c}(\lambda)$ is at most

$$|\lambda|^{m+(d-2)md^n} \le |\lambda|^{m(d-1)d^n}$$

while the order of magnitude of the denominator is $|\lambda|^{md^{n+1}}$. This guarantees the convergence of the product from (7.12) corresponding to $\phi_\lambda(\boldsymbol{c}(\lambda))/\boldsymbol{c}(\lambda)$. We conclude that

$$\phi_\lambda(\boldsymbol{c}(\lambda)) \text{ is an analytic function of } \lambda \text{ (for large } \lambda\text{) and moreover} \tag{7.13}$$

$$\phi_\lambda(\boldsymbol{c}(\lambda)) = q_m\lambda^m + O(\lambda^{m-1}). \tag{7.14}$$

8. Proof of Theorem 2.3: Algebraic case

We work under the hypothesis of Theorem 2.3, and we continue with the notation from the previous sections. Furthermore, we prove Theorem 2.3 under the extra assumptions that

$$\boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{Q}}[\lambda] \quad \text{and} \quad c_i \in \overline{\mathbb{Q}}[\lambda] \quad \text{for each } i = 0, \ldots, d-2. \tag{8.1}$$

Recall that $f_\lambda(x) = x^d + \sum_{i=0}^{d-2} c_i(\lambda)x^i$, where we require that $c_i \in \overline{\mathbb{Q}}[\lambda]$ for $i = 0, \ldots, d-2$. Let $\boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{Q}}[\lambda]$ satisfy the hypotheses (i) and (ii) of Theorem 2.3. Let K be the number field generated by the coefficients of $c_i(\lambda)$ for $i = 0, \ldots, d-2$ and of $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$. Let Ω_K be the set of all inequivalent absolute values on K.

Next, assume there exist infinitely many λ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f_λ. At the expense of replacing $\boldsymbol{a}(\lambda)$ by $f_\lambda^k(\boldsymbol{a}(\lambda))$ and $\boldsymbol{b}(\lambda)$ by $f_\lambda^\ell(\boldsymbol{b}(\lambda))$, we may assume that the polynomials

$$\boldsymbol{a}(\lambda) \text{ and } \boldsymbol{b}(\lambda) \text{ have the same leading coefficient and degree } m \geq m_r. \tag{8.2}$$

Let $h_{\mathbb{M}_{\boldsymbol{a}}}(z)$ and $h_{\mathbb{M}_{\boldsymbol{b}}}(z)$ be the heights of $z \in \overline{K}$ relative to the adelic generalized Mandelbrot sets $\mathbb{M}_{\boldsymbol{a}} := \prod_{v \in \Omega_K} M_{\boldsymbol{a},v}$ and $\mathbb{M}_{\boldsymbol{b}}$ as defined in Section 6. Note that if $\lambda \in \overline{K}$ is a parameter such that $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f_λ, then $h_{\mathbb{M}_{\boldsymbol{a}}}(\lambda) = 0$ by Corollary 6.12. So we may apply the equidistribution result from [Baker and Rumely 2010, Theorem 7.52] (see our Theorem 4.3) and conclude that $M_{\boldsymbol{a},v} = M_{\boldsymbol{b},v}$ for each place $v \in \Omega_K$. Indeed, we know that there exists an infinite sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ of distinct numbers $\lambda \in \overline{K}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f_λ. So for each $n \in \mathbb{N}$, we may take S_n to be the union of the sets of Galois conjugates for λ_m for all $1 \leq m \leq n$. Clearly, $\#S_n \to \infty$ as $n \to \infty$, and also each S_n is $\mathrm{Gal}(\overline{K}/K)$-invariant. Finally, $h_{\mathbb{M}_{\boldsymbol{a}}}(S_n) = h_{\mathbb{M}_{\boldsymbol{b}}}(S_n) = 0$ for all $n \in \mathbb{N}$, and thus, Theorem 4.3 applies in this case. We obtain that $\mu_{\mathbb{M}_{\boldsymbol{a}}} = \mu_{\mathbb{M}_{\boldsymbol{b}}}$, and since they are both supported on $\mathbb{M}_{\boldsymbol{a}}$ and $\mathbb{M}_{\boldsymbol{b}}$, respectively, we also get that $\mathbb{M}_{\boldsymbol{a}} = \mathbb{M}_{\boldsymbol{b}}$. The following lemma applies in the generality of Theorem 2.3, and it will finish our proof. (Note that since K is a number field, it has at least one archimedean valuation.)

Lemma 8.3. *Let $\boldsymbol{f}$, $\boldsymbol{a}$, and $\boldsymbol{b}$ be as in Theorem 2.3; in particular, assume they are all defined over $\mathbb{C}$. Let $|\cdot|$ be the usual archimedean absolute value on $\mathbb{C}$, and let M_a and M_b be the corresponding complex Mandelbrot sets. If $M_a = M_b$, then $\boldsymbol{a} = \boldsymbol{b}$.*

Proof. Since $M_a = M_b$, then the corresponding Green functions are also the same, that is, (using (7.5) and (8.2))

$$|\phi_\lambda(\boldsymbol{a}(\lambda))| = |\phi_\lambda(\boldsymbol{b}(\lambda))| \quad \text{for all } |\lambda| \text{ sufficiently large.}$$

On the other hand, for $|z|$ large, the function $h(z) := \phi_z(\boldsymbol{a}(z))/\phi_z(\boldsymbol{b}(z))$ is an analytic function of constant absolute value. (Note that the denominator does not vanish since ϕ_λ is a homeomorphism for a neighborhood of ∞.) By the open mapping theorem, we conclude that $h(z) := u$ is a constant (for some $u \in \mathbb{C}$ of absolute value equal to 1); that is,

$$\phi_\lambda(\boldsymbol{a}(\lambda)) = u \cdot \phi_\lambda(\boldsymbol{b}(\lambda)). \tag{8.4}$$

Using (7.13) and (7.14) (also note that $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ have the same leading coefficient), we have $u = 1$. Using that ϕ_λ is a homeomorphism on a neighborhood of the infinity, we conclude that $\boldsymbol{a}(\lambda) = \boldsymbol{b}(\lambda)$ for λ sufficiently large in absolute value and thus for *all* λ as desired. (Note that $\boldsymbol{a}$ and $\boldsymbol{b}$ are polynomials.) □

Remark 8.5. Our proof (similar to the proof from [Baker and DeMarco 2011]) only uses in an essential way the information that $M_a = M_b$, that is, that the Mandelbrot sets over the complex numbers corresponding to $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal, even though we know that $M_{\boldsymbol{a},v} = M_{\boldsymbol{b},v}$ for *all* places v.

9. Proof of Theorem 2.3: The converse implication

Now we prove the converse implication in Theorem 2.3 in the general case, that is, for polynomials $c_0, \dots, c_{d-2}$, $\boldsymbol{a}$, and $\boldsymbol{b}$ with arbitrary complex coefficients. Again at the expense of replacing $\boldsymbol{a}(\lambda)$ by $f_\lambda^k(\boldsymbol{a}(\lambda))$ and replacing $\boldsymbol{b}(\lambda)$ by $f_\lambda^\ell(\boldsymbol{b}(\lambda))$, we may assume $\boldsymbol{a}(\lambda) = \boldsymbol{b}(\lambda)$. The following result will finish the converse statement in Theorem 2.3:

Proposition 9.1. *Let $\boldsymbol{c} \in \mathbb{C}[\lambda]$ of degree $m \geq m_r$. Let* $\mathrm{Prep}(\boldsymbol{c})$ *be the set consisting of all $\lambda \in \mathbb{C}$ such that $\boldsymbol{c}(\lambda)$ is preperiodic under f_λ, and let M_c be the set of all $\lambda \in \mathbb{C}$ such that the orbit of $\boldsymbol{c}(\lambda)$ under the action of f_λ is bounded with respect to the usual archimedean metric on $\mathbb{C}$. Then the closure in $\mathbb{C}$ of the set* $\mathrm{Prep}(\boldsymbol{c})$ *contains ∂M_c. In particular,* $\mathrm{Prep}(\boldsymbol{c})$ *is infinite.*

Proof. We first claim that the equation $f_z(\boldsymbol{c}(z)) = \boldsymbol{c}(z)$ has only finitely many solutions. Indeed, according to Lemma 5.2, the degree in z of $f_z(\boldsymbol{c}(z)) - \boldsymbol{c}(z)$ is dm, which means that there are at most dm solutions $z \in \mathbb{C}$ for the equation $f_z(\boldsymbol{c}(z)) = \boldsymbol{c}(z)$.

Let $x_0 \in \partial M_c$, which is *not* a solution z to $f_z(\boldsymbol{c}(z)) = \boldsymbol{c}(z)$; we will show that x_0 is contained in the closure in $\mathbb{C}$ of $\mathrm{Prep}(\boldsymbol{c})$. Since we already know that if $f_z(\boldsymbol{c}(z)) = \boldsymbol{c}(z)$, then $z \in \mathrm{Prep}(\boldsymbol{c})$, we will be done once we prove that each open neighborhood U of x_0 contains at least one point from $\mathrm{Prep}(\boldsymbol{c})$.

Now, let U be an open neighborhood of x_0, and let $h_i : U \to \mathbb{P}^1(\mathbb{C})$ for $i = 1, 2, 3$ be three analytic functions with values taken in the compact Riemann sphere, given by

$$h_1(z) := \infty, \quad h_2(z) := \boldsymbol{c}(z), \quad \text{and} \quad h_3(z) := g_{\boldsymbol{c},1}(z) = f_z(\boldsymbol{c}(z)).$$

Furthermore, since x_0 is not a solution for the equation $h_2(z) = h_3(z)$, then we may assume (at the expense of replacing U with a smaller neighborhood of x_0) that the closures of $h_2(U)$ and $h_3(U)$ are disjoint. Therefore, the closures of $h_1(U)$, $h_2(U)$, and $h_3(U)$ in $\mathbb{P}^1(\mathbb{C})$ are all disjoint.

As before, we let $\{g_{\boldsymbol{c},n}\}_{n\ge 2}$ be the set of polynomials $g_{\boldsymbol{c},n}(z) := f_z^n(\boldsymbol{c}(z))$. Since $x_0 \in \partial M_c$, the family of analytic maps $\{g_{\boldsymbol{c},n}\}_{n\ge 2}$ is not normal on U. Therefore, by Montel's theorem [Beardon 1991, Theorem 3.3.6], there exists $n \ge 2$ and $z \in U$ such that $g_{\boldsymbol{c},n}(z) = \boldsymbol{c}(z)$ or $g_{\boldsymbol{c},n}(z) = f_z(\boldsymbol{c}(z))$. (Clearly, it cannot happen that $g_{\boldsymbol{c},n}(z) = \infty$.) Either way, we obtain that $z \in \mathrm{Prep}(\boldsymbol{c})$ as desired.

Since $\gamma(M_c) > 0$, we know that M_c is an uncountable subset of $\mathbb{C}$, and thus, its boundary is infinite; hence, $\mathrm{Prep}(\boldsymbol{c})$ is also infinite. □

10. Proof of Theorem 2.3: General case

In this section, we finish the proof of Theorem 2.3. With the same notation as in Theorem 2.3, we replace $\boldsymbol{a}$ and $\boldsymbol{b}$ with $f_\lambda^k(\boldsymbol{a}(\lambda))$ and $f_\lambda^\ell(\boldsymbol{b}(\lambda))$, respectively; thus, $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are polynomials with the same degree and same leading coefficient. We assume there exist infinitely many $\lambda \in \mathbb{C}$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for f_λ; we will prove that $\boldsymbol{a} = \boldsymbol{b}$.

Let K denote the field generated over $\overline{\mathbb{Q}}$ by adjoining the coefficients of each c_i (for $i = 1, \ldots, d-2$) and adjoining the coefficients of $\boldsymbol{a}$ and of $\boldsymbol{b}$. According to Corollary 5.3, if there exists $\lambda \in \mathbb{C}$ such that $\boldsymbol{a}(\lambda)$ (or $\boldsymbol{b}(\lambda)$) is preperiodic for f_λ, then $\lambda \in \overline{K}$, where $\overline{K}$ denotes the algebraic closure of K in $\mathbb{C}$. Let Ω_K be the set of inequivalent absolute values of K corresponding to the divisors of a projective $\overline{\mathbb{Q}}$-variety $\mathcal{V}$ regular in codimension 1; then the places in Ω_K satisfy a product formula.

As in Section 8, we let $h_{\mathbb{M}_{\boldsymbol{a}}}(z)$ and $h_{\mathbb{M}_{\boldsymbol{b}}}(z)$ be the heights of $z \in \overline{K}$ relative to the adelic generalized Mandelbrot sets $\mathbb{M}_{\boldsymbol{a}} = \prod_{v\in\Omega_K} \boldsymbol{M}_{\boldsymbol{a},v}$ and $\mathbb{M}_{\boldsymbol{b}}$ as defined in Section 6. Note that if $\lambda \in \overline{K}$ is a parameter such that $\boldsymbol{a}(\lambda)$ is preperiodic for f_λ, then $h_{\mathbb{M}_{\boldsymbol{a}}}(\lambda) = 0$ and $h_{\mathbb{M}_{\boldsymbol{b}}}(\lambda) = 0$, respectively, by Corollary 6.12 again. So arguing as in Section 8, we may apply the equidistribution result from [Baker and Rumely 2010, Theorem 7.52] (Theorem 4.3) and conclude that $\boldsymbol{M}_{\boldsymbol{a},v} = \boldsymbol{M}_{\boldsymbol{b},v}$ for each place $v \in \Omega_K$.

As observed in our proof from Section 8 (see Remark 8.5), in order to finish the proof of Theorem 2.3, it suffices to prove that $M_a = M_b$, where M_a and M_b are the complex Mandelbrot sets corresponding to $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. By *complex* Mandelbrot sets M_a and M_b, we mean the Mandelbrot sets corresponding to $\boldsymbol{a}$ and $\boldsymbol{b}$ constructed with respect to the usual archimedean metric on $\mathbb{C}$.

As before, $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ denote the sets of all $\lambda \in \mathbb{C}$ such that $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$, respectively, are preperiodic for f_λ. As proved in Corollary 5.3, we know that both $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ are subsets of $\overline{K}$. In order to prove that $M_a = M_b$, it suffices to prove that $\mathrm{Prep}(\boldsymbol{a})$ differs from $\mathrm{Prep}(\boldsymbol{b})$ in at most finitely many points.

To ease the notation, we define the symmetric difference of $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ as

$$\mathrm{PrepDiff}(\boldsymbol{a}, \boldsymbol{b}) := (\mathrm{Prep}(\boldsymbol{a}) \setminus \mathrm{Prep}(\boldsymbol{b})) \cup (\mathrm{Prep}(\boldsymbol{b}) \setminus \mathrm{Prep}(\boldsymbol{a})).$$

Proposition 10.1. *If the set* $\mathrm{PrepDiff}(\boldsymbol{a}, \boldsymbol{b})$ *is finite, then* $M_a = M_b$.

Proof. Since M_a contains all points $\lambda \in \mathbb{C}$ such that $\lim_{n\to\infty} \log^+ |f_\lambda^n(a)|/d^n = 0$, the maximum modulus principle yields that the complement of M_a in $\mathbb{C}$ is connected; that is, M_a is a full subset of $\mathbb{C}$; see also [Baker and DeMarco 2011]. So both M_a and M_b are full subsets of $\mathbb{C}$ containing the sets $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ whose closures contain the boundary of M_a and M_b, respectively (according to Proposition 9.1). As $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ differ by at most finitely many elements, we conclude that $M_a = M_b$. □

To prove that $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ differ by at most finitely many elements, we observe first that if $\lambda \in \mathrm{Prep}(\boldsymbol{a})$, then $\hat{h}_{f_\lambda}(\boldsymbol{a}(\lambda)) = 0$, and thus, $\lambda^\sigma \in M_{\boldsymbol{a},v}$ for all v and all $\sigma \in \mathrm{Gal}(\overline{K}/K)$. (See (3.3); note that $\boldsymbol{a}(\lambda)^\sigma = \boldsymbol{a}(\lambda^\sigma)$ since $\boldsymbol{a} \in K[x]$.) Similarly, if $\lambda \in \mathrm{Prep}(\boldsymbol{b})$, then $\lambda^\sigma \in M_{\boldsymbol{b},v}$ for each place $v \in \Omega_K$ and each Galois morphism σ. We would like to use the reverse implication, that is, characterize the elements $\mathrm{Prep}(\boldsymbol{a})$ as the set of all $\lambda \in \overline{K}$ such that $\lambda^\sigma \in M_{\boldsymbol{a},v}$ for each place v and for each Galois morphism σ. This is true if f_λ is not isotrivial over $\overline{\mathbb{Q}}$ by Benedetto's result [2005]. In this case, $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ are exactly the sets of $\lambda \in \overline{K}$ such that $h_{\mathbb{M}_{\boldsymbol{a}}}(\lambda) = 0$ and $h_{\mathbb{M}_{\boldsymbol{a}}}(\lambda) = 0$, respectively. However, notice that if $f_\lambda \in \overline{\mathbb{Q}}[x]$, then

$$\lambda^\sigma \in M_{\boldsymbol{a},v} \text{ for all } v \in \Omega_K \text{ and } \sigma \in \mathrm{Gal}(\overline{K}/K) \text{ if and only if } \boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}.$$

We see that in this case, $\mathrm{Prep}(\boldsymbol{a})$ is strictly smaller than the set of $\lambda \in \overline{K}$ such that $h_{\mathbb{M}_{\boldsymbol{a}}}(\lambda) = 0$. So we will prove that $\mathrm{Prep}(\boldsymbol{a})$ and $\mathrm{Prep}(\boldsymbol{b})$ differ by at most finitely many elements by splitting our analysis into two cases depending on whether there exist infinitely many $\lambda \in \mathbb{C}$ such that f_λ is conjugate to a polynomial with coefficients in $\overline{\mathbb{Q}}$. The following easy result is key for our argument:

Lemma 10.2. *For any* $\lambda \in \mathbb{C}$, *the polynomial* $f_\lambda(x)$ *is conjugate to a polynomial with coefficients in* $\overline{\mathbb{Q}}$ *if and only if* $c_i(\lambda) \in \overline{\mathbb{Q}}$ *for each* $i = 1, \ldots, d-2$.

Proof. One direction is obvious. Now, assume f_λ is conjugate to a polynomial with coefficients in $\overline{\mathbb{Q}}$. Let $\delta(x) := ax + b$ be a linear polynomial so $\delta^{-1} \circ f_\lambda \circ \delta \in \overline{\mathbb{Q}}[x]$. Since f_λ is in normal form, we note that $a, b \in \overline{\mathbb{Q}}$ for otherwise the leading coefficient or the next-to-leading coefficient is not algebraic. Now, it is clear that each $c_i(\lambda) \in \overline{\mathbb{Q}}$ as desired. □

Let S be the set of all $\lambda \in \mathbb{C}$ such that f_λ is conjugate to a polynomial in $\overline{\mathbb{Q}}[x]$. Using Lemma 10.2, $S \subset \overline{K}$ since each polynomial c_i has coefficients in K and $\overline{\mathbb{Q}} \subset K$. Also, S is $\mathrm{Gal}(\overline{K}/K)$-invariant since each coefficient of each c_i is in K.

Proposition 10.3. $\mathrm{PrepDiff}(\boldsymbol{a}, \boldsymbol{b}) \subset S$.

Proof. Let $\lambda \in \overline{K} \setminus S$. Since f_λ is not conjugate to a polynomial in $\overline{\mathbb{Q}}$, using Benedetto's result (see also (3.3)) we obtain that $\boldsymbol{a}(\lambda)$ is preperiodic for f_λ if and only if for each $v \in \Omega_K$ and $\sigma \in \mathrm{Gal}(\overline{K}/K)$, the local canonical height of $\boldsymbol{a}(\lambda)^\sigma = \boldsymbol{a}(\lambda^\sigma)$ computed with respect to f_λ^σ equals 0. Since each coefficient of $c_i(\lambda)$ is defined over K, we get that $f_\lambda^\sigma = f_{\lambda^\sigma}$. Therefore, for each $\lambda \in \overline{K} \setminus S$, we see that $\boldsymbol{a}(\lambda)$ or $\boldsymbol{b}(\lambda)$ is preperiodic for f_λ if and only if for all $v \in \Omega_K$ and all $\sigma \in \mathrm{Gal}(\overline{K}/K)$, we have $\lambda^\sigma \in M_{\boldsymbol{a},v}$ or $\lambda^\sigma \in M_{\boldsymbol{b},v}$, respectively. Using the fact that $M_{\boldsymbol{a},v} = M_{\boldsymbol{b},v}$ for all $v \in \Omega_K$, we conclude that if $\lambda \in \overline{K} \setminus S$, then $\lambda \in \mathrm{Prep}(a)$ if and only if $\lambda \in \mathrm{Prep}(b)$. Hence, $\mathrm{PrepDiff}(\boldsymbol{a}, \boldsymbol{b}) \subset S$ as desired. □

Lemma 10.4. *If $\lambda \in S$ and $\boldsymbol{a}(\lambda) \notin \overline{\mathbb{Q}}$, then $\boldsymbol{a}(\lambda)$ is not preperiodic for f_λ.*

Proof. The assertion is immediate since for $\lambda \in S$ we have $f_\lambda \in \overline{\mathbb{Q}}[x]$ by the definition of S (see also Lemma 10.2); hence, the set of preperiodic points of f_λ is contained in $\overline{\mathbb{Q}}$. By assumption $\boldsymbol{a}(\lambda) \notin \overline{\mathbb{Q}}$; therefore, $\boldsymbol{a}(\lambda)$ is not preperiodic for f_λ. □

Proposition 10.5. *The set* $\mathrm{PrepDiff}(\boldsymbol{a}, \boldsymbol{b})$ *is finite.*

Proof. If S is a finite set, then the assertion follows from Proposition 10.3. So in the remaining part of the proof, we assume that S is an infinite set. By Lemma 10.2 we know that there exist infinitely many $\lambda \in \overline{K}$ such that $c_i(\lambda) \in \overline{\mathbb{Q}}$ for each $i = 0, \ldots, d-2$. The following lemma will be key for our proof:

Lemma 10.6. *Let $L_1 \subset L_2$ be algebraically closed fields of characteristic* 0, *and let $f_1, \ldots, f_n \in L_2[x]$. If there exist infinitely many $z \in L_2$ such that $f_i(z) \in L_1$ for each $i = 1, \ldots, n$, then there exists $h \in L_2[x]$, and there exist $g_1, \ldots, g_n \in L_1[x]$ such that $f_i = g_i \circ h$ for each $i = 1, \ldots, n$.*

Proof. Let $C \subset \mathbb{A}^n$ be the Zariski closure of the set

$$\{\, (f_1(z), \ldots, f_n(z)) : z \in L_2 \,\}. \tag{10.7}$$

Then C is a rational curve that (by our hypothesis) contains infinitely many points over L_1. Therefore, C is defined over L_1, and thus, it has a rational parametrization over L_1. Let

$$(g_1, \ldots, g_n) : \mathbb{A}^1 \to C$$

be a birational morphism defined over L_1; we denote by $\psi : C \to \mathbb{A}^1$ its inverse. (For more details, see [Shafarevich 1994, Chapter 1]). Since the closure of C in $\mathbb{P}^n$ (by considering the usual embedding of $\mathbb{A}^n \subset \mathbb{P}^n$) has only one point at infinity (due to the parametrization (10.7) of C), we conclude that (perhaps after a change of coordinates) we may assume each g_i is also a polynomial; more precisely, $g_i \in L_1[x]$. We let $h : \mathbb{A}^1 \to \mathbb{A}^1$ be the rational map (defined over L_2) given by the composition

$$h := \psi \circ (f_1, \ldots, f_n).$$

Therefore, for each $i = 1, \dots, n$, we have $f_i = g_i \circ h$, and since both f_i and g_i are polynomials, we conclude that h is also a polynomial, as desired. $\square$

As an immediate consequence of Lemma 10.6, we have the following result:

Corollary 10.8. *Let $L_1 \subset L_2$ be algebraically closed fields of characteristic* 0, *and let $f_1, \dots, f_n \in L_2[x]$. If there exist infinitely many $z \in L_2$ such that $f_i(z) \in L_1$ for $i = 1, \dots, n$, then for any $i, j \in \{1, \dots, n\}$ and any $z \in L_2$, we have $f_i(z) \in L_1$ if and only if $f_j(z) \in L_1$.*

There are two possibilities: Either there exist infinitely many $\lambda \in S$ such that $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$ or not.

Lemma 10.9. *If there exist infinitely many $\lambda \in S$ such that $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$, then $\boldsymbol{a} = \boldsymbol{b}$. In particular*, $\mathrm{Prep}(\boldsymbol{a}) = \mathrm{Prep}(\boldsymbol{b})$.

Proof. Using Corollary 10.8 we obtain that actually for *all* $\lambda \in S$ we have that $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$. So in this case *each* λ^σ belongs to each $M_{\boldsymbol{a},v}$ for each place v of the function field $K/\overline{\mathbb{Q}}$ and for each $\sigma \in \mathrm{Gal}(\overline{K}/K)$. (Note that for such $\lambda \in S$ we have that both $f_\lambda \in \overline{\mathbb{Q}}[x]$ and $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$, and also note that S is $\mathrm{Gal}(\overline{K}/K)$-invariant.) Since $M_{\boldsymbol{a},v} = M_{\boldsymbol{b},v}$ for each place v, we conclude that $\lambda^\sigma \in M_{\boldsymbol{b},v}$ for each $\lambda \in S$, for each $v \in \Omega_K$, and for each $\sigma \in \mathrm{Gal}(\overline{K}/K)$. Since $f_\lambda \in \overline{\mathbb{Q}}[x]$, we conclude that $\boldsymbol{b}(\lambda) \in \overline{\mathbb{Q}}$ as well. Indeed, otherwise $|\boldsymbol{b}(\lambda)^\sigma|_v > 1$ for some place v and some Galois morphism σ, and thus, $|f_\lambda^n(\boldsymbol{b}(\lambda^\sigma))|_v \to \infty$ as $n \to \infty$, contradicting the fact that $\lambda^\sigma \in M_{\boldsymbol{b},v}$. Hence, both $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$ and $\boldsymbol{b}(\lambda) \in \overline{\mathbb{Q}}$ for $\lambda \in S$.

Therefore, applying Lemma 10.6 to the polynomials $c_0, \dots, c_{d-2}$, $\boldsymbol{a}$, and $\boldsymbol{b}$, we conclude that there exist polynomials $c'_0, \dots, c'_{d-2}, \boldsymbol{a}', \boldsymbol{b}' \in \overline{\mathbb{Q}}[x]$ and $h \in \overline{K}[x]$ such that

$$c_i = c'_i \circ h \text{ for each } i = 0, \dots, d-2 \text{ and} \tag{10.10}$$

$$\boldsymbol{a} = \boldsymbol{a}' \circ h \text{ and } \boldsymbol{b} = \boldsymbol{b}' \circ h. \tag{10.11}$$

We let $\delta := h(\lambda)$ and define the family of polynomials

$$f'_\delta(x) := x^d + \sum_{i=0}^{d-2} c'_i(\delta) x^i.$$

So we reduced the problem to the case studied in Section 8 for the family of polynomials $f'_\delta \in \overline{\mathbb{Q}}[x]$ and to the starting points $\boldsymbol{a}', \boldsymbol{b}' \in \overline{\mathbb{Q}}[\delta]$. Note that using hypotheses (i) and (ii) from Theorem 2.3 and also relations (10.10) and (10.11), $\boldsymbol{a}'(\delta)$ and $\boldsymbol{b}'(\delta)$ have the same leading coefficient and the same degree, which is larger than the degrees of the c'_i. So since we know there exist infinitely many $\delta \in \mathbb{C}$ such that $\boldsymbol{a}'(\delta)$ and $\boldsymbol{b}'(\delta)$ are both preperiodic for f'_δ, we conclude that $\boldsymbol{a}' = \boldsymbol{b}'$ as proved in Section 8. Hence, $\boldsymbol{a} = \boldsymbol{b}$, and thus, $\mathrm{Prep}(\boldsymbol{a}) = \mathrm{Prep}(\boldsymbol{b})$. $\square$

Lemma 10.12. *If finitely many $\lambda \in S$ exist such that $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$, then* $\operatorname{PrepDiff}(\boldsymbol{a}, \boldsymbol{b})$ *is finite.*

Proof. First, note that there must be at most finitely many $\lambda \in S$ such that $\boldsymbol{b}(\lambda) \in \overline{\mathbb{Q}}$. Otherwise, arguing as in the proof of Lemma 10.9, we would obtain that for all the infinitely many $\lambda \in S$, both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are in $\overline{\mathbb{Q}}$, which violates the lemma's hypothesis. So let T be the finite subset of S containing all λ such that either $\boldsymbol{a}(\lambda) \in \overline{\mathbb{Q}}$ or $\boldsymbol{b}(\lambda) \in \overline{\mathbb{Q}}$.

Let $\lambda \in (\overline{K} \setminus T) \cap \operatorname{Prep}(\boldsymbol{a})$. If $\lambda \in S$, then by Lemma 10.4 we know that $\lambda \notin \operatorname{Prep}(\boldsymbol{a})$, a contradiction. Therefore, $\lambda \notin S$, so by Proposition 10.3, we have $\lambda \notin \operatorname{PrepDiff}(\boldsymbol{a}, \boldsymbol{b})$. Similarly, if $\lambda \in (\overline{K} \setminus T) \cap \operatorname{Prep}(\boldsymbol{b})$, then $\lambda \notin \operatorname{PrepDiff}(\boldsymbol{a}, \boldsymbol{b})$. Thus, $\operatorname{PrepDiff}(\boldsymbol{a}, \boldsymbol{b})$ is contained in the finite set T. □

Lemmas 10.9 and 10.12 finish the proof of Proposition 10.5. □

Therefore, Proposition 10.5 yields that $\operatorname{Prep}(\boldsymbol{a})$ and $\operatorname{Prep}(\boldsymbol{b})$ differ by at most finitely many elements. Then it follows from Proposition 10.1 that the corresponding complex Mandelbrot sets M_a and M_b are equal, so we conclude our proof of Theorem 2.3 using Lemma 8.3.

11. Connections to the dynamical Manin–Mumford conjecture

We first prove Corollary 2.7, and then we present further connections between our Question 1.3 and the dynamical Manin–Mumford conjecture formulated by Ghioca, Tucker, and Zhang [2011].

Proof of Corollary 2.7. At the expense of replacing f by a conjugate $\delta^{-1} \circ f \circ \delta$ and replacing $\boldsymbol{a}$ and $\boldsymbol{b}$ by $\delta^{-1} \circ \boldsymbol{a}$ and $\delta^{-1} \circ \boldsymbol{b}$, respectively, we may assume f is in normal form. By the hypothesis of Corollary 2.7, we know that there are infinitely many $\lambda_n \in \overline{\mathbb{Q}}$ such that

$$\lim_{n\to\infty} \hat{h}_f(\boldsymbol{a}(\lambda_n)) + \hat{h}_f(\boldsymbol{b}(\lambda_n)) = 0.$$

We let $\boldsymbol{f} := f_\lambda := f$ be the constant family of polynomials f indexed by $\lambda \in \overline{\mathbb{Q}}$. As before, we let K be the field generated by coefficients of f, $\boldsymbol{a}$, and $\boldsymbol{b}$ and let $h_{\mathbb{M}_{\boldsymbol{a}}}(z)$ and $h_{\mathbb{M}_{\boldsymbol{b}}}(z)$ be the heights of $z \in \overline{K}$ relative to the adelic generalized Mandelbrot sets $\mathbb{M}_{\boldsymbol{a}} := \prod_{v\in\Omega_K} M_{\boldsymbol{a},v}$ and $\mathbb{M}_{\boldsymbol{b}}$, respectively, as defined in Section 6. So we may apply the equidistribution result from [Baker and Rumely 2010, Theorem 7.52] (see our Theorem 4.3) and conclude that $M_{\boldsymbol{a},v} = M_{\boldsymbol{b},v}$ for each place $v \in \Omega_K$. Indeed, for each $n \in \mathbb{N}$, we may take S_n to be the set of Galois conjugates of λ_n. Clearly $\#S_n \to \infty$ as $n \to \infty$ (since the points λ_n are distinct and their heights are bounded because the heights of $\boldsymbol{a}(\lambda_n)$ and $\boldsymbol{b}(\lambda_n)$ are bounded). Finally, $\lim_{n\to\infty} h_{\mathbb{M}_{\boldsymbol{a}}}(S_n) = \lim_{n\to\infty} h_{\mathbb{M}_{\boldsymbol{b}}}(S_n) = 0$ (by Corollary 6.12), and thus, Theorem 4.3 applies in this case.

Using that $M_{\boldsymbol{a},v} = M_{\boldsymbol{b},v}$ for an archimedean place v, the same argument as in the proof of Theorem 2.3 yields that $\boldsymbol{a} = \boldsymbol{b}$ as desired. $\square$

Next we discuss the connection between our Question 1.3 and the dynamical Manin–Mumford conjecture [Ghioca et al. 2011, Conjecture 1.4]. First we recall that for a projective variety X and an endomorphism Φ of X, we say that Φ is *polarizable* if there exists an integer $d > 1$ and there exists an ample line bundle $\mathscr{L}$ on X such that $\Phi^*(\mathscr{L}) = \mathscr{L}^{\otimes d}$.

Conjecture 11.1 (Ghioca, Tucker, Zhang). *Let X be a projective variety, define $\varphi : X \to X$ to be a polarizable endomorphism defined over $\mathbb{C}$, and let Y be a subvariety of X that has no component included into the singular part of X. Then Y is preperiodic under φ if and only if there exists a Zariski dense subset of smooth points $x \in Y \cap \mathrm{Prep}_\varphi(X)$ such that the tangent subspace of Y at x is preperiodic under the induced action of φ on the Grassmannian $\mathrm{Gr}_{\dim(Y)}(T_{X,x})$. (Here $T_{X,x}$ denotes the tangent space of X at the point x.)*

Ghioca, Tucker, and Zhang [2011] prove that Conjecture 11.1 holds whenever Φ is a polarizable algebraic group endomorphism of the abelian variety X and also when $X = \mathbb{P}^1 \times \mathbb{P}^1$, Y is a line, and $\Phi(x, y) = (f(x), g(y))$ for any rational maps f and g. We claim that a positive answer to Question 1.3 yields the following special case of Conjecture 11.1 that is not covered by the results from [Ghioca et al. 2011]. Note that we do not need the condition on preperiodicity of tangent spaces in the Grassmannian, only an infinite family of preperiodic points; hence, what one would obtain here is really a special case of Zhang's original dynamical Manin–Mumford conjecture (which did not require the extra hypothesis on tangent spaces).

Proposition 11.2. *If Question 1.3 holds in the affirmative, then for any endomorphism Φ of $\mathbb{P}^1 \times \mathbb{P}^1$ given by $\Phi(x, y) := (f(x), f(y))$ for some rational map $f \in \mathbb{C}(x)$ of degree at least 2, a curve $Y \subset \mathbb{P}^1 \times \mathbb{P}^1$ will contain infinitely many preperiodic points if and only if Y is preperiodic under Φ. In particular, Question 1.3 implies Conjecture 11.1 for such Y and Φ.*

Proof. Let $Y \subset \mathbb{P}^1 \times \mathbb{P}^1$ be a curve containing infinitely many points (x, y) such that both x and y are preperiodic for f. Furthermore, we may assume Y projects dominantly on each coordinate of $\mathbb{P}^1 \times \mathbb{P}^1$ since otherwise it is immediate to conclude that Y contains infinitely many preperiodic points for Φ if and only if $Y = \{c\} \times \mathbb{P}^1$ or $Y = \mathbb{P}^1 \times \{c\}$, where c is a preperiodic point for f.

We let $\boldsymbol{f} = f_\lambda := f$ be the constant family of rational functions (equal to f) indexed by all points $\lambda \in Y$ and let K be the function field of Y. Let $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{P}^1(K) \times \mathbb{P}^1(K)$ be a generic point for Y. By our assumption, there exist infinitely many $\lambda \in Y$ such that both $\boldsymbol{a}(\lambda)$ and $\boldsymbol{b}(\lambda)$ are preperiodic for $f_\lambda = f$. Since Y projects dominantly on each coordinate of $\mathbb{P}^1 \times \mathbb{P}^1$, we get that neither $\boldsymbol{a}$ nor $\boldsymbol{b}$

is preperiodic under the action of f. (Otherwise, $\boldsymbol{a}$ or $\boldsymbol{b}$ would be constant.) So assuming the answer to Question 1.3 is "yes", we obtain that the curve $Y(\mathbb{C}) = \{(\boldsymbol{a}(\lambda), \boldsymbol{b}(\lambda)) : \lambda \in Y\} \subset \mathbb{P}^1_K \times_K \mathbb{P}^1_K$ lies on a preperiodic proper subvariety Z of $\mathbb{P}^1 \times \mathbb{P}^1$ defined over a finite extension of K. More precisely, we get that $Z = Y \otimes_{\mathbb{C}} K$, so Y must be itself preperiodic under the action of (f, f) on $\mathbb{P}^1 \times \mathbb{P}^1$.

Conversely, suppose that Y is preperiodic under Φ. Then some iterate of Y contains a dense set of periodic points by [Fakhruddin 2003], so Y contains an infinite set of preperiodic points. □

Remarks 11.3. (a) In the proof of Proposition 11.2, we did not use the full strength of the hypothesis from Conjecture 11.1. Instead we used the weaker hypothesis of [Zhang 1995, Conjecture 2.5] or [Zhang 2006, Conjecture 1.2.1, Conjecture 4.1.7] (which was the original formulation of the dynamical Manin–Mumford conjecture). This is not surprising since for curves contained in $\mathbb{P}^1 \times \mathbb{P}^1$, the *only* counterexamples to the original formulation of the dynamical Manin–Mumford conjecture are expected to occur when $\Phi := (f, g)$ for two *distinct* Lattès maps.

(b) Finally, we note that a positive answer to Conjecture 11.1 does not yield a positive answer to Question 1.3. Instead, Question 1.3 goes in a different direction that is likely to shed more light on the dynamical Manin–Mumford conjecture especially in the case when Y is a curve in Conjecture 11.1.

12. Acknowledgments

Hsia acknowledges the support from the 2010–2011 France–Taiwan Orchid Program, which enabled him to attend the Summer School on Berkovich spaces held at Institut de Mathématiques de Jussieu, where this project was initiated in the summer of 2010. We thank Matthew Baker, Laura DeMarco, and Joseph Silverman for several useful conversations and also thank the referee for comments.

References

[Baker 2009] M. Baker, "A finiteness theorem for canonical heights attached to rational maps over function fields", *J. Reine Angew. Math.* **626** (2009), 205–233. MR 2011c:14075 Zbl 1187.37133

[Baker and DeMarco 2011] M. Baker and L. DeMarco, "Preperiodic points and unlikely intersections", *Duke Math. J.* **159**:1 (2011), 1–29. MR 2012h:37170 Zbl 1242.37062

[Baker and Hsia 2005] M. H. Baker and L.-C. Hsia, "Canonical heights, transfinite diameters, and polynomial dynamics", *J. Reine Angew. Math.* **585** (2005), 61–92. MR 2006i:11071 Zbl 1071.11040

[Baker and Rumely 2010] M. Baker and R. Rumely, *Potential theory and dynamics on the Berkovich projective line*, Mathematical Surveys and Monographs **159**, American Mathematical Society, Providence, RI, 2010. MR 2012d:37213 Zbl 1196.14002

[Beardon 1991] A. F. Beardon, *Iteration of rational functions*, Graduate Texts in Mathematics **132**, Springer, New York, 1991. MR 92j:30026 Zbl 0742.30002

[Benedetto 2005] R. L. Benedetto, "Heights and preperiodic points of polynomials over function fields", *Int. Math. Res. Not.* **2005**:62 (2005), 3855–3866. MR 2006m:11095 Zbl 1114.14018

[Berkovich 1990] V. G. Berkovich, *Spectral theory and analytic geometry over non-Archimedean fields*, Mathematical Surveys and Monographs **33**, American Mathematical Society, Providence, RI, 1990. MR 91k:32038 Zbl 0715.14013

[Bertrand 2011] D. Bertrand, "Special points and Poincaré bi-extensions, with an Appendix by Bas Edixhoven", preprint, 2011. arXiv 1104.5178

[Bombieri and Gubler 2006] E. Bombieri and W. Gubler, *Heights in Diophantine geometry*, New Mathematical Monographs **4**, Cambridge University Press, 2006. MR 2007a:11092 Zbl 1115.11034

[Bombieri et al. 1999] E. Bombieri, D. Masser, and U. Zannier, "Intersecting a curve with algebraic subgroups of multiplicative groups", *Internat. Math. Res. Notices* 20 (1999), 1119–1140. MR 2001c:11081 Zbl 0938.11031

[Branner and Hubbard 1988] B. Branner and J. H. Hubbard, "The iteration of cubic polynomials. I. The global topology of parameter space", *Acta Math.* **160**:3-4 (1988), 143–206. MR 90d:30073 Zbl 0668.30008

[Call and Silverman 1993] G. S. Call and J. H. Silverman, "Canonical heights on varieties with morphisms", *Compositio Math.* **89**:2 (1993), 163–205. MR 95d:11077 Zbl 0826.14015

[Carleson and Gamelin 1993] L. Carleson and T. W. Gamelin, *Complex dynamics*, Springer, New York, 1993. MR 94h:30033 Zbl 0782.30022

[Fakhruddin 2003] N. Fakhruddin, "Questions on self maps of algebraic varieties", *J. Ramanujan Math. Soc.* **18**:2 (2003), 109–122. MR 2004f:14038 Zbl 1053.14025

[Ghioca and Tucker 2008] D. Ghioca and T. J. Tucker, "Equidistribution and integral points for Drinfeld modules", *Trans. Amer. Math. Soc.* **360**:9 (2008), 4863–4887. MR 2009c:11095 Zbl 1178.11046

[Ghioca et al. 2011] D. Ghioca, T. J. Tucker, and S. Zhang, "Towards a dynamical Manin–Mumford conjecture", *Int. Math. Res. Not.* **2011**:22 (2011), 5109–5122. MR 2012i:37142 Zbl 05988299

[Habegger 2009] P. Habegger, "Intersecting subvarieties of abelian varieties with algebraic subgroups of complementary dimension", *Invent. Math.* **176**:2 (2009), 405–447. MR 2010f:11109 Zbl 1176.14008

[Hsia 2008] L.-C. Hsia, "On the reduction of a non-torsion point of a Drinfeld module", *J. Number Theory* **128**:6 (2008), 1458–1484. MR 2009f:11065 Zbl 1144.11046

[Lang 1983] S. Lang, *Fundamentals of Diophantine geometry*, Springer, New York, 1983. MR 85j:11005 Zbl 0528.14013

[Masser and Zannier 2010] D. Masser and U. Zannier, "Torsion anomalous points and families of elliptic curves", *Amer. J. Math.* **132**:6 (2010), 1677–1691. MR 2012d:11133 Zbl 1225.11078

[Masser and Zannier 2012] D. Masser and U. Zannier, "Torsion points on families of squares of elliptic curves", *Math. Ann.* **352**:2 (2012), 453–484. MR 2012k:11076 Zbl 06006368

[Pink 2005] R. Pink, "A combination of the conjectures of Mordell–Lang and André–Oort", pp. 251–282 in *Geometric methods in algebra and number theory*, edited by F. Bogomolov and Y. Tschinkel, Progr. Math. **235**, Birkhäuser, Boston, MA, 2005. MR 2006f:11068 Zbl 1200.11041

[Ransford 1995] T. Ransford, *Potential theory in the complex plane*, London Mathematical Society Student Texts **28**, Cambridge University Press, 1995. MR 96e:31001 Zbl 0828.31001

[Raynaud 1983a] M. Raynaud, "Courbes sur une variété abélienne et points de torsion", *Invent. Math.* **71**:1 (1983), 207–233. MR 84c:14021 Zbl 0564.14020

[Raynaud 1983b] M. Raynaud, "Sous-variétés d'une variété abélienne et points de torsion", pp. 327–352 in *Arithmetic and geometry*, vol. I, edited by M. Artin and J. Tate, Progr. Math. **35**, Birkhäuser, Boston, MA, 1983. MR 85k:14022 Zbl 0581.14031

[Shafarevich 1994] I. R. Shafarevich, *Basic algebraic geometry, I*, 2nd ed., Springer, Berlin, 1994. MR 95m:14001 Zbl 0797.14001

[Silverman 1994a] J. H. Silverman, "Variation of the canonical height on elliptic surfaces, II: Local analyticity properties", *J. Number Theory* **48**:3 (1994), 291–329. MR 95h:11054 Zbl 0807.14019

[Silverman 1994b] J. H. Silverman, "Variation of the canonical height on elliptic surfaces, III: Global boundedness properties", *J. Number Theory* **48**:3 (1994), 330–352. MR 95h:11055 Zbl 0807.14020

[Zhang 1995] S. Zhang, "Small points and adelic metrics", *J. Algebraic Geom.* **4**:2 (1995), 281–300. MR 96e:14025 Zbl 0861.14019

[Zhang 2006] S.-W. Zhang, "Distributions in algebraic dynamics", pp. 381–430 in *Surveys in differential geometry*, vol. 10, edited by S. T. Yau, Surv. Differ. Geom. **10**, Int. Press, Somerville, MA, 2006. MR 2009k:32016 Zbl 1207.37057

Communicated by Bjorn Poonen
Received 2011-09-21 Revised 2012-03-25 Accepted 2012-03-29

dghioca@math.ubc.ca *Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada*

hsia@math.ntnu.edu.tw *Department of Mathematics, National Taiwan Normal University, Taipei, Taiwan*

ttucker@math.rochester.edu *Department of Mathematics, University of Rochester, Rochester, NY 14627, United States*

ALGEBRA AND NUMBER THEORY 7:3 (2013)
dx.doi.org/10.2140/ant.2013.7.733

F-blowups of normal surface singularities

Nobuo Hara, Tadakazu Sawada and Takehiko Yasuda

We study F-blowups of non-F-regular normal surface singularities. Especially the cases of rational double points and simple elliptic singularities are treated in detail.

1. Introduction

The F-blowup introduced in [Yasuda 2012] is a canonical birational modification of a variety in positive characteristic. For a nonnegative integer e, the e-th F-blowup of a variety X is defined as the blowup at $F^e_*\mathcal{O}_X$, that is, the universal birational flattening of $F^e_*\mathcal{O}_X$. Here $F^e_*\mathcal{O}_X$ is the pushforward of the structure sheaf by the e-iterated Frobenius morphism. It turns out that the F-blowup of a quotient singularity has a connection with the G-Hilbert scheme [Toda and Yasuda 2009; Yasuda 2012]. However, the F-blowup has the advantage that it is canonically defined for arbitrary singularity in positive characteristic, whereas the G-Hilbert scheme is defined only for a quotient singularity. Actually, it is proved in [Yasuda 2012] that the e-th F-blowup of any curve singularity with $e \gg 0$ is normal, and hence resolves singularities in dimension one.

As is naturally expected, the F-blowup is also connected to F-singularities in positive characteristic such as F-pure and F-regular singularities. It is proved that the sequence of F-blowups for an F-pure singularity is monotone [Yasuda 2009] and that the e-th F-blowup of an F-regular surface singularity coincides with the minimal resolution for $e \gg 0$ [Hara 2012]. However, it is too much to ask for F-blowups of normal surface singularities to be the minimal resolution or even smooth in general. Actually, there exist (non-F-regular) rational double points whose F-blowups are singular [Hara and Sawada 2011].

Although some good aspects as well as pathologies of F-blowups have recently been discovered as above, their behavior is a mystery yet, even in dimension two. In this paper, we explore the behavior of F-blowups of certain normal surface

Hara and Yasuda are supported by Grants-in-Aid for Scientific Research, numbers 22540039 and 22740020, respectively.
MSC2010: primary 14B05; secondary 14G17, 14E15.
Keywords: F-blowups, Frobenius maps, rational double points, simple elliptic singularities, `Macaulay2`.

singularities more in detail. We are mainly concerned with two classes of surface singularities, that is, non-F-regular rational double points (which exist only in characteristics up to five) and simple elliptic singularities. We will discuss F-blowups of these singularities, focusing on the normality, smoothness and stabilization of F-blowup sequences.

For this purpose, we do utilize not only the classical theory of surface singularities, but also computations with `Macaulay2` [Grayson and Stillman 2012], which are complementary to each other. The key to our computations is two `Macaulay2` functions that we will write down. Given a module, the first function computes an ideal such that the blowups at the ideal and module coincide, following Villamayor's description of such an ideal [2006]. Using this together with a built-in function to compute Rees algebras, one can explicitly compute a graded ring describing the blowup at a module. The second function we will write computes the Frobenius pushforward F_*M of a given module M. These functions enable us to investigate F-blowups in detail, especially for hypersurface surface singularities in characteristic two or three.

In the case of rational double points, one can apply general theory of rational surface singularities to show that F-blowups are normal and dominated by the minimal resolution. Then a version of McKay's correspondence [Artin and Verdier 1985] enables us to determine the e-th F-blowup by the direct sum decomposition of $F_*^e M$ into indecomposable modules. For F-regular surface singularities $R = \mathcal{O}_{X,x}$, all indecomposable reflexive R-modules appear as a direct summand of $F_*^e R$ with $e \gg 0$, so that the e-th F-blowup coincides with the minimal resolution [Hara and Sawada 2011; Hara 2012]. Contrary to this we have the following result for non-F-regular Frobenius sandwiches in characteristic $p \le 5$.

Theorem 1.1 (see [Hara and Sawada 2011, Example 4.8]). *Let (X, x) be a rational double point of type D_{2n}^0 for $n \ge 2$, E_7^0, E_8^0 in $p = 2$, E_6^0, E_8^0 in $p = 3$ or E_8^0 in $p = 5$; see [Artin 1977] for the notation. Then for any $e \ge 1$, the e-th F-blowup $\mathrm{FB}_e(X)$ coincides with the normal surface obtained by contracting all but one exceptional curve on the minimal resolution $\widetilde{X}$. The unique exceptional curve on $\mathrm{FB}_e(X)$ is indicated by the solid circle in Theorem 3.5.*

We can analyze other types of non-F-regular rational double points by computer-aided calculation. In these cases, computations of the blowups at modules are again useful. For instance, one can see with such computation whether two obtained indecomposable modules are isomorphic. A particularly interesting result is that for $e \ge 2$, the e-th F-blowup of D_4^1- and D_5^1-singularities in characteristic two is the minimal resolution, though D_4^1- and D_5^1-singularities are not F-regular. In our computations so far, there is no other non-F-regular rational double point such that any of its F-blowups is the minimal resolution.

We will investigate F-blowups of simple elliptic singularities in detail as well. Since a simple elliptic singularity (X, x) is quasihomogeneous in general, its minimal resolution $\widetilde{X}$ has the same structure as the conormal bundle over the elliptic exceptional curve E, which is identified with the negative section. We can use this fact to determine the structure of the F-blowups up to normalization, which turns out to be different according to the self-intersection number E^2 and whether the singularity (X, x) is F-pure or not. We summarize the results obtained in Theorems 4.5, 4.7, 4.13 and Proposition 4.18 in the following.

Theorem 1.2. *Let (X, x) be a simple elliptic singularity in characteristic $p > 0$ with the elliptic exceptional curve E on the minimal resolution $\widetilde{X}$. Let $\widetilde{\mathrm{FB}}_e(X)$ be the normalization of the e-th F-blowup $\mathrm{FB}_e(X)$ of (X, x) for any $e \geq 1$.*

(1) *If (X, x) is F-pure with $E^2 = -1$, then $\widetilde{\mathrm{FB}}_e(X)$ coincides with the blowup of $\widetilde{X}$ at $p^e - 1$ nontrivial p^e-torsion points on E.*

(2) *If (X, x) is not F-pure with $E^2 = -1$, then $\widetilde{\mathrm{FB}}_e(X)$ coincides with the blowup of $\widetilde{X}$ at an ideal supported at a point $P_0 \in E$ with local expression (t, u^{p^e-1}), where t and u are local coordinates at $P_0 \in \widetilde{X}$.*

(3) *If $E^2 \leq -2$ and $-E^2$ is not a power of p, then $\widetilde{\mathrm{FB}}_e(X) \cong \widetilde{X}$ for all $e \geq 1$. Moreover, if (X, x) is F-pure and $E^2 \leq -3$, then $\mathrm{FB}_e(X) \cong \widetilde{X}$.*

We cannot determine whether or not an F-blowup is normal in general, but we see that an F-blowup is nonnormal in some cases with `Macaulay2` computation. The theorem above tells us that an F-blowup coincides with the minimal resolution in some cases, but in general, F-blowups of simple elliptic singularities behave badly: They are nonnormal, not dominated by the minimal resolution and the sequence of F-blowups does not stabilize. The study of F-blowups for simple elliptic singularities will be pushed further and completed in [Hara 2013].

2. Preliminaries

2a. *Blowups at modules.* Let X be a Noetherian integral scheme and $\mathcal{M}$ a coherent sheaf on X. For a modification $f : Y \to X$, we denote the torsion-free pullback $(f^*\mathcal{M})/\mathrm{tors}$ by $f^\star\mathcal{M}$, where tors denotes the subsheaf of torsions.

Definition 2.1. A modification $f : Y \to X$ is called a *flattening* of $\mathcal{M}$ if $f^\star\mathcal{M}$ is flat, or equivalently locally free. A flattening f is said to be *universal* if every flattening $g : Z \to X$ of $\mathcal{M}$ factors as

$$g : Z \to Y \xrightarrow{f} X.$$

(The universal flattening exists and is unique. It can be constructed as a subscheme of a Quot scheme. See for instance [Oneto and Zatini 1991; Villamayor U. 2006].) The universal flattening is also called the *blowup of X at $\mathcal{M}$* and denoted by $\mathrm{Bl}_{\mathcal{M}}(X)$.

The following are basic properties of the blowup at a module, which directly follow from the definition:

(1) The modification $\mathrm{Bl}_{\mathcal{M}}(X) \to X$ is an isomorphism exactly over the locus where $\mathcal{M}$ is flat.

(2) If $\mathcal{N} \subset \mathcal{M}$ is a torsion subsheaf, then $\mathrm{Bl}_{\mathcal{M}}(X) = \mathrm{Bl}_{\mathcal{M}/\mathcal{N}}(X)$.

(3) If $\mathcal{M}$ is an ideal sheaf, then the blowup at $\mathcal{M}$ defined above coincides with the usual blowup with the center $\mathcal{M}$.

The following are examples of blowups at modules. Therefore one can compute them in the method explained below.

Example 2.2. If X is an algebraic variety over a field k, then its Nash blowup is the blowup at $\Omega_{X/k}$, the sheaf of differentials. The higher version of the Nash blowup is also an example of the blowup at a module; see [Yasuda 2007].

Example 2.3. Let Y be a quasiprojective algebraic variety, G a finite group of automorphisms of Y and $X := Y/G$ the quotient variety. Then the G-Hilbert scheme $\mathrm{Hilb}^G(Y)$ is defined to be the closure of the set of free G-orbits in the Hilbert scheme of Y; see [Ito and Nakamura 1996]. One can show that $\mathrm{Hilb}^G(Y)$ is isomorphic to the blowup at $\pi_*\mathcal{O}_Y$, where $\pi : Y \to X$ is the quotient map.

Let r be the rank of $\mathcal{M}$, K the function field of X and fix an isomorphism $\bigwedge^r \mathcal{M} \otimes K \cong K$. Then define a fractional ideal sheaf

$$\mathcal{I}_{\mathcal{M}} := \mathrm{Im}\big(\textstyle\bigwedge^r \mathcal{M} \to \bigwedge^r \mathcal{M} \otimes K \cong K\big).$$

Proposition 2.4 (see [Oneto and Zatini 1991; Villamayor U. 2006]). *The blowup at $\mathcal{M}$ is isomorphic to the blowup at $\mathcal{I}_{\mathcal{M}}$,*

$$\mathrm{Bl}_{\mathcal{I}_{\mathcal{M}}}(X) = \mathrm{Proj}_X\Big(\bigoplus_{n\geq 0} \mathcal{I}_{\mathcal{M}}^n\Big).$$

Note that although $\mathcal{I}_{\mathcal{M}}$ depends on the choice of the isomorphism $\bigwedge^r \mathcal{M} \otimes K \cong K$, the isomorphism class of $\mathcal{I}_{\mathcal{M}}$ and so $\mathrm{Bl}_{\mathcal{I}}(X)$ are independent of it.

We will now recall Villamayor's method [2006] for computing $\mathcal{I}_{\mathcal{M}}$ in the affine case. Suppose that $X = \mathrm{Spec}\, R$. Abusing the notation, we identify the sheaf $\mathcal{M}$ with the corresponding R-module M, the fractional ideal sheaf $\mathcal{I}_{\mathcal{M}}$ with the fractional ideal $I_M \subset K$, and so forth. Let

$$R^m \xrightarrow{A} R^n \to M \to 0$$

be a presentation of M given by an $n \times m$ matrix A. Here and hereafter we think of elements of free modules as column vectors and the map $A : R^m \to R^n$ is given by left multiplication with A, that is, $v \mapsto Av$. We call A a *presentation matrix* of M. Then there exist $n-r$ columns of A such that if A' denotes the submatrix of A formed by these columns, then

$$M' := \mathrm{Coker}(R^{n-r} \xrightarrow{A'} R^n)$$

has rank r. Then M is a quotient of M' by some torsion submodule of M'. Therefore the blowups at M and M' are equal.

Proposition 2.5 [Villamayor U. 2006]. *The ideal generated by $(n-r)$-minors of A', which is by definition the r-th Fitting ideal of M', is equal to I_M for a suitable choice of isomorphism $\bigwedge^r M \otimes K \cong K$.*

The computation of this ideal is implemented in `Macaulay2` as

```
villamayorIdeal = M -> (
  r := rank M;
  P := presentation M;
  s := rank source P;
  t := rank target P;
  I := {};
  for j to s-1 when #I < t-r do (
    J := append(I,j);
    if rank coker P_J == t - #J then I = J;
  );
  fittingIdeal(r,coker P_I);
);
```

Once the ideal I_M was computed, then the blowup at M is computed as the projective spectrum of the Rees algebra of the ideal:

$$\mathrm{Bl}_M(X) = \mathrm{Proj}\, R[I_M t], \qquad R[I_M t] := \bigoplus_{i\geq 0} I_M^i t^i \subset R[t].$$

The computation of Rees algebras has been already implemented in `Macaulay2` as `reesAlgebra`.

The computation of blowups at modules is useful for studying modules themselves. For instance, one can see that two given modules are not isomorphic if the associated blowups are not isomorphic.

2b. *F-blowups.* Suppose now that X is a Noetherian integral scheme of characteristic $p > 0$ and that its (absolute) Frobenius morphism $F : X \to X$ is finite.

Definition 2.6 [Yasuda 2012]. For a nonnegative integer e, we define the *e-th F-blowup* of X to be the blowup of X at $F_*^e \mathcal{O}_X$ and denote it by $\mathrm{FB}_e(X)$.

From [Kunz 1969], if $e > 0$, then the flat locus of $F^e_* \mathcal{O}_X$ coincides with the regular locus of X. Therefore the e-th F-blowup is an isomorphism exactly over the regular locus.

If X is an algebraic variety over an algebraically closed field k, then there is a more moduli-theoretic construction of F-blowups, which was actually the original definition of F-blowups in [Yasuda 2012]: The e-th F-blowup is isomorphic (over Z) to the closure of the set

$$\{[(F^e)^{-1}(x)] \mid \text{nonsingular point } x \in X(k)\}$$

in the Hilbert scheme of zero-dimensional subschemes. Here $(F^e)^{-1}(x)$ is the scheme-theoretic inverse image and a closed subscheme of X with length $p^{e \dim X}$, and $[(F^e)^{-1}(x)]$ is the corresponding point in the Hilbert scheme.

2c. ***Computing the Frobenius pushforward.*** Let us now suppose that X is affine, say $X = \operatorname{Spec} R$. In order to compute F-blowups of X along the lines explained above, we need to first compute a presentation of $F^e_* R$. For later use, we will explain more generally how to compute $F^e_* M$ for any finitely generated R-module M in the case where R is finitely generated over the prime field $\mathbb{F}_p$.

2c1. *The case of a polynomial ring.* Set $S = \mathbb{F}_p[x_1, \dots, x_n]$ and $q = p^e$. A monomial $x^a = x_1^{a_1} \cdots x_n^{a_n}$ defines an S-linear map

$$\mu_{x^a} : S \to S, \quad f \mapsto x^a f.$$

Then we reinterpret this map according to another S-module structure on S by $g \cdot f := g^q f$. We denote this new S-module by S', which is a free S-module of rank q^n and nothing but $F^e_* S$. We also denote the map μ_{x^a} regarded as an endomorphism of S' by μ'_{x^a}, which is nothing but $F^e_* \mu_{x^a}$.

Let $\Lambda := \{0, 1, \dots, q-1\}^n$. Then q^n monomials x^b for $b \in \Lambda$ form a standard basis of S'. For such a monomial x^b, we have

$$\mu_{x^a}(x^b) = x^{a+b} = x^{q((a+b) \div q)} x^{(a+b) \% q}.$$

Here $\div q$ and $\% q$ respectively denote the quotient and the remainder by the component-wise division by q. We rewrite it as

$$\mu'_{x^a}(x^b) = x^{(a+b) \div q} \cdot x^{(a+b) \% q}.$$

Thus we obtain:

Lemma 2.7. *The defining matrix, $U(a, e) = (u_{ij})_{i, j \in \Lambda}$, of μ'_{x^a} with respect to the standard basis is given by*

$$u_{ij} = \begin{cases} x^{(a+j) \div q} & i = (a+j) \% q, \\ 0 & \textit{otherwise}. \end{cases}$$

Then for a polynomial $f = \sum_a c_a x^a \in S$, if $\mu_f : S \to S$ denotes the multiplication with f, then $\mu'_f = F^e_* \mu_f$ is defined by the matrix

$$U(f, e) := \sum_a c_a \cdot U(a, e).$$

Note that since the coefficient field is $\mathbb{F}_p$ and the Frobenius map of $\mathbb{F}_p$ is the identity map, we do not have to change the coefficients c_a.

Let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{l1} & \cdots & a_{lm} \end{pmatrix}$$

be an $l \times m$ matrix with entries in S, which defines an S-linear map $S^m \to S^l$ denoted again by A. Then the $F^e_* A : (S')^{\oplus m} \to (S')^{\oplus l}$ is given by $q^n l \times q^n m$ matrix

$$U(A, e) = \begin{pmatrix} U(a_{11}, e) & \cdots & U(a_{1m}, e) \\ \vdots & \ddots & \vdots \\ U(a_{l1}, e) & \cdots & U(a_{lm}, e) \end{pmatrix}.$$

Therefore:

Proposition 2.8. *If A is a presentation matrix of an S-module M, then $U(A, e)$ is a presentation matrix of $F^e_* M$.*

2c2. *The general case.* Suppose that R is the quotient ring S/I, $I = (f_1, \dots, f_l)$, and M is a finitely generated R-module. Then we first have to compute a presentation of M as an S-module. Let A be a matrix with entries in S and let $\bar{A}$ be the matrix with entries in R induced from A. Suppose that $\bar{A}$ is a presentation matrix of M:

$$R^m \xrightarrow{\bar{A}} R^n \to M \to 0.$$

Let $\tilde{M}$ be the S-module with the presentation matrix A:

$$S^m \xrightarrow{A} S^n \to \tilde{M} \to 0.$$

Then $M = R \otimes_S \tilde{M}$. The S-module R has a standard presentation

$$S^l \xrightarrow{(f_1, \dots, f_l)} S \to R \to 0.$$

Now a presentation of M as an S-module can be computed from those of R and $\tilde{M}$.

If B is a presentation matrix of M as an S-module, then $U(B, e)$ is one of $F^e_* M$ as an S-module. If $\overline{U(B, e)}$ denotes the matrix with entries in R induced from $U(B, e)$, then $\overline{U(B, e)}$ is a presentation matrix of $F^e_* M$ as an R-module.

2c3. *Implementation in* `Macaulay2`. The following `Macaulay2` function returns the pushforward $F^e_* M$ of the given module M, following the recipe explained above:

```
frobeniusPushForward = (M, e) -> (
  R := ring M;
  p := char R;
  assert(p > 0); q := p^e;
  I := ideal R;
  l := numgens I;
  B := gens ideal R;
  S := ambient R;
  n := numgens S;
  qSequence := i ->
    apply(0..n-1, j -> (i % q^(n-j)) // q^(n-j-1));
  toNumber := i -> sum(n, j ->  i_j * q^(n-j-1) );
  qQuotient := i -> apply(i, j -> j // q);
  qRemainder := i -> apply(i, j -> j % q);
  monoToMatrix := m ->
    (coefficients m)_1_(0,0)
      * map(S^(q^n),S^(q^n),
          (i,j) -> (e = (toList qSequence i) + (exponents m)_0;
     if(toNumber qRemainder e) == j
       then S_(toList qQuotient e)
       else 0));
  polyToMatrix := f ->
    if f == 0_S
      then map(S^(q^n),S^(q^n),0_S)
      else sum(terms f, i -> monoToMatrix i);
  basisToMatrix := b ->
    fold((i, j)->(i | j),
      apply((flatten entries b), polyToMatrix));
  matrixToMatrix := m ->
    fold((i, j)->(i || j),
      apply(apply(entries m, i -> matrix{i}), basisToMatrix));
  ROverS := coker map(S^1,S^l, entries B);
  PresenOverR := presentation minimalPresentation M;
  PresenOverS := presentation minimalPresentation(
coker(sub(PresenOverR,S))**ROverS);
  L := matrixToMatrix PresenOverS;
  minimalPresentation coker sub(L,R)
);
```

Note that in the computations with `Macaulay2`, columns and rows of matrices should be indexed by single indices rather than multiindices. For this aim, the inner functions `qSequence` and `toNumber` above define bijections between the sets $\{0, 1, \dots, q^n - 1\}$ and Λ that are inverses to each other.

Note that one can compute $F^e_* M$ also with the built-in function `PushForward` in the case where the ring and the module are (weighted) homogeneous.

2d. ***Computing the singular and nonnormal loci of a blowup.*** We often would like to know if a given blowup is smooth or normal, or to know where the singular locus or the nonnormal locus is. One way to compute the singular locus of $\mathrm{Bl}_I(X)$ is to compute the singular locus of $\operatorname{Spec} R[It]$. For instance, suppose that we have an expression of $R[It]$ as a quotient of a polynomial ring over R,

$$R[It] = R[t_1, \dots, t_n]/J.$$

Then $\mathrm{Bl}_I(X)$ is smooth if and only if the singular locus of $\operatorname{Spec} R[It]$ is contained in the closed subset $V(t_1, \dots, t_n) \subset \operatorname{Spec} R[It]$. This method is useful when the Rees algebra is relatively simple. Otherwise, the computation may not finish in a reasonable time.

In that case, an alternative way is to compute the singular loci of affine charts. With the notation above, the blowup $\mathrm{Bl}_I(X)$ is covered by n affine charts corresponding to the variables $t_1, \dots, t_n$. Their coordinate rings are

$$R[t_1, \dots, t_n]/(J + (t_i - 1)) \quad \text{for } i = 1, \dots, n.$$

These rings are likely to become simpler than $R[It]$ and easier to compute the singular loci. Computation of these rings is implemented as follows:

```
affineCharts = S -> (
  T := (flattenRing S)_0;
  varsOfS := apply(flatten entries vars S, i->sub(i, T));
  apply(varsOfS, i -> minimalPresentation(T / ideal(i - 1)))
);
```

The same method can apply to find the nonnormal locus.

2e. ***Embedding F-blowups into the Grassmannian and the projective space.*** As already mentioned above, F-blowups are constructed as a subscheme of the Grassmannian. Then further composing with the Plücker embedding, we obtain an embedding into a projective space over X.

To describe this embedding, let $X = \operatorname{Spec} R$ be of dimension n, let K be the function field of X, and let the fractional ideal $I = \operatorname{Im}(\bigwedge^{p^n} R^{1/p^e} \to K)$ be generated by $m+1$ elements $s_0, \dots, s_m$. Then, being the blowup of X at I, the e-th F-blowup $\mathrm{FB}_e(X)$ of X is embedded into the projective space $\mathbb{P}^m_X$ over X.

Suppose now that $f\colon Y \to X$ is any flattening of $R^{1/p^e} \cong F^e_*\mathcal{O}_X$. Then we have a surjection $\mathcal{O}_Y^{\oplus m+1} \to f^\star I = \det f^\star R^{1/p^e}$ induced by $s_0, \dots, s_m$, which gives rise to a morphism $\Phi_e\colon Y \to \mathbb{P}^m_X$ such that $\Phi_e^*\mathcal{O}_{\mathbb{P}}(1) \cong \det f^\star R^{1/p^e}$, and the image $\Phi_e(Y)$ of this morphism is nothing but $\mathrm{FB}_e(X) = \mathrm{Bl}_I(X)$.

In dimension two where the existence of resolution of singularities is established in arbitrary characteristic, we can study F-blowups downwards from a resolution that flattens the $\mathcal{O}_X$-module $F^e_*\mathcal{O}_X \cong \mathcal{O}_X^{1/p^e}$. The following is an immediate consequence of the observation above.

Proposition 2.9. *Let X be a surface over k and let $f\colon \widetilde{X} \to X$ be a resolution with irreducible exceptional curves $E_1, \dots, E_s$. Suppose that $f^\star\mathcal{O}_X^{1/p^e}$ is flat, so that we have a birational morphism $\Phi_e\colon \widetilde{X} \to \mathrm{FB}_e(X)$. Then $\Phi_e(E_i)$ is a curve on $\mathrm{FB}_e(X)$ if $c_1(f^\star\mathcal{O}_X^{1/p^e})E_i > 0$, and E_i contracts to a point on $\mathrm{FB}_e(X)$ if $c_1(f^\star\mathcal{O}_X^{1/p^e})E_i = 0$.*

3. F-blowups of rational surface singularities

Throughout this section we work under the following notation:

k an algebraically closed field of characteristic $p > 0$,

(X, x) a rational surface singularity defined over k with local ring $R = \mathcal{O}_{X,x}$,

$f\colon \widetilde{X} \to X$ the minimal resolution of (X, x) with $\mathrm{Exc}(f) = \bigcup_{i=1}^s E_i$.

The situation is quite simple in this case because of the following fact [Artin and Verdier 1985]: If M is a reflexive $\mathcal{O}_X$-module,[1] then its torsion-free pullback $\widetilde{M} = f^\star M = f^*M/\text{torsion}$ is an f-generated locally free $\mathcal{O}_{\widetilde{X}}$-module such that $f_*\widetilde{M} = M$ and $R^1 f_*\widetilde{M} = 0$. Note that this vanishing of the higher direct image is an easy consequence of the rationality of the singularity (X, x) and the f-generation of $\widetilde{M}$, which gives rise to a surjection $\mathcal{O}_{\widetilde{X}}^{\oplus n} \twoheadrightarrow \widetilde{M}$.

Lemma 3.1 [Hara 2012, Lemma 1.8]. *If M is a reflexive $\mathcal{O}_X$-module of rank r, then the natural map $\bigwedge^r M \to f_*(\det \widetilde{M})$ is surjective.*

Proposition 3.2. *The e-th F-blowup $\mathrm{FB}_e(X)$ of a rational surface singularity (X, x) is dominated by the minimal resolution $\widetilde{X}$ and has only rational singularities for all $e \geq 0$.*

Proof. Because $M := R^{1/p^e}$ is a reflexive R-module, its torsion-free pullback $\widetilde{M} = f^\star R^{1/p^e}$ to $\widetilde{X}$ is flat, so that the minimal resolution $f\colon \widetilde{X} \to X$ factors through the universal flattening $\mathrm{FB}_e(X)$ of R^{1/p^e}. On the other hand, the ideal $I = I_M$ for $M = R^{1/p^e}$ is $I = H^0(\widetilde{X}, \det \widetilde{M})$ by Lemma 3.1, so that we can take I to be an integrally closed ideal in R, or *complete* ideal in the sense of Lipman [1969]. Then the Rees algebra $R[It]$ is normal by [ibid., Proposition 8.1], so

[1] We always assume that M is a finitely generated $\mathcal{O}_X$-module.

$\mathrm{FB}_e(X) = \operatorname{Proj} R[It]$ is normal. It then follows from [Artin 1962] that $\mathrm{FB}_e(X)$ has only rational singularities. $\square$

Corollary 3.3. *Let (X, x) be a rational surface singularity over k.*

(1) *For any $e \geq 0$, the e-th F-blowup $\mathrm{FB}_e(X)$ is obtained by contracting some of the exceptional curves $E_1, \dots, E_s$ on the minimal resolution $\widetilde{X}$ to normal points with at most rational singularities.*

(2) *The minimal resolution $\widetilde{X}$ of (X, x) is obtained by finitely many iteration of F-blowups. More explicitly, for any sequence of positive integers $e_1, \dots, e_s$, we have $\widetilde{X} = \mathrm{FB}_{e_s}(\mathrm{FB}_{e_{s-1}}(\cdots \mathrm{FB}_{e_2}(\mathrm{FB}_{e_1}(X))\cdots))$.*

The behavior of F-blowups is especially nice for F-regular surface singularities. Namely, the e-th F-blowup of any F-regular surface singularity is the minimal resolution for $e \gg 0$ [Hara 2012]. We next consider F-blowups of non-F-regular rational double points more in detail. In this case we can use the classification of rational double points in characteristic $p > 0$ [Artin 1977], as well as the following:

Lemma 3.4 [Artin and Verdier 1985]. *Let (X, x) be a rational double point and let $Z_0 = \sum_{i=1}^{s} r_i E_i$ be the fundamental cycle on the minimal resolution $\widetilde{X}$. Then there is a one-to-one correspondence between the exceptional curves E_i of f and the isomorphism classes of nontrivial indecomposable reflexive $\mathcal{O}_X$-modules M_i, satisfying the following properties.*

(1) $\operatorname{rank} M_i = r_i$ *for* $1 \leq i \leq s$.

(2) $c_1(\widetilde{M_i})E_j = \delta_{ij}$ *for* $1 \leq i, j \leq s$.

In what follows, we use the notation of [Artin 1977] for rational double points in positive characteristic.

Among non-F-regular rational double points, Frobenius sandwiches have particularly easy to analyze F-blowups. Let X be a Frobenius sandwich of a smooth surface S, that is, the Frobenius morphism of S factors as $F\colon S \xrightarrow{\pi} X \to S$. Then F-blowups of X are also the universal flattening of the reflexive $\mathcal{O}_X$-module $\pi_*\mathcal{O}_S$ [Hara and Sawada 2011, Proposition 4.3]. Thanks to this observation, we can study F-blowups of the Frobenius sandwich X via $\pi_*\mathcal{O}_S$ instead of $F^e_*\mathcal{O}_X$. For example, we find whether the irreducible exceptional curve E_i appears on $\mathrm{FB}_e(X)$ or not by evaluating the intersection number $c_1(f^\star(\pi_*\mathcal{O}_S))E_i$ in Proposition 2.9.

3a. D^0_{2n}-singularities. Here we consider a D^0_{2n}-singularity for $n \geq 2$ in $p = 2$ as a Frobenius sandwich. Let $\mathbb{A}^2 = \operatorname{Spec} k[x, y]$ and $\pi : \mathbb{A}^2 \to X = \mathbb{A}^2/\delta$ the quotient map by a vector field $\delta = (x^2 + nxy^{n-1})\partial/\partial x + y^n \partial/\partial y \in \mathrm{Der}_k\, \mathcal{O}_{\mathbb{A}^2}$. Here

$$\mathcal{O}_X = k[x, y]^\delta = k[x^2, y^2, x^2y + xy^n] \cong k[X, Y, Z]/(Z^2 + X^2Y + XY^n)$$

and X has a D^0_{2n}-singularity. Then $R = \mathcal{O}_X = k[X, Y, Z]/(Z^2 + X^2Y + XY^n)$ is a graded ring with $\deg X = 2(n-1)$, $\deg Y = 2$ and $\deg Z = 2n-1$. The $4(n-1)$-st Veronese ring of R is

$$R^{(4(n-1))} = k[X^2, Y^{2(n-1)}, XY^{n-1}] \cong k[u, v, w]/(w^2 - uv).$$

Set $x_0 = u^{1/2} = X = x^2$ and $x_1 = v^{1/2} = Y^{n-1} = y^{2(n-1)}$. Then

$$R^{(4(n-1))} \cong k[x_0^2, x_1^2, x_0x_1] = k[x_0, x_1]^{(2)},$$

so that $\operatorname{Proj} R \cong \mathbb{P}^1$ with homogeneous coordinates $(x_0 : x_1) = (x^2 : y^{2(n-1)})$. Let $s = x_1/x_0 = y^{2(n-1)}/x^2$ be the affine coordinate of $U_0 = D_+(x_0) \subset \operatorname{Proj} R \cong \mathbb{P}^1$ and pick a homogeneous element $t = Z/X = y(x + y^{n-1})/x \in R$ of degree 1. Since

$$t^{2(n-1)} = \frac{x_1(x_1 - x_0)^{n-1}}{x_0^{n-1}},$$

the $\mathbb{Q}$-divisor

$$D = \frac{1}{2(n-1)}(0) + \tfrac{1}{2}(1) - \tfrac{1}{2}(\infty)$$

on $\mathbb{P}^1$ gives $R = \bigoplus_{n\geq 0} H^0(\mathbb{P}^1, nD)t^n$ (the Pinkham–Demazure construction).

Let $g\colon X' \to X = \operatorname{Spec} R$ be the weighted blowup with respect to the weight $(2(n-1), 2, 2n-1)$. Then $X' \cong \operatorname{Spec}_{\mathbb{P}^1}(\bigoplus_{n\geq 0} \mathcal{O}_{\mathbb{P}^1}(nD)t^n)$ admits an affine morphism $\rho\colon X' \to \mathbb{P}^1$ that is an $\mathbb{A}^1$-bundle over $\mathbb{P}^1\backslash\{0, 1, \infty\}$, and the exceptional curve of g is the negative section $E \cong \mathbb{P}^1$ of ρ. Let $X'_0 = \rho^{-1}U_0$. Then

$$\mathcal{O}_{X'_0} = k\left[s, t, \frac{t^2}{s-1}, \frac{t^3}{s-1}, \dots, \frac{t^{2(n-1)-2}}{(s-1)^{n-2}}, \frac{t^{2(n-1)-1}}{(s-1)^{n-2}}, \frac{t^{2(n-1)}}{s(s-1)^{n-1}}\right]$$

and X' has an A_{2n-3}-singularity on $E|_{X'_0} \cong \operatorname{Spec} k[s]$ at $s = 0$.

To resolve the A_{2n-3}-singularity, we may replace

$$X'_0 = \rho^{-1}U_0 \quad \text{by} \quad V = \rho^{-1}(U_0\backslash\{1\}).$$

The affine coordinate ring of V is

$$\mathcal{O}_V = \mathcal{O}_{X'_0}\left[\frac{1}{s-1}\right] = k[s, t, t^{2(n-1)}/s]_{s-1}.$$

The minimal resolution $h\colon \widetilde{V} \to V$ of V is given by

$$\widetilde{V} = \bigcup_{i=1}^{2(n-1)} \widetilde{V}_i, \quad \text{where } \widetilde{V}_i = \operatorname{Spec} k[s/t^{i-1}, t^i/s]_{s-1}.$$

Let $\widetilde{E} \cong \mathbb{P}^1$ be the h-exceptional curve lying on $\widetilde{V}_{n-2} \cup \widetilde{V}_{n-1}$:

$$\widetilde{E} = \operatorname{Spec} k[t^{n-2}/s] \cup \operatorname{Spec} k[s/t^{n-2}] \subset \widetilde{V}_{n-2} \cup \widetilde{V}_{n-1}.$$

Now suppose that n is even; $n = 2k$ with $k \geq 1$. Let $\varphi = t^{n-2}/s(s-1)^{k-1}$ and $\psi = t^{n-1}/s(s-1)^{k-1}$. Then $\varphi, \psi \in \mathcal{O}_{\widetilde{V}_{n-2}}$ and $x = \psi + y\varphi$. Thus

$$(h \circ g)^{\star}(\pi_* \mathcal{O}_{\mathbb{A}^2})|_{\widetilde{V}_{n-2}} = \mathrm{Im}(\mathcal{O}_{\widetilde{V}_{n-2}} \otimes_{\mathcal{O}_X} \mathcal{O}_{\mathbb{A}^2} \to k(\mathbb{A}^2)) = k[s/t^{n-3}, t^{n-2}/s, x, y]_{s-1}$$

is a free $\mathcal{O}_{\widetilde{V}_{n-2}}$-module with basis $1, y$. Similarly it follows that $(h \circ g)^{\star}(\pi_* \mathcal{O}_{\mathbb{A}^2})|_{\widetilde{V}_{n-1}}$ is a free $\mathcal{O}_{\widetilde{V}_{n-1}}$-module with basis $1, x$. The transition matrix of the two bases on $\widetilde{V}_{n-2} \cap \widetilde{V}_{n-1}$ is given by

$$(1\ x) = (1\ y) \begin{pmatrix} 1 & t^{n-1}/s(s-1)^{k-1} \\ 0 & t^{n-2}/s(s-1)^{k-1} \end{pmatrix}.$$

Since $s - 1$ is a unit on V, the intersection number of $L = c_1((h \circ g)^{\star}(\pi_* \mathcal{O}_{\mathbb{A}^2}))$ with $\widetilde{E}$ is $L\widetilde{E} = 1$. In light of Lemma 3.4, this means that the reflexive $\mathcal{O}_X$-module $\pi_* \mathcal{O}_{\mathbb{A}^2}$ of rank 2 is the indecomposable one corresponding to $\widetilde{E}$, which is identified with the exceptional curve E_{n+1} on the minimal resolution $\widetilde{X}$ indicated in the figure below:

$$\begin{array}{c} \quad\; E_1 \\ \quad\; | \\ E_2 - E_3 - E_4 - \cdots - E_{n+1} - \cdots - E_{2n} \end{array}$$

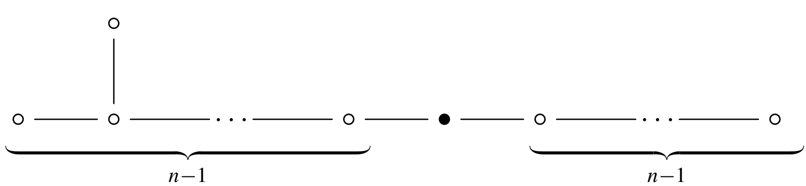
In the case where $n = 2k + 1$ with $k \geq 1$, we obtain the same conclusion that $c_1((h \circ g)^{\star}(\pi_* \mathcal{O}_{\mathbb{A}^2})) \cdot E_i = \delta_{i,n+1}$.

Thus we conclude that for all $e \geq 1$, the e-th F-blowup $\mathrm{FB}_e(X)$ coincides with the normal surface obtained by contracting all exceptional curves on $\widetilde{X}$ except E_{n+1}.

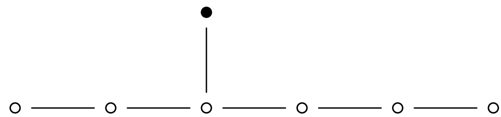
Putting the result above together with [Hara and Sawada 2011, Example 4.8], we obtain the following.

Theorem 3.5. *Let (X, x) be a rational double point of type D_{2n}^0 for $n \geq 2$, E_7^0, E_8^0 in $p = 2$, E_6^0, E_8^0 in $p = 3$ or E_8^0 in $p = 5$. Then for any $e \geq 1$, the e-th F-blowup* $\mathrm{FB}_e(X)$ *coincides with the normal surface obtained by contracting the exceptional curves on the minimal resolution $\widetilde{X}$ corresponding to the blank circles in the figure below*:

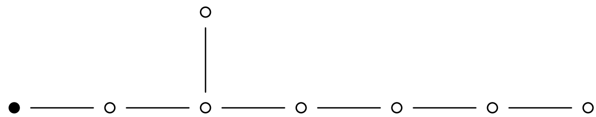
(1) *D_{2n}^0-singularity for $n \geq 2$ in $p = 2$:*



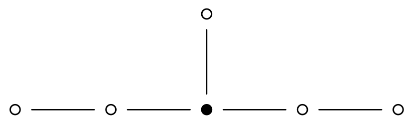
(2) E_7^0*-singularity in* $p=2$:



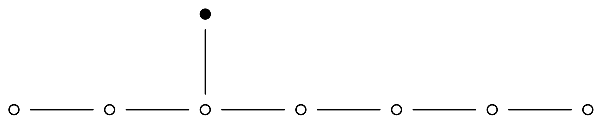
(3) E_8^0*-singularity in* $p=2$:



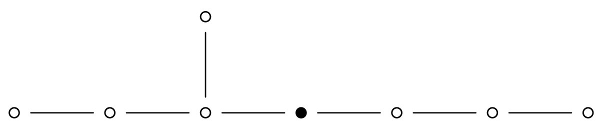
(4) E_6^0*-singularity in* $p=3$:



(5) E_8^0*-singularity in* $p=3$:



(6) E_8^0*-singularity in* $p=5$:



We want to emphasize that all rational double points listed in Theorem 3.5 are non-F-regular Frobenius sandwiches[2] and their F-blowups $\mathrm{FB}_e(X)$ with $e \geq 1$ have only a single exceptional curve corresponding to the solid circle. In particular, their F-blowups do not coincide with the minimal resolution.

We are also able to apply `Macaulay2` to study F-blowups of a few non-F-regular rational double points that are not supposed to be Frobenius sandwiches.

3b. ***D_4^1- and D_5^1-singularities in $p=2$.*** First we consider the case of a D_4^1-singularity in $p=2$: Let $X = \operatorname{Spec} R$ with $R = k[x, y, z]/(z^2 + x^2 y + x y^2 + x y z)$. Using the `Macaulay2` function `frobeniusPushForward` in Section 2c3, we see that the presentation matrix of $F_* R$ is equivalent to

$$\begin{pmatrix} z & x+y+z \\ xy & z \end{pmatrix} \oplus \begin{pmatrix} z & y \\ x(x+y+z) & z \end{pmatrix} \oplus \begin{pmatrix} z & y(x+y+z) \\ x & z \end{pmatrix} \oplus 0,$$

where 0 is the zero matrix of size 1. Then the cokernel of each matrix of size 2 defines a nontrivial reflexive R-module of rank 1 and those reflexive R-modules are different from each other. Thus $\mathrm{FB}_1(X)$ coincides with the normal surface obtained

[2] We expect that all non-F-regular Frobenius sandwich rational double points are exhausted in Theorem 3.5, although we have not proved it yet. On the other hand, any F-regular Frobenius sandwich double point is an A_{p^e-1}-singularity and its e-th F-blowup is the minimal resolution for $e \gg 0$; see [Hara and Sawada 2011; Yasuda 2012].

by contracting the exceptional curve E_1 on the minimal resolution $\widetilde{X}$ indicated in the figure below:

$$\begin{array}{ccccc} & & E_3 & & \\ & & | & & \\ E_2 & — & E_1 & — & E_4 \end{array}$$

Furthermore, we see that the reflexive R-module corresponding to the central curve E_1 appears as a direct summand of the Frobenius pushforward of each nontrivial rank 1 reflexive module corresponding to E_i with $i = 2, 3, 4$. Thus $\mathrm{FB}_e(X)$ is the minimal resolution for $e \geq 2$, since the D_4^1-singularity is F-pure. A similar result holds for the case of a D_5^1-singularity. Note that D_4^1- and D_5^1-singularities are not F-regular.

Remark 3.6. The D_4^1-singularity in $p = 2$ is a *wild* quotient singularity, that is, there exists a group G of order 2 acting on $Y = \operatorname{Spec} k[[x, y]]$ such that the quotient $X = Y/G$ has the D_4^1-singularity. Although F-blowups of a *tame* quotient singularity are always dominated by the G-Hilbert scheme [Yasuda 2012], this example shows that the same does not hold for wild quotients. Let $R = k[[x, y]]^G \subset S = k[[x, y]]$ be the invariant subring. Then S is an R-module of rank 2. Thus the blowup of X at the R-module S, which coincides with the G-Hilbert scheme $\mathrm{Hilb}^G(Y)$, has at most two irreducible exceptional curves. On the other hand, the F-blowups $\mathrm{FB}_e(X)$ of the D_4^1-singularity have more than three irreducible exceptional curves. Hence the e-th F-blowup $\mathrm{FB}_e(X)$ of the D_4^1-singularity is not dominated by the G-Hilbert scheme $\mathrm{Hilb}^G(Y)$ for all $e \geq 1$.

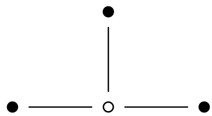
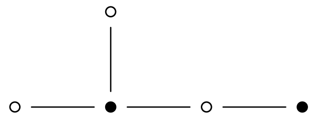
3c. *E_6^0-singularity in $p = 2$.* Let $R = k[x, y, z]/(z^2 + x^3 + y^2 z)$ and $X = \operatorname{Spec} R$. Then X has an E_6^0-singularity in characteristic $p = 2$. Write

$$A_1 = \begin{pmatrix} z & y & x & 0 \\ yz & z & 0 & x \\ x^2 & 0 & z & y \\ 0 & x^2 & yz & z \end{pmatrix}, \quad A_2 = \begin{pmatrix} x & y^2+z & y & 0 \\ z & x^2 & 0 & xy \\ 0 & 0 & x & y^2+z \\ 0 & 0 & z & x^2 \end{pmatrix} \quad \text{and} \quad A_3 = {}^t A_2.$$

Then their cokernels define nontrivial reflexive R-modules of rank 2 and those R-modules are different from each other. Now we see that presentation matrices of $F_* R$ and $F_*^2 R$ are equivalent to $A_1^{\oplus 2}$ and $A_1^{\oplus 4} \oplus A_2^{\oplus 2} \oplus A_3^{\oplus 2}$, respectively. Furthermore, a direct summand other than A_1, A_2 and A_3 does not appear in the presentation matrices of $F_*^e R$ for $e \geq 2$. Since the blowup of X at $\operatorname{Coker} A_1$ has only one singular point, we can specify the exceptional curve on the minimal resolution corresponding to $\operatorname{Coker} A_1$. The resulting descriptions of $\mathrm{FB}_e(X)$ are summarized in the following.

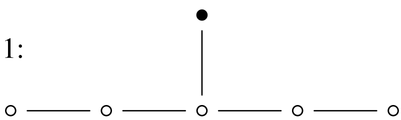
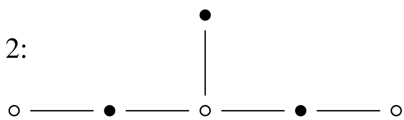
Proposition 3.7. *Let (X, x) be a rational double point of type D_4^1, D_5^1 or E_6^0 in characteristic $p = 2$. Then the e-th F-blowup $\mathrm{FB}_e(X)$ of (X, x) coincides with* *the normal surface obtained by contracting the exceptional curves on the minimal resolution $\widetilde{X}$ corresponding to the blank circles in the figure below*:

(1) *D_4^1 and D_5^1-singularity in $p=2$*:

$e=1$:  *and* 

For $e \geq 2$, the F-blowups $\mathrm{FB}_e(X)$ of both singularities coincide with the minimal resolution.

(2) *E_6^0-singularity in $p=2$*:

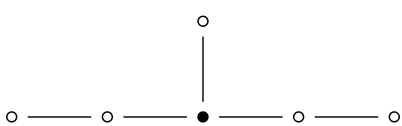
$e=1$:  $e \geq 2$: 

We can also compute the first F-blowup $\mathrm{FB}_1(X)$ of a few other rational double points with `Macaulay2`.

Example 3.8. (1) E_6^1-singularity in $p=2$: Let $R=k[x,y,z]/(z^2+x^2y+xy^2+xyz)$ and $X=\operatorname{Spec} R$. Then X has an E_6^1-singularity. Write

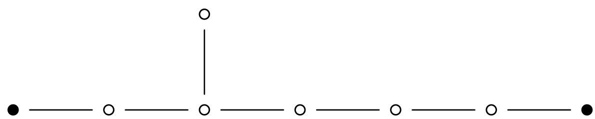
$$A=\begin{pmatrix} z & 0 & 0 & 0 & x & z \\ 0 & z & y & 0 & y & x \\ xy & yz & z & x^2+yz & 0 & 0 \\ 0 & 0 & x & x & y & 0 \\ x^2 & xz & 0 & yz & z & 0 \\ xy+y^2 & x^2 & 0 & xy & 0 & z \end{pmatrix}.$$

Then the cokernel of A defines an indecomposable reflexive R-module of rank 3. The presentation matrix of F_*R is equivalent to $A\oplus 0$, where 0 is the zero matrix of size 1. Thus $\mathrm{FB}_1(X)$ has a unique exceptional curve corresponding to the solid circle in the figure below and has three singular points (an A_1- and two A_2-singularities) on it:



(2) E_8^3-singularity in $p=2$: Let $R=k[x,y,z]/(z^2+x^3+y^5+y^3z)$ and $X=\operatorname{Spec} R$. Then X has an E_8^3-singularity. In this case, F_*R has two kinds of indecomposable reflexive R-modules. Since $\operatorname{rank} F_*R=4$, we see that F_*R is a direct sum of

indecomposable reflexive R-modules of rank 2 corresponding to the solid circles in the figure below:



Thus $\mathrm{FB}_1(X)$ has two exceptional curves corresponding to the solid circles meeting at the unique singular point of type D_6.

4. F-blowups of simple elliptic singularities

In this section (X, x) will denote a simple elliptic singularity defined over an algebraically closed field k of characteristic $p > 0$ unless otherwise noted. Then by a result of Hirokado [2004], (X, x) is quasihomogeneous. So we may assume that $X = \operatorname{Spec} R$ for a graded k-algebra

$$R = R(E, L) = \bigoplus_{n\geq 0} H^0(E, L^n)t^n,$$

where E is an elliptic curve over k, L is an ample line bundle on E and $\deg t = 1$. The minimal resolution $f : \widetilde{X} \to X$ of X is described as follows: $\widetilde{X}$ has an $\mathbb{A}^1$-bundle structure $\pi : \widetilde{X} = \operatorname{Spec}_E(L^n t^n) \to E$ over E, and its zero-section, which we also denote by E, is the exceptional curve of f. Its self-intersection number is $E^2 = -\deg L$. Our situation is summarized in the following diagram:

$$\begin{tikzcd} E \arrow[r, hook] \arrow[dr, "\mathrm{id}"'] & \widetilde{X} \arrow[r, "f"] \arrow[d, "\pi"] & X \\ & E & \end{tikzcd}$$

To compute the F-blowup $\mathrm{FB}_e(X)$ of X, we will look at the structure of the torsion-free pullback $f^\star R^{1/q}$ of $R^{1/q} \cong F^e_* \mathcal{O}_X$, where $q = p^e$. For this purpose we decompose

$$R^{1/q} = \bigoplus_{n\geq 0} H^0(E, F^e_* L^n)t^{n/q} \quad \text{as} \quad R^{1/q} = \bigoplus_{i=0}^{q-1} [R^{1/q}]_{i/q \bmod \mathbb{Z}},$$

where

$$[R^{1/q}]_{i/q \bmod \mathbb{Z}} = \bigoplus_{0\leq n\equiv i \bmod q} H^0(E, F^e_* L^n)t^{n/q} \cong \bigoplus_{m\geq 0} H^0(E, L^m \otimes F^e_* L^i)$$

is an R-summand of $R^{1/q}$ for $i = 0, 1, \ldots, q-1$; see [Smith and Van den Bergh 1997].

In what follows we put $q = p^e$ and $d = \deg L = -E^2$.

Lemma 4.1. *If* $1 \le i \le q-1$ *and* $q \ne di$, *then* $\widetilde{X}$ *is a flattening of* $[R^{1/q}]_{i/q \bmod \mathbb{Z}}$.

Proof. First of all, the locally free sheaf $L^m \otimes F^e_* L^i$ on E is generated by its global sections if $m \ge 1$, or $m = 0$ and $q < di$. To see this, let $P \in E$ and consider the exact sequence

$$0 \to L^m(-P) \otimes F^e_* L^i \to L^m \otimes F^e_* L^i \to \kappa(P) \otimes L^m \otimes F^e_* L^i \to 0. \tag{1}$$

Since $h^1(L^m(-P) \otimes F^e_* L^i) = h^1(L^{qm+i}(-qP)) = h^0(L^{-qm-i}(qP)) = 0$ by the assumption, the induced map $H^0(E, L^m \otimes F^e_* L^i) \to H^0(E, \kappa(P) \otimes L^m \otimes F^e_* L^i)$ is surjective, that is, $L^m \otimes F^e_* L$ is generated by its global sections at $P \in E$. Hence

$$\begin{aligned} f^\star[R^{1/q}]_{i/q \bmod \mathbb{Z}} &= \operatorname{Im}([R^{1/q}]_{i/q \bmod \mathbb{Z}} \otimes_R \mathcal{O}_{\widetilde{X}} \to F^e_* \mathcal{O}_{\widetilde{X}}) \\ &= \operatorname{Im}\Big(\bigoplus_{m \ge 0} H^0(E, L^m \otimes F^e_* L^i) \otimes_k \mathcal{O}_E \xrightarrow{\alpha} \bigoplus_{m \ge 0} L^m \otimes F^e_* L^i\Big) \\ &= \operatorname{Im}(H^0(E, F^e_* L^i) \otimes \mathcal{O}_E \xrightarrow{\alpha_0} F^e_* L^i) \oplus \bigoplus_{m \ge 1} L^m \otimes F^e_* L^i \\ &\subset \bigoplus_{m \ge 0} L^m \otimes F^e_* L^i \cong \pi^* F^e_* L^i, \end{aligned}$$

where α_m ($m \ge 0$) is the graded part of the map α of degree m, and in particular, $f^\star[R^{1/q}]_{i/q \bmod \mathbb{Z}} \cong \pi^* F^e_* L^i$ if $q < di$. Since $\pi^* F^e_* L^i$ is a locally free $\mathcal{O}_{\widetilde{X}}$-module, we consider the case $q > di$. Since α_m is surjective for $m \ge 1$, the $\mathcal{O}_{\widetilde{X}}$-module $\operatorname{Coker}(\alpha) = \operatorname{Coker}(\alpha_0)$ is regarded as a coherent sheaf on the exceptional curve $E \subset \widetilde{X}$ of f.

Claim. $\operatorname{Coker}(\alpha) = \operatorname{Coker}(\alpha_0)$ *is a locally free sheaf on* E, *so that it has depth* 1 *as an* $\mathcal{O}_{\widetilde{X}}$*-module at each point on* $E \subset \widetilde{X}$.

To prove the claim, note that $h^0(F^e_* L^i) = h^0(L^i) = di$ by Riemann–Roch and that $F^e_* L^i$ is a locally free sheaf on E of rank q, so that the rank of $\operatorname{Coker}(\alpha) = \operatorname{Coker}(\alpha_0)$ as an $\mathcal{O}_E$-module is at least $q - di$. On the other hand, since

$$H^0(E, \mathcal{O}_E(-P) \otimes F^e_* L^i) = H^0(E, L^i(-qP)) = 0$$

by our assumption, the cohomology long exact sequence of (3) for $m = 0$ turns out to be

$$0 \to H^0(E, F^e_* L^i) \to \kappa(P) \otimes F^e_* L^i \to H^1(E, \mathcal{O}_E(-P) \otimes F^e_* L^i) \to 0,$$

from which we see that the minimal number of local generators of $\operatorname{Coker}(\alpha)$ is $\dim \operatorname{Coker}(\alpha_0) \otimes \kappa(P) = q - di$. Comparing the rank and the minimal number of local generators, we conclude that $\operatorname{Coker}(\alpha) = \operatorname{Coker}(\alpha_0)$ is a locally free sheaf on E of rank $q - di$.

Now we have an exact sequence of $\mathcal{O}_{\widetilde{X}}$-modules

$$0 \to f^{\star}[R^{1/q}]_{i/q \bmod \mathbb{Z}} \to \pi^* F^e_* L^i \to \operatorname{Coker}(\alpha) \to 0,$$

in which $\pi^* F^e_* L^i$ and $\operatorname{Coker}(\alpha)$ have depth 2 and 1, respectively. Thus the depth of $f^{\star}[R^{1/q}]_{i/q \bmod \mathbb{Z}}$ is 2, so that it is locally free on $\widetilde{X}$. □

Remark 4.2. In the case where $1 \le i \le q-1$ and $q = di$, an argument similar to that in the proof of Lemma 4.1 shows that $f^{\star}[R^{1/q}]_{i/q \bmod \mathbb{Z}}$ is *not* flat at $P \in E \subset \widetilde{X}$ if and only if $L^i \cong \mathcal{O}_E(qP)$.

Corollary 4.3. *If $q = p^e > 1$ and $d = -E^2$ is not a power of the characteristic p, then $\widetilde{X}$ is the normalization of the blowup $\mathrm{Bl}_{N_q}(X)$ of $X = \operatorname{Spec} R$ at the R-module*

$$N_q = \bigoplus_{i=1}^{q-1} [R^{1/q}]_{i/q \bmod \mathbb{Z}}.$$

Proof. First we will see that N_q is not flat if $q = p^e > 1$. For, if N_q is flat, then the $\mathcal{O}_{X,x}$-module $\mathcal{O}_{X,x}^{1/q}$ has a free summand of rank at least $q(q-1)$. However, the rank of the free summand of $\mathcal{O}_{X,x}^{1/q}$ is exactly equal to 1, since $\mathcal{O}_{X,x}$ is a Gorenstein F-pure local ring with isolated non-F-regular locus; see [Aberbach and Enescu 2005; Sannai and Watanabe 2011, Theorem 5.1].

Now by Lemma 4.1, the minimal resolution $f : \widetilde{X} \to X$ is a flattening of N_q, so it factors as

$$f : \widetilde{X} \xrightarrow{g} \mathrm{Bl}_{N_q}(X) \xrightarrow{h} X.$$

Since N_q is not flat and X is normal, h is not an isomorphism and has an exceptional curve, which is equal to $g(E)$. Hence g is finite (and birational), so that $\widetilde{X}$ is the normalization of $\mathrm{Bl}_{N_q}(X)$. □

Next we consider the structure of $f^{\star}[R^{1/q}]_{0 \bmod \mathbb{Z}}$, which depends on whether R is F-pure or not. This is equivalent to saying whether the elliptic curve E is ordinary or supersingular, since the section ring $R = R(E, L)$ is F-pure if and only if $E = \operatorname{Proj} R$ is F-split.

4a. *The F-pure case.* We first consider the case where R is F-pure, or equivalently, E is an ordinary elliptic curve. In this case, given a fixed point $P_0 \in E$ as the identity element of the group law of E, there are exactly $q = p^e$ distinct q-torsion points $P_0, \dots, P_{q-1}$. In other words, there are exactly q nonisomorphic q-torsion line bundles $L_0, \dots, L_{q-1} \in \operatorname{Pic}^{\circ}(E)$ given by $L_i = \mathcal{O}_E(P_i - P_0)$. Then $F^e_* \mathcal{O}_E$ splits into line bundles as

$$F^e_* \mathcal{O}_E \cong \bigoplus_{i=0}^{q-1} L_i. \tag{2}$$

Indeed, since $\mathcal{O}_E$ is a direct summand of $F^e_*\mathcal{O}_E$ by F-splitting, each L_i is a direct summand of $L_i \otimes F^e_*\mathcal{O}_E \cong F^e_* F^{e*} L_i \cong F^e_*(L_i^q) \cong F^e_*\mathcal{O}_E$; see [Atiyah 1957].

Lemma 4.4. *Let E be an ordinary elliptic curve.*

(1) *Suppose $d = 1$ and choose the identity element $P_0 \in E$ so that $L \cong \mathcal{O}_E(P_0)$. Then $f^\star[R^{1/q}]_{0 \bmod \mathbb{Z}}$ is not flat exactly at the $q-1$ distinct q-torsion points $P_1, \dots, P_{q-1} \in E \subset \widetilde{X}$ other than P_0. Moreover, $[R^{1/q}]_{0 \bmod \mathbb{Z}}$ is flattened by blowing up the points $P_1, \dots, P_{q-1}$.*

(2) *If $d \geq 2$, then $\widetilde{X}$ is a flattening of $[R^{1/q}]_{0 \bmod \mathbb{Z}}$.*

Proof. Corresponding to the splitting of $F^e_*\mathcal{O}_E$ as in the formula (2) above, the R-module $[R^{1/q}]_{0 \bmod \mathbb{Z}}$ has a splitting $[R^{1/q}]_{0 \bmod \mathbb{Z}} \cong \bigoplus_{i=0}^{q-1} J_i$ into q nonisomorphic reflexive R-modules $R = J_0, J_1, \dots, J_{q-1}$ of rank 1, where

$$J_i = \Gamma_*(L_i) := \bigoplus_{m \in \mathbb{Z}} H^0(E, L_i \otimes L^m) = \bigoplus_{m \geq 0} H^0(E, L_i \otimes L^m).$$

In case (1) where $d = 1$, it is sufficient to show the following:

Claim. *For $i = 1, \dots, q-1$, $f^\star J_i$ is not flat exactly at the single point $P_i \in E \subset \widetilde{X}$. If $\sigma_i : \widetilde{X}_i \to \widetilde{X}$ is the blowup at P_i, then $(f \circ \sigma_i)^\star J_i$ is invertible.*

To prove the claim, note that $\deg L = 1$ and $\deg L_i = 0$. Then the following holds for the linear system $|L_i \otimes L^m|$ on E: $|L_i| = \varnothing$, $|L_i \otimes L| = \operatorname{Bs}|L_i \otimes L| = \{P_i\}$ and $|L_i \otimes L^m|$ is base point free for $m \geq 2$. Hence, as in the proof of the previous lemma,

$$\begin{aligned} f^\star J_i &= \operatorname{Im}(J_i \otimes_R \mathcal{O}_{\widetilde{X}} \to F^e_*\mathcal{O}_{\widetilde{X}}) \\ &= \operatorname{Im}\Big(\bigoplus_{m \geq 0} H^0(E, L_i \otimes L^m) \otimes_k \mathcal{O}_E \to \bigoplus_{m \geq 0} L_i \otimes L^m\Big) \\ &= L_i \otimes L(-P_i) \oplus \bigoplus_{m \geq 2} L_i \otimes L^m \subset \bigoplus_{m \geq 1} L_i \otimes L^m \cong \mathcal{O}_{\widetilde{X}}(-E) \otimes \pi^* L_i, \end{aligned}$$

where $L_i \otimes L(-P_i) \cong \mathcal{O}_E \subset L_i \otimes L$ is the graded part of degree $m = 1$. We therefore have the following exact sequence of $\mathcal{O}_{\widetilde{X}}$-modules:

$$0 \to f^\star J_i \to \mathcal{O}_{\widetilde{X}}(-E) \otimes \pi^* L_i \to \kappa(P_i) \to 0,$$

which tells us that $f^\star J_i = \mathcal{I}_{P_i} \cdot \mathcal{O}_{\widetilde{X}}(-E) \otimes \pi^* L_i$, where $\mathcal{I}_{P_i}$ is the ideal sheaf defining the closed point $P_i \in \widetilde{X}$. Now the claim follows immediately.

(2) If $\deg L \geq 2$, then the same argument as in (1) shows that $f^\star J_i$ is isomorphic to $\mathcal{O}_{\widetilde{X}}(-E) \otimes \pi^* L_i$, which is invertible. $\square$

We now state a structure theorem for F-blowups of F-pure $\widetilde{E}_8$-singularities, that is, F-pure simple elliptic singularities with $E^2 = -1$.

Theorem 4.5. *Let (X, x) be an F-pure simple elliptic singularity with the elliptic exceptional curve E on the minimal resolution $\widetilde{X}$ such that $E^2 = -1$. Let $P_0, \dots, P_{q-1} \in E$ be the $q = p^e$ distinct q-torsion points on $E \subset \widetilde{X}$, where the identity element P_0 is chosen so that*

$$\mathcal{O}_{\widetilde{X}}(-E) \otimes \mathcal{O}_E \cong \mathcal{O}_E(P_0),$$

and let $Z = \{P_1, \dots, P_{q-1}\} \subset \widetilde{X}$. Then for any $e \geq 1$, the normalization of the e-th F-blowup $\mathrm{FB}_e(X)$ coincides with the blowup $\mathrm{Bl}_Z(\widetilde{X})$ of $\widetilde{X}$ at the nontrivial q-torsion points.

In particular, the e-th F-blowup of X is not dominated by the minimal resolution of the singularity (X, x), and the monotonic sequence of F-blowups (*see* [Yasuda 2009]),

$$\cdots \to \mathrm{FB}_e(X) \to \cdots \to \mathrm{FB}_2(X) \to \mathrm{FB}_1(X) \to X,$$

does not stabilize.

Proof. Since $N_q = \bigoplus_{i=1}^{q-1} [R^{1/q}]_{i/q \bmod \mathbb{Z}}$ is a direct summand of $R^{1/q}$ as an R-module, we have a morphism $\mathrm{FB}_e(X) \to \mathrm{Bl}_{N_q}(X)$ over X. If we denote the normalization of $\mathrm{FB}_e(X)$ by $\widetilde{\mathrm{FB}}_e(X)$, then we have a morphism $\varphi\colon \widetilde{\mathrm{FB}}_e(X) \to \widetilde{X}$ by Corollary 4.3. On the other hand, since $\mathrm{Bl}_Z(\widetilde{X})$ is a flattening of $R^{1/q}$ by Lemmas 4.1 and 4.4, we have a morphism $\mathrm{Bl}_Z(\widetilde{X}) \to \mathrm{FB}_e(X)$ over X, which induces $\psi\colon \mathrm{Bl}_Z(\widetilde{X}) \to \widetilde{\mathrm{FB}}_e(X)$. Thus the blowup $\pi\colon \mathrm{Bl}_Z(\widetilde{X}) \to \widetilde{X}$ at $Z \subset \widetilde{X}$ factors as

$$\pi = \varphi \circ \psi\colon \mathrm{Bl}_Z(\widetilde{X}) \xrightarrow{\psi} \widetilde{\mathrm{FB}}_e(X) \xrightarrow{\varphi} \widetilde{X}.$$

Since $f^\star R^{1/q}$ is not flat exactly at $Z = \{P_1, \dots, P_{q-1}\}$ by Lemma 4.4, φ has an exceptional curve over every P_i and ψ is finite (and birational), by the same argument as in the proof of Corollary 4.3. Since $\widetilde{\mathrm{FB}}_e(X)$ is normal, ψ is an isomorphism, that is, $\mathrm{Bl}_Z(\widetilde{X}) \cong \widetilde{\mathrm{FB}}_e(X)$ as required. □

The theorem above has nothing to say about the normality of the F-blowups. Let us take a look at a `Macaulay2` computation.

Example 4.6. From [Hirokado 2004, Corollary 4.3], the variety

$$X = \operatorname{Spec} \mathbb{F}_2[x, y, z]/(y^2 + x^3 + xyz + z^6)$$

has a simple elliptic singularity of type $\tilde{E}_8$. Moreover from Fedder's criterion [1983], this is F-pure. Note that since F-blowups are compatible with extensions of perfect fields [Yasuda 2012], the fact that the base field is not algebraically closed does not pose a problem. By `Macaulay2` computation, one can check the following: The first F-blowup $\mathrm{FB}_1(X)$ is nonnormal and its exceptional set consists of two projective lines E_1 and E_2, which intersect transversally at one point. The normalization $\widetilde{\mathrm{FB}}_1(X)$ of $\mathrm{FB}_1(X)$ is smooth. The inverse image of E_1 in $\widetilde{\mathrm{FB}}_1(X)$

is a smooth elliptic curve, which agrees with Theorem 4.5. In particular, this experimental result shows that the normalization in the theorem is really necessary.

Next we consider the case where $E^2 \le -2$.

Theorem 4.7. *Let (X, x) be an F-pure simple elliptic singularity with the elliptic exceptional curve E on the minimal resolution $\widetilde{X}$ such that $E^2 \le -2$. Assume further that $d = -E^2$ is not a power of the characteristic p. Then $\widetilde{X}$ is the normalization of the e-th F-blowup $\mathrm{FB}_e(X)$ for all $e \ge 1$. Moreover, if $E^2 \le -3$, then $\widetilde{X} \cong \mathrm{FB}_e(X)$ for all $e \ge 1$.*

Proof. Since $\widetilde{X}$ is a flattening of $R^{1/q}$ by Lemmas 4.1 and 4.4, we see that $\widetilde{X}$ is the normalization of $\mathrm{FB}_e(X)$ as in the proof of Corollary 4.3.

To deduce a stronger conclusion in the special case $E^2 \le -3$, we need the following:

Lemma 4.8 [Mumford 1970]. *Let V be a projective variety, $\mathcal{F}$ a coherent sheaf on V and let L be a line bundle on V generated by its global sections. Suppose that $H^i(V, \mathcal{F} \otimes L^{-i}) = 0$ for all $i > 0$. Then the natural map*

$$H^0(V, \mathcal{F}) \otimes H^0(V, L)^{\otimes n} \to H^0(V, \mathcal{F} \otimes L^n)$$

is surjective for all $n \ge 1$.

Lemma 4.9. *Let $H_1, \dots, H_n$ be line bundles on an elliptic curve E of $\deg H_i \ge 3$ for $i = 1, \dots, n$. Then the natural map*

$$H^0(E, H_1) \otimes \cdots \otimes H^0(E, H_n) \to H^0(E, H_1 \otimes \cdots \otimes H_n)$$

is surjective.

Proof. The case $n \ge 3$ is easily reduced to the case $n = 2$ by induction on n, so let $n = 2$. If $\deg H_1 > \deg H_2$, then $H^1(E, H_1 \otimes H_2^{-1}) = 0$, so that the surjectivity of the map $H^0(E, H_1) \otimes H^0(E, H_2) \to H^0(E, H_1 \otimes H_2)$ immediately follows from Mumford's lemma. Suppose that $\deg H_1 = \deg H_2$ and let $L = H_2 - P$ for any fixed point $P \in E$. Then L is globally generated since $\deg L \ge 2$, and $H^1(E, H_1 \otimes L^{-1}) = 0$ since $\deg(H_1 \otimes L^{-1}) = 1 > 0$. Hence the map

$$H^0(E, H_1) \otimes H^0(E, L) \to H^0(E, H_1 \otimes L)$$

is surjective by Mumford's lemma. We now consider the following commutative diagram with exact rows:

$$\begin{array}{ccccccccc}
0 \to & H^0(H_1) \otimes H^0(L) & \to & H^0(H_1) \otimes H^0(H_2) & \to & H^0(H_1) \otimes H^0(H_2 \otimes \kappa(P)) & \to 0 \\
& \downarrow & & \downarrow & & \downarrow & \\
0 \to & H^0(H_1 \otimes L) & \to & H^0(H_1 \otimes H_2) & \to & H^0(H_1 \otimes H_2 \otimes \kappa(P)), &
\end{array}$$

where we have just verified the surjectivity of the vertical map on the left, and the vanishing of the right upper corner comes from $H^1(E, L) = 0$. So, to prove the required surjectivity of the vertical map in the middle, it suffices to show that the vertical map on the right is surjective, by the five-lemma. This map is factorized as

$$H^0(H_1) \otimes H^0(H_2 \otimes \kappa(P)) \xrightarrow{\alpha} H^0(H_1 \otimes \kappa(P)) \otimes H^0(H_2 \otimes \kappa(P)) \\ \xrightarrow{\beta} H^0(H_1 \otimes H_2 \otimes \kappa(P)).$$

Here α is surjective because of the vanishing $H^1(E, H_1(-P)) = 0$, and β is identified with the multiplication map $k^{\otimes 2} \xrightarrow{\sim} k$, which is clearly surjective. Thus $\beta \circ \alpha$ is surjective, and the lemma is proved. □

We continue the proof of Theorem 4.7 in the case $E^2 \leq -3$. Consider the decomposition (2) of $F^e_* \mathbb{O}_E$ into $q = p^e$-torsion line bundles $\mathbb{O}_E = L_0, L_1, \ldots, L_{q-1}$ on E. We fix any i with $0 < i \leq q-1$ and let $I \subset R$ be an ideal isomorphic to the reflexive R-module

$$J_i = \Gamma_*(L_i) = \bigoplus_{n \geq 1} H^0(E, L_i \otimes L^n) t^n$$

of rank 1, which is a nontrivial R-summand of $R^{1/q}$. Then the minimal resolution $f : \widetilde{X} \to X = \operatorname{Spec} R$ is factorized as

$$f : \widetilde{X} \to \mathrm{FB}_e(X) \to \mathrm{Bl}_I(X) \to X,$$

where the blowup $\mathrm{Bl}_I(X) = \operatorname{Proj} R[It]$ of X with respect to the ideal I has an exceptional curve that is the image of $E \subset \widetilde{X}$, since $I \cong J_i$ is not a flat R-module. It follows that $\widetilde{X}$ is the normalization of $\mathrm{Bl}_I(X)$. So, to prove the theorem, it is sufficient to show that the Rees algebra $R[It]$ is normal.

To prove the normality of $R[It] = \bigoplus_{m \geq 0} I^m t^m$, note that its normalization is

$$\widetilde{R[It]} = \bigoplus_{m \geq 0} \overline{I^m} t^m,$$

where $\overline{I^m} \subseteq R$ is the integral closure of the ideal I^m; see [Lipman 1969]. Note also that

$$I\mathbb{O}_{\widetilde{X}} \cong f^\star J_i \cong \bigoplus_{n \geq 1} (L_i \otimes L^n) t^n \cong \mathbb{O}_{\widetilde{X}}(-E) \otimes \pi^* L_i$$

is an invertible sheaf on $\widetilde{X}$ by Lemma 4.4, so that

$$\overline{I^m} \cong H^0(\widetilde{X}, \mathbb{O}_{\widetilde{X}}(-mE) \otimes \pi^* L_i^m) \cong \bigoplus_{n \geq m} H^0(E, L_i^m \otimes L^n) t^n \quad \text{for all } m \geq 1.$$

Now, since $\deg L \geq 3$, we can apply Lemma 4.9 to $H_1 = \cdots = H_m := L_i \otimes L$ and $H_{m+1} = \cdots = H_n := L$ to obtain the surjectivity of the map

$$H^0(E, L_i \otimes L)^{\otimes m} \otimes H^0(E, L)^{\otimes n-m} \to H^0(E, L_i^m \otimes L^n)$$

for all $n \geq m \geq 1$. This implies that the multiplication map $\bar{I}^{\otimes m} \to \overline{I^m}$ is surjective in all degree n. Since $I = \bar{I}$ is integrally closed, we conclude that $I^m = \overline{I^m}$, from which the normality of the Rees algebra $R[It]$ follows. □

Example 4.10. Let

$$X = \operatorname{Spec} \mathbb{F}_2[x, y, z]/(y^2 + xyz + x^3 z + xz^3).$$

Again from [Hirokado 2004, Corollary 4.3] and Fedder's criterion, X has an F-pure simple elliptic singularity of type $\tilde{E}_7$ at the origin. The exceptional set of $\mathrm{FB}_1(X)$ consists of three projective lines. It shows that it is necessary to suppose in Theorem 4.7 that $d = -E^2$ is not a power of p. The normalization of $\mathrm{FB}_1(X)$ is smooth.

Example 4.11. The variety

$$X = \operatorname{Spec} \mathbb{F}_2[x, y, z]/(y^2 z + xyz + x^3 + z^3)$$

has an F-pure simple elliptic singularity of type $\tilde{E}_6$. By `Macaulay2` computations, we can see that $\mathrm{FB}_1(X)$ is smooth and the exceptional set is a smooth elliptic curve, as expected from Theorem 4.7.

4b. ***The non-F-pure case.*** Now we consider the structure of $f^\star[R^{1/q}]_{0 \bmod \mathbb{Z}}$ assuming that R is not F-pure, or equivalently, E is a supersingular elliptic curve. In this case E has no nontrivial q-torsion point under the group law. Then, contrary to the F-pure case, $F^e_* \mathcal{O}_E$ turns out to be indecomposable as we will see below.

For any elliptic curve E and an integer $r > 0$, there exists an indecomposable vector bundle $\mathcal{F}_r$ on E of rank r and degree zero with $h^0(\mathcal{F}_r) = 1$, determined inductively by $\mathcal{F}_1 = \mathcal{O}_E$ and the unique nontrivial extension

$$0 \to \mathcal{F}_{r-1} \to \mathcal{F}_r \to \mathcal{O}_E \to 0. \tag{3}$$

Note that $\mathcal{F}_r$ is self-dual and (3) is the dual sequence of that in [Atiyah 1957, Theorem 5].

Lemma 4.12 (see [Atiyah 1957; Tango 1972]). *If E is a supersingular elliptic curve, then $F^e_* \mathcal{O}_E \cong \mathcal{F}_q$ for all $q = p^e$.*

Proof. Let $F^e_*\mathcal{O}_E = \mathcal{E}_1 \oplus \cdots \oplus \mathcal{E}_n$ be the decomposition of $F^e_*\mathcal{O}_E$ into indecomposable bundles $\mathcal{E}_i$ of rank r_i and degree d_i. Then $d_1 + \cdots + d_n = \chi(F^e_*\mathcal{O}_E) = 0$ by Riemann–Roch. Pick a nontrivial line bundle L of degree zero. Then

$$\sum_{i=1}^{n} h^0(\mathcal{E}_i \otimes L) = h^0(L \otimes F^e_*\mathcal{O}_E) = h^0(L^q) = 0,$$

since there is no nontrivial q-torsion line bundle on a supersingular elliptic curve. Hence $d_i = \deg(\mathcal{E}_i \otimes L) \le 0$ for all $i = 1, \dots, n$. Thus the indecomposable summands $\mathcal{E}_i$ of $F^e_*\mathcal{O}_E$ have degree $d_i = 0$, and exactly one of them, say $\mathcal{E}_1$, has a nonzero global section since $h^0(F^e_*\mathcal{O}_E) = 1$. Then by [Atiyah 1957, Theorem 5], we have $\mathcal{E}_1 \cong \mathcal{F}_{r_1}$ and $\mathcal{E}_i \cong \mathcal{F}_{r_i} \otimes L_i$ for $i = 2, \dots, n$, where $L_2, \dots, L_n$ are nontrivial line bundles of degree zero. Suppose that $n \ge 2$. Then $L_2^{-1} \otimes F^e_*\mathcal{O}_E$ has a nonzero global section since its direct summand $\mathcal{F}_{r_2}$ does. On the other hand, however, $H^0(E, L_2^{-1} \otimes F^e_*\mathcal{O}_E) = H^0(E, L_2^{-q}) = 0$ since L_2 is not a q-torsion line bundle by our assumption. We thus conclude that $n = 1$, that is, $F^e_*\mathcal{O}_E \cong \mathcal{F}_q$. $\square$

Now for each r, we consider the graded R-module

$$M_r = \bigoplus_{n \ge 0} H^0(E, \mathcal{F}_r \otimes L^n) t^n$$

and regard its torsion-free pullback $\widetilde{M_r} = f^\star M_r$ to the minimal resolution $\widetilde{X}$ of $X = \operatorname{Spec} R$ as a subsheaf of

$$\mathcal{M}_r = \bigoplus_{n \ge 0} (\mathcal{F}_r \otimes L^n) t^n.$$

To obtain information on the flattening of $R^{1/q}$, we consider the torsion-free pullback $f^\star M_r$ of M_r to the minimal resolution, because $[R^{1/q}]_{0 \bmod \mathbb{Z}} \cong M_q$ by Lemma 4.12.

4b1. *Non-F-pure $\widetilde{E}_8$-singularities.* We first consider the case of $\widetilde{E}_8$-singularities, that is, the case $\deg L = -E^2 = 1$. In this case, $L \cong \mathcal{O}_E(P_0)$ for a point $P_0 \in E$.

We fix any point $P \in E$ and let $V \subset E$ be a sufficiently small open neighborhood V of P on which L and $\mathcal{F}_r$ trivialize. We choose a local basis $e_1, \dots, e_r$ of $\mathcal{F}_r$ on V inductively as follows. For $r = 1$, let e_1 be a (local) basis of $\mathcal{F}_1 = \mathcal{O}_E$ corresponding to its global section $1 \in H^0(E, \mathcal{O}_E)$. For $r \ge 2$, we think of $\mathcal{F}_{r-1}$ as a subbundle of $\mathcal{F}_r$ via the exact sequence (3), and extend the local basis $e_1, \dots, e_{r-1}$ of $\mathcal{F}_{r-1}$ on V to a local basis $e_1, \dots, e_r$ of $\mathcal{F}_r$.

Let $U = \pi^{-1}V \subset \widetilde{X}$. Then, with the local trivialization $L|_V \cong \mathcal{O}_V$ and

$$\mathcal{F}_r|_V \cong \bigoplus_{i=1}^{r} \mathcal{O}_V e_i \cong \mathcal{O}_V^{\oplus r}$$

as above, we have

$$\mathcal{M}_r|_U \cong \bigoplus_{i=1}^{r} \mathcal{O}_U e_i \cong \mathcal{O}_U^{\oplus r},$$

where $\mathcal{O}_U = \bigoplus_{n\geq 0}(L|_V)^n t^n \cong \bigoplus_{n\geq 0} \mathcal{O}_V t^n = \mathcal{O}_V[t]$. Note that the fiber coordinate t and a regular parameter u at $P \in E$ form a system of coordinates of U. With this notation we shall express generators of the $\mathcal{O}_U$-module $\widetilde{M}_r|_U \subseteq \mathcal{M}_r|_U$, which come from homogeneous elements of the graded R-module M_r.

First note that the degree zero piece $[M_r]_0 = H^0(E, \mathcal{F}_r) = H^0(E, \mathcal{F}_1)$ of M_r is a one-dimensional k-vector space, so that its contribution to the generation of $\widetilde{M}_r|_U$ is just e_1. It is also easy to see that the graded parts of $\widetilde{M}_r|_U$ and $\mathcal{M}_r|_U$ coincide in degree ≥ 2 and are generated by $t^2 e_1, \ldots, t^2 e_r$, since $\mathcal{F}_r \otimes L^n$ is generated by global sections for $n \geq 2$. It remains to consider the contribution of the degree one piece $[M_r]_1 = H^0(E, \mathcal{F}_r \otimes L)t$ to the generation of $\widetilde{M}_r|_U$. To this end, note that we have an exact sequence

$$0 \to H^0(E, \mathcal{F}_i \otimes L) \to H^0(E, \mathcal{F}_{i+1} \otimes L) \to H^0(E, L) \to 0$$

for $1 \leq i \leq r-1$, via which we regard $H^0(E, \mathcal{F}_i \otimes L)$ as a subspace of $H^0(E, \mathcal{F}_r \otimes L)$. Then, since $h^0(\mathcal{F}_i \otimes L) = i$ by Riemann–Roch, we can choose a basis $s_1, \ldots, s_r$ of $H^0(E, \mathcal{F}_r \otimes L)$ so that $s_1, \ldots, s_i$ form a basis of $H^0(E, \mathcal{F}_i \otimes L)$ for $1 \leq i \leq r$. It also follows from exact sequence (3)$\otimes L$ that the global sections $s_1, \ldots, s_i$ generate $\mathcal{F}_i \otimes L$ on $E \setminus \{P_0\}$, so that they give a basis of $\mathcal{F}_i \otimes L \otimes K$ as a vector space over the function field K of E. On the other hand, $e_1, \ldots, e_i$ can also be viewed as a basis of $\mathcal{F}_i \otimes L \otimes K \cong K^{\oplus i}$ under the local trivialization $\mathcal{F}_i \otimes L|_V \cong \bigoplus_{j=1}^{i} \mathcal{O}_V e_i \cong \mathcal{O}_V^{\oplus i}$ induced from $\mathcal{F}_i|_V \cong \mathcal{O}_V^{\oplus i}$ and $L|_V \cong \mathcal{O}_V$. We will compare the basis consisting of $s_i \otimes 1$ and the standard basis $e_1, \ldots, e_r$ of $\mathcal{F}_r \otimes L \otimes K \cong K^{\oplus r}$ using the following commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & H^0(\mathcal{F}_{i-1} \otimes L) \otimes \mathcal{O}_V \arrow[r] \arrow[d] & H^0(\mathcal{F}_i \otimes L) \otimes \mathcal{O}_V \arrow[r] \arrow[d] & H^0(L) \otimes \mathcal{O}_V \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathcal{F}_{i-1} \otimes L|_V \arrow[r] \arrow[d, "\cong"] & \mathcal{F}_i \otimes L|_V \arrow[r] \arrow[d, "\cong"] & L|_V \arrow[r] \arrow[d, "\cong"] & 0 \\
0 \arrow[r] & \mathcal{O}_V^{\oplus i-1} \arrow[r] & \mathcal{O}_V^{\oplus i} \arrow[r] & \mathcal{O}_V \arrow[r] & 0
\end{tikzcd}$$

Suppose now that $P = P_0$. Since $\operatorname{Bs}|L| = \{P_0\}$, we may choose a regular parameter u at $P_0 \in E$ so that $s_1 \otimes 1 = u$. It then follows from the diagram above that

$$s_i \otimes 1 = u e_i + \sum_{j=1}^{i-1} a_{i,j} e_j,$$

where the a_{ij} are local regular functions on V. We claim that we can replace $s_1, \dots, s_r$ so that they satisfy the following condition:

$$u | a_{i,j} \text{ for } 1 \le j \le i-2 \text{ but } a_{i,i-1} \text{ is not divisible by } u. \tag{4}$$

To prove the claim, there is nothing to do for $i = 1$. So let $i = 2$ and suppose $u | a_{2,1}$. We consider a k-linear map $H^0(E, L) \to H^0(E, \mathcal{F}_2 \otimes L)$ given by $s_1 \mapsto s_2$, which gives rise to a K-linear map $K \cong L \otimes K \to \mathcal{F}_2 \otimes L \otimes K \cong K^2$ sending $1 = u^{-1}(s_1 \otimes 1) \mapsto u^{-1}(s_2 \otimes 1) = e_2 + (a_{21}/u)e_1$. Since $a_{21}/u \in \mathcal{O}_V$, this gives a splitting of the surjective map $\mathcal{O}_V^{\oplus 2} \cong \mathcal{F}_2 \otimes L|_V \to L|_V \cong \mathcal{O}_V$ at $P_0 \in V$, as well as at any other point. Then we have a global splitting of the surjective map $\mathcal{F}_2 \otimes L \to L$, contradicting the nontriviality of the extension (3). Thus $a_{2,1}(P_0) \ne 0$. Next let $i \ge 3$. Then by induction, we may replace s_i by $s_i - \sum_{j=1}^{i-2} (a_{i,j}(P_0)/a_{j+1,j}(P_0)) s_{j+1}$ to assume that $u | a_{i,j}$ for $1 \le j \le i-2$. It then follows that $a_{i,i-1}$ is not divisible by u because otherwise, $s_1 \mapsto s_i$ would give a global splitting of $\mathcal{F}_i \otimes L \to L$ as above.

Consequently, local generators of $\widetilde{M}_r$ on a neighborhood U_0 of P_0 are described as

$$\begin{aligned} \widetilde{M}_r|_{U_0} &= \mathcal{O}_{U_0}\langle e_1, tue_i + a_{i,i-1} t e_{i-1}, t^2 e_i \mid 2 \le i \le r \rangle \\ &= \mathcal{O}_{U_0}\langle e_1, tue_i + a_{i,i-1} t e_{i-1}, t^2 e_r \mid 2 \le i \le r \rangle, \end{aligned}$$

where $a_{i,i-1}(P_0) \ne 0$. Accordingly the ideal $\mathcal{I}_{\widetilde{M}_r} \subset \mathcal{O}_{\widetilde{X}}$ defined in Section 2 has the following local expression:

$$\mathcal{I}_{\widetilde{M}_r}|_{U_0} \cong (t^r, t^{r-1}u^{r-1}) \cong (t, u^{r-1}).$$

If $P_0 \ne P \in U$ then $\widetilde{M}_r|_U = \mathcal{O}_U \langle e_1, te_i \mid 2 \le i \le r \rangle \cong \mathcal{O}_U^{\oplus r}$ by a similar argument.

Summarizing the argument so far, we have

Theorem 4.13. *Let (X, x) be a non-F-pure simple elliptic singularity with the elliptic exceptional curve E on the minimal resolution $\widetilde{X}$ such that $E^2 = -1$. Let P_0 be the point on $E \subset \widetilde{X}$ such that $\mathcal{O}_{\widetilde{X}}(-E) \otimes \mathcal{O}_E \cong \mathcal{O}_E(P_0)$ and let $\mathcal{I}_e \subset \mathcal{O}_{\widetilde{X}}$ be the ideal sheaf defining a fat point supported at $P_0 \in \widetilde{X}$ whose local expression at P_0 is*

$$(\mathcal{I}_e)_{P_0} = (t, u^{p^e - 1})$$

as above. Then for any $e \ge 1$, the blowup $\mathrm{Bl}_{\mathcal{I}_e}(\widetilde{X})$ of $\widetilde{X}$ at $\mathcal{I}_e$ coincides with the normalization of the e-th F-blowup $\mathrm{FB}_e(X)$.

Proof. We know that $Y = \mathrm{Bl}_{\mathcal{I}_e}(\widetilde{X})$ is a flattening of R^{1/p^e} from the argument above and Corollary 4.3. It is also easy to see that the exceptional curve of the blowup

$\pi : Y \to \widetilde{X}$ is a single $\mathbb{P}^1$. Then the same argument as in the proof of Theorem 4.5 shows that π factors through the normalized F-blowup $\widetilde{\mathrm{FB}}_e(X)$ as

$$\pi = \varphi \circ \psi : Y = \mathrm{Bl}_{\mathcal{I}_e}(\widetilde{X}) \xrightarrow{\psi} \widetilde{\mathrm{FB}}_e(X) \xrightarrow{\varphi} \widetilde{X}$$

and that ψ gives an isomorphism $Y \cong \widetilde{\mathrm{FB}}_e(X)$. □

Remark 4.14. Theorem 4.13 says that the e-th normalized F-blowup $\widetilde{\mathrm{FB}}_e(X)$ has the exceptional set consisting of an elliptic curve $E_1 \cong E$ and a smooth rational curve $E_2 \cong \mathbb{P}^1$, and has an A_{p^e-2}-singularity on $E_2 \setminus E_1$. The theorem also says that $\mathrm{FB}_e(X)$ does not dominate $\mathrm{FB}_{e'}(X)$ whenever e and e' are distinct positive integers. In other words, the monotonicity of F-blowup sequences breaks down for non-F-pure $\widetilde{E}_8$-singularities; compare to the F-pure case [Yasuda 2009]. On the other hand, it again has nothing to say about the normality of $\mathrm{FB}_e(X)$.

Let us examine our observation with `Macaulay2` computation.

Example 4.15. The variety

$$X = \operatorname{Spec} \mathbb{F}_3[x, y, z]/(x(x - z^2)(x - 2z^2) - y^2)$$

has a non-F-pure simple elliptic singularity of type $\tilde{E}_8$. The exceptional set of $\mathrm{FB}_1(X)$ is the union of a smooth elliptic curve E_1 and a projective line E_2. We could not check the normality of $\mathrm{FB}_1(X)$ by `Macaulay2` computation only, but we could check the following using `Macaulay2`:

> $\mathrm{FB}_1(X)$ is normal at the generic points of E_1 and E_2, and there is a point on $E_2 \setminus E_1$ where $\mathrm{FB}_1(X)$ is normal but singular. The blowup of $\mathrm{FB}_1(X)$ at this point has the projective line as its exceptional locus. (∗)

It agrees with the fact that $\mathrm{FB}_1(X)$ has an A_1-singularity on $E_2 \setminus E_1$ as stated in the remark above.

Proposition 4.16. *For X as in Example 4.15, if (∗) is correct, then* $\mathrm{FB}_1(X)$ *is normal.*

Proof. We may replace the base field $\mathbb{F}_3$ with an algebraically closed field k. Being quasihomogeneous, X has a k^*-action. From the construction or the universality, the action lifts to F-blowups of X. Every point of the divisor $E_1 \subset \mathrm{FB}_1(X)$, which is a smooth elliptic curve, is fixed by the k^*-action. On the other hand, the divisor $E_2 \cong \mathbb{P}^1$ has exactly two fixed points. One is the singular but normal point mentioned above and the other is the intersection $E_1 \cap E_2$. Since the normal locus is open and there is the k^*-action, $\mathrm{FB}_1(X)$ is normal along E_2 possibly except at $E_1 \cap E_2$. Therefore it is now enough to show that $\mathrm{FB}_1(X)$ is normal along E_1. Let $\tilde{E}_1$ and $\tilde{E}_2$ be the preimages of E_1 and E_2 on the normalization $\widetilde{\mathrm{FB}}_1(X)$ of $\mathrm{FB}_1(X)$. Then for each $i = 1, 2$, since E_i is normal and $\mathrm{FB}_1(X)$ is normal at the generic point of E_i, the map $\tilde{E}_i \to E_i$ is an isomorphism.

Let A be the complete local ring of $\mathrm{FB}_1(X)$ at a point z on E_1. Its normalization is $k[\![s,t]\!]$. We choose local coordinates s, t so that the k^*-action on $k[\![s,t]\!]$ is linear and locally $s = 0$ defines $\tilde{E}_1$ and $t = 0$ defines the only one-dimensional orbit closure passing through the point over z. Then the k^*-action on t is trivial and the one on s is nontrivial. Since $\tilde{E}_i \to E_i$ for $i = 1, 2$ are isomorphisms, the composite maps $A \hookrightarrow k[\![s,t]\!] \to k[\![s]\!]$ and $A \hookrightarrow k[\![s,t]\!] \to k[\![t]\!]$ are surjective. Therefore A contains formal power series of the forms

$$f = f_1 s + f_2 t + (\text{higher terms}), \quad (\text{for } f_i \in k,\ f_1 \neq 0),$$
$$g = g_1 s + g_2 t + (\text{higher terms}), \quad (\text{for } g_i \in k,\ g_2 \neq 0).$$

Then by a suitable linear combination of them, we obtain a formal power series

$$h = h_1 s + h_2 t + (\text{higher terms}), \quad (\text{for } h_i \in k,\ h_1 \neq 0, h_2 \neq 0)$$

contained in A. Then for $1 \neq \lambda \in k^*$, $\lambda h \in A$ has a linear part linearly independent of that of h. It follows that $A = k[\![s,t]\!]$ and hence $\mathrm{FB}_1(X)$ is normal. □

Example 4.17. The variety

$$X = \operatorname{Spec} \mathbb{F}_2[x, y, z]/(y^2 + yz^3 + x^3)$$

has a non-F-pure simple elliptic singularity of type $\tilde{E}_8$. The Frobenius pushforward $F_*\mathbb{O}_X$ of the coordinate ring decomposes into the direct sum of two modules, say N_1 and N_2. Then F_*N_i for $i = 1, 2$ further decomposes as $F_*N_i = N_{i1} \oplus N_{i2}$. By `Macaulay2` computation, we saw that the torsion-free pullbacks $\widetilde{N_1}$ and $\widetilde{N_{11}}$ of N_1 and N_{11} are nonflat at a point and those of the others are flat. Moreover the ideals associated to $\widetilde{N_1}$ and $\widetilde{N_{11}}$ as in Proposition 2.5 are respectively of the forms (u, v) and (u, v^3) around the point with respect to some local coordinates u, v. The last result coincides with Theorem 4.13.

4b2. *Non-F-pure simple elliptic singularities with $E^2 \leq -2$.* In this case, we have $\deg L = -E^2 \geq 2$. Then the argument in Section 4b1 shows that $\widetilde{M_r}$ is flat.

Proposition 4.18. *Let (X, x) be a non-F-pure simple elliptic singularity with elliptic exceptional curve E on the minimal resolution $\widetilde{X}$. Suppose $E^2 \leq -2$ and $d = -E^2$ is not a power of the characteristic p. Then $\widetilde{X}$ is the normalization of the e-th F-blowup* $\mathrm{FB}_e(X)$ *for all $e \geq 1$.*

Proof. Since $\widetilde{X}$ is a flattening of $R^{1/q} = M_q \oplus N_q$ by Lemmas 4.1 and 4.12 and Section 4b1, the proof goes similarly to that of Theorem 4.7. Note that $\mathbb{O}_{X,x}^{1/q}$ has no free summand in this case, since $\mathbb{O}_{X,x}$ is not F-pure. □

Example 4.19. The variety

$$X = \operatorname{Spec} \mathbb{F}_2[x, y, z]/(y^2 z + yz^2 + x^3)$$

has a non-F-pure simple elliptic singularity of type $\tilde{E}_6$. We could check that $\mathrm{FB}_1(X)$ is the minimal resolution.

Remark 4.20. The behavior of F-blowups remains unsettled in some cases, that is, (i) the case $E^2 \leq -2$ and $-E^2$ is a power of p; and (ii) the normality of F-blowups of non-F-pure simple elliptic singularities with $E^2 \leq -3$. These cases will be treated in [Hara 2013].

References

[Aberbach and Enescu 2005] I. M. Aberbach and F. Enescu, "The structure of F-pure rings", *Math. Z.* **250**:4 (2005), 791–806. MR 2006m:13009 Zbl 1102.13001

[Artin 1962] M. Artin, "Some numerical criteria for contractability of curves on algebraic surfaces", *Amer. J. Math.* **84** (1962), 485–496. MR 26 #3704 Zbl 0105.14404

[Artin 1977] M. Artin, "Coverings of the rational double points in characteristic p", pp. 11–22 in *Complex analysis and algebraic geometry*, edited by J. W. L. Baily and T. Shioda, Iwanami Shoten, Tokyo, 1977. MR 56 #8559 Zbl 0358.14008

[Artin and Verdier 1985] M. Artin and J.-L. Verdier, "Reflexive modules over rational double points", *Math. Ann.* **270**:1 (1985), 79–82. MR 85m:14006 Zbl 0553.14001

[Atiyah 1957] M. F. Atiyah, "Vector bundles over an elliptic curve", *Proc. London Math. Soc.* (3) **7** (1957), 414–452. MR 24 #A1274 Zbl 0084.17305

[Fedder 1983] R. Fedder, "F-purity and rational singularity", *Trans. Amer. Math. Soc.* **278**:2 (1983), 461–480. MR 84h:13031 Zbl 0519.13017

[Grayson and Stillman 2012] D. R. Grayson and M. E. Stillman, "Macaulay2, a software system for research in algebraic geometry", 2012, available at http://www.math.uiuc.edu/Macaulay2/.

[Hara 2012] N. Hara, "F-blowups of F-regular surface singularities", *Proc. Amer. Math. Soc.* **140**:7 (2012), 2215–2226. MR 2898685

[Hara 2013] N. Hara, "Structure of the F-blowups of simple elliptic singularities", preprint, 2013.

[Hara and Sawada 2011] N. Hara and T. Sawada, "Splitting of Frobenius sandwiches", pp. 121–141 in *Higher dimensional algebraic geometry*, edited by S. Mukai and N. Nakayama, RIMS Kôkyûroku Bessatsu **B24**, Res. Inst. Math. Sci., Kyoto, 2011. MR 2012f:13009 Zbl 1228.13009

[Hirokado 2004] M. Hirokado, "Deformations of rational double points and simple elliptic singularities in characteristic p", *Osaka J. Math.* **41**:3 (2004), 605–616. MR 2005k:14006 Zbl 1065.14004

[Ito and Nakamura 1996] Y. Ito and I. Nakamura, "McKay correspondence and Hilbert schemes", *Proc. Japan Acad. Ser. A Math. Sci.* **72**:7 (1996), 135–138. MR 97k:14003 Zbl 0881.14002

[Kunz 1969] E. Kunz, "Characterizations of regular local rings for characteristic p", *Amer. J. Math.* **91** (1969), 772–784. MR 40 #5609 Zbl 0188.33702

[Lipman 1969] J. Lipman, "Rational singularities, with applications to algebraic surfaces and unique factorization", *Inst. Hautes Études Sci. Publ. Math.* 36 (1969), 195–279. MR 43 #1986 Zbl 0181.48903

[Mumford 1970] D. Mumford, "Varieties defined by quadratic equations", pp. 29–100 in *Questions on Algebraic Varieties* (Ciclo, Varenna, 1969), Edizioni Cremonese, Rome, 1970. MR 44 #209 Zbl 0198.25801

[Oneto and Zatini 1991] A. Oneto and E. Zatini, "Remarks on Nash blowing-up", *Rend. Sem. Mat. Univ. Politec. Torino* **49**:1 (1991), 71–82. MR 94d:14013 Zbl 0802.14001

[Sannai and Watanabe 2011] A. Sannai and K.-i. Watanabe, "F-signature of graded Gorenstein rings", *J. Pure Appl. Algebra* **215**:9 (2011), 2190–2195. MR 2012c:13009 Zbl 1223.13003

[Smith and Van den Bergh 1997] K. E. Smith and M. Van den Bergh, "Simplicity of rings of differential operators in prime characteristic", *Proc. London Math. Soc.* (3) **75**:1 (1997), 32–62. MR 98d:16039 Zbl 0948.16019

[Tango 1972] H. Tango, "On the behavior of extensions of vector bundles under the Frobenius map", *Nagoya Math. J.* **48** (1972), 73–89. MR 47 #3401 Zbl 0239.14007

[Toda and Yasuda 2009] Y. Toda and T. Yasuda, "Noncommutative resolution, F-blowups and D-modules", *Adv. Math.* **222**:1 (2009), 318–330. MR 2011b:14005 Zbl 1175.14005

[Villamayor U. 2006] O. Villamayor U., "On flattening of coherent sheaves and of projective morphisms", *J. Algebra* **295**:1 (2006), 119–140. MR 2006m:13008 Zbl 1087.14011

[Yasuda 2007] T. Yasuda, "Higher Nash blowups", *Compos. Math.* **143**:6 (2007), 1493–1510. MR 2008j:14029 Zbl 1135.14011

[Yasuda 2009] T. Yasuda, "On monotonicity of F-blowup sequences", *Illinois J. Math.* **53**:1 (2009), 101–110. MR 2011b:14007 Zbl 1194.14010

[Yasuda 2012] T. Yasuda, "Universal flattening of Frobenius", *Amer. J. Math.* **134**:2 (2012), 349–378. MR 2905000 Zbl 06029054

Communicated by Kei-Ichi Watanabe
Received 2011-10-12 Revised 2012-02-26 Accepted 2012-04-13

hara@math.tohoku.ac.jp Mathematical Institute, Tohoku University, Sendai 980-8578, Japan

sawada@fukushima-nct.ac.jp Department of General Education, Fukushima National College of Technology, Iwaki 970-8034, Japan

takehikoyasuda@math.sci.osaka-u.ac.jp Department of Mathematics, Graduate School of Science, Osaka University, Toyonaka 560-0043, Japan

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.

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. 3 2013