Algebra & Number Theory

Volume 2
2008
No. 7

mathematical sciences publishers

Algebra & Number Theory
www.jant.org

See inside back cover or www.jant.org for submission instructions.

Regular subscription rate for 2008: $180.00 a year ($120.00 electronic only).
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory, ISSN 1937-0652, at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

PUBLISHED BY
mathematical sciences publishers
http://www.mathscipub.org
A NON-PROFIT CORPORATION
Typeset in LaTeX

The Frobenius structure of local cohomology

Florian Enescu and Melvin Hochster

Given a local ring of positive prime characteristic there is a natural Frobenius action on its local cohomology modules with support at its maximal ideal. In this paper we study the local rings for which the local cohomology modules have only finitely many submodules invariant under the Frobenius action. In particular we prove that F-pure Gorenstein local rings as well as the face ring of a finite simplicial complex localized or completed at its homogeneous maximal ideal have this property. We also introduce the notion of an antinilpotent Frobenius action on an Artinian module over a local ring and use it to study those rings for which the lattice of submodules of the local cohomology that are invariant under Frobenius satisfies the ascending chain condition.

1. Introduction

All given rings in this paper are commutative, associative with identity, and Noetherian. Throughout, p denotes a positive prime integer. For the most part, we shall be studying local rings, that is, Noetherian rings with a unique maximal ideal. Likewise our main interest is in rings of positive prime characteristic p. If (R, m) is local of characteristic p, there is a natural action of the Frobenius endomorphism of R on each of its local cohomology modules $H^j_m(R)$. We call an R-submodule N of one of these local cohomology modules F-*stable* if the action of F maps N into itself.

One of our objectives is to understand when a local ring, (R, m), especially a reduced Cohen–Macaulay local ring, has the property that only finitely many R-submodules of its local cohomology modules are F-stable. When this occurs we say that R is FH-*finite*. We shall also study the problem of determining conditions under which the local cohomology modules of R have finite length in the category of R-modules with Frobenius action. We say that R has *finite FH-length* in this

MSC2000: primary 13A35; secondary 13D45.

Keywords: local cohomology, Frobenius action, Frobenius functor, F-pure ring, Gorenstein ring, antinilpotent module, tight closure, face ring, FH-finite ring, finite FH-length.

Enescu was partially supported by NSA Young Investigator grant H98230-07-1-0034. Hochster was partially supported by NSF grant DMS–0400633.

case. Of course, when the ring is Cohen–Macaulay there is only one nonvanishing local cohomology module, $H^d_m(R)$, where $d = \dim(R)$. The problem of studying the F-stable submodules of $H^d_m(R)$ arises naturally in tight closure theory, taking a point of view pioneered by K. Smith [2003; 1994; 1997a; 1997b]. For example, if R is complete, reduced, and Gorenstein, the largest proper F-stable submodule of $H^d_m(R)$ corresponds to the tight closure of 0 (in the finitistic sense, see [Hochster and Huneke 1990, §8]), and its annihilator is the test ideal of R. Also see Discussion 2.10 here. We would like to note here that other results related to F-stable submodules of local cohomology may be found in [Enescu 2001; 2003; Katzman 2006; Sharp 2007].

The main result of Section 3 is one of general interest. Let M be a module over an excellent local ring R and consider a family of submodules $\{N_\lambda\}_{\lambda\in\Lambda}$ of M closed under sum, intersection and primary decomposition. Our result states, in particular, that if the set $\{\mathrm{Ann}_R(M/N_\lambda) : \lambda \in \Lambda\}$ consists of radical ideals then it is finite.

This theorem has a number of important corollaries. One of them, relevant to our objectives and also proven by Sharp [2007, Theorem 3.10 and Corollary 3.11], states that for any local ring R of prime characteristic $p > 0$, if Frobenius acts injectively on an Artinian R-module M, then the set of annihilators of F-stable submodules of M is a finite set of radical ideals closed under primary decomposition. This leads to the fact that if R is F-pure and Gorenstein (or even quasi-Gorenstein) then $H^d_m(R)$ has only finitely F-stable R-submodules. See Section 3. Another one is the fact that in an excellent local ring a family of radical ideals closed under sum, intersection and primary decomposition is finite.

In Section 4 we explain the relationship between FH-finite rings and rings that have finite FH-length. We introduce the notion of an *antinilpotent* Frobenius action on an Artinian module over a local ring. Using results of [Lyubeznik 1997] and [Hochster 2008], we show that the local cohomology of a local ring R of characteristic p is antinilpotent if and only if the local cohomology of $R[\![x]\!]$ has finite FH-length, in which case the local cohomology of R and every formal power series ring over R is FH-finite.

In Section 5 we show that if R is the face ring of a finite simplical complex localized or completed at its homogeneous maximal ideal, then R is FH-finite. See Theorem 5.1.

2. Notation and terminology

Discussion 2.1 (Some basics about tight closure). Unless otherwise specified, we shall assume throughout that R is a Noetherian ring of positive prime characteristic p, although this hypothesis is usually repeated in theorems and definitions. R°

denotes the complement of the union of the minimal primes of R, and so, if R is reduced, R° is simply the multiplicative system of all nonzero divisors in R. We shall write $\mathbf{F}^e$ (or $\mathbf{F}^e_R$ if we need to specify the base ring) for the *Peskine–Szpiro* or *Frobenius functor* from R-modules to R-modules. Note that $\mathbf{F}^e$ preserves both freeness and finite generation of modules, and is exact precisely when R is regular (see [Herzog 1974; Kunz 1969]). If $N \subseteq M$ we write $N^{[q]}$ for the image of $F^e(N)$ in $F^e(M)$, although it depends on the inclusion $N \to M$, not just on N. If $u \in M$ we write u^{p^e} for the image $1 \otimes u$ of u in $F^e(M)$. With this notation, $(u+v)^q = u^q + v^q$ and $(ru)^q = r^q u^q$ for $u,\ v \in M$ and $r \in R$.

From time to time, we assume some familiarity with basic tight closure theory in prime characteristic $p > 0$. We use the standard notation N^*_M for the tight closure of the submodule N in the module M. If M is understood, the subscript is omitted, which is frequently the case when $M = R$ and $N = I$ is an ideal. We refer the reader to [Hochster and Huneke 1989; 1990; 1994; 2008; Huneke 1996] for background in this area.

In particular, we assume cognizance of certain facts about test elements, including the notion of a *completely stable* test element. We refer the reader to [Hochster and Huneke 1990, §6 and §8; 1989; 1994, §6; Aberbach et al. 1993, §2] for more information about test elements and to [Aberbach et al. 1993, §3] for a discussion of several basic issues related to the localization problem for tight closure.

Discussion 2.2 (Local cohomology and the action of the Frobenius endomorphism). Our basic reference for local cohomology is [Grothendieck 1967]. Let R be an arbitrary Noetherian ring, let I be an ideal of R and let M be any R-module. The i-th *local cohomology module* $H^i_I(M)$ with support in I may be obtained in several ways. It may be defined as $\varinjlim_t \operatorname{Ext}^i_R(R/I^t,\ M)$: here, any sequence of ideals cofinal with the powers of I may be used instead of the sequence of powers, $\{I^t\}_t$. Alternatively, we may define $C^\bullet(f;\ R)$ to be the complex $0 \to C^0 \to C^1 \to 0$ where $C^0 = R$, $C^1 = R_f$ and the map is the canonical map $R \to R_f$, and then if $\underline{f}$ is a sequence of elements $f_1,\ \dots,\ f_n$ we may define $C^\bullet(\underline{f};\ R)$ to be the tensor product over R of the n complexes $C^\bullet(f_i;\ R)$. Finally, let $C^\bullet(\underline{f};\ M)$ denote $C^\bullet(\underline{f};\ R) \otimes_R M$, which has the form:

$$0 \to M \to \bigoplus_i M_{f_i} \to \bigoplus_{i<j} M_{f_i f_j} \to \cdots \to M_{f_1 \cdots f_n} \to 0.$$

The cohomology of this complex turns out to be $H^\bullet_I(M)$, where $I = (f_1,\ \dots,\ f_n)R$, and actually depends only on the radical of the ideal I.

By the standard theory of local duality (see [Grothendieck 1967, Theorem 6.3]) when (S, m_S, L) is Gorenstein with $\dim(S) = n$ and M is a finitely generated S-module,

$$H^i_m(M) \cong \operatorname{Ext}^{n-i}_S(S,\ M)^\vee$$

as functors of M, where $N^\vee = \mathrm{Hom}_R(N,\ E_S(L))$. Here, $E_S(L)$ is an injective hull of L over S. In particular, if $(R,\ m,\ K)$ is local of Krull dimension d and is a homomorphic image of a Gorenstein local ring S of dimension n, then $\omega_R = \mathrm{Ext}^{n-d}(R,\ S)$ whose Matlis dual over S, and, hence, over R as well, is $H^d_m(R)$. We refer to a finitely generated R-module ω_R as a *canonical module* for R if $\omega_R^\vee = H^d_m(R)$. It is unique up to isomorphism, since its completion is dual to $H^d_m(R)$. Our discussion shows that a canonical module exists if R is a homomorphic image of a Gorenstein ring; in particular, ω_R exists if R is complete. When R is Cohen–Macaulay, one has that

$$H^i_m(M) \cong \mathrm{Ext}^{d-i}_R(M,\ \omega_R)$$

functorially for all finitely generated R-modules M.

When R is a normal local domain, ω_R is isomorphic as an R-module with an ideal of pure height one, that is, with a divisorial ideal.

Finally, suppose that $(R,\ m,\ K) \to (S,\ m_S,\ L)$ is a local homomorphism such that S a module-finite extension of R. Let $\omega = \omega_R$ be a canonical module for R. Then $\mathrm{Hom}_R(S,\ \omega)$ is a canonical module for S. Here, the rings need not be Cohen–Macaulay, nor domains. To see this, note that one can reduce at once to the complete case. We have

$$H^d_{m_S}(S) \cong H^d_m(S) \cong S \otimes H^d_m(R).$$

Then $E_S(L)$ may be identified with $\mathrm{Hom}_R(S,\ E_R(K))$; moreover, on S-modules, the functors $\mathrm{Hom}_R(\_,\ E_R(K))$ and $\mathrm{Hom}_S(\_,\ E_S(L))$ are isomorphic. Hence

$$\mathrm{Hom}_S\big(H^d_{m_S}(S),\ E_S(L)\big) \cong \mathrm{Hom}_R\big(S \otimes_R H^d_m(R),\ E_R(K)\big).$$

By the adjointness of tensor and Hom, this becomes

$$\mathrm{Hom}_R\big(S,\ \mathrm{Hom}_R\big(H^d_m(R),\ E_R(K)\big)\big) \cong \mathrm{Hom}_R(S,\ \omega),$$

as required.

When $M = R$ we have an action of the Frobenius endomorphism on the complex $C^\bullet(\underline{f};\ R)$ induced by the Frobenius endomorphisms of the various rings R_g where g is a product of a subset of $f_1,\ \ldots,\ f_n$, and the action on the cohomology is independent of the choice of f_i.

An alternative point of view is that, quite generally, if $M \to M'$ is any map of R-modules then there is an induced map $H^i_I(M) \to H^i_I(M')$. When S is an R-algebra and I an ideal of R we get a map $H^i_I(R) \to H^i_I(S)$ for all i, and $H^i_I(S)$ may be identified with $H^i_{IS}(S)$. In particular, we may take $S = R$ and let the map $R \to S$ be the Frobenius endomorphism. Since $IS = I^{[p]}$ here, this gives a map $H^i_I(R) \to H^i_{I^{[p]}}(R)$. But since $\mathrm{Rad}\,(I^{[p]}) = \mathrm{Rad}\,(I)$, $H^i_{I^{[p]}}(R) \cong H^i_I(R)$ canonically. The map $H^i_I(R) \to H^i_I(R)$ so obtained again gives the *action of the*

Frobenius endomorphism on $H^i_I(R)$. We shall denote this action by F; note that $F(ru) = r^p F(u)$.

Definition 2.3. When R has prime characteristic $p > 0$, we may construct a noncommutative, associative ring $R\{F\}$ from R which is an R-free left module on the symbols $1,\ F,\ F^2,\ \dots,\ F^e,\ \dots$ by requiring that $Fr = r^p F$ when $r \in R$. We shall say that an R-module M is an $R\{F\}$-*module* if there is given an action $F : M \to M$ such that for all $r \in R$ and for all $u \in M$, $F(ru) = r^p u$. This is equivalent to the condition that M be an $R\{F\}$-module so as to extend the R-module structure on M. We then call an R-submodule N of M F-*stable* if $F(N) \subseteq N$, which is equivalent to requiring that N be an $R\{F\}$-submodule of N. If M is any $R\{F\}$-module and S is an R-algebra then there is an $S\{F\}$-module structure on $S \otimes_R M$ determined by the condition that $F(s \otimes u) = s^p \otimes F(u)$.

In particular, since we have an $R\{F\}$-module structure on $H^i_I(R)$, we may refer to the F-stable submodules of $H^i_I(R)$.

If R is local of Krull dimension d and $x_1, \dots, x_d$ is a system of parameters, then $H^d_m(R)$ may be identified with

$$\varinjlim_t R/(x_1^t,\ \dots,\ x_d^t),$$

where the t-th map in the direct limit system

$$R/(x_1^t,\ \dots,\ x_d^t) \to R/(x_1^{t+1},\ \dots,\ x_d^{t+1})$$

is induced by multiplication by $x_1 \cdots x_d$. If R is Cohen–Macaulay the maps in this direct limit system are injective. When $H^d_m(R)$ is thought of as a direct limit in this way, we write $\langle r;\ x_1^t,\ \dots,\ x_d^t\rangle$ for the image in $H^d_m(R)$ of the element represented by r in $R/(x_1^t,\ \dots,\ x_d^t)$. The action of the Frobenius endomorphism on the highest local cohomology module in this case may be described as sending

$$\langle r;\ x_1^t,\ \dots,\ x_d^t\rangle \mapsto \langle r^p;\ x_1^{pt},\ \dots,\ x_d^{pt}\rangle.$$

Discussion 2.4 (Another point of view for F-stable submodules). Let $(R,\ m,\ K)$ be local of Krull dimension d, where R has characteristic $p > 0$. Consider an F-stable submodule $N \subseteq H^d_m(R)$. Suppose that R is reduced. We have an isomorphism of $(R,\ m,\ K)$ with $(S,\ n,\ L)$ where $S = R^{1/p}$ given by $\Phi : R \to R^{1/p}$, where $\Phi(r) = r^{1/p}$. We have a commutative diagram:

$$\begin{array}{ccc} R & \xrightarrow{F} & R \\ {\scriptstyle =}\downarrow & & \downarrow{\scriptstyle \Phi} \\ R & \xrightarrow[\iota]{} & R^{1/p}\,, \end{array}$$

where $\iota : R \subseteq R^{1/p}$ is the inclusion map. In general, when $\Phi : R \to S$ is any ring isomorphism, for each submodule N of $H^i_I(R)$ there is a corresponding submodule N' of $H^i_{\Phi(I)}(S)$. In fact, if $\Psi = \Phi^{-1}$, and we use ${}^{\Psi}Q$ to indicate restriction of scalars from R-modules to S-modules, then $H^i_{\Phi(I)}(S)$ is canonically isomorphic with ${}^{\Psi}(H^i_I(R))$ and N' is the image of ${}^{\Psi}N$ in $H^i_{\Phi(I)}(S)$. Note that ${}^{\Psi}\_$ is an exact functor.

When $S = R^{1/p}$ and $I = m$, the modules

$$H^i_{\Phi(m)}(S),\ H^i_n(S),\ H^i_{mS}(S), \text{ and } H^i_m(S)$$

may all be identified: the first three may be identified because $\Phi(m)$ and mS both have radical n, and the last two because if $f_1, \ldots, f_h$ generate m their images $g_1, \ldots, g_h$ in S generate mS and the complexes $C^\bullet(f;\ S)$ and $C^\bullet(g;\ S)$ are isomorphic. The condition that N is F-stable is equivalent to the condition that N maps into N' in $H^i_m(S) \cong H^i_{\Phi(m)}(S)$. A very important observation is this:

$(**)$ With notation as just above, if N is F-stable and $J \subseteq R$ kills N (for example, if $J = \mathrm{Ann}_R N$) then $\Phi(J)$ kills the image of N in $H^i_m(R^{1/p})$.

The hypothesis that N is F-stable means that N maps into the corresponding submodule N', and N' is clearly killed by $\Phi(J)$.

Definition 2.5. A local ring $(R,\ m)$ of Krull dimension d is FH-*finite* if, for all i, $0 \le i \le d$, only finitely many R-submodules of $H^i_m(R)$ are F-stable. We shall say that R has *finite FH-length* if for all i, $H^i_m(R)$ has finite length in the category of $R\{F\}$-modules.

Our main focus in studying the properties of being FH-finite and of having finite FH-length is when the local ring R is Cohen–Macaulay. Of course, in this case there is only one nonzero local cohomology module, $H^d_m(R)$. However, we show that every face ring has finite FH-length in Section 5.

Since every $H^i_m(R)$ has DCC even in the category of R-modules, we know that $H^i_m(R)$ has finite length in the category of $R\{F\}$-modules if and only if it has ACC in the category of $R\{F\}$-modules. Of course, it is also equivalent to assert that there is a finite filtration of $H^i_m(R)$ whose factors are simple $R\{F\}$-modules.

Discussion 2.6 (Purity). Recall that a map of R-modules $N \to N'$ is *pure* if for every R-module M the map $N \otimes_R M \to N' \otimes_R M$ is injective. Of course, this implies that $N \to N'$ is injective, and may be thought of as a weakening of the condition that $0 \to N \to N'$ split, that is, that N be a direct summand of N'. If N'/N is finitely presented, $N \to N'$ is pure if and only if it is split. For a treatment of the properties of purity, see, for example, [Hochster and Huneke 1995, Lemma 2.1, p. 49].

An R algebra S is called *pure* if $R \to S$ is pure as a map of R-modules, that is, for every R-module M, the map $M = R \otimes_R M \to S \otimes_R M$ is injective. A Noetherian ring R of characteristic p is called F-*pure* (respectively, F-*split*) if the Frobenius endomorphism $F : R \to R$ is pure (respectively, split). Evidently, an F-split ring is F-pure and an F-pure ring is reduced. If R is an F-finite Noetherian ring, F-pure and F-split are equivalent (since the cokernel of $F : R \to R$ is finitely presented as a module over the left hand copy of R), and the two notions are also equivalent when $(R,\ m,\ K)$ is complete local, for in this case, $R \to S$ is split if and only if $R \otimes_R E \to S \otimes_R E$ is injective, where $E = E_R(K)$. An equivalent condition is that the map obtained by applying $\mathrm{Hom}_R(\_,\ E)$ be surjective, and since $R \cong \mathrm{Hom}_R(E,\ E)$, by the adjointness of tensor and Hom that map can be identified with the maps $\mathrm{Hom}_R(S,\ R) \to \mathrm{Hom}_R(R,\ R) \cong R$.

We say that a local ring R is F-*injective* if F acts injectively on all of the local cohomology modules of R with support in m. This holds if R is F-pure.

When R is reduced, the map $F : R \to R$ may be identified with the algebra inclusion $R \subseteq R^{1/p}$, and so R is F-pure (respectively, F-split) if and only if it is reduced and the map $R \subseteq R^{1/p}$ is pure (respectively, split).

Lemma 2.7. *Let (R, m, K) be a Noetherian local ring of positive prime characteristic p and Krull dimension d.*

(a) *R is FH-finite (respectively, has finite FH-length) if and only if its completion $\widehat{R}$ is FH-finite.*

(b) *Suppose that $(R,\ m) \to (S,\ m_S)$ is a local homomorphism of local rings such that mS is primary to the maximal ideal of S, that is, such that the closed fiber S/mS has Krull dimension* 0. *Suppose either that $R \to S$ is flat (hence, faithfully flat), split over R, or that S is pure over R. If S is FH-finite, then R is FH-finite, and if S has finite FH-length, then R has finite FH-length. More generally, the poset of F-stable submodules of any local cohomology module $H^i_m(R)$ injects in order-preserving fashion into the poset of F-stable submodules of $H^i_{m_S}(S)$.*

(c) *R is F-injective if and only if $\widehat{R}$ is F-injective. R is F-pure if and only if $\widehat{R}$ is F-pure.*

Proof. Completion does not affect either what the local cohomology modules are nor what the action of Frobenius is. Since each element of a local cohomology module over R is killed by a power of m, these are already $\widehat{R}$-modules. Thus, (a) is obvious.

Part (b) follows from the fact that the local cohomology modules of S may be obtained by applying $S \otimes_R \_$ to those of R, and that the action of F is then the one discussed in Definition 2.3 for tensor products, that is, $F(s \otimes u) = s^p \otimes F(u)$.

From this one sees that if N is F-stable in $H^i_m(R)$, then $S \otimes_R N$ is F-stable in $S \otimes_R H^i_m(R) \cong H^i_{m_S}(S)$. Thus, we only need to see that if $N \subseteq N'$ are distinct F-stable submodules of $H^i_m(R)$, then the images of $S \otimes N$ and $S \otimes N'$ are distinct in $S \otimes H^i_m(R)$. It suffices to see this when $R \to S$ is pure: the hypothesis of faithful flatness or that $R \to S$ is split over R implies purity. But N'/N injects into $H^i_m(R)/N$, and $S \otimes_R N'/N$ in turn injects into $S \otimes_R (H^i_m(R)/N)$ by purity, so the image of $u \in N' - N$ is nonzero in

$$S \otimes_R (H^i_m(R)/N) \cong H^i_m(S)/\mathrm{Im}\,(S \otimes N).$$

This shows that $1 \otimes u$ is in the image of $S \otimes N'$ in $H^i_m(S)$ but not in the image of $S \otimes N$.

Part (c), in the case of F-injectivity, follows from the fact that it is equivalent to the injectivity of the action of F on the $H^i_m(R)$, and that neither these modules nor the action of F changes when we complete. In the case of F-purity, we prove that if R is F-pure then so is $\widehat{R}$; the other direction is trivial. Consider an ideal I of the completion, and suppose that there is some element u of the completion such that $u \notin I$ but $u^p \in I^{[p]}$. Choose N such that $u \notin I + m^N \widehat{R}$. We see that we may assume that I is primary to the maximal ideal of $\widehat{R}$, which implies that it is the expansion of its contraction J to R. Then we may choose $v \in R$ such that $v - u \in I = J\widehat{R}$. But then $v \notin J$ but $v^p - u^p \in J^{[p]}\widehat{R}$, and since $u^p \in J^p\widehat{R}$ we have that

$$v^p \in J^{[p]}\widehat{R} \cap R = J^{[p]}$$

and so $v \in J$, a contradiction. Thus, ideals of $\widehat{R}$ are contracted with respect to Frobenius, and, consequently, $\widehat{R}$ is reduced. Then $\widehat{R} \to \widehat{R}^{1/p}$ is cyclically pure, by the contractedness of ideals with respect to Frobenius that we just proved, which shows that it is pure; see [Hochster 1977, Theorem 1.7]. It follows that $\widehat{R}$ is F-pure. □

Discussion 2.8 (Gamma construction). Let K be a field of positive characteristic p with a p-base Λ. Let Γ be a fixed cofinite subset of Λ. For $e \in \mathbb{N}$ we denote by $K_{\Gamma,e}$ the purely inseparable field extension of K that is the result of adjoining p^e-th roots of all elements in Γ to K, which is unique up to unique isomorphism over K.

Now suppose that (R, m) is a complete local ring of positive prime characteristic p and that $K \subseteq R$ is a coefficient field, that is, it maps bijectively onto R/m. Let $x_1, \dots, x_d$ be a system of parameters for R, so that R is module-finite over $A = K[[x_1, \dots, x_d]] \subseteq R$. Let A_Γ denote

$$\bigcup_{e \in \mathbb{N}} K_{\Gamma,e}[[x_1, \dots, x_d]],$$

which is a regular local ring that is faithfully flat and purely inseparable over A. Moreover, the maximal ideal of A expands to that of A_Γ. We shall let R_Γ denote $A_\Gamma \otimes_A R$, which is module-finite over the regular ring A_Γ and which is faithfully flat and purely inseparable over R. The maximal ideal of R expands to the maximal ideal of R_Γ. The residue class field of R_Γ is K_Γ.

We note that R_Γ depends on the choice of coefficient field K for R, and the choice of Γ, but does not depend on the choice of system of parameters $x_1, \ldots, x_d$. We refer the reader to [Hochster and Huneke 1994, §6] for more details. It is of great importance that R_Γ is F-finite, that is, finitely generated as a module over $F(R_\Gamma)$. This implies that it is excellent; see [Kunz 1976].

It is shown in [Hochster and Huneke 1994] that, if R is reduced, then for any sufficiently small choice of the cofinite subset Γ of Λ, R_Γ is reduced. It is also shown in [Hochster and Huneke 1994] that if R is Cohen–Macaulay (respectively, Gorenstein), then R_Γ is Cohen–Macaulay (respectively, Gorenstein).

Lemma 2.9. *Let R be a complete local ring of positive prime characteristic p. Fix a coefficient field K and a p-base Λ for K. Let notation be as in Discussion 2.8.*

(a) *Let W be an Artinian R-module with an $R\{F\}$-module structure such that the action of F is injective. Then for any sufficiently small choice of Γ cofinite in Λ, the action of F on $R_\Gamma \otimes_R W$ is also injective.*

(b) *Suppose that F acts injectively on a given local cohomology module of R. Then F acts injectively on the corresponding local cohomology module of R_Γ for all sufficiently small cofinite Γ. In particular, if R is F-injective, then so is R_Γ.*

(c) *Suppose that R is F-pure. Then for any choice of Γ cofinite in the p-base such that R_Γ is reduced, and, hence, for all sufficiently small cofinite Γ, R_Γ is F-pure.*

Proof. For part (a), let V denote the finite-dimensional K-vector space that is the socle of W. Let $W_\Gamma = R_\Gamma \otimes_R W$. Because the maximal ideal m of R expands to the maximal ideal of R_Γ, and R_Γ is R-flat, the socle in W_Γ may be identified with $V_\Gamma = K_\Gamma \otimes V$. If F has a nonzero kernel on W_Γ then that kernel has nonzero intersection with V_Γ, and that intersection will be some K-subspace of V_Γ. Pick Γ such that the dimension of the kernel is minimum. Then the kernel is a nonzero subspace T of V_Γ whose intersection with $V \subseteq V_\Gamma$ is 0. Choose a basis $v_1, \ldots, v_h$ for V and choose a basis for T as well. Write each basis vector for T as $\sum_{j=1}^{h} a_{ij} v_j$, where the a_{ij} are elements of K_Γ. Thus, the rows of the matrix $\alpha = (a_{ij})$ represent a basis for T. Put the matrix α in reduced row echelon form: the leftmost nonzero entries of the rows are each 1, the columns of these entries are distinct, proceeding from left to right as the index of the row increases, and each column containing the leading 1 of a row has its other entries equal to 0. This matrix is uniquely determined by the subspace T. It has at least one coefficient a not in K (in fact, at least one in every row), since T does not meet V.

Now choose $\Gamma' \subseteq \Gamma$ such that $a \notin K_{\Gamma'}$, which is possible by [Hochster and Huneke 1994, Lemma 6.12]. Then the intersection of T with $V_{\Gamma'}$ must be smaller than T, or else T will have a K_Γ basis consisting of linear combinations of the v_j with coefficients in $K_{\Gamma'}$, and this will give a matrix β over $K_{\Gamma'}$ with the same row space over K_Γ as before. When we put β in row echelon form, it must agree with α, which forces the conclusion that $a \in K_{\Gamma'}$, a contradiction.

Part (b) follows immediately from part (a).

To prove part (c), consider a choice of Γ sufficiently small that R_Γ is reduced. Let E be the injective hull of K over R. For each power m^t of the maximal ideal of R, we have that

$$R_\Gamma/(m^t) \cong K_\Gamma \otimes_K R/m^t.$$

Thus, the injective hull of K_Γ over R_Γ may be identified with $K_\Gamma \otimes_K E$. We are given that the map $E \to E \otimes_R R^{1/p}$ is injective. We want to show that the map

$$E_\Gamma \to E_\Gamma \otimes_{R_\Gamma} (R_\Gamma)^{1/p}$$

is injective. Since the image of a socle generator in E is a socle generator in E_Γ, it is equivalent to show the injectivity of the map $E \to E_\Gamma \otimes_{R_\Gamma} (R_\Gamma)^{1/p}$.

The completion of R_Γ may be thought of as the complete tensor product of K_Γ with R over $K \subseteq R$. However, if one tensors with a module in which every element is killed by a power of the maximal ideal we may substitute the ordinary tensor product for the complete tensor product. Moreover, since R_Γ is reduced, we may identify $(R_\Gamma)^{1/p}$ with $(R^{1/p})_{\Gamma^{1/p}}$: the latter notation means that we are using $K^{1/p}$ as a coefficient field for $R^{1/p}$, that we are using the p-th roots $\Lambda^{1/p}$ of the elements of the p-base Λ (chosen for K) as a p-base for $K^{1/p}$, and that we are using the set $\Gamma^{1/p}$ of p-th roots of elements of Γ as the cofinite subset of $\Lambda^{1/p}$ in the construction of $R^{1/p}_{\Gamma^{1/p}}$. But

$$(K^{1/p})_{\Gamma^{1/p}} \cong (K_\Gamma)^{1/p}.$$

Keeping in mind that every element of E_Γ is killed by a power of the maximal ideal, and that $E_\Gamma \cong K_\Gamma \otimes_K E$, we have that

$$E_\Gamma \otimes_{R_\Gamma} R_\Gamma^{1/p} \cong (K_\Gamma \otimes_K E) \otimes_{K_\Gamma \otimes_K R} (K_\Gamma^{1/p} \otimes_{K^{1/p}} R^{1/p})$$

and so, writing L for K_Γ, we have that

$$E_\Gamma \otimes_{R_\Gamma} R_\Gamma^{1/p} \cong (L \otimes_K E) \otimes_{L \otimes_K R} (L^{1/p} \otimes_{K^{1/p}} R^{1/p}).$$

Now, if K is any ring, L and R are any K-algebras, S is any $(L \otimes_K R)$-algebra (in our case, $S = L^{1/p} \otimes_{K^{1/p}} R^{1/p}$), and E is any R-module, there is an isomorphism

$$(L \otimes_K E) \otimes_{L \otimes_K R} S \cong E \otimes_R S,$$

which maps $(c \otimes u) \otimes s$ to $u \otimes cs$. (The inverse map sends $u \otimes s$ to $(1 \otimes u) \otimes s$. Note that $(c \otimes u) \otimes s = (1 \otimes u) \otimes cs$ in $(L \otimes_K E) \otimes_{L \otimes_K R} S$.)

Applying this fact, we find that $E_\Gamma \otimes_{R_\Gamma} R_\Gamma^{1/p}$ is isomorphic with

$$E \otimes_R (L^{1/p} \otimes_{K^{1/p}} R^{1/p}) \cong E \otimes_R (R^{1/p} \otimes_{K^{1/p}} L^{1/p}) \cong (E \otimes_R R^{1/p}) \otimes_{K^{1/p}} L^{1/p}$$

by the commutativity and associativity of tensor product. But E injects into $E \otimes_R R^{1/p}$ by hypothesis and the latter injects into $(E \otimes_R R^{1/p}) \otimes_{K^{1/p}} L^{1/p}$ simply because $K^{1/p}$ is a field and $L^{1/p}$ is a nonzero free module over it. □

Discussion 2.10 (Finiteness conditions on local cohomology as an F-module and tight closure). We want to make some connections between F-submodules of local cohomology and tight closure theory. Let R be a reduced local ring of characteristic $p > 0$. Let us call a submodule of $H = H_m^d(R)$ *strongly proper* if it is annihilated by a nonzero divisor of R. Assume that R has test elements. The *finitistic* tight closure of 0 in a module M is the union of the submodules 0_N^* as N runs through the finitely generated submodules of M. It is not known, in general, whether the tight closure of 0 in an Artinian module over a complete local ring is the same as the finitistic tight closure: a priori, it might be larger. See [Lyubeznik and Smith 1999; 2001; Elitzur 2003; Stubbs 2008] for results in this direction.

However, if (R, m, K) is an excellent reduced equidimensional local ring with $\dim(R) = d$, the two are the same for $H_m^d(R)$: if u is in the tight closure of 0 and represented by f modulo $I_t = (x_1^t, \dots, x_d^t)$ in $H_m^d(R) = \varinjlim_t R/(x_1^t, \dots, x_d^t)R$, then there exists $c \in R^\circ$ such that for all $q = p^e \gg 1$ the class of cu^q maps to 0 under the map $R/I_{qt} \to H_m^d(R)$, that is, for some k_q, $cu^q(x_1 \cdots x_d)^{k_q} \in I_{qt+k+q}$ for all $q \gg 1$. But $I_{qt+k_q} :_R (x_1 \cdots x_d)^{k_q} \subseteq I_{qt}^*$ (since R is excellent, reduced, and local, it has a completely stable test element, and this reduces to the complete case, which follows from [Hochster and Huneke 1990, Theorem 7.15]), and so if d is a test element for R we have that $cdu^q \in I_{qt} = I_t^{[q]}$ for all $q \gg 0$, and so the class of u modulo I_t is in the tight closure of 0 in R/I_t and hence in the image of R/I_t in $H_m^d(R)$, as required.

Let us note that the finitistic tight closure of 0 in H is an F-stable strongly proper submodule of $H_m^d(R)$, as shown in [Lyubeznik and Smith 2001, Proposition 4.2]. The reason is that it is immediate from the definition of tight closure that if $u \in 0_N^*$ then $u^q \in 0_{N^{[q]}}^*$, where $q = p^e$ and $N^{[q]}$ denotes an image of $F^e(N)$ in $F^e(H)$ for some ambient module $H \supseteq N$. In particular, $u^q \in 0_{F^e(N)}^*$. Moreover, if c is a test element for the reduced ring R, then $c \in R^{\circ}$ and so c is a nonzero divisor, and c kills 0^*_N for every finitely generated R-module N.

Conversely, any strongly proper F-stable submodule $N \subseteq H$ is in the tight closure of 0. If c is a nonzero divisor that kills N and $u \in N$, then $cu^q = 0$ for all q: when we identify $F^e(H)$ with H, u^q is identified with $F^e(u)$.

What are the strongly proper submodules of $H = H^d_m(R)$? If (R, m, K) is complete with $E = E_R(K)$ the injective hull of the residue class field and canonical module $\omega := \mathrm{Hom}_R(H^d_m(R), E)$, then submodules of H correspond to the proper homomorphic images of ω: the inclusion $N \subseteq H$ is dual under $\mathrm{Hom}_R(\_, E)$ to a surjection $\omega \twoheadrightarrow \mathrm{Hom}_R(N, E)$. If R is a domain, for every proper $N \subseteq H$ we have that $\omega \twoheadrightarrow \mathrm{Hom}_R(N, E)$ is a proper surjection, and therefore is killed by a nonzero divisor. Therefore, we have the following results of K. E. Smith [1997a]; see Proposition 2.5 on page 169 and the remark on page 170 immediately following the proof of Proposition 2.5. See also [Smith 2003, Theorem 3.1.4], where the restricted generality is not needed.

Proposition 2.11 (K. E. Smith). *If R is a reduced equidimensional excellent local ring of characteristic p, then the tight closure 0^* of 0 in $H^d_m(R)$ (which is the same in the finitistic and ordinary senses) is the largest strongly proper F-stable submodule of $H^d_m(R)$. If R is a complete local domain, it is the largest proper F-stable submodule of $H^d_m(R)$.*

A Noetherian local ring is called F-*rational* if some (equivalently, every) ideal generated by parameters is tightly closed. An excellent F-rational local ring is a Cohen–Macaulay normal domain. The completion of an excellent F-rational local ring is again F-rational. See [Hochster and Huneke 1994, Proposition 6.27a]. From this and Discussion 2.10 we have at once:

Proposition 2.12 [Smith 1997a, Theorem 2.6, p. 170]. *Let R be an excellent Cohen–Macaulay local ring of characteristic p and Krull dimension d. Then R is F-rational if and only if $H^d_m(R)$ is a simple $R\{F\}$-module.*

Example 2.13. The ring obtained by killing the size t minors of a matrix of indeterminates in the polynomial ring in those indeterminates is an example of an F-rational ring. In fact this ring is weakly F-regular, that is, every ideal is tightly closed. The local ring at the origin is therefore FH-finite by the above result: the unique nonvanishing local cohomology module is $R\{F\}$-simple.

Proposition 2.14 [Smith 1997b, 4.17.1]. *Let R be a Cohen–Macaulay local domain and suppose that there is an m-primary ideal $\mathfrak{A}$ such that $\mathfrak{A}I^* \subseteq I$ for every ideal I of R generated by part of a system of parameters. Then R has finite FH-length.*

Proof. Let d be the dimension of R. By the discussion above, every proper F-stable submodule of $H = H^d_m(R)$ is contained in 0^*_H. But the discussion above shows that 0^*_H is a union of submodules of the form I^*/I where I is a parameter ideal, and so 0^*_H is killed by $\mathfrak{A}$, and has finite length even as an R-module. $\square$

See Theorem 4.22, which gives a stronger conclusion when the residue class field is perfect and R is F-injective.

Example 2.15. Let $R = K[\![X, Y, Z]\!]/(X^3 + Y^3 + Z^3)$, where K is a field of positive characteristic different from 3. Then R is a Gorenstein domain, and the tight closure of 0 in $H^2_m(R)$ is just the socle, a copy of K: the tight closure of every parameter ideal is known to contain just one additional element, a representative of the generator of the socle modulo the parameter ideal. Evidently R is FH-finite. It is known (see, for example, [Hochster and Roberts 1976, Proposition 5.21c, p. 157]) that R is F-injective if and only if the characteristic of K is congruent to 1 modulo 3. If the characteristic is congruent to 2 modulo 3, $R[\![t]\!]$ does not have FH-finite length by Theorem 4.16 of Section 4.

Example 2.16. We construct a complete local F-injective domain of dimension one (hence, it is Cohen–Macaulay) that is not FH-finite. Note that Theorem 4.22 implies that there are no such examples when the residue class field of the ring is perfect.

Let K be an infinite field of characteristic $p > 0$ (it will be necessary that K not be perfect) and let L be a finite algebraic extension field of K such that

(1) $[L : L^p[K]] > 2$ (all one needs is that the dimension of $L/L^p[K]$ over K is at least 2) and

(2) L does not contain any element of $K^{1/p} - K$ (equivalently, $L^p \cap K = K^p$).

Then the quotient L/K has infinitely many $K\{F\}$-submodules but F acts injectively on it. Moreover, if $R = K + xL[\![x]\!] \subseteq L[\![x]\!]$ then R is a complete local one-dimensional domain that is F-injective but not FH-finite.

The conditions in (1) and (2) above may be satisfied as follows: if k is infinite perfect of characteristic $p > 2$, $K = k(u, v)$, where u and v are indeterminates, and

$$L = K[y]/(y^{2p} + uy^p - v),$$

then (1) and (2) above are satisfied.

Proof. The image of L under F is L^p — this need not be a K-vector space, but $L_1 = L^p[K]$ is a K-vector space containing the image of F. All of the K-vector subspaces of L strictly between L_1 and L are F-stable, and there are infinitely many. The statement that F acts injectively on L/K is exactly the statement that $L^p \cap K = K^p$.

With R as above, the exact sequence

$$0 \to R \to L[\![x]\!] \to L/K \to 0$$

yields a long exact sequence for local cohomology:

$$0 \to L/K \to H^1_m(R) \to H^1_m(L[\![x]\!]) \to 0.$$

Since L/K embeds in $H^1_m(R)$ as an F-stable submodule and m kills it, its R-module structure is given by its K-vector space structure. Moreover, since F is injective on $H^1_m(L[\![x]\!])$, F-injectivity holds for R if and only if it holds for L/K.

This establishes all assertions except that the given example satisfies (1) and (2). Note that the expression $y^{2p} + uy^p - v$ is irreducible over $k[y, u, v]$ (the quotient is $k[y, u]$). Suppose that L contains an element w of $K^{1/p}$ not in K. Then

$$[K[w] : K] = p,$$

and so $[L : K[w]] = 2$. It follows that y satisfies a monic quadratic equation over $K[w]$.

But if we enlarge $K[w]$ to all of $K^{1/p}$ we know the quadratic equation that y satisfies:

$$y^2 + u^{1/p}y - v^{1/p} = 0,$$

which is clearly irreducible over $K^{1/p} = k(u^{1/p}, v^{1/p})$. This quadratic is unique, so we must have $u^{1/p}$, $v^{1/p}$ are both in $K[w]$, a contradiction, since adjoining both produces an extension of K of degree p^2.

It remains to determine $[L : L^p[K]] = [K[y] : K[y^p]]$. Since y^p satisfies an irreducible quadratic equation over K, $[K[y^p] : K] = 2$, and so $[K[y] : K[y^p]] = \frac{2p}{2} = p > 2$, by assumption. □

3. Annihilators of F-stable submodules and the FH-finite property for F-pure Gorenstein local rings

In this section we shall prove a theorem of independent interest which can be used to establish that certain families of radical ideals in excellent local rings are finite. As an immediate corollary to it, we obtain that if R is local, excellent, then any family of radical ideals closed under sum, intersection and primary decomposition is finite. Another consequence is that if R is a local ring of positive prime characteristic p and M is an Artinian $R\{F\}$-module such that F acts injectively on M, then the set of annihilator ideals in R of F-stable submodules of M is a finite set of radical ideals closed under primary decomposition (R. Y. Sharp [2007, Theorem 3.10 and Corollary 3.11] proved this result independently). In fact, it consists of a finite set of prime ideals and their intersections. From this we deduce that an F-pure Gorenstein local ring is FH-finite. We say that a family of radical ideals of a Noetherian ring is *closed under primary decomposition* if for every ideal I in the family and every minimal prime P of I, the ideal P is also in the family.

The following result is the main theorem of this section.

Theorem 3.1. *Let M be a Noetherian module over an excellent local ring $(R,\ m)$. Then there is no family $\{N_\lambda\}_{\lambda\in\Lambda}$ of submodules of M satisfying all four of the conditions below*:

(1) *The family is closed under finite sum.*

(2) *The family is closed under finite intersection.*

(3) *All of the ideals $\operatorname{Ann}_R(M/N)$ for N in the family are radical.*

(4) *There exist infinitely many modules in the family such that if N, N' are any two of them, the minimal primes of N are mutually incomparable with the minimal primes of N'.*

Hence, if a family of submodules $\{N_\lambda\}_{\lambda\in\Lambda}$ of M satisfies conditions (1), (2), and (3) above and the set

$$\{\operatorname{Ann}_R(M/N_\lambda) : \lambda \in \Lambda\}$$

is closed under primary decomposition, then this set of annihilators is finite.

Proof. Assume that one has a counterexample. We use both induction on the dimension of R and Noetherian induction on M. Take a counterexample in which the ring has minimum dimension. One can pass to the completion. Radical ideals stay radical, and (4) is preserved (although there may be more minimal primes). The key point is that if P, Q are incomparable primes of R, and $\widehat{R}$ is the completion of R, then $P\widehat{R}$, $Q\widehat{R}$ are radical with no minimal prime in common. A common minimal prime would contain $(P+Q)\widehat{R}$, a contradiction, since the minimal primes of $P\widehat{R}$ lie over P. This, applied together with the fact that the minimal primes of $\widehat{M/N}$ are minimal over $P\widehat{R}$ for some minimal prime P of M/N, enables us to pass to the completion.

Take infinitely many N_i as in condition (4). Let M_0 be maximal among submodules of M contained in infinitely many of the N_i. Then the set of modules in the family containing M_0 gives a new counterexample, and we may pass to all quotients by M_0 (M_0 need not be in the family to make this reduction). Thus, by Noetherian induction on M we may assume that every infinite subset of the N_i has intersection 0.

Consider the set of all primes of R in the support of an M/N_i. If $Q \neq m$ is in the support of infinitely many we get a new counterexample over R_Q. The $(N_i)_Q$ continue to have the property that no two have a minimal prime in common (in particular, they are distinct). Since R had minimum dimension for a counterexample, we can conclude that every Q other than m is in the support of just finitely many M/N_i.

Choose h as large as possible such that there are infinitely many primes of height h occurring among the minimal primes of an M/N_i. Then there are only finitely many primes of height $h+1$ or more occurring as a minimal prime of an M/N_i, and, by the preceding paragraph, each one occurs for only finitely many N_i. Delete sufficiently many N_i from the sequence so that no prime of height bigger than h occurs among the minimal primes of the M/N_i.

Let $D_1(i) = \bigcap_{s=1}^{i} N_s$. By Chevalley's lemma, $D_1(i_1)$ is contained in m^2M for i_1 sufficiently large; fix such a value of i_1. Let $W_1 = D_1(i_1)$. Let $D_2(j) = \bigcap_{s=i_1+1}^{j} N_s$. Then $D_2(j)$ is contained in m^2M for sufficiently large j; fix such a value i_2. Recursively, we can choose a strictly increasing sequence of integers $\{i_t\}_t$ with $i_0 = 0$ such that every

$$W_t = \bigcap_{s=i_{t-1}+1}^{i_t} N_s$$

is contained in m^2M. In this way we can construct a sequence $W_1, W_2, W_3, \ldots$ with the same properties as the N_i but such that all of them are in m^2M. Now $W_1 + W_2 + \cdots + W_t$ stabilizes for $t \gg 0$, since M is Noetherian, and the stable value W is contained in m^2M. There cannot be any prime other than m in the support of M/W, or it will be in the support of M/W_j for all j and this will put it in the support of infinitely many of the original N_j. Hence, the annihilator of M/W is an m-primary ideal, and, by construction, it is contained in m^2 and, therefore, not radical, a contradiction.

It remains only to prove the final statement. If the set of annihilators were infinite it would contain infinitely many prime ideals. Since there are only finitely many possibilities for the height, infinitely many of them would be prime ideals of the same height. The modules in the family having these primes as annihilator satisfy condition (4), a contradiction. □

By applying Theorem 3.1 to the family of ideals of R, we immediately have:

Corollary 3.2. *A family of radical ideals in an excellent local ring closed under sum, intersection, and primary decomposition is finite.*

Discussion 3.3. For any local ring (R, m, K) we let E denote an injective hull of the residue class field, and we write $\_^{\vee}$ for $\mathrm{Hom}_R(\_, E)$. Note that E is also a choice for $E_{\widehat{R}}(K)$, and that its submodules over R are the same as its submodules over $\widehat{R}$. E is determined up to nonunique isomorphism: the obvious map $\widehat{R} \to \mathrm{Hom}_R(E, E)$ is an isomorphism, and so every automorphism of E is given by multiplication by a unit of $\widehat{R}$.

Now suppose that R is complete. Then $R^{\vee} \cong E$ and $E^{\vee} \cong R$, by Matlis duality. Matlis duality gives an antiequivalence between modules with ACC and modules with DCC: in both cases, the functor used is a restriction of $\_^{\vee}$. In particular, the natural map $N \to N^{\vee\vee}$ is an isomorphism whenever N has DCC or ACC. Note that there is an order-reversing bijection between ideals I of R and submodules N of E given by $I \mapsto \mathrm{Ann}_E I$ and $N \mapsto \mathrm{Ann}_R N$: this is a consequence of the fact that the inclusion $N \hookrightarrow E$ is dual to a surjection $N^\vee \twoheadleftarrow R$ so that $N^\vee \cong R/I$ for a unique ideal I of R, and since $I = \mathrm{Ann}_R N^\vee$, we have that $I = \mathrm{Ann}_R N$. Note that $N \cong N^{\vee\vee} \cong (R/I)^\vee \cong \mathrm{Hom}_R(R/I,\ E) \cong \mathrm{Ann}_E I$. When R is regular or even if R is Gorenstein, $E \cong H^d_m(R)$.

When R is complete local and W is Artinian, Matlis duality provides a bijection between the submodules of W and the surjections from $W^\vee = M$, and each such surjection is determined by its kernel N. This gives an order-reversing bijection between the submodules of W and the submodules of M. Specifically, $V \subseteq W$ corresponds to $\ker(W^\vee \twoheadrightarrow V^\vee) = \ker(M \twoheadrightarrow V^\vee)$, and $N \subseteq M$ corresponds to $\ker(M^\vee \twoheadrightarrow (M/N)^\vee)$. Here, $M^\vee = (W^\vee)^\vee \cong W$ canonically. This bijection converts sums to intersections and intersections to sums; the point is that the sum (intersection) of a family of submodules is the smallest (respectively, largest) submodule containing (respectively, contained in) all of them, and the result follows from the fact that the correspondence is an order antiisomorphism. Since the annihilator of a module kills the annihilator of its dual, Matlis duality preserves annihilators: it is obvious that the annihilator of a module kills its dual, and we have that each of the two modules is the dual of the other. In particular, under the order-reversing bijection between submodules V of W and submodules N of M, we have that

$$\mathrm{Ann}_R V = \mathrm{Ann}_R(M/N).$$

Discussion 3.4. Let $(R,\ m,\ K) \to (S,\ n,\ L)$ be local, and suppose that mS is n-primary and that L is finite algebraic over K: both these conditions hold if S is module-finite over R. Let $E = E_R(K)$ and $E_S(L)$ denote choices of injective hulls for K over R and for L over S, respectively. The functor $\mathrm{Hom}_R(\_,\ E)$ from S-modules to S-modules may be identified with

$$\mathrm{Hom}_R(\_ \otimes_S S,\ E) \cong \mathrm{Hom}_S(\_,\ \mathrm{Hom}_R(S,\ E)),$$

which shows that $\mathrm{Hom}_R(S,\ E)$ is injective as an S-module. Every element is killed by a power of the maximal ideal of S, since mS is primary to n, and the value of the functor on $L = S/n$ is $\mathrm{Hom}_R(L,\ E) \cong \mathrm{Hom}_R(L,\ K)$ since the image of L is killed by m. But this is L as an L-module. Thus, $E_S(L) \cong \mathrm{Hom}_R(S,\ E)$, and the functor $\mathrm{Hom}_R(\_,\ E)$, on S-modules, is isomorphic with the functor $\mathrm{Hom}_S(\_,\ E_S(L))$.

The following proposition can be seen as a consequence of the more general Theorem 3.6 and its corollary in [Sharp 2007]. However, its proof is not very difficult and we include it here for the convenience of the reader.

Proposition 3.5. *Let R be a ring of characteristic p and let W be an $R\{F\}$-module.*

(a) *If F acts injectively on W, the annihilator in R of every F-stable submodule is radical.*

(b) *If I is the annihilator of an F-stable submodule V of W, then $I :_R f$ is also the annihilator of an F-stable submodule, namely, fV. Hence, if I is radical with minimal primes $P_1, \ldots, P_k$ then every P_j (and every finite intersection of a subset of the P_j) is the annihilator of an F-stable submodule of M.*

Proof. If V is F-stable and $u \in R$ is such that $u^p \in \mathrm{Ann}_R V$, then

$$F(uV) = u^p F(V) \subseteq u^p V = 0.$$

Since F is injective on W, $uV = 0$. This proves part (a). For part (b), note that fV is F-stable since $F(fV) = f^p F(V) \subseteq f^p V \subseteq fV$ and $u(fV) = 0$ if and only if $(uf)V = 0$ if and only if $uf \in \mathrm{Ann}_R V = I$ if and only if $u \in I :_R f$. For the final statement, choose f in all of the P_j except P_i, and note that $I :_R f = P_i$. More generally, given a subset of the P_j, choose f in all of the minimal primes except those in the specified subset. □

Now we are in position to state an important consequence of our main result in this section. This result has also been obtained by Sharp [2007, Theorem 3.10 and, more precisely, Corollary 3.11]. Our proof is via Theorem 3.1 so we will include here.

Theorem 3.6 (R. Y. Sharp). *Let R be a local ring of positive prime characteristic p and let W be an Artinian $R\{F\}$-module. Suppose that F acts injectively on W. Then*

$$\{\mathrm{Ann}_R V : V \text{ is an F-stable submodule of } W\}$$

is a finite set of radical ideals, and consists of all intersections of the finitely many prime ideals in it.

Proof. By Proposition 3.5, it suffices to prove that family of annihilators is finite. We may replace R by its completion without changing M or the action of F on M. The set of F-stable submodules is unaffected. The annihilator of each such submodule in R is obtained from its annihilator in $\widehat{R}$ by intersection with R. Therefore, it suffices to prove the result when R is complete, and we henceforth assume that R is complete.

As in Discussion 3.3 fix an injective hull E of K and let $\_^\vee = \mathrm{Hom}_R(\_, E)$. Matlis duality gives a bijection of submodules of W with submodules of $M = W^\vee$. The F-stable submodules of W are obviously closed under sum and intersection. Therefore, the submodules N of M that correspond to them are also closed under sum and intersection. We refer to these as the *costable* submodules of M. The annihilators of the modules M/N, where N runs through the costable submodules of M, are the same as the annihilators of the F-stable submodules of W. We may now apply the final statement of Theorem 3.1. □

It is now easy to prove the second main result of this section. Recall that a local ring (R, m, K) of Krull dimension d is *quasi-Gorenstein* if $H^d_m(R)$ is an injective hull of K; equivalently, this means that R is a canonical module for R in the sense that its Matlis dual is $H^d_m(R)$.

Theorem 3.7. *Let (R, m, K) be a local ring of prime characteristic $p > 0$. If R is F-pure and quasi-Gorenstein, then $H^d_m(R)$ has only finitely many F-stable submodules. Hence, if R is F-pure and Gorenstein, then R is FH-finite.*

Proof. There is no loss of generality in replacing R by its completion. We apply Theorem 3.6 to the action of F on $H^d_m(R) = E$. The point is that because $E^\vee = R$, the dual of the F-stable module V has the form R/I, where I is the annihilator of V, and so V is uniquely determined by its annihilator.

Since there are only finitely many possible annihilators, there are only finitely many F-stable submodules of $H^d_m(R)$. □

In relation to this theorem the following observation is interesting.[1] Since we do not know a reference for it in the case of F-rational rings, we will also sketch a proof of it. For the definition of tight closure in the case of modules and its variant of finitistic tight closure, we refer the reader to [Hochster and Huneke 1990]. The result on F-rational rings is not actually related to rest of the paper, but we include it for the sake of completness.

Remark 3.8. Let (R, m, K) be a local ring of prime characteristic $p > 0$. Assume that R is quasi-Gorenstein. Then R is F-pure if and only if R is F-injective. If R is excellent as well, then R is weakly F-regular if and only if R is F-rational.

Proof. If R is F-pure (respectively, weakly F-regular), then it is immediate that R is F-injective (respectively F-rational).

Assume that R is F-injective. To prove that R is F-pure, one can proceed exactly as in [Fedder 1983, Lemma 3.3].

Now assume that R is F-rational. We can assume that R is complete. Let $E = E_R(K)$. To prove that R is weakly F-regular we need to show that the finitistic tight closure of zero in E equals zero, that is,

$$0 = 0_E^{*,fg}.$$

But E is isomorphic to $H = H^d_m(R)$, and $0_H^{*,fg} = 0_H^* = 0$ since R is F-rational. This finishes the sketch of the proof. □

[1] We thank Karl Schwede for suggesting that we incorporate this remark in our paper.

4. F-purity, finite length, and antinilpotent modules

In this section we prove that certain quotients by annihilators are F-split, and we study the family of F-stable submodules of the highest local cohomology both in the F-pure Cohen–Macaulay case, and under less restrictive hypotheses. We do not know an example of an F-injective ring which does not have finite FH-length, but we have not been able to prove that one has finite FH-length even in the F-split Cohen–Macaulay case. We also give various characterizations of when a local ring has finite FH-length.

Theorem 4.1. *Let (R, m, K) be a local ring of prime characteristic $p > 0$ of Krull dimension d. Suppose that R is F-split. Let N be an F-stable submodule of $H^d_m(R)$, and let $J = \operatorname{Ann}_R N$. Then R/J is F-split. In fact, let $\Phi : R \to R^{1/p}$ such that $\Phi(r) = r^{1/p}$. If $T : R^{1/p} \to R$ is any R-linear splitting, then for every such annihilator ideal J, $T(\Phi(J)) \subseteq J$, and so T induces a splitting*

$$(R/J)^{1/p} \cong R^{1/p}/\Phi(J) \twoheadrightarrow R/J.$$

Proof. Let $H = H^d_m(R)$. When we apply $\_\otimes_R H$ to $\iota : R \subseteq R^{1/p}$ and to $T : R^{1/p} \to R$, we get maps

$$\alpha : H \to R^{1/p} \otimes H \cong H^d_m(R^{1/p}) \cong H^d_n(R^{1/p}),$$

where n is the maximal ideal of $R^{1/p}$, and also a map $\tilde{T} : H^d_m(R^{1/p}) \to H$. Therefore α equals $\iota \otimes_R \mathrm{id}_H$ and $\tilde{T}$ equals $T \otimes_R \mathrm{id}_H$.

Let $u \in H$ and $s \in R^{1/p}$. We have

$$\tilde{T}(s\alpha(u)) = (T \otimes_R \mathrm{id}_H)(s(1 \otimes u)) = (T \otimes_R \mathrm{id}_H)(s \otimes u) = T(s) \otimes u,$$

and since $T(s) \in R$, this is simply $T(s)u$. To show that $T(J^{1/p}) \subseteq J$, we need to prove that $T(J^{1/p})$ kills N in $H^d_m(R)$. Take $u \in N$ and $j \in J$. Then

$$T(j^{1/p})u = \tilde{T}(j^{1/p})(\alpha(u)),$$

taking $s = j^{1/p}$. But now, since N is an F-stable submodule of $H^d_m(R)$, α maps N into the corresponding submodule N' of $H^d_m(R^{1/p})$, whose annihilator in $R^{1/p}$ is $\Phi(J)$. We therefore have that $j^{1/p}$ kills $\alpha(N) \subseteq N'$. This is the displayed fact $(**)$ in Discussion 2.4. Therefore, $T(j^{1/p})u = 0$, and $T(j^{1/p}) \in J$. □

We now want to discuss the condition of having finite FH-length.

Proposition 4.2. *Let R be a characteristic p local ring with nilradical J, and let M be an Artinian $R\{F\}$-module. Then M has finite $R\{F\}$-length if and only if JM has finite length as an R-module and M/JM has finite $(R/J)\{F\}$-length.*

Proof. JM has a finite filtration by submodules $J^t M$, and F acts trivially on each factor. □

Because of Proposition 4.2, we shall mostly limit our discussion of finite FH-length to the case where R is reduced.

Discussion 4.3. Let (R, m, K) be a local ring of characteristic $p > 0$ and let M be an Artinian $R\{F\}$-module. We note that M is also an Artinian $\widehat{R}\{F\}$-module with the same action of F, and we henceforth assume that R is complete in this discussion. We shall also assume that R is reduced. (In the excellent case, completing will not affect whether the ring is reduced.) Fix an injective hull $E = E_R(K)$ for the residue field and let $\_^\vee$ denote the functor $\mathrm{Hom}_R(\_, E)$.

Lemma 4.4. *Let (R, m, K) be a local ring of characteristic p. Then we may construct S local and faithfully flat over R with maximal ideal mS such that S is complete and faithfully flat over R, such that S is Cohen–Macaulay if R is Cohen–Macaulay, such that S is F-injective if R is, and such that S is F-split if R is F-pure. For every i, the poset of F-stable modules of $H^i_m(R)$ injects by a strictly order-preserving map into the poset of F-stable modules of $H^i_{mS}(S) = H^i_m(S)$. Hence, R is FH-finite* (*respectively, has finite FH-length*) *if S is FH-finite* (*respectively, has FH-finite length*).

Proof. By Lemma 2.7, we may first replace R by its completion $\widehat{R}$. We then choose a coefficient field K and a p-base Λ for K, and replace $\widehat{R}$ by $\widehat{R}_\Gamma$ for Γ a sufficiently small cofinite subset of Λ, using Lemma 2.9. Finally, we replace $\widehat{R}_\Gamma$ by its completion. The map on posets is induced by applying $S \otimes_R \_$. □

Discussion 4.5 (Reductions in the Cohen–Macaulay F-pure case). Consider the following three hypotheses on a local ring R of prime characteristic $p > 0$:

(1) R is F-injective and Cohen–Macaulay, with perfect residue class field.

(2) R is F-pure.

(3) R is F-pure and Cohen–Macaulay.

In all three cases we do not know, for example, whether the top local cohomology module has only finitely many F-stable submodules. The point we want to make is that Lemma 4.4 permits us to reduce each of these questions to the case where R is complete and F-finite. Moreover, because F-pure then implies F-split, in cases (2) and (3) the hypothesis that R be F-pure may be replaced by the hypothesis that R be F-split.

If $V \subseteq V' \subseteq W$ are F-stable R-submodules of the $R\{F\}$-module W, we refer to V'/V as a *subquotient* of W. Also, we let $\widetilde{F}^k(V)$ denote the R-span of $F^k(V)$ in W. We next note:

Proposition 4.6. *Let R be a ring of positive prime characteristic p and let W be an $R\{F\}$-module. The following conditions on W are equivalent*:

(a) *If V is an F-stable submodule of W, then F acts injectively on W/V.*

(b) *F acts injectively on every subquotient of W.*

(c) *The action of F on any subquotient of W is not nilpotent.*

(d) *The action of F on any nonzero subquotient of W is not zero.*

(e) *If $V \subseteq V'$ are F-stable submodules of W such that $F^k(V') \subseteq V$ for some $k \geq 1$ then $V' = V$.*

Proof. (a) and (b) are equivalent because a subquotient V'/V is an $R\{F\}$-submodule of W/V. If the action of F on a subquotient $V''/V \subseteq W/V$ is not injective, the kernel has the form V'/V where V' is F-stable. Hence (b) and (d) are equivalent. If F is nilpotent on V''/V it has a nonzero kernel of the form V'/V. This shows that (c) is also equivalent. (e) follows because F^k kills V'/V if and only if $F^k(V') \subseteq V$. $\square$

Definition 4.7. With R and W as in Proposition 4.6 we shall say that W is *antinilpotent* if it satisfies the equivalent conditions (a)–(e).

Theorem 4.8 [Lyubeznik 1997, Theorem 4.7, p. 108]. *Let R be a local ring of prime characteristic $p > 0$ and let W be an Artinian R-module that has an action F of Frobenius on it. Then W has a finite filtration*

$$0 = L_0 \subseteq N_0 \subseteq L_1 \subseteq N_1 \subseteq \cdots \subseteq L_s \subseteq N_s = M$$

by F-stable submodules such that every N_j/L_j is nilpotent, that is, killed by a single power of F, while every L_j/N_{j-1} is simple in the category of $R\{F\}$-modules, with a nonzero action of F on it. The integer s and the isomorphism classes of the modules L_j/N_{j-1} are invariants of W.

Note that the assumption that the action of F on a simple module $L \neq 0$ is nonzero is equivalent to the assertion that the action of F is injective, for if F has a nontrivial kernel it is an $R\{F\}$-submodule and so must be all of L.

Proposition 4.9. *Let the hypothesis be as in Theorem 4.8 and let W have a filtration as in that theorem. Then*:

(a) *W has finite length as an $R\{F\}$-module if and only if each of the factors N_j/L_j has finite length in the category of R-modules.*

(b) *W is antinilpotent if and only if in some* (*equivalently, every filtration*) *as in Theorem 4.8*, *the nilpotent factors are all zero.*

Proof. (a) This comes down to the assertion that if a power of F kills W then W has finite length in the category of $R\{F\}$-modules if and only if it has finite length in the category of R-modules. But W has a finite filtration with factors $\widetilde{F}^j(W)/\widetilde{F}^{j+1}(W)$ on which F acts trivially, and the result is obvious when F acts trivially.

(b) If W is antinilpotent, then the nilpotent factors in any finite filtration must be 0, since they are subquotients of W. Now suppose that W has a finite filtration by simple $R\{F\}$-modules on which F acts injectively. Suppose that we have $0 \subseteq V \subseteq V' \subseteq W$ such that F acts trivially on V'/V. This filtration and the filtration by simple $R\{F\}$-modules on which F acts injectively have a common refinement in the category of $R\{F\}$-modules. This implies that V'/V has a finite filtration in which all the factors are simple $R\{F\}$-modules on which F acts injectively. Since F must be zero on the smallest nonzero submodule in the filtration, this is a contradiction. □

Corollary 4.10. *Let R be a local ring of positive prime characteristic p. If M is an Artinian R-module that is antinilpotent as an $R\{F\}$-module, then so is every submodule, quotient module, and every subquotient of M in the category of $R\{F\}$-modules.*

Proof. It suffices to show this for submodules and quotients. But if N is any $R\{F\}$-submodule, the filtration $0 \subseteq N \subseteq M$ has a common refinement with the filtration of M with factors that are simple $R\{F\}$-modules with nontrivial F-action. □

We also note the following, which is part of [Lyubeznik 1997, Theorem 4.2].

Theorem 4.11 (Lyubeznik). *Let $T \to R$ be a surjective map from a complete regular local ring T of prime characteristic $p > 0$ onto a local ring (R, m, K). Then there exists a contravariant additive functor $\mathcal{H}_{T,R}$ from the category of $R\{F\}$-modules that are Artinian over R to the category of F_T-finite modules in the sense of Lyubeznik such that:*

(a) *$\mathcal{H}_{T,R}$ is exact.*

(b) *$\mathcal{H}_{T,R}(M) = 0$ if and only if the action of some power of F on M is zero.*

Theorem 4.12. *Let R be a local ring of positive prime characteristic p. Let M be an Artinian R-module that is antinilpotent as an $R\{F\}$-module. Then M has only finitely many $R\{F\}$-submodules.*

Proof. We may replace R by its completion and write R as T/J where T is a complete regular local ring of characterisitc p. By Theorem 4.11 above, we have a contravariant exact functor $\mathcal{H}_{T,R}$ on $R\{F\}$-modules Artinian over R to F_T-finite modules in the sense of Lyubeznik. This functor is faithfully exact when restricted to antinilpotent modules, and all subquotients of M are antinilpotent. It follows that if M_1 and M_2 are distinct $R\{F\}$-submodules of M, then N_1 and N_2 are distinct, where

$$N_i = \ker\big(\mathcal{H}(M) \to \mathcal{H}(M_i)\big)$$

for $i = 1, 2$. By the main result of [Hochster 2008], an F_T-finite module in the sense of Lyubeznik over a regular local ring T has only finitely many F_T-submodules, from which the desired result now follows at once. □

Discussion 4.13 (Local cohomology after adjunction of a formal indeterminate). Let R be any ring and M an R-module. Let x be a formal power series indeterminate over R. We shall denote by $M\langle x^{-1}\rangle$ the $R[\![x]\!]$-module

$$M \otimes_{\mathbb{Z}} (\mathbb{Z}[x,\, x^{-1}]/\mathbb{Z}[x]).$$

This is evidently an $R[x]$-module, and since every element is killed by a power of x, it is also a module over $R[\![x]\!]$. Note that if R is an A-algebra, this module may also be described as $M \otimes_A (A[x,\, x^{-1}]/A[x])$. In particular,

$$M\langle x^{-1}\rangle \cong M \otimes_R (R[x,\, x^{-1}]/R[x]),$$

and if R contains a field K, $M\langle x^{-1}\rangle \cong M \otimes_K (K[x,\, x^{-1}]/K[x])$. We have that $M \cong M \otimes x^{-n}$ for all $n \geq 1$ via the map $u \mapsto u \otimes x^{-n}$, and we write Mx^{-n} for $M \otimes x^{-n}$. As an R-module, $M\langle x^{-1}\rangle \cong \bigoplus_{n=1}^{\infty} Mx^{-n}$, a countable direct sum of copies of M. The action of x kills Mx^{-1} and for $n > 1$ maps Mx^{-n} to $Mx^{-(n-1)}$ in the obvious way, sending ux^{-n} to $ux^{-(n-1)}$. Then $M \to M\langle x^{-1}\rangle$ is a faithfully exact functor from R-modules to $R[\![x]\!]$-modules. If R has prime characteristic $p > 0$ and M is an $R\{F\}$-module, then we may also extend this to an $R[\![x]\!]\{F\}$-module structure on $M\langle x^{-1}\rangle$ by letting F send $ux^{-n} \mapsto F(u)x^{-pn}$. This gives a convenient way of describing what happens to local cohomology when we adjoin a formal power series indeterminate to a local ring R.

Proposition 4.14. *Let R be a Noetherian ring, let I be a finitely generated ideal of R, and let x be a formal power series indeterminate over R. Let J denote the ideal $(I,\, x)R[\![x]\!]$ of $R[\![x]\!]$.*

(a) *For every i,*
$$H^i_J(R[\![x]\!]) \cong H^i_I(R)\langle x^{-1}\rangle.$$
If R has prime characteristic $p > 0$ then the action of Frobenius on $H^i_J(R[\![x]\!])$ agrees with the action on $H^i_I(R)\langle x^{-1}\rangle$ described above.

(b) *In particular, if $(R,\, m,\, K)$ is local and $J = \mathfrak{n}$, the maximal ideal of $R[\![x]\!]$, then for every i,*
$$H^i_{\mathfrak{n}}(R[\![x]\!]) \cong H^i_m(R)\langle x^{-1}\rangle.$$

(c) *If $(R,\, m,\, K)$ is local and M is Artinian, then $M\langle x^{-1}\rangle$ is Artinian over $R[\![x]\!]$.*

(d) *If $(R,\, m,\, K)$ is local of prime characteristic $p > 0$ and M is a simple $R\{F\}$-module on which the action of F is not 0, then $M\langle x^{-1}\rangle$ is a simple $R[\![x]\!]\{F\}$-module.*

(e) *If $(R,\, m,\, K)$ is local of prime characteristic $p > 0$ and M is an antinilpotent $R\{F\}$-module, then $M\langle x^{-1}\rangle$ is an antinilpotent $R[\![x]\!]\{F\}$-module.*

Proof. (a) Let $f_1, \ldots, f_n \in R$ generate I. Then $H^i_I(R)$ is the i-th cohomology of the complex $C^\bullet(\underline{f};\ R)$, and $H^i_J(R[\![x]\!])$ is the i-th cohomology module of the complex

$$C^\bullet(\underline{f}, x;\ R[\![x]\!]) \cong C^\bullet(\underline{f};\ R) \otimes_R C^\bullet(x;\ R[\![x]\!]).$$

The complex $C^\bullet(x;\ R[\![x]\!])$ is simply

$$0 \to R[\![x]\!] \to R[\![x]\!]_x \to 0$$

and has augmentation $R\langle x^{-1}\rangle$. Since $R[\![x]\!]$, $R[\![x]\!]_x$, and $R\langle x^{-1}\rangle$ are all R-flat, we have that $H^i_J(R[\![x]\!])$ is the i-th cohomology module of the mapping cone of the injection of complexes

$$C^\bullet(\underline{f};\ R[\![x]\!]) \hookrightarrow C^\bullet(\underline{f};\ R[\![x]\!]_x),$$

which may be identified with the cohomology of the quotient complex, and so with the cohomology of $C^\bullet(\underline{f};\ R\langle x^{-1}\rangle) \cong C^\bullet(\underline{f};\ R)\langle x^{-1}\rangle$. Since $R\langle x^{-1}\rangle$ is R-flat (in fact, R-free), applying $\_ \otimes_R R\langle x^{-1}\rangle$ commutes with formation of cohomology, from which the result follows. Part (b) is immediate from part (a).

To prove (c), note that every element of $M\langle x^{-1}\rangle$ is killed by a power of m and of x, and so by a power of $\mathfrak{n}$. It therefore suffices to see that the annihilator of $\mathfrak{n}$ is a finite-dimensional vector space over K. But the annihilator of x is Mx^{-1}, and the annihilator of m in Mx^{-1} is isomorphic with the annihilator of m in M.

We next prove (d). Since the kernel of the action of F on M is an F-stable R-submodule of M, the fact that M is a simple $R\{F\}$-module implies that F acts injectively on M. Suppose that N is a nonzero $R[\![x]\!]\{F\}$-submodule of M, and that $u_1x^{-1} + \cdots + u_kx^{-k} \in N$. By multiplying by x^{k-1} we see that $u_kx^{-1} \in N$. Hence, N has nonzero intersection N_1x^{-1} with Mx^{-1}. N_1 is an R-submodule of M. It is also F-stable, since if $ux^{-1} \in N$ then

$$x^{p-1}F(ux^{-1}) = F(u)x^{-1} \in N \cap Mx^{-1}.$$

Thus, N contains Mx^{-1}. In every degree h, let $N_hx^{-h} = N \cap Mx^{-h}$. Then $N_h \neq 0$, for if $u \in M - \{0\}$ and $q = p^e > h$,

$$x^{q-h}F^e(ux^{-1}) = F^e(u)x^{-h},$$

and $F(u) \neq 0$. Moreover, the R-submodule $N_h \subseteq M$ is F-stable, for if $vx^{-h} \in N_hx^{-h}$, then $x^{ph-h}F(vx^{-h}) = F(v)x^{-h} \in N_hx^{-h}$. Thus, $N_h = M$ for all $h \geq 1$, and so $N = M$.

For part (e), if M has a finite filtration by simple $R\{F\}$-modules M_j on which F has nonzero action, then applying $N \mapsto N\langle x^{-1}\rangle$ gives a finite filtration of $M\langle x^{-1}\rangle$ with factors $M_j\langle x^{-1}\rangle$, each of which is a simple $R[\![x]\!]\{F\}$-module by part (d) on which F has nonzero action. $\square$

Theorem 4.15. *Let (R, m, K) be a local ring of prime characteristic $p > 0$. Let x be a formal power series indeterminate over R. Let M be an $R\{F\}$-module that is Artinian as an R-module. Then the following are equivalent*:

(1) *M is antinilpotent.*

(2) *$M\langle x^{-1}\rangle$ has finite length over $R[\![x]\!]\{F\}$.*

(3) *$M\langle x^{-1}\rangle$ has only finitely many F-stable submodules over $R[\![x]\!]$.*

When these equivalent conditions hold, M has only finitely many F-stable modules over R.

Proof. We show that (2) ⇒ (1) ⇒ (3). Assume (2). If M is not antinilpotent, it has a subquotient $N \neq 0$ on which the action of F is 0. Then $N\langle x^{-1}\rangle$ is a subquotient of $M\langle x^{-1}\rangle$, and so has finite length as an $R[\![x]\!]\{F\}$-module. Since F kills it, it must have finite length as an $R[\![x]\!]$-module. But this is clearly false, since no power of x kills $N\langle x^{-1}\rangle$.

To see that (1) ⇒ (3), note that by Proposition 4.14 (e), the fact that M is antinilpotent implies that $M\langle x^{-1}\rangle$ is antinilpotent over $R[\![x]\!]$, and the result now follows from Theorem 4.12. The implication (3) ⇒ (2) is obvious. □

The next result is an immediate corollary.

Theorem 4.16. *Let (R, m, K) be a local ring of prime characteristic $p > 0$ and let $x = x_1$ and $x_2, \dots, x_n$ be formal power series indeterminates over R. Let $R_n = R[\![x_1, \dots, x_n]\!]$, where $R_0 = R$, and let m_n be its maximal ideal. Then the following conditions on R are equivalent*:

(1) *The local cohomology modules $H^i_m(R)$ are antinilpotent.*

(2) *The ring $R[\![x]\!]$ has FH-finite length.*

(3) *The ring R_n is FH-finite for every $n \in \mathbb{N}$.*

(4) *The local cohomology modules $H^i_{m_n}(R_n)$ are antinilpotent over R_n for all $n \in \mathbb{N}$.*

(5) *The ring R_n has FH-finite length for all $n \in \mathbb{N}$.*

When these conditions hold, R is F-injective.

Proof. We have that (1) ⇒ (4) by Proposition 4.14 (a) and (e) and a straightforward induction on n. This implies that R_n is FH-finite for all n by Theorem 4.12. Thus, (4) ⇒ (3) ⇒ (5) ⇒ (2), and it suffices to prove that (2) ⇒ (1), which is a consequence of Theorem 4.15.

The statement that R is then F-injective is obvious, since F acts injectively on any antinilpotent module. □

Corollary 4.17. *Let $(R,\ m,\ K)$ be an F-pure Gorenstein local ring of prime characteristic $p > 0$ and Krull dimension d. Then $H^d_m(R)$ is antinilpotent, and so F acts injectively on every subquotient of $H^d_m(R)$.*

Proof. The hypothesis also holds for $R[\![x]\!]$, and so the result follows from Theorem 3.7 and Theorem 4.16. □

Corollary 4.18. *Let $(R,\ m,\ K)$ be an F-pure Gorenstein local ring of prime characteristic $p > 0$ and of Krull dimension d. Let J be an ideal of R such that* $\dim(R/J) = d$. *Then $H^d_m(R/J)$ is antinilpotent, and so F acts injectively on $H^d_m(R/J)$ (here, F is induced naturally from the Frobenius action F on R/J). Hence, if R/J is Cohen–Macaulay, it is F-injective.*

Proof. Since R and R/J have the same dimension, the long exact sequence for local cohomology gives an $R\{F\}$-module surjection $H^d_m(R) \twoheadrightarrow H^d_m(R/J)$, which shows that $H^d_m(R/J)$ is antinilpotent as an $R\{F\}$-module, and therefore as an $(R/J)\{F\}$ module as well. □

Next we shall need the following result from [Watanabe 1991]: the F-pure case is attributed there to Srinivas. See [Watanabe 1991, Theorem 2.7] and the comment that precedes it.

Theorem 4.19 (Watanabe and Srinivas). *Let $h : (R,\ m,\ K) \to (S,\ \mathfrak{n},\ L)$ be a local homomorphism of local normal domains of prime characterisitic $p > 0$ such that S is module-finite over R and the map h is étale in codimension one. If R is strongly F-regular, then so is S. If R is F-pure, then so is S.*

The explicit statement in [Watanabe 1991] is for the F-regular case, by which the author means the weakly F-regular case. However, the proof given uses the criterion (i) of [Watanabe 1991, Proposition 1.4], that the local ring $(R,\ m,\ K)$ is weakly F-regular if and only if 0 is tightly closed in $E_R(K)$, which is correct for finitistic tight closure but not for the version of tight closure being used in [Watanabe 1991]. In fact, condition (i) as used in [Watanabe 1991] characterizes strong F-regularity in the F-finite case, and we take it as the definition of strong F-regularity here.

On the other hand, there are no problems whatsoever in proving the final statement about F-purity. The action of Frobenius $F_S : E_S(L) \to F_S(E_S(L))$ is shown to be the same as the action of Frobenius when $E_S(L)$ is viewed as R-module. Since R is F-pure, the Frobenius action $M \to F_R(M)$ is injective for any R-module.

Corollary 4.20 (Watanabe). *Let $(R,\ m)$ be a normal local domain of characteristic $p > 0$. Let I be an ideal of pure height one, and suppose that I has finite order $k > 1$ in the divisor class group of R. Choose a generator u of $I^{(k)}$. We let*

$$S = R \oplus It \oplus \cdots \oplus I^{(j)} \oplus \cdots \oplus I^{(k-1)}$$

with $I^{(k)}$ identified with R using the isomorphism $R \cong I^{(k)}$ such that $1 \mapsto u$. (If t is an indeterminate, we can give the following more formal description: form $T = \bigoplus_{j=0}^{\infty} I^{(j)} t^j \subseteq R[t]$, and let $S = T/(ut^k - 1)$.) This ring is module-finite over R, and if k is relatively prime to p, it is étale over R in codimension one.

Hence, if k is relatively prime to p, then S is strongly F-regular if and only if R is strongly F-regular, and S is F-pure if and only if R is F-pure.

Moreover, if $I \cong \omega$ is a canonical module for R, then S is quasi-Gorenstein.

The final statement is expected because, by the discussion of canonical modules for module-finite extensions in Discussion 2.2, we have that $\omega_S \cong \mathrm{Hom}_R(S, \omega)$ and $\mathrm{Hom}_R(\omega^{(i)}, \omega) \cong \omega^{-(i-1)} \cong \omega^{(k-(i-1))}$. See [Watanabe 1991; Tomari and Watanabe 1992; Watanabe and Yoshida 2004, §3; Singh 2003] for further details and background on this technique. The result above will enable us to use this "canonical cover trick" to prove the theorem below by reduction to the quasi-Gorenstein case. A word of caution is in order: even if R is Cohen–Macaulay, examples in [Singh 2003] show that the auxiliary ring S described in Corollary 4.20 need not be Cohen–Macaulay, and so one is forced to consider the quasi-Gorenstein property. There are examples (see [Singh 2003, Theorem 6.1]) where R is F-rational but S is not Cohen–Macaulay. On the other hand, if R is strongly F-regular, the result of [Watanabe 1991] shows that S is as well; in particular, S is Cohen–Macaulay in this case.

Theorem 4.21. *Let R be a Cohen–Macaulay F-pure normal local domain of Krull dimension d such that R has canonical module $\omega = \omega_R$ of finite order k relatively prime to p in the divisor class group of R. Then $H^d_m(R)$ is antinilpotent, so R is FH-finite.*

Proof. Since the hypotheses are stable under adjunction of a power series indeterminate, it follows from Theorem 4.16 that it is sufficient to show that R is FH-finite. We may identify ω with a pure height one ideal of I of R. We form the ring S described in Corollary 4.20. Then S is F-pure and quasi-Gorenstein, and so $H^d_m(S)$ has only finitely many F-stable submodules by Theorem 3.7. The same holds for $H^d_m(R)$ by Lemma 2.7 (b), while the lower local cohomology modules of R with support in m vanish. □

The following improves the conclusion of Proposition 2.14 with some additional hypotheses.

Theorem 4.22. *Let (R, m, K) be an F-injective Cohen–Macaulay local ring with of prime characteristic $p > 0$ such that K is perfect. Suppose that R has an m-primary ideal $\mathfrak{A}$ such that $\mathfrak{A}I^* \subseteq I$ for every ideal I generated by a system of parameters. Let $d = \dim(R)$. Then $H^d_m(R)$ is antinilpotent, so that R and every formal power series ring over R is FH-finite.*

Proof. Let $H = H^d_m(R)$ and let $V = 0^*_H$, which, as in the proof of Proposition 2.14, is killed by $\mathfrak{A}$ and has finite length. Then $R/\mathfrak{A}$ is a complete local ring with a perfect residue class field, and contains a unique coefficient field K. This gives V the structure of a K-module, that is, it is a finite-dimensional K-vector space, and $F : V \to V$ is K-linear if we let the action of K on the second copy of V be such that $c \cdot v = c^p v$ for $c \in K$. Since K is perfect, the dimension of V does not change when we restrict F in this way. Since R is F-injective, the action of F on V is then a vector space isomorphism, and is then also an isomorphism when restricted to subquotients that are $K\{F\}$-modules. It follows that V is antinilpotent over $R\{F\}$, and to complete the proof it will suffice to show that F cannot act trivially on the simple $R\{F\}$-module H/V.

Choose a system of parameters $x_1, \ldots, x_d$ for R. Let

$$I_t = (x_1^t, \ldots, x_d^t)R.$$

For any sufficiently large value of t, we may identify V with I^*/I. If F acts trivially on H/V, then for all large t, the image of $1 \in R/I_t \subseteq H$ under F is 0 in H/V, which means that $(x_1 \cdots x_d)^{p-1} \in I^*_{pt}$, and then $\mathfrak{A}(x_1 \cdots x_d)^{p-1} \subseteq I_{pt}$. This implies that

$$\mathfrak{A} \subseteq I_{pt} :_R (x_1 \cdots x_d)^{p-1} = I_{pt-p+1},$$

for all $t \gg 0$, which is clearly false. $\square$

5. Face rings

We give a brief treatment of the decomposition of the local cohomology of face rings over a field with support in the homogeneous maximal ideal. This is discussed in [Bruns and Herzog 1993, §5.3], although not in quite sharp enough a form for our needs here, and there are sharp results in substantially greater generality in [Brun et al. 2007, Theorem 5.5, p. 218]. However, neither result discusses the $R\{F\}$-structure.

Let K be a fixed field of positive characteristic p and let Δ be an abstract finite simplicial complex with vertices $x_1, \ldots, x_n$. Let I_Δ denote the ideal in the polynomial ring $S = K[x_1, \ldots, x_n]$ generated by all monomials in the x_j such that the set of variables occurring in the monomial is not a face of Δ. This ideal is evidently generated by the square-free monomials in the x_j corresponding to minimal subsets of the variables that are not faces of Δ. Let $K[\Delta] = S/I_\Delta$, the *face ring* (or *Stanley–Reisner ring*) of Δ over K. The minimal primes of $K[\Delta]$ correspond to the maximal faces σ of Δ; the quotient by the minimal prime corresponding to σ is a polynomial ring in the variables occurring in σ. The Krull dimension of $K[\Delta]$ is therefore one more than the dimension of the simplicial complex Δ.

If σ is any face of Δ, the link, denoted $\mathrm{link}(\sigma)$, of σ in Δ is the abstract simplicial complex consisting of all faces τ of Δ disjoint from σ such that $\sigma \cup \tau \in \Delta$. The link of the empty face is Δ itself. By a theorem of G. Reisner [1976], $K[\Delta]$ is Cohen–Macaulay if and only if the reduced simplicial cohomology of every link vanishes except possibly in the top dimension, that is, in the dimension of the link itself.

Note that the reduced simplicial cohomology $\widetilde{H}^i(\Delta;\, K)$ of a finite simplicial complex $\Delta \neq \varnothing$ is the same as the simplicial cohomology unless $i = 0$, in which case its dimension as a K-vector space is one smaller. If Δ is an i-simplex, the reduced simplicial cohomology vanishes in all dimensions, unless Δ is empty, in which case we have $\widetilde{H}^{-1}(\varnothing;\, K) \cong K$, $\widetilde{H}^i(\varnothing;\, K) = 0$ for all $i \neq -1$. Note also that $\varnothing$ is the only simplicial complex Δ such that $\widetilde{H}^i(\Delta;\, K) \neq 0$ for a value of $i < 0$.

Let m be the homogeneous maximal ideal of $K[\Delta]$. We shall show that $K[\Delta]_m$ and its completion are FH-finite in all cases, and in fact, the local cohomology modules are antinilpotent. This follows from the following theorem, which also recovers Reisner's result [1976] mentioned above in a finer form; it also gives a completely explicit description of all the $H^i_m(K[\Delta])$, including their structure as $R\{F\}$-modules. We write $|\nu|$ for the cardinality of the set ν. If $\nu \in \Delta$, we let

$$K[\nu] = K[\Delta]/(x_i : x_i \notin \nu),$$

which is a $K[\Delta]$-algebra and is also the polynomial ring over K in the variables x_j that are vertices of ν. Then $H^i_m(S_\nu)$ vanishes except when $i = |\nu|$. When $i = \nu$ it is the highest nonvanishing local cohomology of a polynomial ring, and, if the characteristic of K is $p > 0$, it is a simple $R\{F\}$-module on which F acts injectively.

Note that if $p > 0$ is prime, $\kappa = \mathbb{Z}/p\mathbb{Z}$, R and K are rings of characteristic p, and H is an $R\{F\}$-module, $K \otimes_\kappa H$ has the structure of a $(K \otimes_\kappa R)\{F\}$-module: the action of F is determined by the rule $F(c \otimes u) = c^p \otimes F(u)$ for all $c \in K$ and $u \in H$. This is well defined because the action of F restricts to the identity map on $\mathbb{Z}/p\mathbb{Z}$.

Theorem 5.1. *With $R = K[\Delta]$ as above, let κ denote the prime field of K. Let m and μ be the homogeneous maximal ideals of R and $\kappa[\Delta]$, respectively. Then*

$$H^i_m(R) \cong \bigoplus_{\nu \in \Delta} \widetilde{H}^{i-1-|\nu|}(\mathrm{link}(\nu);\, K) \otimes_\kappa H^{|\nu|}_\mu(\kappa[\nu]). \tag{1}$$

If K has characteristic $p > 0$, this is also an isomorphism of $R\{F\}$-modules, with the action of F described in the paragraph above. Hence, every $H^i_m(R)$ is a finite direct sum of simple $R\{F\}$-modules on which F acts injectively.

If $(R_1,\ m_1)$ is either R_m or its completion, then for all i, $H^i_m(R)$ may be identified with $H^i_{m_1}(R_1)$, and $H^i_{m_1}(R_1)$ is a finite direct sum of simple $R_1\{F\}$-modules on which F acts injectively, and so is antinilpotent and FH-finite over R_1.

Proof. If σ is a subset of the x_j we denote by $x(\sigma)$ the product of the x_j for $j \in \sigma$. Thus, in $K[\Delta]$, the image of $x(\sigma)$ is nonzero if and only if $\sigma \in \Delta$. Our convention is that $\sigma = \varnothing$ is in Δ, is the unique face of dimension -1, and that $x(\varnothing) = 1$. We write $[\Delta]_i$ for the set of faces of Δ of dimension i. Then $H^i_m(K[\Delta])$ is the i-th cohomology module of the complex $C^\bullet$ whose i-th term is displayed below:

$$0 \to K[\Delta] \to \cdots \to \bigoplus_{\sigma \in \Delta_{i-1}} K[\Delta]_{x(\sigma)} \to \cdots .$$

The initial nonzero term $K[\Delta]$ may be thought of as $K[\Delta]_{x(\varnothing)}$ and the highest nonzero terms occur in degree $\dim(\Delta) + 1$. Let $\theta = (\theta_1,\ \ldots, \theta_n) \in \mathbb{Z}^n$ by an n-tuple of integers. We want to calculate that θ-graded piece of $H^i_m(K[\Delta])$. This is the same as the i-th cohomology of the θ-graded piece of the complex: denote by $C^\bullet[\theta]$ the θ-graded piece of the complex $C^\bullet$. Let $\mathrm{neg}(\theta)$ (respectively, $\mathrm{pos}(\theta)$; respectively, $\mathrm{supp}(\theta)$) denote the set of variables x_i such that θ_i is strictly negative (respectively, strictly positive; respectively, nonzero). Thus, $\mathrm{supp}(\theta)$ is the disjoint union of $\mathrm{neg}(\theta)$ and $\mathrm{pos}(\theta)$.

Let $\nu = \mathrm{neg}(\theta)$ and $\pi = \mathrm{pos}(\theta)$. Then $K[\Delta]_{x(\sigma)}$ has a nonzero component in degree θ if and only if $\nu \subseteq \sigma$ and $\sigma \cup \pi \in \Delta$, and then there is a unique copy of K corresponding to θ in the complex. By deleting the variables occurring in $\nu = \mathrm{neg}(\theta)$ from each face, we find that $C^\bullet[\theta]$ corresponds, with a shift in degree by the cardinality $|\nu|$ of ν, to the complex used to calculate the reduced simplicial cohomology of $\Delta_{\nu,\pi}$, where this is the subcomplex of the link of ν consisting of all simplices τ such that $\tau \cup \pi \in \Delta$. If π is nonempty, $\Delta_{\nu,\pi}$ is a cone on any vertex in π. Hence, the graded component of local cohomology in degree θ can be nonzero only when $\mathrm{pos}(\theta) = \varnothing$ and $\nu = \mathrm{neg}(\theta) \in \Delta$. It now follows that

$$[H^i_m(K[\Delta])]_\theta \cong \bigoplus_{\nu} \widetilde{H}^{i-|\nu|-1}(\mathrm{link}\ \nu) x^\theta,$$

if $\pi = \varnothing$ and $\nu \in \Delta$ is the set of variables corresponding to strictly negative entries in θ, and is zero otherwise.

It follows that we may identify

$$H^i_m(K[\Delta]) \cong \bigoplus_{\nu \in \Delta} \Biggl(\bigoplus_{\mathrm{supp}(w) = \mathrm{neg}(w) = \nu} \widetilde{H}^{i-|\nu|-1}(\mathrm{link}\ \nu) w \Biggr),$$

where w runs through all monomials with nonpositive exponents such that the set of variables with strictly negative exponents is ν.

We next want to show that if $\nu \in \Delta$, then

$$\bigoplus_{\text{supp}(\theta)=\text{neg}(\theta)=\nu} [H^i_m(K[\Delta])]_\theta \cong \widetilde{H}^{i-|\nu|-1}(\text{link}(\nu);\ K) \otimes_K H^{|\nu|}_m(K[\nu]).$$

The term on the right may also be written as $\widetilde{H}^{i-|\nu|-1}(\text{link}(\nu);\ K) \otimes_\kappa H^{|\nu|}_\mu(\kappa[\nu])$. We also need to check that the actions of F agree.

The action of F on $C^\bullet(\underline{x};\ K[\Delta])$ is obtained from the action of F, by applying $K \otimes_\kappa \_$, on $C^\bullet(\underline{x};\ \kappa[\Delta])$. Thus, we reduce at once to the case where $K = \kappa$, which we assume henceforth.

Let $\gamma \otimes w$ be an element in the cohomology, where $\gamma \in \widetilde{H}^{i-|\nu|-1}(\text{link}(\nu);\ K)$ and w is a monomial in the variables of ν with all exponents strictly negative. The action of x_i by multiplication is obvious in most cases. If $x_i \notin \nu$, the product is 0. If x_i occurs with an exponent other than -1 in w, one simply gets $\gamma \otimes (x_i w)$. The main nontrivial point is that if x_i occurs with exponent -1 in w, x_i kills $\gamma \otimes w$. To verify this, let $\nu' = \nu - \{x_i\}$. Take a cocycle η that represents γ. After we multiply by x_i, we get an element of $H^{i-|\nu|}(\text{link}(\nu'))$. Note that each simplex remaining when we delete the variables in ν' involves x_i, and so that the cocycle η' we get from η may be viewed as a cocycle of the complex used to compute the reduced simplicial cohomology of the closed star of x_i in $\text{link}(\nu')$. Since this closed star is a cone, that cohomology is 0. This shows that x_i kills every homogeneous component whose degree has -1 in the i-th coordinate.

We have completed the calculation of the structure of the local cohomology as an R-module. On the other hand, given $\nu \in \Delta$, because the field is κ, when F acts on the complex

$$\bigoplus_{\text{supp}(\theta)=\text{neg}(\theta)=\nu} [C^\bullet(\underline{x};\ R)]_\theta,$$

the value of F acting on an element of the form ηw, where η is a cocycle, is simply ηw^p, and so it follows that

$$F(\gamma \otimes w) = \gamma \otimes w^p.$$

This shows that $R\{F\}$-module structure is preserved by the isomorphism (1). □

References

[Aberbach et al. 1993] I. M. Aberbach, M. Hochster, and C. Huneke, "Localization of tight closure and modules of finite phantom projective dimension", *J. Reine Angew. Math.* **434** (1993), 67–114. MR 94h:13005 Zbl 0772.13004

[Brun et al. 2007] M. Brun, W. Bruns, and T. Römer, "Cohomology of partially ordered sets and local cohomology of section rings", *Adv. Math.* **208**:1 (2007), 210–235. MR 2008c:13022 Zbl 1116.13013

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

[Elitzur 2003] H. Elitzur, *Tight closure in Artinian modules*, Ph.D. thesis, University of Michigan, Ann Arbor, MI, 2003.

[Enescu 2001] F. Enescu, *A study of F-rationality and F-injectivity*, Ph.D. thesis, University of Michigan, Ann Arbor, MI, 2001.

[Enescu 2003] F. Enescu, "F-injective rings and F-stable primes", *Proc. Amer. Math. Soc.* **131**:11 (2003), 3379–3386. MR 2004f:13003 Zbl 1039.13006

[Fedder 1983] R. Fedder, "F-purity and rational singularity", *Trans. Amer. Math. Soc.* **278**:2 (1983), 461–480. MR 84h:13031 Zbl 0519.13017

[Grothendieck 1967] A. Grothendieck, *Local cohomology*, Lecture Notes in Mathematics **41**, Springer, Berlin, 1967. MR 37 #219 Zbl 0185.49202

[Herzog 1974] J. Herzog, "Ringe der Charakteristik p und Frobeniusfunktoren", *Math. Z.* **140** (1974), 67–78. MR 50 #4569 Zbl 0278.13006

[Hochster 1977] M. Hochster, "Cyclic purity versus purity in excellent Noetherian rings", *Trans. Amer. Math. Soc.* **231**:2 (1977), 463–488. MR 57 #3111 Zbl 0369.13005

[Hochster 2008] M. Hochster, "Some finiteness properties for Lyubeznik's F-modules", preprint, 2008, Available at http://www.math.lsa.umich.edu/~hochster/FDCC.mgc2.pdf. To appear in *Proceedings of the 2005 Midwest Algebraic Geometry Conference (MAGIC05).*

[Hochster and Huneke 1989] M. Hochster and C. Huneke, "Tight closure and strong F-regularity", pp. 119–133 in *Colloque en l'honneur de Pierre Samuel* (Orsay, 1987), Mém. Soc. Math. France (N.S.) **38**, Soc. Math. France, Paris, 1989. MR 91i:13025 Zbl 0699.13003

[Hochster and Huneke 1990] M. Hochster and C. Huneke, "Tight closure, invariant theory, and the Briançon–Skoda theorem", *J. Amer. Math. Soc.* **3**:1 (1990), 31–116. MR 91g:13010 Zbl 0701.13002

[Hochster and Huneke 1994] M. Hochster and C. Huneke, "F-regularity, test elements, and smooth base change", *Trans. Amer. Math. Soc.* **346**:1 (1994), 1–62. MR 95d:13007 Zbl 0844.13002

[Hochster and Huneke 1995] M. Hochster and C. Huneke, "Applications of the existence of big Cohen–Macaulay algebras", *Adv. Math.* **113**:1 (1995), 45–117. MR 96d:13014 Zbl 0834.13013

[Hochster and Huneke 2008] M. Hochster and C. Huneke, "Tight closure in equal characteristic zero", preprint, 2008.

[Hochster and Roberts 1976] M. Hochster and J. L. Roberts, "The purity of the Frobenius and local cohomology", *Advances in Math.* **21**:2 (1976), 117–172. MR 54 #5230 Zbl 0348.13007

[Huneke 1996] C. Huneke, *Tight closure and its applications*, CBMS Regional Conference Series in Mathematics **88**, American Mathematical Society, Providence, RI, 1996. MR 96m:13001 Zbl 0930.13004

[Katzman 2006] M. Katzman, "F-stable submodules of top local cohomology modules of Gorenstein rings", preprint, 2006. arXiv math/0602550

[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

[Kunz 1976] E. Kunz, "On Noetherian rings of characteristic p", *Amer. J. Math.* **98**:4 (1976), 999–1013. MR 55 #5612 Zbl 0341.13009

[Lyubeznik 1997] G. Lyubeznik, "F-modules: applications to local cohomology and D-modules in characteristic $p > 0$", *J. Reine Angew. Math.* **491** (1997), 65–130. MR 99c:13005 Zbl 0904.13003

[Lyubeznik and Smith 1999] G. Lyubeznik and K. E. Smith, “Strong and weak F-regularity are equivalent for graded rings”, *Amer. J. Math.* **121**:6 (1999), 1279–1290. MR 2000m:13006 Zbl 0970.13003

[Lyubeznik and Smith 2001] G. Lyubeznik and K. E. Smith, “On the commutation of the test ideal with localization and completion”, *Trans. Amer. Math. Soc.* **353**:8 (2001), 3149–3180. MR 2002f: 13010 Zbl 0977.13002

[Reisner 1976] G. A. Reisner, “Cohen–Macaulay quotients of polynomial rings”, *Advances in Math.* **21**:1 (1976), 30–49. MR 53 #10819 Zbl 0345.13017

[Sharp 2007] R. Y. Sharp, “Graded annihilators of modules over the Frobenius skew polynomial ring, and tight closure”, *Trans. Amer. Math. Soc.* **359**:9 (2007), 4237–4258. MR 2008b:13006 Zbl 1130.13002

[Singh 2003] A. K. Singh, “Cyclic covers of rings with rational singularities”, *Trans. Amer. Math. Soc.* **355**:3 (2003), 1009–1024. MR 2003m:13006 Zbl 1021.13007

[Smith 1994] K. E. Smith, “Tight closure of parameter ideals”, *Invent. Math.* **115**:1 (1994), 41–60. MR 94k:13006 Zbl 0820.13007

[Smith 1997a] K. E. Smith, “F-rational rings have rational singularities”, *Amer. J. Math.* **119**:1 (1997), 159–180. MR 97k:13004 Zbl 0910.13004

[Smith 1997b] K. E. Smith, “Vanishing, singularities and effective bounds via prime characteristic local algebra”, pp. 289–325 in *Algebraic geometry — Santa Cruz 1995* (Santa Cruz, 1995), edited by J. Kollár et al., Proc. Sympos. Pure Math. **62**, Amer. Math. Soc., Providence, RI, 1997. MR 99a:14026 Zbl 0913.13004

[Smith 2003] K. E. Smith, *Tight closure of parameter ideals and F-rationality*, Ph.D. thesis, University of Michigan, Ann Arbor, MI, 2003.

[Stubbs 2008] J. Stubbs, *Potent elements and tight closure in artinian modules*, Ph.D. thesis, University of Michigan, Ann Arbor, MI, 2008.

[Tomari and Watanabe 1992] M. Tomari and K. Watanabe, “Normal Z_r-graded rings and normal cyclic covers”, *Manuscripta Math.* **76**:3-4 (1992), 325–340. MR 93j:13002 Zbl 0780.13001

[Watanabe 1991] K. Watanabe, “F-regular and F-pure normal graded rings”, *J. Pure Appl. Algebra* **71**:2-3 (1991), 341–350. MR 92g:13003 Zbl 0738.13005

[Watanabe and Yoshida 2004] K.-i. Watanabe and K.-i. Yoshida, “Minimal relative Hilbert–Kunz multiplicity”, *Illinois J. Math.* **48**:1 (2004), 273–294. MR 2005b:13033 Zbl 1089.13007

Communicated by Craig Huneke
Received 2007-07-27 Revised 2008-07-15 Accepted 2008-08-26

fenescu@gsu.edu *Department of Mathematics and Statistics, Georgia State University, 750 COE, 7th floor, 30 Pryor Street, Atlanta, GA 30303-3083, United States*

hochster@umich.edu *Department of Mathematics, University of Michigan, 2074 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, United States*

ALGEBRA AND NUMBER THEORY 2:7(2008)

The half-integral weight eigencurve

Nick Ramsey
Appendix by Brian Conrad

In this paper we define Banach spaces of overconvergent half-integral weight p-adic modular forms and Banach modules of families of overconvergent half-integral weight p-adic modular forms over admissible open subsets of weight space. Both spaces are equipped with a continuous Hecke action for which U_{p^2} is moreover compact. The modules of families of forms are used to construct an eigencurve parameterizing all finite-slope systems of eigenvalues of Hecke operators acting on these spaces. We also prove an analog of Coleman's theorem stating that overconvergent eigenforms of suitably low slope are classical.

1. Introduction

In [Ramsey 2006], the author set up a geometric theory of modular forms of weight $k/2$ for odd positive integers k, complete with geometrically defined Hecke operators. This approach naturally led to a theory of overconvergent p-adic modular forms of such weights equipped with a Hecke action for which U_{p^2} is compact.

In this paper we define overconvergent half-integral weight p-adic modular forms of general p-adic weights, as well as rigid-analytic families thereof over

MSC2000: primary 11F33; secondary 14G22, 11F37.
Keywords: modular forms of half-integral weight, p-adic modular forms, eigenvarieties.
This research is supported in part by NSF Grant DMS-0503264.

admissible open subsets of weight space. We use the latter spaces and Buzzard's eigenvariety machine [Buzzard 2007] to construct a half-integral weight eigencurve parameterizing all systems of eigenvalues of Hecke operators occurring on spaces of half-integral weight overconvergent eigenforms of finite slope. In contrast to the integral weight situation, this space *does not* parameterize actual forms because a half-integral weight form that is an eigenform for all Hecke operators is not always characterized by its weight and collection of Hecke eigenvalues. We also prove an analog of Coleman's result that overconvergent eigenforms of suitably low slope are classical.

This paper lays the foundation for a forthcoming one in which the author will construct a map from our half-integral weight eigencurve to its integral weight counterpart (at least after passage to the underlying reduced spaces) that rigid-analytically interpolates the classical Shimura lifting introduced in [Shimura 1973].

2. Preliminaries

General notation. Fix a prime number p. The symbol K will always denote a complete and discretely-valued field extension of $\mathbb{Q}_p$. For such K we denote the ring of integers by $\mathcal{O}_K$ and the maximal ideal therein by m_K. The absolute value on K will always be normalized by $|p| = 1/p$.

2.1. ***Modular curves.*** For positive integers N and n, $X_1(N)$ and $X_1(N,n)$ will denote the usual moduli stacks of generalized elliptic curves with level structure. The former classifies generalized elliptic curves with a point P of order N, while the latter classifies generalized elliptic curves with a pair (P,C) consisting of a point P of order N and a cyclic subgroup C of order n meeting the subgroup generated by P trivially (plus a certain ampleness condition for nonsmooth curves). This level structure will always be taken to be the Drinfeld-style level structure found in [Katz and Mazur 1985], [Conrad 2007], and the appendix to this paper, and in all cases the base ring will be a $\mathbb{Z}_{(p)}$-algebra.

Throughout this paper we will make extensive use of certain admissible opens in rigid spaces associated to some of these modular curves. Traditionally these opens were defined using the Eisenstein series E_{p-1}, but this requires that we pose unfavorable restrictions on p and N. Fortunately, more recent papers of Buzzard [2003] and Goren and Kassaei [2006] define these opens and explore their properties in greater generality using alternative techniques. These authors define a "measure of singularity" $v(E) \in \mathbb{Q}^{\geq 0}$ associated to an elliptic curve over a complete extension of $\mathbb{Q}_p$. In case $v(E) \leq p/(p+1)$, one may associate to E a canonical subgroup $H_p(E)$ of order p in an appropriately functorial manner. Moreover, one understands $v(E/C)$ for finite cyclic subgroups $C \subseteq E$ as well as the canonical subgroup of E/C when it exists. Inductively applying this with $C = H_p(E)$, one can define (upon further restricting $v(E)$) canonical subgroups $H_{p^m}(E)$ of higher p-power order. For details regarding these constructions and facts, see [Buzzard 2003, Section 3] and [Goren and Kassaei 2006, Section 4].

We will denote the Tate elliptic curve over $\mathbb{Z}(\!(q)\!)$ by $\underline{\mathrm{Tate}}(q)$; see [Katz 1973]. Our notations concerning the Tate curve differ from those often found in the literature as follows. In the presence of, for example, level N structure, previous authors (for example [Katz 1973]) have preferred to consider the curve $\underline{\mathrm{Tate}}(q^N)$ over the base $\mathbb{Z}(\!(q)\!)$. Points of order N on this curve are used to characterize the behavior of a modular form at the cusps, and are all defined over the fixed ring $\mathbb{Z}(\!(q)\!)[\zeta_N]$ (where ζ_N is some primitive N-th root of 1). We prefer to fix the curve $\underline{\mathrm{Tate}}(q)$ and instead consider *extensions* of the base. Thus, in the presence of level N structure, we introduce the formal variable q_N, and *define* $q = q_N^N$. Then the curve $\underline{\mathrm{Tate}}(q)$ is defined over the subring $\mathbb{Z}(\!(q)\!)$ of $\mathbb{Z}(\!(q_N)\!)$, and all of its N-torsion is defined over the ring $\mathbb{Z}(\!(q_N)\!)[\zeta_N]$. To be precise, the N-torsion is given by

$$\zeta_N^i q_N^j \quad \text{for } 0 \le i, j \le N-1.$$

Cusps will always be referred to by specifying a level structure on the Tate curve.

Suppose that $N \ge 5$ so that we have a fine moduli scheme $X_1(N)_{\mathbb{Q}_p}$, and let $K/\mathbb{Q}_p$ be a finite extension (which will generally be fixed in applications). If $r \in [0, 1] \cap \mathbb{Q}$, then the region in the rigid space $X_1(N)_K^{\mathrm{an}}$ whose points correspond to pairs (E, P) with $v(E) \le r$ is an admissible affinoid open. We denote by $X_1(N)^{\mathrm{an}}_{\ge p^{-r}}$ the connected component of this region that contains the cusp associated to the datum $(\underline{\mathrm{Tate}}(q), \zeta_N)$ for some (equivalently, any) choice of primitive N-th root of unity ζ_N. Similarly, $X_1(N, n)^{\mathrm{an}}_{\ge p^{-r}}$ will denote the connected component of the region defined by $v(E) \le r$ in $X_1(N, n)_K^{\mathrm{an}}$ containing the cusp associated to $(\underline{\mathrm{Tate}}(q), \zeta_N, \langle q_n \rangle)$ for any such ζ_N. For smaller N one defines these spaces by first adding prime-to-p level structure to rigidify the moduli problem and proceeding as above, and then taking invariants. Similarly, the space $X_0(N)^{\mathrm{an}}_{\ge p^{-r}}$ is defined as the quotient of $X_1(N)^{\mathrm{an}}_{\ge p^{-r}}$ by the action of the diamond operators. See [Buzzard 2007, Section 6] for a more detailed discussion of these quotients.

2.2. ***Norms.*** If $\mathfrak{X}$ is an admissible formal scheme over $\mathcal{O}_K$ (in the sense of [Bosch and Lütkebohmert 1993]), we will denote its (Raynaud) generic fiber by $\mathfrak{X}_{\mathrm{rig}}$ and its special fiber by $\mathfrak{X}_0$. In case $\mathfrak{X} = \mathrm{Spf}(\mathcal{A})$ is a formal affine, we have $\mathfrak{X}_{\mathrm{rig}} = \mathrm{Sp}(\mathcal{A} \otimes_{\mathcal{O}_K} K)$ and $\mathfrak{X}_0 = \mathrm{Spec}(\mathcal{A}/\pi\mathcal{A})$, where $\pi \in \mathcal{O}_K$ is a uniformizer. We recall for later use that the natural specialization map $\mathrm{sp} : \mathfrak{X}_{\mathrm{rig}} \to \mathfrak{X}_0$ is surjective on the level of closed points; see [Bosch and Lütkebohmert 1993, Proposition 3.5].

Assume that $\mathfrak{X}$ is reduced, and let $\mathfrak{L}$ be an invertible sheaf on $\mathfrak{X}$ (that is to say, a sheaf of modules on this ringed space that is Zariski-locally free of rank one). For a point $x \in \mathfrak{X}_{\mathrm{rig}}(L)$, let $\hat{x} : \mathrm{Spf}(\mathcal{O}_L) \to \mathfrak{X}$ denote the unique extension of x to the formal model. Then the canonical identification

$$H^0(\mathrm{Sp}(L), x^*\mathfrak{L}_{\mathrm{rig}}) = H^0(\mathrm{Spf}(\mathcal{O}_L), \hat{x}^*\mathfrak{L}) \otimes_{\mathcal{O}_L} L$$

furnishes a norm $|\cdot|_x$ on this one-dimensional vector space by declaring the formal sections on the right to be the unit ball. Now for any admissible open $\mathcal{U} \subseteq \mathfrak{X}_{\mathrm{rig}}$ and any $f \in H^0(\mathcal{U}, \mathfrak{L}_{\mathrm{rig}})$, we define

$$\|f\|_{\mathcal{U}} = \sup_{x \in \mathcal{U}} |x^* f|_x.$$

Note that, in case $\mathfrak{L} = \mathcal{O}_{\mathfrak{X}}$, this is simply the usual supremum norm on functions.

There is no reason for $\|f\|_{\mathcal{U}}$ to be finite in general, but in case $\mathcal{U}$ is affinoid then this is indeed finite and endows $H^0(\mathcal{U}, \mathfrak{L}_{\mathrm{rig}})$ with the structure of a Banach space over K, as we now demonstrate.

Lemma 2.1. *Suppose $\mathfrak{X}$ is a reduced quasicompact admissible formal scheme over $\mathcal{O}_K$, let $\mathfrak{L}$ be an invertible sheaf on $\mathfrak{X}$, and let $\mathcal{U}$ be an admissible affinoid open in $\mathfrak{X}_{\mathrm{rig}}$. Then $H^0(\mathcal{U}, \mathfrak{L}_{\mathrm{rig}})$ is a K-Banach space with respect to $\|\cdot\|_{\mathcal{U}}$.*

Proof. By Raynaud's theorem there is quasicompact admissible formal blowup $\pi : \mathfrak{X}' \to \mathfrak{X}$ and an admissible formal open $\mathfrak{U}$ in $\mathfrak{X}'$ with generic fiber $\mathcal{U}$. For $x \in \mathcal{U}$, let $\hat{x}'$ denote the unique extension to an $\mathcal{O}_L$-valued point of $\mathfrak{U}$, and let $\hat{x}$ denote its image in $\mathfrak{X}$ (which is the same $\hat{x}$ as above by uniqueness). Then we have

$$H^0(\mathrm{Spf}(\mathcal{O}_L), \hat{x}'^*\pi^*\mathfrak{L}) = H^0(\mathrm{Spf}(\mathcal{O}_L), \hat{x}^*\mathfrak{L})$$

as lattices in $H^0(\mathrm{Sp}(L), \mathfrak{L}_{\mathrm{rig}})$. It follows that $|f|_x = |\pi^* f|_x$, and we may compute $\|f\|_{\mathcal{U}}$ using the models $\mathfrak{X}'$ and $\pi^*\mathfrak{L}$. Hence we may as well assume that $\mathcal{U}$ is the generic fiber of an admissible formal open $\mathfrak{U}$ in $\mathfrak{X}$. Furthermore, we may just as well replace $\mathfrak{X}$ by $\mathfrak{U}$ and assume that $\mathfrak{U}$ is the generic fiber of $\mathfrak{X}$ itself.

Cover $\mathfrak{X}$ by a finite collection of admissible formal affine opens $\mathfrak{U}_i$ trivializing $\mathfrak{L}$. Pick a trivializing section ℓ_i of $\mathfrak{L}$ on $\mathfrak{U}_i$. Let $\mathcal{U}_i = (\mathfrak{U}_i)_{\mathrm{rig}}$, so that the $\mathcal{U}_i$ form an admissible cover of $\mathcal{U}$ by admissible affinoid opens. Then, for any section $f \in H^0(\mathcal{U}, \mathfrak{L}_{\mathrm{rig}})$, we may write $f|_{\mathcal{U}_i} = a_i \ell_i$ for a unique $a_i \in \mathcal{O}(\mathcal{U}_i)$, and one easily checks that $\|f\|_{\mathcal{U}} = \max_i \|a_i\|_{\sup}$. The desired assertion now follows easily from the analogous assertion about the supremum norm on a reduced affinoid. □

The following lemma and its corollary establish a sort of maximum modulus principle for these norms.

Lemma 2.2. *Suppose $\mathfrak{X} = \mathrm{Spf}(\mathcal{A})$ is a reduced admissible affine formal scheme over $\mathcal{O}_K$, and let $U \subseteq \mathfrak{X}_0$ be a Zariski-dense open subset of the special fiber. Suppose that the generic fiber $X = \mathrm{Sp}(\mathcal{A} \otimes_{\mathcal{O}_K} K)$ is equidimensional. Then, for any $a \in \mathcal{A} \otimes_{\mathcal{O}_K} K$, the supremum norm of a over X is achieved on $\mathrm{sp}^{-1}(U)$.*

Proof. Let us first prove the lemma in the case that $\mathcal{A}$ is normal. First note that if $\|a\|_{\sup} = 0$, then the result is obvious. Otherwise, since the supremum norm is power-multiplicative we may assume that $\|a\|_{\sup}$ is a norm from K and scale to reduce to the case $\|a\|_{\sup} = 1$. By [de Jong 1995, Theorem 7.4.1] it follows that $a \in \mathcal{A}$ (this is where normality is used). If the reduction $a_0 \in \mathcal{A}_0 = \mathcal{A}/\pi\mathcal{A}$ vanishes at every closed point of U, then it vanishes everywhere by density, so $a_0^n = 0$ in $\mathcal{A}_0$ for some n, which is to say that $\pi \mid a^n$ in $\mathcal{A}$. But this is impossible because by power-multiplicativity we have $\|a^n\|_{\sup} = 1$ for all $n \geq 1$. Thus a_0 must be nonvanishing at some point of U. By the surjectivity of the specialization map we can find a point x reducing to this point. Clearly then $|a(x)| = 1$, which establishes the normal case.

Suppose that X is equidimensional of dimension d. We claim that it follows that the special fiber $\mathfrak{X}_0$ must be equidimensional of dimension d as well. Indeed, inside each irreducible component of this special fiber we can find a nonempty Zariksi-open subset V that does not meet any of the other irreducible components. The generic fiber V_{rig} is an admissible open in X and therefore has dimension d. It follows that V has dimension d, and the claim follows.

Let $f : \widetilde{\mathfrak{X}} \to \mathfrak{X}$ be the normalization map (meaning Spf applied to the normalization map on algebras), and note that this map is finite by general excellence considerations. By [Conrad 1999, Theorem 2.1.3] the generic fiber of this map coincides with the normalization of X. Thus $\widetilde{\mathfrak{X}}_{\mathrm{rig}}$ is also equidimensional of dimension d, and the argument above shows that $\widetilde{\mathfrak{X}}_0$ is equidimensional of dimension d as well. Now since f is finite it follows that f_0 carries generic points to generic points. In particular we see that $f_0^{-1}(U)$ is Zariski-dense in $\widetilde{\mathfrak{X}}_0$. Thus by the normal case proved above, there exists an $x \in \widetilde{\mathfrak{X}}_{\mathrm{rig}}$ reducing to $f_0^{-1}(U)$ at which a (thought of as an element of $\widetilde{\mathcal{A}} \otimes_{\mathcal{O}_K} K$) attains its supremum norm. But then $f(x)$ is a point in X reducing to U with the same property, since the supremum norm of a is the same thought of on X or on $\widetilde{X}$ (since $\widetilde{X} \to X$ is surjective). □

Remark 2.3. Note that the proof in the normal case did not use the equidimensionality hypothesis. This hypothesis may not be required in the general case, but the above proof breaks down without it since it is not clear how to control the special fiber under normalization in general, especially if $\mathfrak{X}_0$ is nonreduced (as is often the case for us).

Corollary 2.4. *Suppose $\mathfrak{X}$ is a reduced quasicompact admissible formal scheme over $\mathcal{O}_K$, let $U \subseteq \mathfrak{X}_0$ be a Zariski-dense open, and let $\mathfrak{L}$ be an invertible sheaf on $\mathfrak{X}$. Assume that $\mathfrak{X}_{\mathrm{rig}}$ is equidimensional. Then, for any $f \in H^0(\mathfrak{X}_{\mathrm{rig}}, \mathfrak{L}_{\mathrm{rig}})$ we have*

$$\|f\|_{\mathfrak{X}_{\mathrm{rig}}} = \sup_{x \in \mathrm{sp}^{-1}(U)} |x^* f|_x = \max_{x \in \mathrm{sp}^{-1}(U)} |x^* f|_x.$$

Proof. Cover $\mathfrak{X}$ with a finite collection of admissible formal affine opens trivializing $\mathfrak{L}$, and apply Lemma 2.2 on each such affine separately. □

The invertible sheaves whose sections we will be taking norms of in this paper will all be of the form $\mathcal{O}_X(D)$ for some divisor D on $X = X_1(N)_K$ or $X_1(N, n)_K$ supported on the cusps. In the end, the main consequence of Corollary 2.4 (namely, Lemma 2.5) will be that these norms are equal to the supremum norm of the restriction of the section in question to the complement of the residue disks around the cusps (where it is simply an analytic function). We feel it worthwhile to give more natural definitions using the above norm machinery in the cases that it applies to (those where we have nice moduli schemes to work with), in the hope that the techniques used and Corollary 2.4 will be useful in other similar situations. The reader content with this equivalent "ad hoc" definition (that is, the supremum norm on the complement of the residue disks around the cusps) can skip to Section 2.3 and ignore the appendix altogether.

In order to endow spaces of sections of a line bundle as in the previous paragraph with norms using the techniques above, we need formal models of the spaces X and sheaves $\mathcal{O}(D)$. For technical reasons (involving regularity of certain moduli stacks), we are forced to work over $\mathbb{Z}_p$ in going about this. The formal models over $\mathcal{O}_K$ will then be obtained by extension of scalars. The general procedure for obtaining formal models over $\mathbb{Z}_p$ goes as follows. Let X denote one the stacks $X_1(N)$ or $X_1(N, n)$ over $\mathbb{Z}_p$, and assume that the generic fiber $X_{\mathbb{Q}_p}$ is a scheme. Let D be a divisor on $X_{\mathbb{Q}_p}$ that is supported on the cusps. If the closure $\overline{D}$ of D in X lies in the maximal open subscheme X^{sch} of X and this subscheme is regular along $\overline{D}$, then this closure is Cartier and we may associate to it the invertible sheaf $\mathcal{O}(\overline{D})$ on X^{sch}. Let $(X^{\mathrm{sch}})\widehat{\ }$ and $\mathcal{O}(\overline{D})\widehat{\ }$ denote the formal completions of these objects along the special fiber.

In case $X = X_1(N)$ or $X_1(N, n)$ with $p \nmid n$, assume that N has a divisor that is prime to p and at least 5. Then $X^{\mathrm{sch}} = X$ by [Conrad 2007, Theorem 4.2.1], and X is regular (at least over $\mathbb{Z}_{(p)}$) by [Conrad 2007, Theorem 4.1.1]. That passage to $\mathbb{Z}_p$ preserves regularity follows by excellence considerations from the fact that $\mathbb{Z}_{(p)} \to \mathbb{Z}_p$ is geometrically regular. Strictly speaking, the results of [Conrad 2007] do not apply to $X_1(N, n)$ as stated, but since $p \nmid n$, the proofs of these results are still valid over $\mathbb{Z}_{(p)}$, as is observed in the appendix. Since X is proper over $\mathbb{Z}_p$, we have $\widehat{X}_{\mathrm{rig}} = X^{\mathrm{an}}_{\mathbb{Q}_p}$ (the analytification of the algebraic generic fiber of X) and hence we have a formal model $(\widehat{X}, \mathcal{O}(\overline{D})\widehat{\ })$ of $(X^{\mathrm{an}}_{\mathbb{Q}_p}, \mathcal{O}(D))$.

Suppose that $X = X_1(Mp, p^2)$ for an integer $M \geq 5$ prime to p. Let D be any divisor supported on the cusps in the connected component $X_1(Mp, p^2)^{\mathrm{an}}_{\geq 1}$ of the ordinary locus. By Theorem A.11, the closure $\overline{D}$ of D in X lies in X^{sch} and is Cartier. Thus we obtain a formal model $((X^{\mathrm{sch}})\widehat{\ }, \mathcal{O}(\overline{D})\widehat{\ })$ of $((X^{\mathrm{sch}})\widehat{\ }_{\mathrm{rig}}, \mathcal{O}(D))$.

Observe that, by Lemma A.9 and the comments that follow it, X^{sch} is simply the complement of a finite collection of cusps on the characteristic p fiber (namely, the ones with nontrivial automorphisms). It follows that the open immersion

$$(\widehat{X^{\mathrm{sch}}})_{\mathrm{rig}} \hookrightarrow (X^{\mathrm{sch}}_{\mathbb{Q}_p})^{\mathrm{an}} \cong X^{\mathrm{an}}_{\mathbb{Q}_p} \tag{1}$$

identifies the Raynaud generic fiber on the left with the complement of the residue disks around the cusps in the analytification on the right that reduce to the missing points in characteristic p. Thus (1) is an isomorphism when restricted to any connected component of the locus defined by $v(E) \le r$ that contains no such cusps. In particular, by Theorem A.11 it is an isomorphism when restricted to $X_1(Mp, p^2)^{\mathrm{an}}_{\ge p^{-r}}$.

Given a complete discretely valued extension $K/\mathbb{Q}_p$, we may extend scalars on our formal models of $\mathcal{O}(D)$ to arrive at norms on the following spaces:

- sections of $\mathcal{O}(D)$ over any admissible open $\mathcal{U}$ in $X = X_1(N)^{\mathrm{an}}_K$ (respectively $X_1(N, n)^{\mathrm{an}}_K$ with $p \nmid n$), where D is (the scalar extension of) a divisor on $X_1(N)_{\mathbb{Q}_p}$ (respectively $X_1(N, n)_{\mathbb{Q}_p}$) and N is divisible by an integer that is prime to p and at least 5; and
- sections of $\mathcal{O}(D)$ over any admissible open $\mathcal{U}$ in $X = X_1(Mp, p^2)^{\mathrm{an}}_{\ge p^{-r}}$, where D is (the scalar extension of) a divisor that is supported on the cusps in $X_1(Mp, p^2)^{\mathrm{an}}_{\mathbb{Q}_p}$ and M is an integer that is prime to p and at least 5.

Lemma 2.5. *Let X, D, and $\mathcal{U}$ be as in either of the two cases above, and assume that $\mathcal{U}$ contains every component of the ordinary locus that it meets. Let $\mathcal{U}'$ denote the complement of the residue disks around the cusps in $\mathcal{U}$. Then, for any $f \in H^0(\mathcal{U}, \mathcal{O}(D))$, we have $\|f\|_{\mathcal{U}} = \|f|_{\mathcal{U}'}\|_{\mathrm{sup}}$.*

Proof. We will treat the case of $X = X_1(N)^{\mathrm{an}}_K$; the other cases are proved in exactly the same manner. First note that, since points in $\mathcal{U}'$ reduce to points outside of the support of $\overline{D}$, the claim is equivalent to the claim that $\|f\|_{\mathcal{U}} = \|f|_{\mathcal{U}'}\|_{\mathcal{U}'}$. That is, the norm on $\mathcal{U}'$ that we have defined using formal models happens to be equal to the supremum norm on $\mathcal{U}'$.

Note that the supersingular loci of $\mathcal{U}$ and $\mathcal{U}'$ coincide, so the contributions to the above norms over this locus are equal, and it suffices to check the assertion upon restriction to the ordinary locus. By assumption, the ordinary locus in $\mathcal{U}$ is a finite union of connected components of the ordinary locus in $X_1(N)^{\mathrm{an}}_K$. Each such component corresponds via reduction to an irreducible component of the special fiber. Let $\mathfrak{X}$ denote the admissible formal open in $\widehat{X_1(N)}$ given by the union of the components so obtained with the supersingular points removed. Then $\mathfrak{X}_{\mathrm{rig}}$ is precisely the ordinary locus in $\mathcal{U}$, and the result now follows from Corollary 2.4 with U equal to the complement of the cusps in $\mathfrak{X}_0$. $\square$

Remark 2.6. There remain some curves on which we will need to have norms for sections of $\mathcal{O}(D)$ but to which the norm machinery as set up here does not apply. Namely, for $p \neq 2$ we have the curves $X_1(4p^m)_K^{\mathrm{an}}$ and $X_1(4p^m, p^2)_K^{\mathrm{an}}$, while for $p = 2$ we have $X_1(2^{m+1}N)_K^{\mathrm{an}}$ and $X_1(2^{m+1}N, 4)_K^{\mathrm{an}}$, where $m \geq 1$ and $N \in \{1, 3\}$. The previous lemma suggests an ad hoc workaround to this problem. In case we are working with sections of $\mathcal{O}(D)$ for a cuspidal divisor on one of these curves, we simply *define* the norm to be the supremum norm of the restriction of our section to the complement of the residue disks about the cusps. A more natural definition would likely result from considerations of "formal stacks", but this norm would surely turn out to be equal to ours by an analogue of Lemma 2.5.

2.3. *Weight space.* Throughout most of this paper, $\mathcal{W}$ will denote p-adic weight space (everywhere except for the beginning of Section 7, where it is allowed to be a general reduced rigid space for the purpose of reviewing a general construction). That is, $\mathcal{W}$ is a rigid space over $\mathbb{Q}_p$ whose points with values in an extension $K/\mathbb{Q}_p$ are $\mathcal{W}(K) = \mathrm{Hom}_{\mathrm{cont}}(\mathbb{Z}_p^\times, K^\times)$. Define $\mathbf{q} = p$ if $p \neq 2$ and $\mathbf{q} = 4$ if $p = 2$. Let $\tau : \mathbb{Z}_p^\times \to (\mathbb{Z}/\mathbf{q}\mathbb{Z})^\times \to \mathbb{Q}_p^\times$ denote reduction composed with the Teichmüller lifting, and let $\langle x \rangle = x/\tau(x) \in 1 + \mathbf{q}\mathbb{Z}_p$. For a weight κ we have

$$\kappa(x) = \kappa(\langle x \rangle)\kappa(\tau(x)) = \kappa(\langle x \rangle)\tau(x)^i$$

for a unique integer i with $0 \leq i < \varphi(\mathbf{q})$ (where φ denotes Euler's function). Moreover, this breaks up the space $\mathcal{W}$ as the admissible disjoint union of $\varphi(\mathbf{q})$ admissible opens $\mathcal{W}^i$, each of which is isomorphic to a one-dimensional open ball.

For each positive integer n, let $\mathcal{W}_n$ denote the admissible open subspace of $\mathcal{W}$ whose points are those κ with

$$|\kappa(1 + \mathbf{q})^{p^{n-1}} - 1| \leq |\mathbf{q}|.$$

Then $\mathcal{W}_n^i := \mathcal{W}^i \cap \mathcal{W}_n$ is an affinoid disk in $\mathcal{W}^i$, and the $\{\mathcal{W}_n^i\}_n$ form a nested admissible cover of $\mathcal{W}^i$.

To each integer λ we may associate the weight $x \mapsto x^\lambda$. This weight, which by abuse of notation we simply refer to as λ, lies in $\mathcal{W}^i$ for the unique $i \equiv \lambda \pmod{\varphi(\mathbf{q})}$. Also, if λ is an integer and $\psi : (\mathbb{Z}/\mathbf{q}p^{n-1}\mathbb{Z})^\times \to \mathbb{C}_p^\times$ is a character, then $x \mapsto x^\lambda\psi(x)$ is a point in $\mathcal{W}$ (with values in $\mathbb{Q}_p(\mu_{p^{n-1}})$) that lies in $\mathcal{W}_n$, as standard estimates for $|\zeta - 1|$ for roots of unity ζ demonstrate.

3. Some modular functions

Our definition of the spaces of half-integral weight modular forms will follow the general approach of [Coleman and Mazur 1998] (in the integral weight p-adic situation) and [Ramsey 2006] (in the half-integral weight situation). The motivating idea behind this approach is to reduce to weight zero by dividing by a well-understood form of the same weight. For example, if f is a half-integral weight p-adic modular form of weight $k/2$, θ is the usual Jacobi theta function of weight $1/2$, and E_λ is the weight $\lambda = (k-1)/2$ Eisenstein series introduced below, then $f/(E_\lambda\theta)$ should certainly be a meromorphic modular function of weight zero. As we have no working notion of "half-integral weight p-adic modular form", we simply use the weight zero forms so obtained as the *definition* of this notion. One must of course work out issues such as exactly what kind of poles are introduced, how dividing by θE_λ affects the nebentypus character, and how to translate the classical Hecke action into an action on these new forms. The precise definition will be given in Section 4.

This was carried out in [Ramsey 2006] by dividing by θ^k instead of θE_λ. That approach had the disadvantage of limiting us to *classical* weights $k/2$, whereas the current approach will work for more general p-adic weights (and indeed, for families of modular forms) since E_λ interpolates nicely in the variable λ.

This technique of division to reduce to weight zero in order to define modular forms forces us to modify the usual construction of the Hecke operators using the Hecke correspondences on the curve $X_1(N)$ by multiplying by certain functions on the source spaces of these correspondences. Our first task is to define these functions and to establish their overconvergence properties. Since we are dividing by $E_\lambda\theta$ to reduce to weight zero, we will require, for each prime number ℓ, a modular function whose q-expansion (at the appropriate cusp and on the appropriate space, which depends on whether or not $\ell = p$) is

$$\frac{E_\lambda(q_{\ell^2})\theta(q_{\ell^2})}{E_\lambda(q)\theta(q)}.$$

Factoring this into its Eisenstein part and theta part, we split the problem into two problems, the first of which is nearly done in the integral-weight literature (see [Buzzard 2007; Coleman and Mazur 1998]), and the second of which is done in [Ramsey 2006]. We briefly review both problems here, but see these references for details. Note that all analytic spaces in this section are taken over $\mathbb{Q}_p$.

Let $\mathbf{c}$ denote the cusp on $X_1(4)_{\mathbb{Q}}$ corresponding to the point $\zeta_4 q_2$ of order 4 on the Tate curve. Define a $\mathbb{Q}$-divisor Σ_{4N} on the curve $X_1(4N)_{\mathbb{Q}}$ by

$$\Sigma_4 := \tfrac{1}{4}\pi^*[\mathbf{c}], \quad \text{where } \pi : X_1(4N)_{\mathbb{Q}} \to X_1(4)_{\mathbb{Q}}$$

is the obvious degeneracy map. This divisor is set up to look like the divisor of zeros of the pullback of the Jacobi theta function θ to $X_1(4N)_{\mathbb{Q}}$ and will later be used to control poles introduced in dividing by $E_\lambda\theta$.

In [Ramsey 2006], we defined a rational function Θ_{ℓ^2} on $X_1(4, \ell^2)_{\mathbb{Q}}$ with divisor

$$\operatorname{div}(\Theta_{\ell^2}) = \pi_2^* \Sigma_4 - \pi_1^* \Sigma_4$$

such that

$$\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle q_{\ell^2} \rangle) = \sum_{n \in \mathbb{Z}} q_{\ell^2}^{n^2} \Big/ \sum_{n \in \mathbb{Z}} q^{n^2} = \theta(q_{\ell^2})/\theta(q).$$

Here π_1 and π_2 are the maps comprising the ℓ^2 Hecke correspondence on $X_1(4)$ and are defined in Section 5.1. Strictly speaking, we had assumed $\ell \neq 2$ in the arguments in [Ramsey 2006], but if one is only interested in the result above, then one can easily check that the arguments work for $\ell = 2$ verbatim.

Let us now turn to the Eisenstein part of the above functions. For further details and proofs of the claims in this paragraph, see [Buzzard 2007, Sections 6 and 7]. Let

$$E(q) := 1 + \frac{2}{\zeta_p(\kappa)} \sum_n \Big(\sum_{d \mid n,\, p \nmid d} \kappa(d) d^{-1} \Big) q^n \in \mathcal{O}(\mathcal{W}^0)[\![q]\!]$$

be the q-expansion of the p-deprived Eisenstein family over $\mathcal{W}^0$. Note that there are no problems with zeros of ζ_p since we are restricting our attention to $\mathcal{W}^0$. For a particular choice of $\kappa \in \mathcal{W}^0$, we denote by $E_\kappa(q)$ the expansion obtained by evaluating all of the coefficients at κ. In particular, for a positive integer λ no less than 2 and divisible by $\varphi(\mathbf{q})$, $E_\lambda(q)$ is the q-expansion of the usual p-deprived classical Eisenstein series of weight λ and level p.

Let ℓ be a prime number. If $\ell \neq p$, then there exists a rigid analytic function $\mathbf{E}_\ell$ on $X_0(p\ell)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0$ whose q-expansion at $(\underline{\mathrm{Tate}}(q), \mu_{p\ell})$ is $E(q)/E(q^\ell)$. If $\ell = p$, then the same holds with $X_0(p\ell)^{\mathrm{an}}_{\geq 1}$ replaced by $X_0(p)^{\mathrm{an}}_{\geq 1}$ and $\mu_{p\ell}$ replaced by μ_p. Buzzard [2007] shows that there exists a sequence of rational numbers

$$1/(p+1) > r_1 \geq r_2 \geq \cdots \geq r_n \geq \cdots > 0$$

with $r_i < p^{2-i}/\mathbf{q}(1+p)$ such that, when restricted to $X_0(p\ell)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0_n$ (respectively, $X_0(p)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0_n$ if $\ell = p$), $\mathbf{E}_\ell$ analytically continues to an invertible function on $X_0(p\ell)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n$ (respectively, $X_0(p)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n$ if $\ell = p$). Fix such a sequence once and for all. Let us first extend these results to square level.

Lemma 3.1. *Let $\ell \neq p$ be a prime number. There exists an invertible function $\mathbf{E}_{\ell^2}$ on $X_0(p\ell^2)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0$ whose q-expansion at $(\underline{\mathrm{Tate}}(q), \mu_{p\ell^2})$ is $E(q)/E(q^{\ell^2})$. Moreover, the function $\mathbf{E}_{\ell^2}$, when restricted to $\mathcal{W}^0_n$, analytically continues to an invertible function on $X_0(p\ell^2)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n$.*

There exists an invertible function $\mathbf{E}_{p^2}$ on $X_0(p)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0$ whose q-expansion at $(\underline{\mathrm{Tate}}(q), \mu_p)$ is $E(q)/E(q^{p^2})$. Moreover, the function $\mathbf{E}_{p^2}$, when restricted to $\mathcal{W}^0_n$, analytically continues to an invertible function on $X_0(p)^{\mathrm{an}}_{\geq p^{-r_n/p}} \times \mathcal{W}^0_n$.

Proof. Let ℓ be a prime different from p. There are two natural maps

$$X_0(p\ell^2)^{\mathrm{an}}_{\mathbb{Q}_p} \to X_0(p\ell)^{\mathrm{an}}_{\mathbb{Q}_p},$$

namely those given on noncuspidal points by

$$(E, C) \xmapsto{d_{\ell,1}} (E, \ell C) \quad \text{and} \quad (E, C) \xmapsto{d_{\ell,2}} (E/p\ell C, C/p\ell C).$$

Both of these restrict to maps

$$d_{\ell,1}, d_{\ell,2} : X_0(p\ell^2)^{\mathrm{an}}_{\geq p^{-r_n}} \to X_0(p\ell)^{\mathrm{an}}_{\geq p^{-r_n}}.$$

We define $\mathbf{E}_{\ell^2}$ to be the invertible function

$$\mathbf{E}_{\ell^2} := d^*_{\ell,1}\mathbf{E}_\ell \cdot d^*_{\ell,2}\mathbf{E}_\ell \in \mathcal{O}(X_0(p\ell^2)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n)^\times. \tag{2}$$

The q-expansion of $\mathbf{E}_{\ell^2}$ at $(\underline{\mathrm{Tate}}(q), \mu_{p\ell^2})$ is

$$\begin{aligned}
&\mathbf{E}_\ell(d_{\ell,1}(\underline{\mathrm{Tate}}(q), \mu_{p\ell^2}))\mathbf{E}_\ell(d_{\ell,2}(\underline{\mathrm{Tate}}(q), \mu_{p\ell^2}))\\
&\qquad = \mathbf{E}_\ell(\underline{\mathrm{Tate}}(q), \mu_{p\ell})\mathbf{E}_\ell(\underline{\mathrm{Tate}}(q)/\mu_\ell, \mu_{p\ell^2}/\mu_\ell)\\
&\qquad = \mathbf{E}_\ell(\underline{\mathrm{Tate}}(q), \mu_{p\ell})\mathbf{E}_\ell(\underline{\mathrm{Tate}}(q^\ell), \mu_{p\ell})\\
&\qquad = \frac{E(q)}{E(q^\ell)}\frac{E(q^\ell)}{E(q^{\ell^2})} = \frac{E(q)}{E(q^{\ell^2})}.
\end{aligned}$$

One must take additional care if $\ell = p$. Then there is a well-defined map

$$d : X_0(p)^{\mathrm{an}}_{\geq p^{-r_n/p}} \to X_0(p)^{\mathrm{an}}_{\geq p^{-r_n}}, \quad (E, C) \mapsto (E/C, H_{p^2}/C),$$

where H_{p^2} is the canonical subgroup of E of order p^2. This follows from the fact that $X_0(p)^{\mathrm{an}}_{\geq p^{-r_n/p}}$ consists of pairs (E, C) with C equal to the canonical subgroup of E of order p, and standard facts about quotienting by such subgroups; see for example [Buzzard 2003, Theorem 3.3]. We define an invertible function by

$$\mathbf{E}_{p^2} := \mathbf{E}_p \cdot d^*\mathbf{E}_p \in \mathcal{O}(X_0(p)^{\mathrm{an}}_{\geq p^{-r_n/p}} \times \mathcal{W}^0_n)^\times,$$

where we have implicitly restricted $\mathbf{E}_p$ to

$$X_0(p)^{\mathrm{an}}_{\geq p^{-r_n/p}} \times \mathcal{W}^0_n \subseteq X_0(p)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n.$$

The q-expansion of $\mathbf{E}_{p^2}$ at $(\underline{\mathrm{Tate}}(q), \mu_p)$ is

$$\begin{aligned}
&\mathbf{E}_p(\underline{\mathrm{Tate}}(q), \mu_p)\mathbf{E}_p(d(\underline{\mathrm{Tate}}(q), \mu_p))\\
&\qquad = \mathbf{E}_p(\underline{\mathrm{Tate}}(q), \mu_p)\mathbf{E}_p(\underline{\mathrm{Tate}}(q)/\mu_p, \mu_{p^2}/\mu_p)\\
&\qquad = \mathbf{E}_p(\underline{\mathrm{Tate}}(q), \mu_p)\mathbf{E}_p(\underline{\mathrm{Tate}}(q^p), \mu_p)\\
&\qquad = \frac{E(q)}{E(q^p)}\frac{E(q^p)}{E(q^{p^2})} = \frac{E(q)}{E(q^{p^2})}.
\end{aligned}$$

□

Let

$$\pi : X_1(p,\ell^2)^{\mathrm{an}}_{\mathbb{Q}_p} \to \begin{cases} X_0(p\ell^2)^{\mathrm{an}}_{\mathbb{Q}_p} & \text{if } \ell \neq p, \\ X_0(p)^{\mathrm{an}}_{\mathbb{Q}_p} & \text{if } \ell = p \end{cases}$$

denote the map given on noncuspidal points by

$$(E,P,C) \longmapsto \begin{cases} (E/C, (\langle P\rangle + E[\ell^2])/C) & \text{if } \ell \neq p, \\ (E/C, \langle P\rangle/C) & \text{if } \ell = p. \end{cases}$$

Note that we have

$$\pi(\underline{\mathrm{Tate}}(q), \zeta_p, \langle q_{\ell^2}\rangle) = \begin{cases} (\underline{\mathrm{Tate}}(q_{\ell^2}), \mu_{p\ell^2}) & \text{if } \ell \neq p, \\ (\underline{\mathrm{Tate}}(q_{p^2}), \mu_p) & \text{if } \ell = p. \end{cases} \tag{3}$$

This observation suggests that perhaps the components $X_1(p,\ell^2)^{\mathrm{an}}_{\geq p^{-r}}$ should be related (via π) to the components $X_0(p\ell^2)^{\mathrm{an}}_{\geq p^{-r}}$.

Lemma 3.2. *If $\ell \neq p$, then the map π restricts to*

$$\pi : X_1(p,\ell^2)^{\mathrm{an}}_{\geq p^{-r}} \to X_0(p\ell^2)^{\mathrm{an}}_{\geq p^{-r}} \quad \textit{for all } r < p/(1+p).$$

In case $\ell = p$, the map π restricts to

$$X_1(p,p^2)^{\mathrm{an}}_{\geq p^{-p^2 r}} \to X_0(p)^{\mathrm{an}}_{\geq p^{-r}} \quad \textit{for all } r < 1/p(1+p).$$

Proof. First suppose $\ell \neq p$. Let $\mathfrak{U}$ denote the entirety of the locus in $X_0(p\ell^2)^{\mathrm{an}}_{\mathbb{Q}_p}$ defined by $v(E) \leq r$. First note that, since quotienting by a subgroup of order prime to p does not change its measure of singularity, the map π restricts to a map

$$X_1(p,\ell^2)^{\mathrm{an}}_{\geq p^{-r}} \to \mathfrak{U}.$$

The inverse images of the two connected components of $\mathfrak{U}$ under this map are disjoint admissible opens that admissibly cover a connected space, and, by (3), $\pi^{-1}(X_0(p\ell^2)^{\mathrm{an}}_{\geq p^{-r}})$ is nonempty, so this must be all of $X_1(p,\ell^2)^{\mathrm{an}}_{\geq p^{-r}}$. The result follows.

Now suppose that $\ell = p$. Let $\mathfrak{U}$ denote the entirety of the locus in $X_0(p)^{\mathrm{an}}_{\mathbb{Q}_p}$ defined by $v(E) \leq r$. Once we verify that π restricts to

$$X_1(p,p^2)^{\mathrm{an}}_{\geq p^{-p^2 r}} \to \mathfrak{U},$$

the argument may proceed exactly as above. We claim, moreover, that if (E,P,C) is a point in $X_0(p,p^2)^{\mathrm{an}}_{\geq p^{-p^2 r}}$, then $v(E/C) = v(E)/p^2$. This would follow if we knew that C met the canonical subgroup of E trivially (again by standard facts about quotienting by canonical and noncanonical subgroups of order p, as in [Buzzard 2003, Section 3]), so it suffices to prove that $\langle P\rangle$ *is* the canonical subgroup of E.

The natural map

$$X_1(p, p^2) \to X_0(p), \quad (E, P, C) \mapsto (E, \langle P \rangle)$$

restricts to $X_1(p, p^2)^{\mathrm{an}}_{\geq p^{-r}} \to X_0(p)^{\mathrm{an}}_{\geq p^{-r}}$ by the same connectivity argument used in the $\ell \neq p$ case (since this map clearly doesn't change $v(E)$). But it is well known that the locus $X_0(p)^{\mathrm{an}}_{\geq p^{-r}}$ consists of pairs (E, C) with C equal to the canonical subgroup of E. □

We may pull back the Eisenstein family of Lemma 3.1 for $\ell \neq p$ through the map π to arrive at an invertible function on $X_1(p, \ell^2)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n$. By the previous lemma, we may also pull back the family for $\ell = p$ through π to arrive at an invertible function on $X_0(p, p^2)^{\mathrm{an}}_{\geq p^{-pr_n}} \times \mathcal{W}^0_n$. For any ℓ, it follows from (3) that the function $\pi^*\mathbf{E}_{\ell^2}$ satisfies

$$\pi^*\mathbf{E}_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_p, \langle q_{\ell^2} \rangle) = \frac{E(q_{\ell^2})}{E((q_{\ell^2})^{\ell^2})} = \frac{E(q_{\ell^2})}{E(q)}.$$

To arrive at the functions we need, we simply multiply $\pi^*\mathbf{E}_{\ell^2}$ and Θ_{ℓ^2} (which is constant in the weight). Of course, to do so we must first pull these functions back so that they lie on a common curve. The natural ("smallest") curve to use depends on whether or not $p = 2$, since 2 already lies in the Γ_1 part of the level of Θ_{ℓ^2}. The next proposition summarizes the properties of the resulting functions.

Proposition 3.3. *Let p be and ℓ be primes. There exists an element $\mathbf{H}_{\ell^2}$ of*

$$\begin{cases} H^0(X_1(4p, \ell^2)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0, \mathcal{O}(\pi_1^*\Sigma_{4p} - \pi_2^*\Sigma_{4p})) & \text{if } p \neq 2, \\ H^0(X_1(4, \ell^2)^{\mathrm{an}}_{\geq 1} \times \mathcal{W}^0, \mathcal{O}(\pi_1^*\Sigma_4 - \pi_2^*\Sigma_4)) & \text{if } p = 2 \end{cases}$$

whose q-expansion at

$$\begin{cases} (\underline{\mathrm{Tate}}(q), \mu_{4p}, \langle q_{\ell^2} \rangle)) & \text{if } p \neq 2, \\ (\underline{\mathrm{Tate}}(q), \mu_4, \langle q_{\ell^2} \rangle)) & \text{if } p = 2 \end{cases} \quad \text{is equal to} \quad \frac{E(q_{\ell^2})\theta(q_{\ell^2})}{E(q)\theta(q)}.$$

Moreover, there exists a sequence of rational numbers r_n such that

$$1/(1+p) > r_1 \geq r_2 \geq \cdots > 0$$

with $r_i < p^{2-i}/\mathbf{q}(1+p)$ such that $\mathbf{H}_{\ell^2}$, when restricted to $\mathcal{W}^0_n$, analytically continues to the region

$$\begin{cases} X_1(4p, \ell^2)^{\mathrm{an}}_{\geq p^{-r_n}} \times \mathcal{W}^0_n & \text{if } p \neq 2,\ \ell \neq p, \\ X_1(4p, p^2)^{\mathrm{an}}_{\geq p^{-pr_n}} \times \mathcal{W}^0_n & \text{if } p \neq 2,\ \ell = p, \\ X_1(4, \ell^2)^{\mathrm{an}}_{\geq 2^{-r_n}} \times \mathcal{W}^0_n & \text{if } p = 2,\ \ell \neq 2, \\ X_1(4, 4)^{\mathrm{an}}_{\geq 2^{-2r_n}} \times \mathcal{W}^0_n & \text{if } p = \ell = 2. \end{cases}$$

Finally, we wish to extend $\mathbf{H}_{\ell^2}$ and $E(q)$ to all of $\mathcal{W}$. To do this, we simply pull back through the natural map

$$\mathcal{W} \to \mathcal{W}^0, \quad \kappa \mapsto \kappa \circ \langle \cdot \rangle. \tag{4}$$

When restricted to $\mathcal{W}^i$, this map is simply the isomorphism $\kappa \mapsto \kappa/\tau^i$.

Remark 3.4. We have chosen in the end to use Γ_1-structure on the curves on which the $\mathbf{H}_{\ell^2}$ lie both to rigidify the associated moduli problems over $\mathbb{Q}_p$, as well as because these are the curves that will actually turn up in the sequel. We note, however, that the $\mathbf{H}_{\ell^2}$ are invariant under all diamond automorphisms.

4. The spaces of forms

In this section we define spaces of overconvergent p-adic modular forms as well as families thereof over admissible open subsets of $\mathcal{W}$. Again, the motivating idea behind these definitions is that we have reduced to weight 0 via division by the well-understood forms $E_\lambda\theta$. By "well-understood" we essentially mean two things. The first is that we understand their zeros once we eliminate part of the supersingular locus (and thereby *remove* the zeros of the Eisenstein part). The second is that, by the previous section, we know that there are modular functions with q-expansions

$$\frac{E_\lambda(q_{\ell^2})\theta(q_{\ell^2})}{E_\lambda(q)\theta(q)}$$

that interpolate rigid-analytically in λ, a fact that we will need to define Hecke operators on families in the next section.

Before defining the spaces of forms, we need to make a couple of remarks about diamond automorphisms. For a positive integer N and an element $d \in (\mathbb{Z}/N\mathbb{Z})^\times$, let $\langle d\rangle$ denote the usual diamond automorphism of $X_1(N)$ given on (noncuspidal) points by $(E, P) \mapsto (E, dP)$. Now suppose we are given a factorization $N = N_1N_2$ into relatively prime factors, so the natural reduction map

$$(\mathbb{Z}/N\mathbb{Z})^\times \xrightarrow{\sim} (\mathbb{Z}/N_1\mathbb{Z})^\times \times (\mathbb{Z}/N_2\mathbb{Z})^\times$$

is an isomorphism. For $a \in (\mathbb{Z}/N_1\mathbb{Z})^\times$ and $b \in (\mathbb{Z}/N_2\mathbb{Z})^\times$ we let $(a, b) \in (\mathbb{Z}/N\mathbb{Z})^\times$ denote the inverse image of the pair (a, b) under the this map. For $a \in (\mathbb{Z}/N_1\mathbb{Z})^\times$, we define $\langle a\rangle_{N_1} := \langle (a, 1)\rangle$, and we refer to these automorphisms as *the diamond automorphisms at* N_1. The diamond automorphisms at N_2 are defined similarly, and we have a factorization $\langle d\rangle = \langle d\rangle_{N_1} \circ \langle d\rangle_{N_2}$. Finally, we observe that the diamond operators on $X_1(4N)^{\mathrm{an}}_K$ preserve the subspaces $X_1(4N)^{\mathrm{an}}_{\geq p^{-r}}$ and the divisor Σ_{4N} in the sense that $\langle d\rangle^{-1}(X_1(4N)^{\mathrm{an}}_{\geq p^{-r}}) = X_1(4N)^{\mathrm{an}}_{\geq p^{-r}}$ and $\langle d\rangle^*\Sigma_{4N} = \Sigma_{4N}$, respectively.

Convention 4.1. By the symbol $\mathcal{O}(\Sigma)$ for a $\mathbb{Q}$-divisor Σ we shall always mean $\mathcal{O}(\lfloor \Sigma \rfloor)$, where $\lfloor \Sigma \rfloor$ is the divisor obtained by taking the floor of each coefficient occurring in Σ.

First we define the spaces of forms of fixed weight. Let N be a positive integer and suppose that either $p \nmid 4N$ or that $p = 2$ and $p \nmid N$.

Definition 4.2. Let $\kappa \in \mathcal{W}^i(K)$ and pick n such that $\kappa \in \mathcal{W}^i_n$. Then, for any rational number r with $0 \leq r \leq r_n$, we define the space of p-adic half-integral weight modular forms of weight κ, tame level $4N$ (or rather N if $p = 2$), and growth condition p^{-r} over K to be

$$\widetilde{M}_\kappa(4N, K, p^{-r}) := \begin{cases} H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}, \mathcal{O}(\Sigma_{4Np}))^{\tau^i} \times \{\kappa\} & \text{if } p \neq 2, \\ H^0(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}, \mathcal{O}(\Sigma_{4N}))^{(-1/\,\cdot\,)^i \tau^i} \times \{\kappa\} & \text{if } p = 2, \end{cases}$$

where $(\,\cdot\,)^{\tau^i}$ denotes the τ^i eigenspace for the action of the diamond automorphisms at p, and similarly for $(-1/\,\cdot\,)^i \tau^i$ if $p = 2$.

Remarks 4.3.

- For $p \neq 2$, we have chosen to remove p from the level and only indicate the tame level in the notation because, as we will see, these spaces contain forms of all p-power level. However, for $p = 2$, we have left the 4 in as a reminder that the forms have at least a 4 in the level, as well as for some uniformity in notation.
- Note that this space has been "tagged" with the weight κ because the actual space has only a rather trivial dependence on κ (κ serves only to restrict the admissible K and r and to determine i). The point is that, as we will see, the Hecke action on this space is very sensitive to κ. The tag will generally be ignored in what follows as the weight will be clear from the context.
- This space is endowed with a norm which is defined as in Section 2.2 and is a Banach space over K with respect to this norm.
- We call the forms belonging to spaces with $r > 0$ *overconvergent*. The space of all overconvergent forms (of this weight and level) is the inductive limit

$$\widetilde{M}^\dagger_\kappa(4N, K) = \lim_{r \to 0} \widetilde{M}_\kappa(4N, K, p^{-r}).$$

- In case κ is the character associated to an integer $\lambda \geq 0$, the space of forms defined above would classically be thought of as having weight $\lambda + 1/2$. Our choice of p-adic weight character bookkeeping seems to be the most natural one (the Shimura lifting has the effect of squaring the weight character, for example).
- In case κ is the weight associated to an integer $\lambda \geq 0$, then the definition here is somewhat less general than the definition of the space of forms of weight $\lambda + 1/2$ contained in [Ramsey 2006], due to the need to eliminate enough of

the supersingular locus to get rid of the Eisenstein zeros. The two definitions are (Hecke-equivariantly) isomorphic whenever they are both defined, as we will see in Proposition 6.2.

- The tilde is an homage to the metaplectic literature and will be used henceforth on all half-integral weight objects in order to distinguish them from their integral weight counterparts.

We now turn to the spaces of families of modular forms.

Definition 4.4. Let X be a connected affinoid subdomain of $\mathcal{W}$. Then $X \subseteq \mathcal{W}^i$ for some i since X is connected, and $X \subseteq \mathcal{W}^i_n$ for some n since X is affinoid. For any rational number r with $0 \leq r \leq r_n$, we define the space of families of half-integral weight modular forms of tame level $4N$ and growth condition p^{-r} on X to be

$$\widetilde{M}_X(4N, K, p^{-r}) := \begin{cases} H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}, \mathcal{O}(\Sigma_{4Np}))^{\tau^i} \widehat{\otimes}_K \mathcal{O}(X) & \\ & \text{if } p \neq 2, \\ H^0(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}, \mathcal{O}(\Sigma_{4N}))^{(-1/\cdot)^i \tau^i} \widehat{\otimes}_K \mathcal{O}(X) & \\ & \text{if } p = 2. \end{cases}$$

Remarks 4.5.

- We endow $\widetilde{M}_X(4N, K, p^{-r})$ with the completed tensor product norm obtained from the norms defined in Section 2.2 and from the supremum norm on $\mathcal{O}(X)$. The space $\widetilde{M}_X(4N, K, p^{-r})$ with this norm is a Banach module over the Banach algebra $\mathcal{O}(X)$.
- As in the case of fixed weight, the definition depends rather trivially on X, but the Hecke action will be very sensitive to X.
- In general, if X is an affinoid subdomain of $\mathcal{W}$, we define $\widetilde{M}_X$ to be the direct sum of the spaces corresponding to the connected components of X. Also, just as for particular weights, we can talk about the space of all overconvergent families of forms on X, namely $\widetilde{M}^\dagger_X(4N, K) = \lim_{r\to 0} \widetilde{M}_X(4N, K, p^{-r})$.
- Using a simple projector argument, one sees easily that we have a canonical identification

$$H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}, \mathcal{O}(\Sigma_{4Np}))^{\tau^i} \widehat{\otimes}_K \mathcal{O}(X) \\ \cong (H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}, \mathcal{O}(\Sigma_{4Np})) \widehat{\otimes}_K \mathcal{O}(X))^{\tau^i},$$

  and similarly at level $4N$ if $p = 2$. This will prove useful in the next section.

For each X as above and each L-valued point $\kappa \in X$, evaluation at x induces a specialization map $\widetilde{M}_X(4N, K, p^{-r}) \to \widetilde{M}_\kappa(4N, L, p^{-r})$. In the next section we will define a Hecke action on both of these spaces for which such specialization maps are equivariant and which recover the usual Hecke operators on the right side above (in the sense that they are given by the usual formulas on q-expansions).

Each of the spaces of forms that we have defined has a cuspidal subspace consisting of forms that "vanish at the cusps." This notion is a little subtle in half-integral weight because there are often cusps at which *all* forms are *forced* to vanish. To explain this comment and motivate the subsequent definition of the space of cusp forms, let us go back to the motivation behind our definitions of the spaces of forms. If F is a form of half-integral weight in our setting, then $F\theta E$ (where E is an appropriate Eisenstein series) is what we would "classically" like to think of as a half-integral weight form. Indeed, if F is classical (this notion is defined in Section 6), then $F\theta E$ can literally be identified with a classical holomorphic modular form of half-integral weight over $\mathbb{C}$. The condition $\operatorname{div}(F) \geq -\Sigma_{4Np}$ (we are assuming $p \neq 2$ for the sake of this motivation) in our definition is exactly the condition that $F\theta E$ be holomorphic at all cusps. Likewise, the condition that this inequality be strict at all cusps is the condition that $F\theta E$ be cuspidal. But since $\operatorname{div}(F)$ has integral coefficients, the nonstrict inequality *implies* the strict inequality at all cusps where Σ_{4Np} has nonintegral coefficients.

With this in mind, we are led to the following definition of cusp forms. For an integer M, let C_{4M} be the divisor on $X_1(4M)^{\mathrm{an}}_{\mathbb{Q}_p}$ given by the sum of the cusps at which Σ_{4M} has integral coefficients. To define the cuspidal subspace of any of the above spaces of forms, we replace the divisor Σ_{4Np} (respectively Σ_{4N} if $p=2$) by the divisor $\Sigma_{4Np} - C_{4Np}$ (respectively $\Sigma_{4N} - C_{4N}$ if $p=2$). We will denote the cuspidal subspaces by the letter S instead of M. Thus, for example, if $\kappa \in \mathcal{W}^i_n(K)$ and $0 \leq r \leq r_n$, we define

$$\widetilde{S}_\kappa(4N, K, p^{-r}) = \begin{cases} H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}, \mathcal{O}(\Sigma_{4Np} - C_{4Np}))^{\tau^i} \times \{\kappa\} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } p \neq 2, \\ H^0(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}, \mathcal{O}(\Sigma_{4N} - C_{4N}))^{(-1/\cdot)^i \tau^i} \times \{\kappa\} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } p = 2. \end{cases}$$

Remarks 4.3 and 4.5 apply equally well to the corresponding spaces of cusp forms.

5. Hecke operators

Before we construct Hecke operators, we need to make some remarks on diamond operators and nebentypus. Since the p-part of the nebentypus character is encoded as part of the p-adic weight character, we need to separate out the tame part of the diamond action. Fix a weight $\kappa \in \mathcal{W}^i(K)$. To define the tame diamond operators compatibly with the classical definitions and those in [Ramsey 2006], we must twist (at least in the case $p \neq 2$) those obtained via pullback from the automorphism

$\langle\,\cdot\,\rangle_{4N}$ by $(-1/\,\cdot\,)^i$. That is, for $F \in \widetilde{M}_\kappa(4N, K, p^{-r})$, we define

$$\langle d\rangle_{4N,\kappa} F = \left(\frac{-1}{d}\right)^i \langle d\rangle^*_{4N} F \quad \text{if } p \neq 2,$$
$$\langle d\rangle_{N,\kappa} F = \langle d\rangle^*_N F \qquad\qquad \text{if } p = 2.$$

Without this twist in the $p \neq 2$ case, the definition would not agree with the classical one because of the particular nature of the automorphy factor of the form θ used in the identification of our forms with classical forms. The same formulas define operators $\langle\,\cdot\,\rangle_{4N,\,X}$ and $\langle\,\cdot\,\rangle_{N,\,X}$ on the space of families of modular forms over $X \subseteq \mathcal{W}^i$. For a more general $X \subseteq \mathcal{W}$, we break into the components in $\mathcal{W}^i$ for each i and define $\langle\,\cdot\,\rangle_{4N,\,X}$ and $\langle\,\cdot\,\rangle_{N,\,X}$ component by component. For a character χ modulo $4N$ (respectively modulo N if $p = 2$), we define the space of forms of tame nebentypus χ to be the χ-eigenspace of $\widetilde{M}_\kappa(4N, K, p^{-r})$ for the operators $\langle\,\cdot\,\rangle_{4N,\,\kappa}$ (respectively $\langle\,\cdot\,\rangle_{N,\,\kappa}$ if $p = 2$). The same definition applies to families of forms. These subspaces are denoted by appending a χ to the list of arguments (for example, $\widetilde{M}_\kappa(4N, K, p^{-r}, \chi)$).

Let $\mathcal{X}$ and $\mathcal{Y}$ be rigid spaces equipped with a pair of maps $\pi_1, \pi_2 : \mathcal{X} \to \mathcal{Y}$, and let D be a $\mathbb{Q}$-divisor on $\mathcal{Y}$ such that $\pi_1^*D - \pi_2^*D$ has integral coefficients. Let $\mathcal{Z} \subseteq \mathcal{X}$ be an admissible affinoid open, and let $H \in H^0(\mathcal{Z}, \mathcal{O}(\pi_1^*D - \pi_2^*D))$. Let $\mathcal{U}, \mathcal{V} \subseteq \mathcal{Y}$ be admissible affinoid opens such that $\pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z} \subseteq \pi_2^{-1}(\mathcal{U}) \cap \mathcal{Z}$, and suppose that $\pi_1 : \pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z} \to \mathcal{V}$ is finite and flat. Then there is a well-defined map $H^0(\mathcal{U}, \mathcal{O}(D)) \to H^0(\mathcal{V}, \mathcal{O}(D))$ given by the composition

$$\begin{tikzcd}
H^0(\mathcal{U}, \mathcal{O}(D)) \arrow[r, "\pi_2^*"] & H^0(\pi_2^{-1}(\mathcal{U}) \cap \mathcal{Z}, \mathcal{O}(\pi_2^*D)) \arrow[r, "\mathrm{res}"] & H^0(\pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z}, \mathcal{O}(\pi_2^*D)) \arrow[d, "\cdot H"'] \\
& H^0(\mathcal{V}, \mathcal{O}(D)) & H^0(\pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z}, \mathcal{O}(\pi_1^*D)) \arrow[l, "\pi_{1*}"']
\end{tikzcd}$$

in which π_{1*} is the trace map corresponding to the finite and flat map π_1.

5.1. ***Hecke operators for a fixed weight.*** Let N be as above, let ℓ be any prime number, and let

$$\pi_1, \pi_2 : \begin{cases} X_1(4Np, \ell^2)^{\mathrm{an}}_K \to X_1(4Np)^{\mathrm{an}}_K & \text{if } p \neq 2, \\ X_1(4N, \ell^2)^{\mathrm{an}}_K \to X_1(4N)^{\mathrm{an}}_K & \text{if } p = 2 \end{cases}$$

be the maps defined on noncuspidal points of the underlying moduli problem by

$$\pi_1 : (E, P, C) \mapsto (E, P) \quad \text{and} \quad \pi_2 : (E, P, C) \mapsto (E/C, P/C).$$

Suppose that $\ell \neq p$. Then

$$\begin{cases} \pi_1^{-1}(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}) = \pi_2^{-1}(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}) & \text{if } p \neq 2, \\ \pi_1^{-1}(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}) = \pi_2^{-1}(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}) & \text{if } p = 2 \end{cases}$$

for any $r < p/(1+p)$, since quotienting an elliptic curve by a subgroup of order prime to p does not change its measure of singularity. Fix a weight $\kappa \in \mathcal{W}^i(K)$, and let $\mathbf{H}_{\ell^2}(\kappa)$ denote the specialization of $\mathbf{H}_{\ell^2}$ to $\kappa \in \mathcal{W}$ (which, recall, is defined to be the specialization of $\mathbf{H}_{\ell^2}$ to $\kappa/\tau^i \in \mathcal{W}^0$). Pick n such that $\kappa \in \mathcal{W}^i_n$, and suppose $0 \le r \le r_n$. Apply the general construction above with the following table:

	$p \neq 2$	$p = 2$
$\mathcal{X}$	$X_1(4Np, \ell^2)^{\mathrm{an}}_K$	$X_1(4N, \ell^2)^{\mathrm{an}}_K$
$\mathcal{Y}$	$X_1(4Np)^{\mathrm{an}}_K$	$X_1(4N)^{\mathrm{an}}_K$
$\mathcal{Z}$	$X_1(4Np, \ell^2)^{\mathrm{an}}_{\ge p^{-r}}$	$X_1(4N, \ell^2)^{\mathrm{an}}_{\ge 2^{-r}}$
D	Σ_{4Np}	Σ_{4N}
H	$\mathbf{H}_{\ell^2}(\kappa)$	$\mathbf{H}_{\ell^2}(\kappa)$
$\mathcal{U} = \mathcal{V}$	$X_1(4Np)^{\mathrm{an}}_{\ge p^{-r}}$	$X_1(4N)^{\mathrm{an}}_{\ge 2^{-r}}$

Then we arrive at an endomorphism of the K-vector space

$$\begin{cases} H^0(X_1(4Np)^{\mathrm{an}}_{\ge p^{-r}}, \mathcal{O}(\Sigma_{4Np})) & \text{if } p \neq 2, \\ H^0(X_1(4N)^{\mathrm{an}}_{\ge 2^{-r}}, \mathcal{O}(\Sigma_{4N})) & \text{if } p = 2. \end{cases}$$

We may easily check that, since the diamond operators act trivially on $\mathbf{H}_{\ell^2}$ (see Remark 3.4), this endomorphism commutes with the action of the diamond operators, and therefore induces an endomorphism of $\widetilde{M}_\kappa(4N, K, p^{-r})$. We define T_{ℓ^2} (or U_{ℓ^2} if $\ell \mid 4N$) to be the quotient of this endomorphism by ℓ^2.

Now suppose that $\ell = p$. Note that

$$\begin{cases} \pi_1^{-1}(X_1(4Np)^{\mathrm{an}}_{\ge p^{-p^2 r}}) \subseteq \pi_2^{-1}(X_1(4Np)^{\mathrm{an}}_{\ge p^{-r}}) & \text{if } p \neq 2, \\ \pi_1^{-1}(X_1(4N)^{\mathrm{an}}_{\ge 2^{-2^2 r}}) \subseteq \pi_2^{-1}(X_1(4N)^{\mathrm{an}}_{\ge 2^{-r}}) & \text{if } p = 2 \end{cases}$$

for any $r < 1/p(1+p)$. This follows from repeated application of the observation (made, for example, in [Buzzard 2003, Theorem 3.3(v)]) that if $v(E) < p/(1+p)$ and C is a subgroup of order p other than the canonical subgroup, then $v(E/C) = v(E)/p$ and the canonical subgroup of E/C is $E[p]/C$.

If $\kappa \in \mathcal{W}^i_n$ and r is chosen so that $0 \le r \le r_n$, then we may apply the construction above with the table

	$p \neq 2$	$p = 2$
$\mathcal{X}$	$X_1(4Np, p^2)^{\mathrm{an}}_K$	$X_1(4N, 4)^{\mathrm{an}}_K$
$\mathcal{Y}$	$X_1(4Np)^{\mathrm{an}}_K$	$X_1(4N)^{\mathrm{an}}_K$
$\mathcal{Z}$	$X_1(4Np, p^2)^{\mathrm{an}}_{\ge p^{-pr}}$	$X_1(4N, 4)^{\mathrm{an}}_{\ge 2^{-2r}}$
D	Σ_{4Np}	Σ_{4N}
H	$\mathbf{H}_{p^2}(\kappa)$	$\mathbf{H}_4(\kappa)$
$\mathcal{U}$	$X_1(4Np)^{\mathrm{an}}_{\ge p^{-r}}$	$X_1(4N)^{\mathrm{an}}_{\ge 2^{-r}}$
$\mathcal{V}$	$X_1(4Np)^{\mathrm{an}}_{\ge p^{-pr}}$	$X_1(4N)^{\mathrm{an}}_{\ge 2^{-2r}}$

to arrive at a linear map

$$\begin{cases} H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}, \mathcal{O}(\Sigma_{4Np})) \to H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-pr}}, \mathcal{O}(\Sigma_{4Np})) & \text{if } p \neq 2, \\ H^0(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}, \mathcal{O}(\Sigma_{4N})) \to H^0(X_1(4N)^{\mathrm{an}}_{\geq 2^{-2r}}, \mathcal{O}(\Sigma_{4N})) & \text{if } p = 2. \end{cases}$$

This map commutes with the diamond operators and restricts to a map

$$\widetilde{M}_\kappa(4N, K, p^{-r}) \to \widetilde{M}_\kappa(4N, K, p^{-pr}).$$

When composed with the natural restriction map

$$\widetilde{M}_\kappa(4N, K, p^{-pr}) \to \widetilde{M}_\kappa(4N, K, p^{-r}) \tag{5}$$

and divided by p^2, we arrive at an endomorphism of $\widetilde{M}_\kappa(4N, K, p^{-r})$, which we denote by U_{p^2}.

Proposition 5.1. *The Hecke operators defined above are continuous.*

Proof. Each of the spaces arising in the construction is a Banach space over K, so it suffices to show that each of the constituent maps of which our Hecke operators are the composition has finite norm. By Lemma 2.5 we may ignore the residue disks around the cusps when computing norms, thereby reducing ourselves to the supremum norm on functions. It follows easily that the pullback, restriction, and trace maps have norm not exceeding 1 and that multiplication by H has norm not exceeding the supremum norm of H on the complement of the residue disks around the cusps. The latter is finite since this complement is affinoid. □

Remarks 5.2.
- In the overconvergent case, that is, when we have $r > 0$, the restriction map (5) is compact; see [Coleman 1997, Proposition A5.2]. It follows that U_{p^2} is compact since it is the composition of a continuous map with a compact map.
- The Hecke operators T_{ℓ^2} and U_{ℓ^2} preserve the space of cusp forms, as can be seen by simply constructing them directly on this space in the same manner as above. The operator U_{p^2} is compact on a space of overconvergent cusp forms.

5.2. ***Hecke operators in families.*** Let $X \subseteq \mathcal{W}$ be a connected admissible affinoid open. We wish to define endomorphisms of $\widetilde{M}_X(4N, K, p^{-r})$ that interpolate the endomorphisms T_{ℓ^2} and U_{ℓ^2} constructed above for fixed weights $\kappa \in X$.

Suppose that $\ell \neq p$, and adopt the table

	$p \neq 2$	$p = 2$
$\mathcal{U} = \mathcal{V}$	$X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}$	$X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}$
$\mathcal{Z}$	$X_1(4Np, \ell^2)^{\mathrm{an}}_{\geq p^{-r}}$	$X_1(4N, \ell^2)^{\mathrm{an}}_{\geq 2^{-r}}$
Σ	Σ_{4Np}	Σ_{4N}

For more compact notation, let us for the rest of this section define

$$
\begin{aligned}
M &= H^0(\mathcal{U}, \mathcal{O}(\Sigma)), & P &= H^0(\pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z}, \mathcal{O}(\pi_1^*\Sigma - \pi_2^*\Sigma)),\\
N &= H^0(\pi_2^{-1}(\mathcal{U}) \cap \mathcal{Z}, \mathcal{O}(\pi_2^*\Sigma)), & Q &= H^0(\pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z}, \mathcal{O}(\pi_1^*\Sigma)).\\
L &= H^0(\pi_1^{-1}(\mathcal{V}) \cap \mathcal{Z}, \mathcal{O}(\pi_2^*\Sigma)), & &
\end{aligned}
$$

The Hecke operator T_{ℓ^2} (or U_{ℓ^2} if $\ell \mid 4N$) at a fixed weight was constructed in the previous section by first taking the composition of the following continuous maps: a pullback $M \to N$, a restriction $N \to L$, multiplication by an element of $H \in P$ to arrive at an element of Q, and a trace $Q \to M$. The construction was completed by restricting to an eigenspace of the diamond operators at p and dividing by ℓ^2.

The module of families of forms on X is an eigenspace of $M\widehat{\otimes}_K\mathcal{O}(X)$ (by the final remark in Remarks 4.5). To define T_{ℓ^2} (or U_{ℓ^2}) we begin as in the fixed weight case by defining an endomorphism of $M\widehat{\otimes}_K\mathcal{O}(X)$ and then observing that it commutes with the diamond automorphisms and therefore restricts to an operator on families of modular forms. To define this endomorphism, we modify the above sequence of maps by first applying $\widehat{\otimes}_K\mathcal{O}(X)$ to all of the spaces and taking the unique continuous $\mathcal{O}(X)$-linear extension of each map, with the exception of the multiplication step, where we opt instead to multiply by $\mathbf{H}_{\ell^2}|_X \in P\widehat{\otimes}_K\mathcal{O}(X)$. In so doing, we arrive at an $\mathcal{O}(X)$-linear endomorphism of $M\widehat{\otimes}_K\mathcal{O}(X)$ that is easily seen to commute with the diamond automorphisms, thereby inducing an endomorphism of the module $\widetilde{M}_X(4N, K, p^{-r})$.

Lemma 5.3. *The Hecke operators defined above for families are continuous.*

Proof. By definition, each map arising in the construction is continuous except perhaps for the multiplication map. The proof of the continuity of this map requires several simple facts about completed tensor products, all of which can be found in [Bosch et al. 1984, Section 2.1.7].

It follows trivially from Lemma 2.5 that the multiplication map $L \times P \to Q$ is a bounded K-bilinear map and therefore extends uniquely to a bounded K-linear map $L\widehat{\otimes}_K P \to Q$. Extending scalars to $\mathcal{O}(X)$ and completing, we arrive at a bounded $\mathcal{O}(X)$-linear map $(L\widehat{\otimes}_K P)\widehat{\otimes}_K\mathcal{O}(X) \to Q\widehat{\otimes}_K\mathcal{O}(X)$. There is an isometric isomorphism $(L\widehat{\otimes}_K P)\widehat{\otimes}_K\mathcal{O}(X) \cong (L\widehat{\otimes}_K\mathcal{O}(X))\widehat{\otimes}_{\mathcal{O}(X)}(P\widehat{\otimes}_K\mathcal{O}(X))$, so we conclude that the $\mathcal{O}(X)$-bilinear multiplication map

$$(L\widehat{\otimes}_K\mathcal{O}(X))\widehat{\otimes}_{\mathcal{O}(X)}(P\widehat{\otimes}_K\mathcal{O}(X)) \to Q\widehat{\otimes}_K\mathcal{O}(X)$$

is bounded. In particular, multiplication by $H \in P\widehat{\otimes}_K\mathcal{O}(X)$ is a bounded (and hence continuous) map $\cdot H : L\widehat{\otimes}_K\mathcal{O}(X) \to Q\widehat{\otimes}_K\mathcal{O}(X)$, as desired. □

Remarks 5.4.
- The construction of a continuous endomorphism U_{p^2} is entirely analogous, and once again we find that U_{p^2} is compact in the overconvergent case, that is, whenever $r > 0$.

- The endomorphisms T_{ℓ^2} and U_{ℓ^2} can be extended to $\widetilde{M}_X(4N, K, p^{-r})$ for general admissible affinoid opens X in the usual manner, working component by component.
- All of the Hecke operators defined on families preserve the cuspidal subspaces, as a direct construction on these spaces demonstrates. Again, the operator U_{p^2} is compact on a module of overconvergent cusp forms.

Effect on q-expansions. In this section we will work out the effect of the Hecke operators that we have defined on q-expansions. As in [Ramsey 2006], we must adjust the naive q-expansions obtained by literally evaluating our forms on Tate curves with level structure to get at the classical q-expansions. In particular, by the q-expansion of a form $F \in \widetilde{M}_\kappa(4N, K, p^{-r})$ at the cusp associated to $(\underline{\mathrm{Tate}}(q), \zeta)$, where ζ is a primitive $4Np$-th root of unity if $p \neq 2$ and a primitive $4N$-th root of unity if $p = 2$, we mean $F(\underline{\mathrm{Tate}}(q), \zeta)\theta(q)E_\kappa(q)$. Similarly, for a family $F \in M_X(4N, K, p^{-r})$ the corresponding q-expansion is $F(\underline{\mathrm{Tate}}(q), \zeta)\theta(q)E(q)|_X$ and has coefficients in the ring of analytic functions on X.

Proposition 5.5. *Let F be an element of $\widetilde{M}_\kappa(4N, K, p^{-r})$ or $\widetilde{M}_X(4N, K, p^{-r})$, and let $\sum a_n q^n$ be the q-expansion of F at $(\underline{\mathrm{Tate}}(q), \zeta)$. Then the corresponding q-expansion of $U_{p^2}F$ is $\sum a_{p^2n}q^n$.*

Proof. We prove the theorem for U_{p^2} acting on $\widetilde{M}_\kappa(4N, K, p^{-r})$. To obtain the result for families, one could either proceed in the same manner or deduce the result for families over X from the result for fixed weight by specializing to weights in X. Let $F \in \widetilde{M}_\kappa(4N, K, p^{-r})$, and suppose that $F(\underline{\mathrm{Tate}}(q), \zeta)\theta(q)E_\kappa(q) = \sum a_n q^n$. The expansion we seek is $(1/p^2)\pi_{1*}(\pi_2^* F \cdot \mathbf{H}_{p^2}(\kappa))(\underline{\mathrm{Tate}}(q), \zeta) \cdot \theta(q)E_\kappa(q)$. The cyclic subgroups of order p^2 that intersect the subgroup generated by ζ trivially are exactly those of the form $\langle \zeta_{p^2}^i q_{p^2}\rangle$ for $0 \le i \le p^2 - 1$. Thus we have

$$
\begin{aligned}
&\pi_{1*}(\pi_2^* F \cdot \mathbf{H}_{p^2}(\kappa))(\underline{\mathrm{Tate}}(q), \zeta) \\
&\quad = \sum_{i=0}^{p^2-1} (\pi_2^* F \cdot \mathbf{H}_{p^2}(\kappa))(\underline{\mathrm{Tate}}(q), \zeta, \langle \zeta_{p^2}^i q_{p^2}\rangle) \\
&\quad = \sum_{i=0}^{p^2-1} F(\underline{\mathrm{Tate}}(q)/\langle \zeta_{p^2}^i q_{p^2}\rangle, \zeta/\langle \zeta_{p^2}^i q_{p^2}\rangle)\mathbf{H}_{p^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta, \langle \zeta_{p^2}^i q_{p^2}\rangle) \\
&\quad = \sum_{i=0}^{p^2-1} F(\underline{\mathrm{Tate}}(\zeta_{p^2}^i q_{p^2}), \zeta)\mathbf{H}_{p^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta, \langle \zeta_{p^2}^i q_{p^2}\rangle) \\
&\quad = \sum_{i=0}^{p^2-1} \frac{\sum a_n(\zeta_{p^2}^i q_{p^2})^n}{\theta(\zeta_{p^2}^i q_{p^2})E_\kappa(\zeta_{p^2}^i q_{p^2})} \frac{\theta(\zeta_{p^2}^i q_{p^2})E_\kappa(\zeta_{p^2}^i q_{p^2})}{\theta(q)E_\kappa(q)} = p^2 \frac{\sum a_{p^2n}q^n}{\theta(q)E_\kappa(q)}. \qquad \square
\end{aligned}
$$

The same analysis also proves the following.

Proposition 5.6. *Suppose $\ell \mid 4N$. Let F be an element of $\widetilde{M}_\kappa(4N, K, p^{-r})$ or $\widetilde{M}_X(4N, K, p^{-r})$, and let $\sum a_n q^n$ be the q-expansion of F at $(\underline{\mathrm{Tate}}(q), \zeta)$. Then the corresponding q-expansion of $U_{\ell^2} F$ is then $\sum a_{\ell^2 n} q^n$.*

To work out the effect of T_{ℓ^2} for $\ell \nmid 4Np$ on q-expansions, we will need several more q-expansions of Θ_{ℓ^2} and $\mathbf{E}_{\ell^2}$. For the former, see [Ramsey 2006]. The latter will follow from the following lemma. For $x \in \mathbb{Z}_p^\times$, we denote by $[x]$ the analytic function on $\mathcal{W}$ defined by $[x](\kappa) = \kappa(x)$.

Lemma 5.7. *For $\ell \neq p$, we have*

$$\mathbf{E}_\ell(\underline{\mathrm{Tate}}(q), \mu_p + \langle q_\ell \rangle) = [\langle \ell \rangle] \frac{E(q)}{E(q_\ell)} \quad \textit{and} \quad \mathbf{E}_\ell(\underline{\mathrm{Tate}}(q), \mu_{p\ell}) = \frac{E(q)}{E(q^\ell)}.$$

Proof. The second equality is how we chose to characterize $\mathbf{E}_\ell$ in the first place. We will use it to give an alternative characterization, which we will in turn use to prove the first equality.

By definition, $\mathbf{E}_\ell$ and the coefficients of $E(q)$ are pulled back from their restrictions to $\mathcal{W}^0$ through the map (4). Clearly $[\langle \ell \rangle]$ is the pullback of $[\ell]$ through this map, so it suffices to prove that $\mathbf{E}_\ell(\underline{\mathrm{Tate}}(q), \mu_p + \langle q_\ell \rangle) = [\ell](E(q)/E(q_\ell))$, where the coefficients are now thought of as functions only on $\mathcal{W}^0$. Moreover, it suffices to prove the equality after specialization to integers $\lambda \geq 2$ divisible by $\varphi(\mathbf{q})$, as such integers are Zariski-dense in $\mathcal{W}^0$. Let $E_\lambda(\tau)$ denote the classical analytic p-deprived Eisenstein series of weight λ and level p (normalized to have q-expansion $E_\lambda(q)$). Then

$$\mathbf{E}_\ell^{\mathrm{an}}(\lambda) := E_\lambda(\tau)/E_\lambda(\ell\tau)$$

is a meromorphic function on $X_0(p\ell)_{\mathbb{C}}^{\mathrm{an}}$ with rational q-expansion coefficients, and by GAGA and the q-expansion principle, it yields a rational function on the algebraic curve $X_0(p\ell)_{\mathbb{Q}_p}$. By comparing q-expansions it is evident that the restriction of this function to the region $X_0(p\ell)_{\geq 1}^{\mathrm{an}}$ is equal to the specialization, $\mathbf{E}_\ell(\lambda)$, of $\mathbf{E}_\ell$ to $\lambda \in \mathcal{W}^0$.

It follows that $\mathbf{E}_\ell(\lambda)(\underline{\mathrm{Tate}}(q), \mu_p + \langle q_\ell \rangle) = \mathbf{E}_\ell^{\mathrm{an}}(\lambda)(\underline{\mathrm{Tate}}(q), \mu_p + \langle q_\ell \rangle)$. The right side can be computed using the usual yoga where one pretends to specialize q to $e^{2\pi i \tau}$ and then computes with analytic transformation formulas (see [Ramsey 2006, Section 5] for a rigorous explanation of this yoga). So specializing, we get

$$\mathbf{E}_\ell^{\mathrm{an}}(\lambda)(\underline{\mathrm{Tate}}(q), \mu_p + \langle q_\ell \rangle)(\tau) = \mathbf{E}_\ell^{\mathrm{an}}(\lambda)(\mathbb{C}/\langle 1, \tau \rangle, \langle 1/p \rangle + \langle \tau/\ell \rangle).$$

Choosing a matrix

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \quad \text{such that } p \mid c \text{ and } \ell \mid d,$$

we arrive at an isomorphism

$$(\mathbb{C}/\langle 1,\tau\rangle, \langle 1/p\rangle + \langle \tau/\ell\rangle) \xrightarrow{\sim} (\mathbb{C}/\langle 1,\gamma\tau\rangle, \langle 1/p\ell\rangle), \quad z \mapsto \frac{z}{c\tau+d}.$$

Thus

$$\mathbf{E}_\ell^{\mathrm{an}}(\lambda)(\mathbb{C}/\langle 1,\tau\rangle, \langle 1/p\rangle + \langle \tau/\ell\rangle) = \mathbf{E}_\ell^{\mathrm{an}}(\lambda)(\mathbb{C}/\langle 1,\gamma\tau\rangle, \langle 1/p\ell\rangle) = \frac{E_\lambda(\gamma\tau)}{E_\lambda(\ell\gamma\tau)}.$$

Now $\ell\gamma\tau = ((a\ell)(\tau/\ell)+b)/(c(\tau/\ell)+d/\ell)$, so we have

$$\frac{E_\lambda(\gamma\tau)}{E_\lambda(\ell\gamma\tau)} = \frac{(c\tau+d)^\lambda E_\lambda(\tau)}{((c\tau+d)/\ell)^\lambda E_\lambda(\tau/\ell)} = \ell^\lambda \frac{E_\lambda(\tau)}{E_\lambda(\tau/\ell)}.$$

The result follows. □

Proposition 5.8. *Let $F \in \widetilde{M}_\kappa(4N, K, p^{-r}, \chi)$ with $\kappa \in \mathcal{W}^i$, and let $\sum a_n q^n$ be the q-expansion of F at $(\underline{\mathrm{Tate}}(q), \zeta)$. Then the corresponding q-expansion of $T_{\ell^2}F$ is $\sum b_n q^n$, where*

$$b_n = a_{\ell^2 n} + \kappa(\ell)\chi(\ell)\ell^{-1}\Big(\frac{(-1)^i n}{\ell}\Big)a_n + \kappa(\ell)^2\chi(\ell)^2\ell^{-1}a_{n/\ell^2}.$$

Let $F \in \widetilde{M}_X(4N, K, p^{-r}, \chi)$ with X a connected affinoid in $\mathcal{W}^i$, and let the q-expansion of F be $\sum a_n q^n$ as above. Then the corresponding q-expansion of $T_{\ell^2}F$ is $\sum b_n q^n$, where

$$b_n = a_{\ell^2 n} + [\ell]\chi(\ell)\ell^{-1}\Big(\frac{(-1)^i n}{\ell}\Big)a_n + [\ell]^2\chi(\ell)^2\ell^{-1}a_{n/\ell^2}.$$

Proof. We prove the first assertion. The second may either be proved directly in the same manner or simply deduced from the first via specialization to individual weights in X. Let $\kappa \in \mathcal{W}(K)$, let $F \in \widetilde{M}_\kappa(4N, K, p^{-r}, \chi)$, and let

$$F(\underline{\mathrm{Tate}}(q), \zeta)\theta(q)E_\kappa(q) = \sum a_n q^n$$

be the q-expansion of F at $(\underline{\mathrm{Tate}}(q), \zeta)$. The corresponding q-expansion of $T_{\ell^2}F$ is

$$\frac{1}{\ell^2}\pi_{1*}(\pi_2^* F \cdot \mathbf{H}_{\ell^2}(\kappa)) \cdot \theta(q)E_\kappa(q). \tag{6}$$

The cyclic subgroups of $\underline{\mathrm{Tate}}(q)$ of order ℓ^2 are the subgroups

$$\mu_{\ell^2}, \quad \langle \zeta_{\ell^2}^i q_{\ell^2}\rangle_{0\le i\le \ell^2-1}, \quad \text{and} \quad \langle \zeta_{\ell^2}^j q_\ell\rangle_{1\le j\le \ell-1}.$$

We examine separately the contribution of each of these types of subgroups to $\pi_{1*}(\pi_2^* F \cdot \mathbf{H}_{\ell^2}(\kappa))$.

First, we have

$$
\begin{aligned}
&F(\underline{\mathrm{Tate}}(q)/\mu_{\ell^2}, \zeta/\mu_{\ell^2})\mathbf{H}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta, \mu_{\ell^2})\\
&\quad = F(\underline{\mathrm{Tate}}(q^{\ell^2}), \zeta^{\ell^2})\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \mu_{\ell^2})\pi^*\mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta_p, \mu_{\ell^2})\\
&\quad = F(\underline{\mathrm{Tate}}(q^{\ell^2}), \zeta^{\ell^2})\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \mu_{\ell^2})\\
&\qquad\qquad\qquad \times \mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q)/\mu_{\ell^2}, (\mu_p + \underline{\mathrm{Tate}}(q)[\ell^2])/\mu_{\ell^2})\\
&\quad = F(\underline{\mathrm{Tate}}(q^{\ell^2}), \zeta^{\ell^2})\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \mu_{\ell^2})\mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q^{\ell^2}), \mu_p + \langle q\rangle).
\end{aligned}
$$

From the definition (2) and Lemma 5.7, we have

$$
\begin{aligned}
&\mathbf{E}_{\ell^2}(\underline{\mathrm{Tate}}(q^{\ell^2}), \mu_p + \langle q\rangle)\\
&\quad = \mathbf{E}_{\ell}(\underline{\mathrm{Tate}}(q^{\ell^2}), \mu_p + \langle q^{\ell}\rangle)\mathbf{E}_{\ell}(\underline{\mathrm{Tate}}(q^{\ell^2})/\langle q^{\ell}\rangle, (\mu_p + \langle q\rangle)/\langle q^{\ell}\rangle)\\
&\quad = \mathbf{E}_{\ell}(\underline{\mathrm{Tate}}(q^{\ell^2}), \mu_p + \langle q^{\ell}\rangle)\mathbf{E}_{\ell}(\underline{\mathrm{Tate}}(q^{\ell}), \mu_p + \langle q\rangle)\\
&\quad = [\langle \ell\rangle]\frac{E(q^{\ell^2})}{E(q^{\ell})}\cdot[\langle \ell\rangle]\frac{E(q^{\ell})}{(q)} = [\langle \ell\rangle]^2\frac{E(q^{\ell^2})}{E(q)}.
\end{aligned}
$$

When specialized to κ, this becomes $\kappa(\langle \ell\rangle)^2 E_\kappa(q^{\ell^2})/E_\kappa(q)$. Referring to [Ramsey 2006], we find

$$\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \mu_{\ell^2}) = \ell\theta(q^{\ell^2})/\theta(q).$$

Thus the contribution of this first subgroup is

$$\frac{\chi(\ell^2)\tau(\ell^2)^i \sum a_n q^{\ell^2 n}}{\theta(q^{\ell^2})E_\kappa(q^{\ell^2})}\,\ell\,\frac{\theta(q^{\ell^2})}{\theta(q)}\kappa(\langle \ell\rangle)^2\frac{E_\kappa(q^{\ell^2})}{E_\kappa(q)} = (\kappa(\langle \ell\rangle)\chi(\ell)\tau(\ell)^i)^2\frac{\ell\sum a_n q^{\ell^2 n}}{\theta(q)E_\kappa(q)}.$$

The subgroups $\langle \zeta_{\ell^2}^a q_{\ell^2}\rangle$ contribute

$$
\begin{aligned}
&\sum_{a=0}^{\ell^2-1} F(\underline{\mathrm{Tate}}(q)/\langle \zeta_{\ell^2}^a q_{\ell^2}\rangle, \zeta/\langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\mathbf{H}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta, \langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\\
&\quad = \sum_{a=0}^{\ell^2-1} F(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \zeta)\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\\
&\qquad\qquad\qquad \times \pi^*\mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta_p, \langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\\
&\quad = \sum_{a=0}^{\ell^2-1} F(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \zeta)\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\\
&\qquad\qquad \times \mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q)/\langle \zeta_{\ell^2}^a q_{\ell^2}\rangle, (\mu_p + \underline{\mathrm{Tate}}(q)[\ell^2])/\langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\\
&\quad = \sum_{a=0}^{\ell^2-1} F(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \zeta)\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle \zeta_{\ell^2}^a q_{\ell^2}\rangle)\\
&\qquad\qquad\qquad \times \mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \mu_{p\ell^2}).
\end{aligned}
$$

By (2) we have

$$\begin{aligned}\mathbf{E}_{\ell^2}(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \mu_{p\ell^2}) &= \mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \mu_{p\ell})\mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2})/\mu_\ell, \mu_{p\ell^2}/\mu_\ell)\\ &= \mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_{\ell^2}^a q_{\ell^2}), \mu_{p\ell})\mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_\ell^a q_\ell), \mu_{p\ell})\\ &= \frac{E(\zeta_{\ell^2}^a q_{\ell^2})}{E(\zeta_\ell^a q_\ell)}\frac{E(\zeta_\ell^a q_\ell)}{E(q)} = \frac{E(\zeta_{\ell^2}^a q_{\ell^2})}{E(q)}.\end{aligned}$$

Referring to [Ramsey 2006], we find $\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle\zeta_{\ell^2}^a q_{\ell^2}\rangle) = \theta(\zeta_{\ell^2}^a q_{\ell^2})/\theta(q)$. Thus the total contribution of this collection of subgroups is

$$\sum_{a=0}^{\ell^2-1}\frac{\sum a_n(\zeta_{\ell^2}^a q_{\ell^2})^n}{\theta(\zeta_{\ell^2}^a q_{\ell^2})E_\kappa(\zeta_{\ell^2}^a q_{\ell^2})}\frac{\theta(\zeta_{\ell^2}^a q_{\ell^2})}{\theta(q)}\frac{E_\kappa(\zeta_{\ell^2}^a q_{\ell^2})}{E_\kappa(q)} = \ell^2\frac{\sum a_{\ell^2 n}q^n}{\theta(q)E_\kappa(q)}.$$

The subgroups $\langle\zeta_{\ell^2}^b q_\ell\rangle$ contribute

$$\begin{aligned}&\sum_{b=1}^{\ell-1} F(\underline{\mathrm{Tate}}(q)/\langle\zeta_{\ell^2}^b q_\ell\rangle, \zeta/\langle\zeta_{\ell^2}^b q_\ell\rangle)\mathbf{H}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta, \langle\zeta_{\ell^2}^b q_\ell\rangle)\\ &\quad= \sum_{b=1}^{\ell-1} F(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \zeta^\ell)\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle\zeta_{\ell^2}^b q_\ell\rangle)\\ &\qquad\qquad\qquad \times \pi^*\mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q), \zeta_p, \langle\zeta_{\ell^2}^b q_\ell\rangle)\\ &\quad= \sum_{b=1}^{\ell-1} F(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \zeta^\ell)\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle\zeta_{\ell^2}^b q_\ell\rangle)\\ &\qquad\qquad \times \mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(q)/\langle\zeta_{\ell^2}^b q_\ell\rangle, (\mu_p + \underline{\mathrm{Tate}}(q)[\ell^2])/\langle\zeta_{\ell^2}^b q_\ell\rangle)\\ &\quad= \sum_{b=1}^{\ell-1} F(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \zeta^\ell)\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle\zeta_{\ell^2}^b q_\ell\rangle)\\ &\qquad\qquad\qquad \times \mathbf{E}_{\ell^2}(\kappa)(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \mu_p + \langle q_\ell\rangle).\end{aligned}$$

By (2) and Lemma 5.7 we have

$$\begin{aligned}&\mathbf{E}_{\ell^2}(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \mu_p + \langle q_\ell\rangle)\\ &\quad= \mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \mu_p + \langle q\rangle)\mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_\ell^b q)/\mu_\ell, (\mu_p + \langle q_\ell\rangle)/\mu_\ell)\\ &\quad= \mathbf{E}_\ell(\underline{\mathrm{Tate}}(\zeta_\ell^b q), \mu_{p\ell})\mathbf{E}_\ell(\underline{\mathrm{Tate}}(q^\ell), \mu_p + \langle q\rangle)\\ &\quad= \frac{E(\zeta_\ell^b q)}{E(q^\ell)}\cdot[\langle\ell\rangle]\frac{E(q^\ell)}{E(q)} = [\langle\ell\rangle]\frac{E(\zeta_\ell^b q)}{E(q)}.\end{aligned}$$

When specialized to κ, this becomes $\kappa(\langle\ell\rangle)E_\kappa(\zeta_\ell^b q)/E_\kappa(q)$. Referring to [Ramsey 2006] we find

$$\Theta_{\ell^2}(\underline{\mathrm{Tate}}(q), \zeta_4, \langle\zeta_\ell^b q\rangle) = \left(\frac{-1}{\ell}\right)\mathfrak{g}_\ell(\zeta_\ell^b)\frac{\theta(\zeta_\ell^b q)}{\theta(q)}, \quad \text{where } \mathfrak{g}_\ell(\zeta) = \sum_{m=1}^{\ell-1}\left(\frac{m}{\ell}\right)\zeta^m$$

is the Gauss sum associated to the ℓ-th root of unity ζ. Thus the total contribution of this third collection of subgroups is

$$\begin{aligned}\sum_{b=1}^{\ell-1} & \frac{\chi(\ell)(-1/\ell)^i \tau(\ell)^i \sum a_n(\zeta_\ell^b q)^n}{\theta(\zeta_\ell^b q)E_\kappa(\zeta_\ell^b q)} \Big(\frac{-1}{\ell}\Big)\mathfrak{g}_\ell(\zeta_\ell^b)\frac{\theta(\zeta_\ell^b q)}{\theta(q)}\kappa(\langle\ell\rangle)\frac{E_\kappa(\zeta_\ell^b q)}{E_\kappa(q)} \\ &= \kappa(\langle\ell\rangle)\chi(\ell)\Big(\frac{-1}{\ell}\Big)^{i+1}\tau(\ell)^i \frac{\mathfrak{g}_\ell(\zeta_\ell)}{\theta(q)E_\kappa(q)}\sum_n a_n\Big(\sum_{b=1}^{\ell-1}\zeta_\ell^{bn}\Big(\frac{b}{\ell}\Big)\Big)q^n \\ &= \kappa(\langle\ell\rangle)\chi(\ell)\Big(\frac{-1}{\ell}\Big)^{i+1}\tau(\ell)^i \frac{\mathfrak{g}_\ell(\zeta_\ell)}{\theta(q)E_\kappa(q)}\sum_n a_n\Big(\frac{n}{\ell}\Big)\mathfrak{g}_\ell(\zeta_\ell)q^n \\ &= \kappa(\langle\ell\rangle)\chi(\ell)\Big(\frac{-1}{\ell}\Big)^{i}\tau(\ell)^i \frac{\ell\sum\big(\frac{n}{\ell}\big)a_nq^n}{\theta(q)E_\kappa(q)}.\end{aligned}$$

Adding all this up and plugging into (6), we see that the q-expansion of $T_{\ell^2}F$ is $\sum b_nq^n$, where

$$\begin{aligned}b_n &= a_{\ell^2n} + \kappa(\langle\ell\rangle)\ell^{-1}\chi(\ell)\Big(\frac{-1}{\ell}\Big)^i\tau(\ell)^i\Big(\frac{n}{\ell}\Big)a_n + \kappa(\langle\ell\rangle)^2\ell^{-1}\chi(\ell)^2\tau(\ell)^{2i}a_{n/\ell^2} \\ &= a_{\ell^2n} + \kappa(\ell)\ell^{-1}\chi(\ell)\Big(\frac{(-1)^i n}{\ell}\Big)a_n + \kappa(\ell)^2\ell^{-1}\chi(\ell)^2 a_{n/\ell^2}. \qquad \square\end{aligned}$$

6. Classical weights and classical forms

In this section we define classical subspaces of our spaces of modular forms and prove the following analog of Coleman's theorem on overconvergent forms of low slope. Throughout this section k will denote an odd positive integer and we set $\lambda = (k-1)/2$.

Theorem 6.1. *Let m be a positive integer, let $\psi : (\mathbb{Z}/\mathbf{q}p^{m-1}\mathbb{Z})^\times \to K^\times$ be a character, and define $\kappa(x) = x^\lambda\psi(x)$. If $F \in \widetilde{M}_\kappa^\dagger(4N, K)$ satisfies $U_{p^2}F = \alpha F$ with $v(\alpha) < 2\lambda - 1$, then F is classical.*

Our proof follows the approach of Kassaei [2006], which is modular in nature and builds the classical form by analytic continuation and gluing. The term "analytic continuation" has little meaning here since we have only defined our modular forms over restricted regions on the modular curve, owing to the need to avoid Eisenstein zeros. To get around this difficulty, we must invoke the formalism of [Ramsey 2006] for p-adic modular forms of *classical* half-integral weight.

Let N be a positive integer. In [Ramsey 2006] we defined the space of modular forms of weight $k/2$ and level $4N$ over a $\mathbb{Z}[1/4N]$-algebra R to be the R-module

$$\widetilde{M}'_{k/2}(4N, R) := H^0(X_1(4N)_R, \mathcal{O}(k\Sigma_{4N})).$$

Note that this space was denoted $M_{k/2}(4N, R)$ and $k\Sigma_{4N}$ was denoted $\Sigma_{4N,k}$. Roughly speaking, in this space of forms we have divided by θ^k to reduce to weight zero instead of $E_\lambda\theta$. Let $r \in [0, 1] \cap \mathbb{Q}$, and define

$$\widetilde{M}'_{k/2}(4Np^m, K, p^{-r}) = H^0(X_1(4Np^m)^{\text{an}}_{\geq p^{-r}}, \mathcal{O}(k\Sigma_{4Np^m})).$$

It is an easy matter to check that the construction of the Hecke operators T_{ℓ^2} and U_{p^2} in Section 5 (using $H = \Theta^k_{\ell^2}$) adapts to this space of forms and furnishes us with Hecke operators having the expected effect on q-expansions. We will briefly review the construction of U_{p^2} in this context later in this section.

The next proposition relates these spaces of p-adic modular forms to the ones defined in this paper, and will ensure that the latter spaces (and consequently the eigencurve defined later in this paper) see the classical half-integral weight modular forms of arbitrary p-power level. Note that this identification requires knowledge of the action of the diamond operators at p because this data is part of the p-adic weight character.

Proposition 6.2. *Let m be a positive integer, let $\psi : (\mathbb{Z}/\mathbf{q}p^{m-1}\mathbb{Z})^\times \to K^\times$ be a character, and define $\kappa(x) = x^\lambda\psi(x)$. Then, for $0 \leq r \leq r_m$, the space*

$$\widetilde{M}'(4Np^{m+1}/\mathbf{q}, K, p^{-r})^{\langle\cdot\rangle^*_{\mathbf{q}p^{m-1}}=\psi} = \begin{cases} \widetilde{M}'_{k/2}(4Np^m, K, p^{-r})^{\langle\cdot\rangle^*_{p^m}=\psi} & \text{if } p \neq 2 \\ \widetilde{M}'_{k/2}(2^{m+1}N, K, p^{-r})^{\langle\cdot\rangle^*_{2^{m+1}}=\psi} & \text{if } p = 2 \end{cases}$$

is isomorphic to $\widetilde{M}_\kappa(4N, K, p^{-r})$ in a manner compatible with the action of the Hecke operators and tame diamond operators.

Proof. Let i be such that $\kappa \in \mathcal{W}^i$. The complex-analytic modular forms θ^{k-1} and $E_{\kappa\tau^{-i}}$ are each of weight λ. If $p \neq 2$, then θ^{k-1} is invariant under the $\langle d\rangle^*_{\mathbf{q}p^{m-1}}$ while if $p = 2$ it has eigencharacter $(-1/\cdot)^i$. In both cases, $E_{\kappa\tau^{-i}}$ has eigencharacter $\psi\tau^{-i}$ for this action. Standard arguments using GAGA and the q-expansion principle show that the ratio $\theta^{k-1}/E_{\kappa\tau^{-i}}$ furnishes an algebraic rational function on $X_1(4Np^{m+1}/\mathbf{q})_K$. Passing to the p-adic analytification and then restricting to $X_1(4Np^{m+1}/\mathbf{q})^{\text{an}}_{\geq p^{-r}}$ shows this function has divisor $(k-1)\Sigma_{4Np^{m+1}/\mathbf{q}}$, since $E_{\kappa\tau^{-i}}$ is invertible in this region for r as in the statement of the proposition (because $\kappa \in \mathcal{W}_m$).

Suppose $F' \in \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r})$ is a form with eigencharacter ψ for $\langle\cdot\rangle^*_{\mathbf{q}p^{m-1}}$, and let

$$F = F' \cdot \frac{\theta^{k-1}}{E_{\kappa\tau^{-i}}}.$$

Then, for $d \in (\mathbb{Z}/\mathbf{q}p^{m-1}\mathbb{Z})^\times$ we have $\langle d\rangle^*_{\mathbf{q}p^{m-1}} F = \tau(d)^i(-1/\,\cdot\,)^i F$. In particular, F is fixed by $\langle d\rangle^*_{p^m}$ with $d \equiv 1 \pmod{\mathbf{q}}$. Consider now the map

$$X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\geq p^{-r}}/\{\langle d\rangle_{\mathbf{q}p^{m-1}} \mid d \equiv 1 (\operatorname{mod} \mathbf{q})\} \longrightarrow \begin{cases} X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}} & \text{if } p \neq 2, \\ X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}} & \text{if } p = 2 \end{cases} \tag{7}$$

induced by $(E, P) \mapsto (E, aP)$, where the integer a is chosen so that

$$\begin{aligned} a &\equiv p^{m-1} (\operatorname{mod} p^m) \quad \text{and } a \equiv 1 (\operatorname{mod} 4N) \quad \text{if } p \neq 2, \\ a &\equiv 2^{m-1} (\operatorname{mod} 2^{m+1}) \text{ and } a \equiv 1 (\operatorname{mod} N) \qquad \text{if } p = 2. \end{aligned}$$

The construction of the canonical subgroup of order $\mathbf{q}p^{m-1}$ (defined because $r \leq r_m < p^{2-m}/\mathbf{q}(1+p)$) ensures that this map is an isomorphism. For $p \neq 2$, this map pulls the divisor Σ_{4Np} back to Σ_{4Np^m}, so we conclude that F descends to a section of $\mathcal{O}(\Sigma_{4Np})$ on $X_1(4Np)^{\mathrm{an}}_{\geq p^{-r}}$ and that this section satisfies $\langle d\rangle^*_p F = \tau(d)^i F$ for all $d \in (\mathbb{Z}/\mathbf{q}\mathbb{Z})^\times$. For $p = 2$, this map pulls the divisor Σ_{4N} back to $\Sigma_{2^{m+1}N}$, so F descends to a section of $\mathcal{O}(\Sigma_{4N})$ on $X_1(4N)^{\mathrm{an}}_{\geq 2^{-r}}$, and this section satisfies $\langle d\rangle^*_4 F = \tau(d)^i(-1/d)^i F$ for all $d \in (\mathbb{Z}/\mathbf{q}\mathbb{Z})^\times$. Thus we may regard F as an element of $\widetilde{M}_\kappa(4N, K, p^{-r})$. Conversely, for $F \in \widetilde{M}_\kappa(4N, K, p^{-r})$, it is easy to see that

$$F \cdot \frac{E_{\kappa\tau^{-i}}}{\theta^{k-1}} \in \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r})^{\langle\,\cdot\,\rangle_{\mathbf{q}p^{m-1}} = \psi}$$

(where F is implicitly pulled back via the above map (7)) and that this furnishes an inverse to the above map $F' \mapsto F$. That these maps are equivariant with respect to the Hecke action is a formal manipulation with the setup in Section 5 used to define the action on both sides. That it is equivariant with respect to tame diamond operators is trivial, but relies essentially on the "twisted" convention for this action on $\widetilde{M}_\kappa(4N, K, p^{-r})$ (for $p \neq 2$). $\square$

In general, if $\mathcal{U}$ is a connected admissible open in $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_K$ containing $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\geq p^{-r}}$ and if $F \in \widetilde{M}_\kappa(4N, K, p^{-r})$ (with κ as in the previous proposition), we will say that F analytically continues to $\mathcal{U}$ if the corresponding form $F' \in \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r})$ analytically continues to an element of

$$H^0(\mathcal{U}, \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}})). \tag{8}$$

Note that, in case $\mathcal{U}$ is preserved by the diamond operators at p, this analytic continuation automatically lies in the ψ-eigenspace of (8) since $G - \langle d\rangle^*_{\mathbf{q}p^{m-1}} G$ vanishes on the nonempty admissible open $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\geq p^{-r}}$ for all d, and hence must vanish on all of $\mathcal{U}$. In particular, in case $\mathcal{U} = X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_K$ we make the following definition.

Definition 6.3. Let $\kappa(x) = x^\lambda \psi(x)$ be as in Proposition 6.2. We say an element $F \in \widetilde{M}_\kappa(4N, K)^\dagger$ is *classical* if it analytically continues in the sense described above to all of $X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}$, that is, if it is in the image of the (injective) map

$$\begin{aligned} H^0(X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}, \mathcal{O}(k\Sigma_{4Np}))^{\langle\cdot\rangle_{p^m}=\psi} \\ \to \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r_m})^{\langle\cdot\rangle_{p^m}=\psi} \cong \widetilde{M}_\kappa(4N, K, p^{-r_m}) \\ \hookrightarrow \widetilde{M}_\kappa(4N, K)^\dagger. \end{aligned}$$

The analytic continuation used to prove Theorem 6.1 will proceed in three steps. All of them involve the construction of the operator U_{p^2} on

$$\widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r}),$$

which goes as follows. Let

$$\pi_1, \pi_2 : X_1(4Np^{m+1}/\mathbf{q}, p^2)_K^{\mathrm{an}} \to X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}$$

be the usual pair of maps, and let Θ_{p^2} denote the rational function on $X_1(4, p^2)_{\mathbb{Q}}$ from Section 3. For any pair of admissible open $\mathcal{U}$ and $\mathcal{V}$ in $X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}$ with $\pi_1^{-1}\mathcal{V} \subseteq \pi_2^{-1}\mathcal{U}$, we have the map

$$\begin{aligned} H^0(\mathcal{U}, \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}})) &\to H^0(\mathcal{V}, \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}})) \\ F &\mapsto \frac{1}{p^2}\pi_{1*}(\pi_2^* F \cdot \Theta_{p^2}^k). \end{aligned}$$

Note that there is no need to introduce the space $\mathcal{Z}$ as in Section 5 since our "twisting" section $\Theta_{p^2}^k$ is defined on all of $X_1(4Np^{m+1}/\mathbf{q}, p^2)_K^{\mathrm{an}}$. Also, recall from Section 5 that if $0 \le r < 1/p(1+p)$, we have

$$\pi_1^{-1}(X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\ge p^{-p^2 r}}) \subseteq \pi_2^{-1}(X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\ge p^{-r}}).$$

Thus if $F \in \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r})$ with $r < 1/p(1+p)$ then $U_{p^2}F$ analytically continues to $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\ge p^{-p^2 r}}$. From this simple observation we get the first and easiest analytic continuation result.

Proposition 6.4. *Let $r > 0$, and let $F \in \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r})$. Suppose that there exists a polynomial $P(T) \in K[T]$ with $P(0) \ne 0$ such that $P(U_{p^2})F$ analytically continues to $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\ge p^{-1/(1+p)}}$. Then F analytically continues to this region as well.*

Proof. Write $P(T) = P_0(T) + a$ with $P_0(0) = 0$ and $a \ne 0$. Then

$$F = \frac{1}{a}(P(U_{p^2})F - P_0(U_{p^2})F).$$

If we have $0 < r < 1/p(1+p)$, then the right side analytically continues to $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\ge p^{-p^2 r}}$ and hence so does F. Since $r > 0$, we may repeat this process until we have analytically continued F to $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\geq p^{-s}}$ for some $s \geq 1/p(1+p)$. Now restrict F to $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\geq p^{-1/p^2(1+p)}}$ and apply the process once more to get the desired result. $\square$

The second analytic continuation step requires that we introduce some admissible opens in $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\mathbb{Q}_p}$ defined in [Buzzard 2003]. Use of the letter $\mathcal{W}$ in this part of the argument is intended to keep the notation parallel to that of [Buzzard 2003] and should not be confused with weight space. If $p \neq 2$, we let $\mathcal{W}_0 \subseteq X_1(4N, p)^{\mathrm{an}}_{\mathbb{Q}_p}$ denote the admissible open subspace whose points reduce to the irreducible component on the special fiber of $X_1(4N, p)$ in characteristic p that contains the cusp associated to the datum $(\underline{\mathrm{Tate}}(q), P, \mu_p)$ for some (equivalently, any) point of order $4N$ on $\underline{\mathrm{Tate}}(q)$. Alternatively, $\mathcal{W}_0$ can be characterized as the complement of the connected component of the ordinary locus in $X_1(4N, p)^{\mathrm{an}}_{\mathbb{Q}_p}$ containing the cusp associated to $(\underline{\mathrm{Tate}}(q), P, \langle q_p\rangle)$ for some (equivalently, any) choice of P. If $p = 2$, we let $\mathcal{W}_0 \subseteq X_1(N, 2)^{\mathrm{an}}_{\mathbb{Q}_p}$ denote the admissible open subspace whose points reduce to the irreducible component on the special fiber of $X_1(N, 2)$ in characteristic 2 that contains the cusp associated to the datum $(\underline{\mathrm{Tate}}(q), P, \mu_2)$ for some (equivalently, any) point of order N on $\underline{\mathrm{Tate}}(q)$. Alternatively, $\mathcal{W}_0$ can be characterized as the complement of the connected component of the ordinary locus in $X_1(N, 2)^{\mathrm{an}}_{\mathbb{Q}_p}$ containing the cusp associated to $(\underline{\mathrm{Tate}}(q), P, \langle q_2\rangle)$ for some (equivalently, any) choice of P. In particular, $\mathcal{W}_0$ always contains the entire supersingular locus. The reader concerned about problems with small N in these descriptions should focus on the "alternative" versions and the remarks in Section 2.1 about adding level structure and taking invariants.

Buzzard [2003] introduces a map $v' : \mathcal{W}_0 \to \mathbb{Q}$ defined as follows. If $x \in \mathcal{W}_0$ is a cusp, then set $v'(x) = 0$. Otherwise, $x \in \mathcal{W}_0$ corresponds to a triple $(E/L, P, C)$ with E/L an elliptic curve, P a point of order $4N$ (N if $p = 2$) on E, and $C \subset E$ a cyclic subgroup of order p. If E has bad or ordinary reduction, then set $v'(x) = 0$. Otherwise, if $0 < v(E) < p/(1+p)$, then E has a canonical subgroup H of order p, and we define

$$v'(x) = \begin{cases} v(E) & \text{if } H = C, \\ 1 - v(E/C) & \text{if } H \neq C. \end{cases}$$

Finally, if $v(E) \geq p/(1+p)$ we define $v'(x) = p/(1+p)$. Note that v' does not depend on the point P. For a nonnegative integer n, we let V_n denote the region in $\mathcal{W}_0$ defined by the inequality $v' \leq 1 - 1/p^{n-1}(1+p)$. Buzzard proves that V_n is an admissible affinoid open in $\mathcal{W}_0$ for each n, and that $\mathcal{W}_0$ is admissibly covered by the V_n.

Let

$$f : X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\mathbb{Q}_p} \to \begin{cases} X_1(4N, p)^{\mathrm{an}}_{\mathbb{Q}_p} & \text{if } p \neq 2, \\ X_1(N, 2)^{\mathrm{an}}_{\mathbb{Q}_p} & \text{if } p = 2 \end{cases}$$

denote the map characterized by

$$(E,P)\mapsto\begin{cases}(E/\langle 4NpP\rangle,\ p^mP/\langle 4NpP\rangle,\ \langle 4NP/\langle 4NpP\rangle\rangle & \text{if } p\neq 2,\\ (E/\langle 2NP\rangle,\ 2^{m+1}P/\langle 2NP\rangle,\ \langle NP/\langle 2NP\rangle\rangle) & \text{if } p=2\end{cases}$$

on noncuspidal points. Define $\mathcal{W}_1=f^{-1}(\mathcal{W}_0)$ and $Z_n=f^{-1}(V_n)$ for $n\geq 0$. It follows from the above that $\mathcal{W}_1$ is an admissible open in $X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}$ and that $\mathcal{W}_1$ is admissibly covered by the admissible opens Z_n. The latter are affinoid since f is finite.

Lemma 6.5. *The inclusion $\pi_1^{-1}(Z_{n+2})\subseteq\pi_2^{-1}(Z_n)$ holds for all $n\geq 0$.*

Proof. Since the maps π_1 and π_2 are finite, the stated inclusion is between affinoids and can be checked on noncuspidal points. Then the assertion follows immediately from two applications of [Buzzard 2003, Lemma 4.2(2)]. □

We can now state and prove the second analytic continuation result.

Proposition 6.6. *Let $r>0$, and let $F\in\widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q},K,p^{-r})$. Suppose that there exists a polynomial $P(T)\in K[T]$ with $P(0)\neq 0$ such that $P(U_{p^2})F$ extends to $\mathcal{W}_1$. Then F extends to this region as well.*

Proof. Note that

$$X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\geq p^{-1/(1+p)}}=Z_0\subseteq\mathcal{W}_1$$

so that by Proposition 6.4, F extends to Z_0. Now we proceed inductively to extend F to each Z_n. Let $P(T)=P_0(T)+a$ with $P_0(0)=0$ and $a\neq 0$. Then

$$F=\frac{1}{a}(P(U_{p^2})F-P_0(U_{p^2})F).$$

Suppose F extends to Z_n for some $n\geq 0$. By hypothesis, $P(U_{p^2})F$ extends to all of $\mathcal{W}_1$, and by the construction of U_{p^2} and Lemma 6.5, $P_0(U_{p^2})F$ extends to Z_{n+2}, and hence so does F. Thus by induction F extends to Z_n for all n, and since $\mathcal{W}_1$ is admissibly covered by the Z_n, F extends to $\mathcal{W}_1$. □

If $p\neq 2$ and $m=1$ (that is, if there is only one p in the level), then this is the end of the second analytic continuation step. In all other cases, techniques in [Buzzard 2003] allow us to analytically continue to more connected components of the ordinary locus. Define

$$\mathbf{m}=\mathrm{ord}_p(\mathbf{q}p^{m-1})=\begin{cases}m & \text{if } p\neq 2,\\ m+1 & \text{if } p=2.\end{cases}$$

We now follow Buzzard: For $0\leq r\leq\mathbf{m}$ let $\mathcal{U}_r$ denote the admissible open in $X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}$ whose noncuspidal points parameterize pairs (E,P) that are either supersingular or satisfy

$$H_{p^{\mathbf{m}-r}}(E)=\begin{cases}H_{p^{m-r}}(E)=\langle 4Np^rP\rangle & \text{if } p\neq 2,\\ H_{2^{m+1-r}}(E)=\langle N2^rP\rangle & \text{if } p=2.\end{cases}$$

We have

$$\mathcal{W}_1 = \mathcal{U}_0 \subseteq \mathcal{U}_1 \subseteq \cdots \subseteq \mathcal{U}_{\mathbf{m}} = X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}.$$

The last goal of the second step is to analytically continue eigenforms to $\mathcal{U}_{\mathbf{m}-1}$.

Lemma 6.7. *For $0 \le r \le \mathbf{m}-2$, we have $\pi_1^{-1}(\mathcal{U}_{r+1}) \subseteq \pi_2^{-1}(\mathcal{U}_r)$.*

Proof. As usual, it suffices to check this on noncuspidal points. Moreover, it suffices to check it on ordinary points, since the entire supersingular locus is contained in each $\mathcal{U}_r$. For brevity we will assume $p \neq 2$. The case $p = 2$ is proved in exactly the same manner. Let $(E, P, C) \in \pi_1^{-1}(\mathcal{U}_{r+1})$ be such a point. Then $H_{p^{m-r-1}}(E) = \langle 4Np^{r+1}P \rangle$, and since $r + 1 < m$, we conclude that $H_{p^{m-r-1}}(E) \cap C = 0$. Now [Buzzard 2003, Proposition 3.5] implies that $H_{p^r}(E/C)$ is indeed generated by the image of $4Np^r P$ in E/C, so $(E, P, C) \in \pi_2^{-1}(\mathcal{U}_r)$. □

Proposition 6.8. *Let $r > 0$, and let $F \in \widetilde{M}'_{k/2}(4Np^{m+1}/\mathbf{q}, K, p^{-r})$. Suppose that there exists a polynomial $P(T) \in K[T]$ with $P(0) \neq 0$ such that $P(U_{p^2})F$ extends to $\mathcal{U}_{\mathbf{m}-1}$. Then F extends to this region as well.*

Proof. Since $\mathcal{U}_0 = \mathcal{W}_1$, Proposition 6.6 ensures that F analytically continues to $\mathcal{U}_0$. Now we proceed inductively to extend F to each $\mathcal{U}_r$ for $0 \le r \le \mathbf{m} - 1$. Let $P(T) = P_0(T) + a$ with $P_0(0) = 0$ and $a \neq 0$. Then

$$F = \frac{1}{a}(P(U_{p^2})F - P_0(U_{p^2})F).$$

Suppose F extends to $\mathcal{U}_r$ for some $0 \le r \le \mathbf{m}-2$. By hypothesis, $P(U_{p^2})F$ extends to all of $\mathcal{U}_{\mathbf{m}-1}$, and by the construction of U_{p^2} and Lemma 6.7, $P_0(U_{p^2})F$ extends to $\mathcal{U}_{r+1}$, and hence so does F. Proceeding inductively, we see that F can be extended all the way to $\mathcal{U}_{\mathbf{m}-1}$. □

The third and most difficult analytic continuation step is to continue to the rest of the curve $X_1(4Np^{m+1}/\mathbf{q})_K^{\mathrm{an}}$. If $p \neq 2$, we let $\mathcal{V}_0$ denote the admissible open in $X_1(4N, p)_K^{\mathrm{an}}$ whose points reduce to the irreducible component on the special fiber in characteristic p that contains the cusp associated to $(\underline{\mathrm{Tate}}(q), P, \langle q_p \rangle)$ for some (equivalently, any) choice of P. On the other hand, if $p = 2$, we let $\mathcal{V}_0$ denote the admissible open in $X_1(N, 2)_K^{\mathrm{an}}$ whose points reduce to the irreducible component on the special fiber in characteristic 2 that contains the cusp associated to $(\underline{\mathrm{Tate}}(q), P, \langle q_2 \rangle)$ for some (equivalently, any) choice of P. Let $\mathcal{V}$ denote the preimage of $\mathcal{V}_0$ under the finite map

$$g : X_1(4Np^{m+1}/\mathbf{q})_{\mathbb{Q}_p}^{\mathrm{an}} \to \begin{cases} X_1(4N, p)_{\mathbb{Q}_p}^{\mathrm{an}} & \text{if } p \neq 2, \\ X_1(N, 2)_{\mathbb{Q}_p}^{\mathrm{an}} \text{if} & p = 2, \end{cases}$$

$$(E, P) \mapsto \begin{cases} (E, p^m P, \langle 4Np^{m-1}P \rangle) & \text{if } p \neq 2, \\ (E, 2^{m+1}P, \langle 2^m NP \rangle) & \text{if } p = 2. \end{cases}$$

Note that the preimage under g of the locus that reduces to the other component of $X_1(4N, p)_{\mathbb{F}_p}$ (or $X_1(N, 2)_{\mathbb{F}_2}$ if $p = 2$) is $\mathcal{U}_{\mathbf{m}-1}$, so in particular $\{\mathcal{U}_{\mathbf{m}-1}, \mathcal{V}\}$ is an admissible cover of $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_{\mathbb{Q}_p}$ and $\mathcal{U}_{\mathbf{m}-1} \cap \mathcal{V}$ is the supersingular locus.

For any subinterval $I \subseteq (p^{-p/(1+p)}, 1]$, let $\mathcal{V}I$ (respectively $\mathcal{U}_{\mathbf{m}-1}I$) denote the admissible open in $\mathcal{V}$ (respectively $\mathcal{U}_{\mathbf{m}-1}$) defined by the condition $p^{-v(E)} \in I$. Note that the complement of $\mathcal{U}_{\mathbf{m}-1}$ in $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_K$ is $\mathcal{V}[1, 1]$. Given a U_{p^2}-eigenform of suitably low slope, we will define a function on $\mathcal{V}[1, 1]$ and use the gluing techniques of [Kassaei 2006] to glue it to the analytic continuation of our eigenform to $\mathcal{U}_{\mathbf{m}-1}$ guaranteed by Proposition 6.6. These techniques rely heavily on the norms introduced in Section 2.2. The use of Lemma 2.5 to reduce these norms to the supremum norm on the complement of the residue disks around the cusps will be implicit in many of the estimates that follow.

Over $\mathcal{V}(p^{-1/p(1+p)}, 1]$ we have a section h to π_1 given on noncuspidal points by

$$\begin{aligned} h : \mathcal{V}(p^{-1/p(1+p)}, 1] &\to X_1(4Np^{m+1}/\mathbf{q}, p^2)^{\mathrm{an}}_K \\ (E, P) &\mapsto (E, P, H_{p^2}). \end{aligned}$$

By standard results on quotienting by the canonical subgroup [Buzzard 2003, Theorem 3.3], the composition $\pi_2 \circ h$ restricts to a map

$$Q : \mathcal{V}(p^{-r}, 1] \to \mathcal{V}(p^{-p^2 r}, 1] \tag{9}$$

for any $0 \le r \le 1/p(1 + p)$. Note that since Q preserves the property of having ordinary or supersingular reduction, Q restricts to a map $\mathcal{V}(p^{-r}, 1) \to \mathcal{V}(p^{-p^2 r}, 1)$. Define a meromorphic function ϑ on $\mathcal{V}(p^{-1/p(1+p)}, 1]$ by $\vartheta = h^*\Theta_{p^2}$, and note that

$$\begin{aligned} \operatorname{div}(\vartheta) &= h^*(\pi_2^* \Sigma_{4Np^{m+1}/\mathbf{q}} - \pi_1^* \Sigma_{4Np^{m+1}/\mathbf{q}}) \\ &= Q^* \Sigma_{4Np^{m+1}/\mathbf{q}} - \Sigma_{4Np^{m+1}/\mathbf{q}}. \end{aligned} \tag{10}$$

Let $F \in H^0(\mathcal{U}_{\mathbf{m}-1}, \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}}))$ and suppose that

$$U_{p^2} F = \alpha F + H$$

on $\mathcal{U}_{\mathbf{m}-1}$ for some *classical* form H and some $\alpha \neq 0$. Note that this condition makes sense because $\pi_1^{-1}(\mathcal{U}_{\mathbf{m}-1}) \subseteq \pi_2^{-1}(\mathcal{U}_{\mathbf{m}-1})$ by Lemma 6.7. For a pair $(E, P) \in \mathcal{U}_{\mathbf{m}-1}$ corresponding to a noncuspidal point, we have

$$F(E, P) = \frac{1}{\alpha p^2} \sum_C F(E/C, P/C)\Theta^k_{p^2}(E, P, C) - \frac{1}{\alpha} H(E, P), \tag{11}$$

where the sum is over the cyclic subgroups of order p^2 having trivial intersection with the group generated by P. Suppose that (E, P) corresponds to a point in $\mathcal{V}(p^{-1/p(1+p)}, 1)$. Then the subgroup generated by P has trivial intersection with the canonical subgroup H_{p^2}, and thus the canonical subgroup is among the subgroups occurring in the sum above. One can check using [Buzzard 2003, Theorem 3.3] that $(E/H_{p^2}, P/H_{p^2})$ corresponds to a point of $\mathcal{V}(p^{-p/(1+p)}, 1)$, while if $C \neq H_{p^2}$ is a cyclic subgroup of order p^2 with trivial intersection with $\langle P\rangle$, then $(E/C, P/C)$ corresponds to a point of $\mathcal{U}_{\mathbf{m}-1}(p^{-1/p(1+p)}, 1]$. Define F_1 on $\mathcal{V}(p^{-1/p(1+p)}, 1)$ by

$$F_1 = F - \frac{1}{\alpha p^2}\vartheta^k Q^*(F|_{\mathcal{V}(p^{-p/(1+p)},1)}).$$

Lemma 6.9. *The function F_1 on $\mathcal{V}(p^{-1/p(1+p)}, 1)$ extends to an element of*

$$H^0(\mathcal{V}(p^{-1/p(1+p)}, 1], \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}})).$$

Proof. Equation (11) and the comments that follow it show how to define the extension $\widetilde{F}_1$ of F_1, at least on noncuspidal points. For a pair (E, P) corresponding to a noncuspidal point of $\mathcal{V}(p^{-1/p(1+p)}, 1]$, we would like

$$\widetilde{F}_1(E, P) = \frac{1}{\alpha p^2}\sum_C F(E/C, P/C)\Theta^k_{p^2}(E, P, C) - \frac{1}{\alpha}H(E, P),$$

where the sum is over the cyclic subgroups of order p^2 of E not meeting $\langle P\rangle$ and not equal to $H_{p^2}(E)$. We can formalize this as follows.

The canonical subgroup of order p^2 furnishes a section to the finite map

$$\pi_1^{-1}(\mathcal{V}(p^{-1/p(1+p)}, 1]) \xrightarrow{\pi_1} \mathcal{V}(p^{-1/p(1+p)}, 1],$$

and section is an isomorphism onto a connected component of

$$\pi_1^{-1}(\mathcal{V}(p^{-1/p(1+p)}, 1]).$$

Let $\mathcal{Z}$ denote the complement of this connected component. Then π_1 restricts to a finite and flat map $\mathcal{Z} \to \mathcal{V}(p^{-1/p(1+p)}, 1]$. Note that

$$\mathcal{Z} = \pi_1^{-1}(\mathcal{V}(p^{-1/p(1+p)}, 1]) \cap \mathcal{Z} \subseteq \pi_2^{-1}(\mathcal{U}_{\mathbf{m}-1}(p^{-1/p(1+p)}, 1]) \cap \mathcal{Z},$$

as can be checked on noncuspidal points (see the comments following Equation (11)). Now we may apply the general construction of Section 5 with this $\mathcal{Z}$ and define

$$\widetilde{F}_1 = \frac{1}{\alpha p^2}\pi_{1*}(\pi_2^* F \cdot \Theta^k_{p^2}) - \frac{1}{\alpha}H.$$

Then $\widetilde{F}_1 \in H^0(\mathcal{V}(p^{-1/p(1+p)}, 1], \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}}))$, and Equation (11) shows that $\widetilde{F}_1$ extends F_1. $\square$

For $n \geq 1$, we define inductively an element F_n of

$$H^0(\mathcal{V}(p^{-1/p^{2n-1}(1+p)}, 1], \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}})),$$

where F_1 is as above and for $n \geq 1$, we set

$$F_{n+1} = F_1 + \frac{1}{\alpha p^2} \vartheta^k Q^*(F_n|_{\mathcal{V}(p^{-1/p^{2n+1}(1+p)},1]}).$$

Note that (9) and (10) show that the F_n do indeed lie in the spaces indicated. Our goal is to show that the sequence $\{F_n\}$, when restricted to $\mathcal{V}[1, 1]$, converges to an element of G of $H^0(\mathcal{V}[1, 1], \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}}))$ that glues to F in the sense that there exists a global section of $\mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}})$ that restricts to F and G on $\mathcal{U}_{\mathbf{m}-1}$ and $\mathcal{V}[1, 1]$, respectively. To do this we will use Kassaei's gluing lemma [2006]. The following lemmas furnish some necessary norm estimates.

Lemma 6.10. *The function Θ_{p^2} on $Y_1(4, p^2)_{\mathbb{Q}_p}$ is integral. That is, it extends to a regular function on the fine moduli scheme $Y_1(4, p^2)_{\mathbb{Z}_p}$.*

Proof. Each $\Gamma_1(4) \cap \Gamma^0(p^2)$ structure on the elliptic curve $\underline{\text{Tate}}(q)/\mathbb{Q}_p((q))$ lifts trivially to one over the Tate curve thought of as over $\mathbb{Z}_p((q))$. Since the Tate curve is ordinary, such a structure specializes to a unique component of the special fiber $Y_1(4, p^2)_{\mathbb{F}_p}$. Since $Y_1(4, p^2)_{\mathbb{Z}_p}$ is Cohen–Macaulay, the usual argument used to prove the q-expansion principal (as in the proof of [Katz 1973, Corollary 1.6.2]) shows that Θ_{p^2} is integral as long as it has integral q-expansion associated to a level structure specializing to each component of the special fiber. In fact, all q-expansions of Θ_{p^2} are computed explicitly in [Ramsey 2006, Section 5], and all are integral. □

Lemma 6.11. *Let R be an $\mathbb{F}_p$-algebra, let E be an elliptic curve over R, and let $E^{(p)}$ denote the base change of E via the absolute Frobenius morphism on* Spec(R). *Let* $\mathrm{Fr} : E \to E^{(p)}$ *denote the relative Frobenius morphism. Then for any point P of order* 4 *on E, we have $\Theta_{p^2}(E, P, \ker(\mathrm{Fr}^2)) = 0$.*

Proof. In characteristic p, the forgetful map $Y_1(4, p^2)_{\mathbb{F}_p} \to Y_1(4)_{\mathbb{F}_p}$ has a section given on noncuspidal points by $s : (E, P) \mapsto (E, P, \ker(\mathrm{Fr}^2))$. By Lemma 6.10, we may pull back (the reduction of) Θ_{p^2} through this section to arrive at a regular function on the smooth curve $Y_1(4)_{\mathbb{F}_p}$.

The q-expansion of $s^*\Theta_{p^2}$ at the cusp associated to $(\underline{\text{Tate}}(q), \zeta_4)$ is

$$s^*\Theta_{p^2}(\underline{\text{Tate}}(q), \zeta_4) = \Theta_{p^2}(\underline{\text{Tate}}(q), \zeta_4, (\ker(\mathrm{Fr}^2))).$$

Recall that the map $\underline{\text{Tate}}(q) \to \underline{\text{Tate}}(q^p)$ given by quotienting by μ_p is a lifting of Fr to characteristic zero (more specifically, to the ring $\mathbb{Z}((q))$). Thus the q-expansion we seek is the reduction of

$$\Theta_{p^2}(\underline{\text{Tate}}(q), \zeta_4, \mu_{p^2}) = p \frac{\sum_{n\in\mathbb{Z}} q^{p^2n^2}}{\sum_{n\in\mathbb{Z}} q^{n^2}}$$

modulo p, which is clearly zero. See [Ramsey 2006, Section 5] for the computation of the above q-expansion in characteristic zero. It follows from the q-expansion principle that $s^*\Theta_{p^2} = 0$, which implies our claim. $\square$

Lemma 6.12. *Let* $0 \le r < 1/p(1+p)$. *Then the section* ϑ *of*

$$\mathcal{O}(\Sigma_{4Np^{m+1}/\mathbf{q}} - Q^*\Sigma_{4Np^{m+1}/\mathbf{q}})$$

satisfies $\|\vartheta\|_{\mathcal{V}[p^{-r},1]} \le p^{pr-1}$.

Proof. By Lemma 2.5, we may, in computing the norm, ignore points reducing to cusps. Let $x \in \mathcal{V}[p^{-r}, 1]$ be outside of this collection of points, so x corresponds to a pair (E, P) with good reduction. Let H_{p^i} denote the canonical subgroup of E of order p^i (for whichever i this is defined). Let $\mathbf{E}$ be a smooth model of E over $\mathcal{O}_L$, and let $\mathbf{P}$ and $\mathbf{H}_{p^2}$ be the extensions of P and H_{p^2} to $\mathbf{E}$, respectively (these $\mathbf{E}$ and $\mathbf{H}$ should not be confused with the functions by the same name from Section 3).

By [Goren and Kassaei 2006, Theorem 3.10], $\mathbf{H}_p$ reduces modulo $p/p^{v(E)}$ to $\ker(\mathrm{Fr})$. Applying this to $\mathbf{E}/\mathbf{H}_p$, we see that $\mathbf{H}_{p^2}/\mathbf{H}_p$ reduces modulo $p/p^{v(E/H_p)}$ to $\ker(\mathrm{Fr})$ on the corresponding reduction of E/H_p. Then from [Buzzard 2003, Theorem 3.3], we know that $v(E/H_p) = pv(E)$, so $p^{1-v(E/H_p)} \mid p^{1-v(E)}$ and we may combine these statements to conclude that $\mathbf{H}_{p^2}$ reduces modulo $p^{1-pv(E)}$ to $\ker(\mathrm{Fr}^2)$ on the reduction of E.

Combining this with the integrality of Θ_{p^2} (from Lemma 6.10), we have

$$h(x) = \Theta_{p^2}(E, P, H_{p^2}) \equiv \Theta_{p^2}(E, P, \ker(\mathrm{Fr}^2)) \pmod{p^{1-pv(E)}}.$$

This is zero by Lemma 6.11, so $|h(x)| \le |p^{1-pv(E)}| = p^{pv(E)-1} \le p^{pr-1}$. $\square$

Proposition 6.13. *Let* $F \in H^0(\mathcal{U}_{\mathbf{m}-1}, \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}}))$. *Suppose* $U_{p^2}F - \alpha F$ *is classical for some* $\alpha \in K$ *with* $v(\alpha) < 2\lambda - 1$. *Then* F *is classical as well.*

Proof. Define F_n as above. We first show that the sequence $F_n|_{\mathcal{V}[1,1]}$ converges. Note that over $\mathcal{V}[1, 1]$ we have

$$\begin{aligned} F_{n+2} - F_{n+1} &= \Big(F_1 + \frac{1}{\alpha p^2}\vartheta^k Q^* F_{n+1}\Big) - \Big(F_1 + \frac{1}{\alpha p^2}\vartheta^k Q^* F_n\Big) \\ &= \frac{1}{\alpha p^2}\vartheta^k Q^*(F_{n+1} - F_n). \end{aligned}$$

By Lemma 6.12 (with $r = 0$), we have

$$\|F_{n+2} - F_{n+1}\|_{\mathcal{V}[1,1]} \le \frac{p^{2-k}}{|\alpha|}\|F_{n+1} - F_n\|_{\mathcal{V}[1,1]}.$$

The hypothesis on α ensures that $(p^{2-k}/|\alpha|)^n \to 0$ as $n \to \infty$ and hence that the sequence has successive differences that tend to zero. Since, by Lemma 2.1,

$H^0(\mathcal{V}[1,1], \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}}))$ is a Banach algebra with respect to $\|\cdot\|_{\mathcal{V}[1,1]}$, it follows that the sequence converges. Set

$$G = \lim_{n\to\infty} F_n|_{\mathcal{V}[1,1]}.$$

Next we apply Kassaei's gluing lemma [Kassaei 2006, Lemma 2.3] to glue G to F as sections of the line bundle $\mathcal{O}(\lfloor k\Sigma_{4Np^{m+1}/\mathbf{q}}\rfloor)$. So that we are gluing over an affinoid as required in the hypotheses of the gluing lemma, we first restrict F to $\mathcal{V}[p^{-1/p(1+p)}, 1)$ and glue G to this restriction to get a section over the smooth affinoid $\mathcal{V}[p^{-1/p(1+p)}, 1]$. Since the pair $\{\mathcal{V}[p^{-1/p(1+p)}, 1], \mathcal{U}_{\mathbf{m}-1}\}$ is an admissible cover of $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_K$, this section glues to F to give a global section.

The "auxiliary" approximating sections that are required in the hypotheses of this lemma (denoted F_n in [Kassaei 2006]) are the F_n introduced above. So that the F_n live on affinoids (as in the hypotheses of the gluing lemma), we simply restrict F_n to $\mathcal{V}[p^{-1/p^{2n}(1+p)}, 1]$. The two conditions to be verified are

$$\|F_n - F\|_{\mathcal{V}[p^{-1/p^{2n}(1+p)},1)} \to 0 \quad \text{and} \quad \|F_n - G\|_{\mathcal{V}[1,1]} \to 0.$$

The second of these is simply the definition of G. As for the first, it is not even clear that the indicated norms are finite (since the norms are over non-affinoids). To see that these norms are finite and that the ensuing estimates make sense, we must show that F has finite norm over $\mathcal{V}[p^{-1/p^2(1+p)}, 1)$. It suffices to show that the norms of F over the affinoids $\mathcal{V}_n = \mathcal{V}[p^{-1/p^{2n}(1+p)}, p^{-1/p^{2n+2}(1+p)}]$ are uniformly bounded for $n \geq 1$. The key is that the map Q restricts to a map $Q : \mathcal{V}_n \to \mathcal{V}_{n+1}$ for each $n \geq 1$. Since F_1 extends to the affinoid $\mathcal{V}[p^{-1/p^2(1+p)}, 1]$, its norms over the $\mathcal{V}_n$ are certainly uniformly bounded, say, by M. We have

$$\begin{aligned}
\|F\|_{\mathcal{V}_n} &\leq \max\Bigl(\|F_1\|_{\mathcal{V}_n}, \Bigl\|\frac{1}{\alpha p^2}\vartheta^k Q^* F\Bigr\|_{\mathcal{V}_n}\Bigr)\\
&\leq \max\Bigl(M, \frac{p^2}{|\alpha|}\|\vartheta^k\|_{\mathcal{V}_n}\|Q^*F\|_{\mathcal{V}_n}\Bigr)\\
&\leq \max\Bigl(M, \frac{p^2}{|\alpha|}\bigl(p^{1/(p^{2n-1}(1+p))-1}\bigr)^k\|Q^*F\|_{\mathcal{V}_n}\Bigr)\\
&\leq \max\Bigl(M, \frac{p^{2-k}}{|\alpha|}p^{k/(p^{2n-1}(1+p))}\|F\|_{\mathcal{V}_{n-1}}\Bigr).
\end{aligned}$$

Iterating this, we see that $\|F\|_{\mathcal{V}_n}$ does not exceed the maximum of

$$\max_{0\leq m\leq n-2}\Bigl(M\Bigl(\frac{p^{2-k}}{|\alpha|}\Bigr)^m p^{\frac{k}{1+p}\,(1/p^{2n-1}+\cdots+1/p^{2(n-m)+1})}\Bigr)$$

and

$$\Bigl(\frac{p^{2-k}}{|\alpha|}\Bigr)^{n-1} p^{\frac{k}{1+p}\,(1/p^{2n-1}+\cdots+1/p^3)}\|F\|_{\mathcal{V}_1}.$$

The sums in the exponents of are geometric and do not exceed $1/(p^3-p)$. Moreover, the hypothesis on α ensures that $p^{2-k}/|\alpha|<1$. Thus we have

$$\|F\|_{\mathcal{V}_n}\leq\max\Bigl(Mp^{\frac{k}{1+p}\frac{1}{p^3-p}},\, p^{\frac{k}{1+p}\frac{1}{p^3-p}}\|F\|_{\mathcal{V}_1}\Bigr),$$

which is independent of n, as desired. This ensures that all of the norms encountered below are indeed finite.

From the definition of the F_n, we have

$$\begin{aligned}F_{n+1}-F&=F_1+\frac{1}{\alpha p^2}\vartheta^k Q^*F_n-F\\&=F-\frac{1}{\alpha p^2}\vartheta^k Q^*F+\frac{1}{\alpha p^2}\vartheta^k Q^*F_n-F=\frac{1}{\alpha p^2}\vartheta^k Q^*(F_n-F).\end{aligned}$$

Taking supremum norms over the appropriate admissible opens, we see

$$\begin{aligned}&\|F_{n+1}-F\|_{\mathcal{V}[p^{-1/(p^{2n+2}(1+p))},1)}\\&\qquad\leq\frac{p^2}{|\alpha|}\|\vartheta\|^k_{\mathcal{V}[p^{-1/(p^{2n+2}(1+p))},1)}\|Q^*(F_n-F)\|_{\mathcal{V}[p^{-1/(p^{2n+2}(1+p))},1)}\\&\qquad\leq\frac{p^2}{|\alpha|}\bigl(p^{1/(p^{2n+1}(1+p))-1}\bigr)^k\|F_n-F\|_{\mathcal{V}[p^{-1/(p^{2n}(1+p)},1))}\\&\qquad=\frac{p^{2-k}}{|\alpha|}p^{k/(p^{2n+1}(1+p))}\|F_n-F\|_{\mathcal{V}[p^{-1/(p^{2n}(1+p))},1)}.\end{aligned}$$

Iterating this we find that

$$\begin{aligned}&\|F_n-F\|_{\mathcal{V}[p^{-1/(p^{2n}(1+p))},1)}\\&\qquad\leq\Bigl(\frac{p^{2-k}}{|\alpha|}\Bigr)^{n-1}p^{\frac{k}{1+p}\,(1/p^3+1/p^5+\cdots+1/p^{2n-1})}\|F_1-F\|_{\mathcal{V}[p^{-1/p^2(1+p)},1)}.\end{aligned}$$

Again the sum in the exponent is less than $1/(p^3-p)$ for all n, so the hypothesis on α ensures that the above norm tends to zero as $n\to\infty$, as desired □

We are now ready to prove the main result of this section, which is a mild generalization of Theorem 6.1.

Theorem 6.14. *Let m be a positive integer, let $\psi:(\mathbb{Z}/\mathbf{q}p^{m-1}\mathbb{Z})^\times\to K^\times$ be a character, and define $\kappa(x)=x^\lambda\psi(x)$. Let $P(T)\in K[T]$ be a monic polynomial all roots of which have valuation less than $2\lambda-1$. If $F\in\widetilde{M}^\dagger_\kappa(4N,K)$ and $P(U_{p^2})F$ is classical, then F is classical as well.*

Proof. Pick r in $0<r<r_m$ such that $F\in\widetilde{M}_\kappa(4N,K,p^{-r})$, and let $F'\in\widetilde{M}_{k/2}(4Np^{m+1}/\mathbf{q},K,p^{-r})$ be the form corresponding to F under the isomorphism of Proposition 6.2. We must show that F' is classical in the sense that it analytically continues to all of $X_1(4Np^{m+1}/\mathbf{q})^{\mathrm{an}}_K$. Note that $P(0)\neq 0$ for such a

polynomial, so by Proposition 6.8, F' analytically continues to an element of $H^0(\mathcal{U}_{\mathbf{m}-1}, \mathcal{O}(k\Sigma_{4Np^{m+1}/\mathbf{q}}))$. Now we proceed by induction on the degree d of P. The case $d = 1$ is Proposition 6.13. Suppose the result holds for some degree $d \geq 1$, and let $P(T)$ be a polynomial of degree $d + 1$ as above. We may pass to a finite extension and write $P(T) = (T - \alpha_1) \cdots (T - \alpha_{d+1})$. The condition that $P(U_{p^2})F'$ is classical implies by the inductive hypothesis that $(U_{p^2} - \alpha_{d+1})F'$ is classical. This implies that F' is classical by the case $d = 1$. □

Remark 6.15. The results of this section likely also follow from the very general classicality machinery developed in [Kassaei 2005], though we have not checked the details.

7. The half-integral weight eigencurve

To construct our eigencurve, we will use the axiomatic version of Coleman and Mazur's Hecke algebra construction, as set up in [Buzzard 2007]. We briefly recall some relevant details.

Let us for the moment allow $\mathcal{W}$ to be any reduced rigid space over K. Let $\mathbf{T}$ be a set with a distinguished element ϕ. Suppose that, for each admissible affinoid open $X \subseteq \mathcal{W}$, we are given a Banach module M_X over $\mathcal{O}(X)$ satisfying a certain technical hypothesis (called *(Pr)* in [Buzzard 2007]), and we are also given a map

$$\mathbf{T} \to \mathrm{End}_{\mathcal{O}(X)}(M_X), \quad t \mapsto t_X$$

whose image consists of commuting endomorphisms and such that ϕ_X is compact for each X. Assume that, for admissible affinoids $X_1 \subseteq X_2 \subseteq \mathcal{W}$, we are given a continuous injective $\mathcal{O}(X_1)$-linear map

$$\alpha_{12} : M_{X_1} \to M_{X_2} \widehat{\otimes}_{\mathcal{O}(X_2)} \mathcal{O}(X_1)$$

that is a "link" in the sense of [Buzzard 2007] and such that $(t_{X_2} \widehat{\otimes} 1) \circ \alpha_{12} = \alpha_{12} \circ t_{X_1}$. Assume moreover that, if $X_1 \subseteq X_2 \subseteq X_3 \subseteq \mathcal{W}$ are admissible affinoids, then $\alpha_{13} = \alpha_{23} \circ \alpha_{12}$ with the obvious notation. Note that the link condition ensures that the characteristic power series $P_X(T)$ of ϕ_X acting on M_X is independent of X in the sense that the image of $P_{X_2}(T)$ under the natural map $\mathcal{O}(X_2)[\![T]\!] \to \mathcal{O}(X_1)[\![T]\!]$ is $P_{X_1}(T)$; see [Buzzard 2007].

Out of this data, Buzzard constructs rigid analytic spaces D and Z, called the *eigenvariety* and *spectral variety*, respectively, equipped with canonical maps

$$D \to Z \to \mathcal{W}. \tag{12}$$

The points of D parameterize systems of eigenvalues of $\mathbf{T}$ acting on the $\{M_X\}$ for which the eigenvalue of ϕ is nonzero, in a sense that will be made precise in Lemma 7.3, while the image of such a point in Z simply records the inverse of the

ϕ eigenvalue and a point of $\mathcal{W}$. If $\mathcal{W}$ is equidimensional of dimension d, then the same is true of both of the spaces D and Z.

As the details of this construction will be required in the next section, we recall them here. The following is the deepest part of the construction.

Theorem 7.1 [Buzzard 2007, Theorem 4.6]. *Let R be a reduced affinoid algebra over K, let $P(T)$ be a Fredholm series over R, and let $Z \subset \mathrm{Sp}(R) \times \mathbb{A}^1$ denote the hypersurface cut out by $P(T)$ equipped with the projection $\pi : Z \to \mathrm{Sp}(R)$. Define $\mathcal{C}(Z)$ to be the collection of admissible affinoid opens Y in Z such that*

- *$Y' = \pi(Y)$ is an admissible affinoid open in* $\mathrm{Sp}(R)$,
- *$\pi : Y \to Y'$ is finite, and*
- *there exists $e \in \mathcal{O}(\pi^{-1}(Y'))$ such that $e^2 = e$ and Y is the zero locus of e.*

Then $\mathcal{C}(Z)$ is an admissible cover of Z.

We will generally take Y' to be connected in what follows. This is not a serious restriction, since Y is the disjoint union of the parts lying over the various connected components of Y'. We also remark that the third of the above conditions follows from the first two (this is observed in [Buzzard 2007], where references to the proof are supplied).

To construct D, first fix an admissible affinoid open $X \subseteq \mathcal{W}$. Let Z_X denote the zero locus of $P_X(T) = \det(1 - \phi_X T \mid M_X)$ in $X \times \mathbb{A}^1$, and let $\pi : Z_X \to X$ denote the projection onto the first factor. Let $Y \in \mathcal{C}(Z_X)$, let $Y' = \pi(Y)$ as above, and assume that Y' is connected. We wish to associate to Y a polynomial factor of $P_{Y'}(T) = \det(1 - (\phi_X \widehat{\otimes} 1)T \mid M_X \widehat{\otimes}_{\mathcal{O}(X)} \mathcal{O}(Y'))$. Since the algebra $\mathcal{O}(Y)$ is a finite and locally free module over $\mathcal{O}(Y')$, we may consider the characteristic polynomial Q' of $T \in \mathcal{O}(Y)$. Since T is a root of its characteristic polynomial, we have a map

$$\mathcal{O}(Y')[T]/(Q'(T)) \to \mathcal{O}(Y). \tag{13}$$

It is shown in [Buzzard 2007, Section 5] that this map is surjective and therefore an isomorphism since both sides are locally free of the same rank.

Now since the natural map $\mathcal{O}(Y')[T]/(Q'(T)) \to \mathcal{O}(Y')\{\{T\}\}/(Q'(T))$ is an isomorphism, it follows that $Q'(T)$ divides $P_{Y'}(T)$ in $\mathcal{O}(Y')\{\{T\}\}$. If a_0 is the constant term of $Q'(T)$, then this divisibility implies that a_0 is a unit. We set $Q(T) = a_0^{-1} Q'(T)$. The spectral theory of compact operators on Banach modules (see [Buzzard 2007, Theorem 3.3]) furnishes a unique decomposition

$$M_X \widehat{\otimes}_{\mathcal{O}(X)} \mathcal{O}(Y') \cong N \oplus F$$

into closed ϕ-invariant $\mathcal{O}(Y')$-submodules such that $Q^*(\phi)$ is zero on N and invertible on F. Moreover, N is projective of rank equal to the degree of Q, and the characteristic power series of ϕ on N is $Q(T)$. The projector $M_X \widehat{\otimes}_{\mathcal{O}(X)} \mathcal{O}(Y') \to N$

is in the closure of $\mathcal{O}(Y')[\phi]$, so N is stable under all of the endomorphisms associated to elements of $\mathbf{T}$. Let $\mathbf{T}(Y)$ denote the $\mathcal{O}(Y')$-subalgebra of $\mathrm{End}_{\mathcal{O}(Y')}(N)$ generated by these endomorphisms. Then $\mathbf{T}(Y)$ is finite over $\mathcal{O}(Y')$ and hence affinoid, so we may set $D_Y = \mathrm{Sp}(\mathbf{T}(Y))$. Because the leading coefficient of Q (that is, the constant term of Q^*) is a unit, there is an isomorphism

$$\mathcal{O}(Y')[T]/(Q(T)) \to \mathcal{O}(Y')[S]/(Q^*(S)), \quad T \mapsto S^{-1}.$$

Thus we obtain a canonical map $D_Y \to Y$, namely, the one corresponding to the map

$$\mathcal{O}(Y) \cong \mathcal{O}(Y')[T]/(Q(T)) \cong \mathcal{O}(Y')[S]/(Q^*(S)) \xrightarrow{S \mapsto \phi} \mathbf{T}(Y)$$

of affinoid algebras.

For general $Y \in \mathcal{C}(Z_X)$, we define D_Y to be the disjoint union of the affinoids defined above from the various connected components of Y'. We then glue the affinoids D_Y for $Y \in \mathcal{C}(Z_X)$ to obtain a rigid space D_X equipped with maps

$$D_X \to Z_X \to X.$$

Finally, we vary X and glue the desired spaces and maps above to obtain the spaces and maps in (12). This final step is where the links α_{ij} above come into play. See [Buzzard 2007] for details.

Definition 7.2. Let L be a complete extension of K. An L-valued system of eigenvalues of $\mathbf{T}$ acting on $\{M_X\}_X$ is a pair (κ, γ) consisting of a map of sets $\gamma : \mathbf{T} \to L$ and a point $\kappa \in \mathcal{W}(L)$ such that there exists an affinoid $X \subseteq \mathcal{W}$ containing κ and a nonzero element $m \in M_X \widehat{\otimes}_{\mathcal{O}(X),\kappa} L$ such that $(t_X \widehat{\otimes} 1)m = \gamma(t)m$ for all $t \in \mathbf{T}$. Such a system of eigenvalues is called ϕ-finite if $\gamma(\phi) \neq 0$.

Let x be an L-valued point of D. Then x lies over a point in $\kappa_x \in \mathcal{W}(L)$ that lies in X for some affinoid X, and x moreover lies in $D_Y(L)$ for some $Y \in \mathcal{C}(Z_X)$. Thus to x and the choice of X and Y corresponds a map $\mathbf{T}(Y) \to L$, and in particular a map of sets $\lambda_x : \mathbf{T} \to L$. Buzzard [2007] proves the following characterization of the points of D.

Lemma 7.3. *The correspondence $x \mapsto (\kappa_x, \lambda_x)$ is a well-defined bijective correspondence between L-valued points of D and ϕ-finite L-valued systems of eigenvalues of $\mathbf{T}$ acting on the $\{M_X\}$.*

In our case, we let $\mathcal{W}$ be weight space over $\mathbb{Q}_p$ as in Section 2.3, and let $\mathbf{T}$ be the set of symbols

$$\begin{cases} \{T_{\ell^2}\}_{\ell \nmid 4Np} \cup \{U_{\ell^2}\}_{\ell \mid 4Np} \cup \{\langle d \rangle_{4N}\}_{d \in (\mathbb{Z}/4N\mathbb{Z})^\times} & \text{if } p \neq 2, \\ \{T_{\ell^2}\}_{\ell \nmid 4N} \cup \{U_{\ell^2}\}_{\ell \mid 4N} \cup \{\langle d \rangle_{N}\}_{d \in (\mathbb{Z}/N\mathbb{Z})^\times} & \text{if } p = 2. \end{cases}$$

For an admissible affinoid open $X \subseteq \mathcal{W}$, we let $M_X = \widetilde{M}_X(4N, \mathbb{Q}_p, p^{-r_n})$, where n is the smallest positive integer such that $X \subseteq \mathcal{W}_n$. This module is a direct summand of the $\mathbb{Q}_p$-Banach space

$$\begin{cases} H^0(X_1(4Np)^{\mathrm{an}}_{\geq p^{-r_n}}, \mathcal{O}(\Sigma_{4Np}))\widehat{\otimes}_{\mathbb{Q}_p}\mathcal{O}(X) & \text{if } p \neq 2, \\ H^0(X_1(4N)^{\mathrm{an}}_{\geq 2^{-r_n}}, \mathcal{O}(\Sigma_{4N}))\widehat{\otimes}_{\mathbb{Q}_p}\mathcal{O}(X) & \text{if } p = 2 \end{cases}$$

and therefore satisfies property (*Pr*) since this latter space is potentially orthonormalizable in the terminology of [Buzzard 2007] by the discussion in [Serre 1962, Section 1]. We take the map $\mathbf{T} \to \mathrm{End}_{\mathcal{O}(X)}(M_X)$ to be the one sending each symbol to the endomorphism by that name defined in Section 5.

Let $X_1 \subseteq X_2 \subseteq \mathcal{W}$ be admissible affinoids, and let n_i be the smallest positive integer with $X_i \subseteq \mathcal{W}_{n_i}$. Then $n_1 \leq n_2$ so that $r_{n_2} \leq r_{n_1}$, and we have an inclusion

$$\widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-r_{n_1}}) \to \widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-r_{n_2}})$$

given by restriction. We define the required continuous injection α_{12} via the diagram

$$\begin{tikzcd} \widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-r_{n_1}}) \arrow[r] \arrow[dr, "\alpha_{12}"'] & \widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-r_{n_2}}) \\ & \widetilde{M}_{X_2}(4N, \mathbb{Q}_p, p^{-r_{n_2}})\widehat{\otimes}_{\mathcal{O}(X_2)}\mathcal{O}(X_1) \arrow[u, "\sim"'] \end{tikzcd}$$

and note that the required compatibility condition is satisfied. To see that these maps are links, choose numbers $r_{n_1} = s_0 \geq s_1 > s_2 > \cdots > s_{k-1} \geq s_k = r_{n_2}$ with the property that $p^2 s_{i+1} > s_i$ for all i. Then the map α_{12} factors as the composition of the maps

$$\widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-s_i}) \to \widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-s_{i+1}})$$

for $0 \leq i \leq k-2$ and the map

$$\widetilde{M}_{X_1}(4N, \mathbb{Q}_p, p^{-s_{k-1}}) \to \widetilde{M}_{X_2}(4N, \mathbb{Q}_p, p^{-s_k})\widehat{\otimes}_{\mathcal{O}(X_2)}\mathcal{O}(X_1).$$

The construction of U_{p^2} shows easily that each of these maps is a primitive link.

The result is that we obtain rigid analytic spaces $\widetilde{D}$ and $\widetilde{Z}$, which we call *the half-integral weight eigencurve* and *the half-integral weight spectral curve*, respectively. We also obtain canonical maps $\widetilde{D} \to \widetilde{Z} \to \mathcal{W}$. As usual, the tilde distinguishes these spaces from their integral weight counterparts first constructed in level 1 by Coleman and Mazur and later constructed for general level by Buzzard [2007].

If instead of using the full spaces of forms we use only the cuspidal subspaces everywhere, then we obtain cuspidal versions of all of the above spaces, which we will delineate with a superscript 0. Thus we have $\widetilde{D}^0$ and $\widetilde{Z}^0$ with the usual maps, and the points of these spaces parameterize systems of eigenvalues of the Hecke operators acting on the spaces of cusp forms by Lemma 7.3. We remark that there is a commutative diagram

$$\begin{tikzcd} \widetilde{D}^0 \arrow[rr] \arrow[d] & & \widetilde{D} \arrow[d] \\ \widetilde{Z}^0 \arrow[rr] \arrow[dr] & & \widetilde{Z} \arrow[dl] \\ & \mathcal{W} & \end{tikzcd}$$

where the horizontal maps are injections that identify the cuspidal spaces on the left with unions of irreducible components of the spaces on the right. Proving this is an exercise in the linear algebra that goes into the construction of these eigenvarieties and basic facts about irreducible components of rigid spaces found in [Conrad 1999], and is left to the reader.

For $\kappa \in \mathcal{W}(K)$, let $\widetilde{D}_\kappa$ and $\widetilde{D}^0_\kappa$ denote the fibers $\widetilde{D}$ and $\widetilde{D}^0$ over κ. The following theorem summarizes the basic properties of these eigencurves.

Theorem 7.4. *Let $\kappa \in \mathcal{W}(K)$. For a complete extension L/K, the correspondence $x \mapsto \lambda_x$ is a bijection between the L-valued points of the fiber $\widetilde{D}_\kappa(L)$ and the set of finite-slope systems of eigenvalues of the Hecke operators and tame diamond operators occurring on the space $\widetilde{M}^\dagger_\kappa(4N, L)$ of overconvergent forms of weight κ defined over L. The same statement holds with $\widetilde{D}$ replaced by $\widetilde{D}^0$ and $\widetilde{M}^\dagger_\kappa(4N, L)$ replaced by $\widetilde{S}^\dagger_\kappa(4N, L)$.*

Proof. We prove the statement for the full space of forms. The proof for cuspidal forms is identical. Fix $\kappa \in \mathcal{W}(K)$. Once we establish that the L-valued systems of eigenvalues of the form (κ, γ) occurring on the $\{M_X\}_X$ as defined above are exactly the systems of eigenvalues of the Hecke and tame diamond operators that occur on $\widetilde{M}^\dagger_\kappa(4N, L)$, the result is simply Lemma 7.3 "collated by weight." To see this one simply notes that, for any $f \in \widetilde{M}^\dagger_\kappa(4N, L)$, we have both $f \in \widetilde{M}_\kappa(4N, L, p^{-r_n})$ and $\kappa \in \mathcal{W}_n$ for n sufficiently large. In particular, if f is a nonzero eigenform for the Hecke and tame diamond operators, then the system of eigenvalues associated to f occurs in the module $M_{\mathcal{W}_n}$ for n sufficiently large. $\square$

We remark that the classicality result of Section 6 has the expected consequence that the collection of points of $\widetilde{D}$ corresponding to systems of eigenvalues occurring on classical forms is Zariski-dense in $\widetilde{D}$. This result is contained in [Ramsey 2007].

Appendix: Properties of the stack $X_1(Mp, p^2)$ over $\mathbb{Z}_{(p)}$

by Brian Conrad

In this appendix, we establish some geometric properties concerning the cuspidal locus in compactified moduli spaces for level structures on elliptic curves. We are especially interested in the case of nonétale p-level structures in characteristic p, so it is not sufficient to cite the work in [Deligne and Rapoport 1973] (which requires étale level structures in the treatment of moduli problems for generalized elliptic curves) or [Katz and Mazur 1985] (which works with Drinfeld structures over arbitrary base schemes but avoids nonsmooth generalized elliptic curves). The viewpoints of these works were synthesized in the study of moduli stacks for Drinfeld structures on generalized elliptic curves in [Conrad 2007], and we will use that reference — abbreviated as [C] — as our foundation in what follows.

Motivated by needs in the main text, for a prime p and an integer $M \geq 4$ not divisible by p, we consider the moduli stack $X_1(Mp^r, p^e)$ over $\mathbb{Z}_{(p)}$ that classifies triples (E, P, C), where E is a generalized elliptic curve over a $\mathbb{Z}_{(p)}$-scheme S, $P \in E^{\mathrm{sm}}(S)$ is a Drinfeld $\mathbb{Z}/Mp^r\mathbb{Z}$-structure on E^{sm}, and $C \subseteq E^{\mathrm{sm}}$ is a cyclic subgroup with order p^e such that some reasonable ampleness and compatibility properties for P and C are satisfied. (See Definition A.1 for a precise formulation of these additional properties.) The relevant case for applications to p-adic modular forms with half-integer weight is $e = 2$, but unfortunately such moduli stacks were only considered in [C] when either $r \geq e$ or $r = 0$. (This is sufficient for applications to Hecke operators, and avoids some complications.) We now need to allow $1 \leq r < e$, and the purpose of this appendix is to explain how to include such r and to record some consequences concerning the cusps in these cases. The consequence relevant in the main text is Theorem A.11. To carry out the proofs in this appendix we simply have to adapt some proofs in [C] rather than develop any essentially new ideas. For the convenience of the reader we will usually use [C] as a reference, though it must be stressed that many of the key notions were first introduced in the earlier works [Deligne and Rapoport 1973] and [Katz and Mazur 1985]. In the context of subgroups of the smooth locus on a generalized elliptic curve, we will refer to a Drinfeld $\mathbb{Z}/N\mathbb{Z}$-structure (respectively a Drinfeld $\mathbb{Z}/N\mathbb{Z}$-basis) as a $\mathbb{Z}/N\mathbb{Z}$-structure (respectively $\mathbb{Z}/N\mathbb{Z}$-basis) unless some confusion is possible.

A.1. ***Definitions.*** See [C, Section 2.1] for the definitions of a generalized elliptic curve $f : E \to S$ over a scheme S and of the closed subscheme $S^\infty \subseteq S$ that is the "locus of degenerate fibers" for such an object. (It would be more accurate to write $S^{\infty,f}$, but the abuse of notation should not cause confusion.) Roughly speaking, $E \to S$ is a proper flat family of geometrically connected and semistable curves of arithmetic genus 1 that are either smooth or are so-called Néron polygons, and the relative smooth locus E^{sm} is endowed with a commutative S-group structure that extends (necessarily uniquely) to an action on E such that whenever E_s is a polygon, the action of E_s^{sm} on E_s is via rotations of the polygon. Also, S^∞ is a scheme structure on the set of $s \in S$ such that E_s is not smooth. The definition of the degeneracy locus S^∞ (given in [C, 2.1.8]) makes sense for any proper flat and finitely presented map $C \to S$ with fibers of pure dimension 1. If S' is any S-scheme, then there is an inclusion $S' \times_S S^{\infty} \subseteq S'^{\infty}$ as closed subschemes of S' (with S'^{∞} corresponding to the S'-curve $C \times_S S'$), but this inclusion can fail to be an equality even when each geometric fiber C_s is smooth of genus 1 or a Néron polygon [C, Example 2.1.11]. Fortunately, if C admits a structure of generalized elliptic curve over S, then this inclusion is always an equality [C, 2.1.12], so the degeneracy locus makes sense on moduli stacks for generalized elliptic curves (where it defines the cusps).

We wish to study moduli spaces for generalized elliptic curves $E_{/S}$ equipped with certain ample level structures defined by subgroups of E^{sm}. Of particular interest are those subgroup schemes $G \subseteq E^{\mathrm{sm}}$ that are not only finite locally free over the base with some constant order n but are even *cyclic* in the sense that *fppf*-locally on the base we can write $G = \langle P \rangle := \sum_{j \in \mathbb{Z}/n\mathbb{Z}} [jP]$ in E^{sm} as Cartier divisors for some n-torsion point P of E^{sm}. By [C, 2.3.5], if P and P' are two such points for the same G, then for any $d \mid n$ the points $(n/d)P$ and $(n/d)P'$ are $\mathbb{Z}/(n/d)\mathbb{Z}$-generators of the same S-subgroup of G, so by descent this naturally defines a cyclic S-subgroup $G_d \subseteq G$ of order d even if P does not exist over the given base scheme S. We call G_d the *standard* cyclic subgroup of G with order d. For example, if $d = d'd''$ with $d', d'' \geq 1$ and $\gcd(d', d'') = 1$, then $G_{d'} \times G_{d''} \simeq G_d$ via the group law on G.

Definition A.1. Let $N, n \geq 1$ be integers.

A $\Gamma_1(N)$-*structure* on a generalized elliptic curve $E_{/S}$ is an S-ample $\mathbb{Z}/N\mathbb{Z}$-structure on E^{sm}, which is to say an N-torsion point $P \in E^{\mathrm{sm}}(S)$ such that the relative effective Cartier divisor $D = \sum_{j \in \mathbb{Z}/N\mathbb{Z}} [jP]$ on E^{sm} is an S-subgroup and D_s is ample on E_s for all $s \in S$.

A $\Gamma_1(N, n)$-*structure* on $E_{/S}$ is a pair (P, C), where P is a $\mathbb{Z}/N\mathbb{Z}$-structure on E^{sm} and $C \subseteq E^{\mathrm{sm}}$ is a cyclic S-subgroup with order n such that the relative effective Cartier divisor $D = \sum_{j \in \mathbb{Z}/N\mathbb{Z}} (jP + C)$ on E is S-ample and there is an equality of closed subschemes

$$\sum_{j \in \mathbb{Z}/p^{e_p}\mathbb{Z}} (j(N/p^{e_p})P + C_{p^{e_p}}) = E^{\mathrm{sm}}[p^{e_p}] \tag{1}$$

for all primes $p \mid \gcd(N, n)$, with $e_p = \mathrm{ord}_p(\gcd(N, n)) \geq 1$.

Example A.2. Obviously a $\Gamma_1(N, 1)$-structure is the same as a $\Gamma_1(N)$-structure. If $N = 1$, then we refer to $\Gamma_1(1)$-structures as $\Gamma(1)$-structures, and such a structure on a generalized elliptic curve $E_{/S}$ must be the identity section. Thus, by the ampleness requirement, the geometric fibers E_s must be irreducible. Hence, the moduli stack $\mathcal{M}_{\Gamma(1)}$ of $\Gamma(1)$-structures on generalized elliptic curves classifies generalized elliptic curves with geometrically irreducible fibers.

In [C, 2.4.3], the notion of $\Gamma_1(N, n)$-structure is defined as above, but with the additional requirement that $\mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ for all primes p such that $p \mid \gcd(N, n)$. This requirement always holds when $n = 1$, and whenever it holds, the standard subgroup $C_{p^{e_p}}$ in (1) is the p-part of C, but it turns out to be unnecessary for the proofs of the basic properties of $\Gamma_1(N, n)$-structures and their moduli, as we shall explain in Section A.2. For example, the proof of [C, 2.4.4] carries over to show that we can replace (1) with the requirement that $\sum_{j \in \mathbb{Z}/d\mathbb{Z}}(j(N/d)P + C_d) = E^{\mathrm{sm}}[d]$ in E for $d = \gcd(N, n)$. Another basic property that carries over to the general case is that if (P, C) is a $\Gamma_1(N, n)$-structure on E, then the relative effective Cartier divisor $\sum_{j \in \mathbb{Z}/N\mathbb{Z}}(jP + C)$ on E^{sm} is an S-subgroup; the proof is given in [C, 2.4.5] under the assumption $\mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ for every prime $p \mid \gcd(N, n)$, but the argument works in general once it is observed that after making an *fppf* base change to acquire a $\mathbb{Z}/n\mathbb{Z}$-generator Q of C we can use symmetry in P and Q in the rest of the argument so as to reduce to the case considered in [Conrad 2007].

A.2. *Moduli stacks.* As in [C, 2.4.6], for $N, n \geq 1$ we define the moduli stack $\mathcal{M}_{\Gamma_1(N,n)}$ in order to classify $\Gamma_1(N, n)$-structures on generalized elliptic curves over arbitrary schemes, and we let $\mathcal{M}^{\infty}_{\Gamma_1(N,n)} \hookrightarrow \mathcal{M}_{\Gamma_1(N,n)}$ denote the closed substack given by the degeneracy locus for the universal generalized elliptic curve. The arguments in [C, Sections 3.1 and 3.2] carry over verbatim (that is, without using the condition $\mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ for all primes $p \mid \gcd(N, n)$) to prove the following result.

Theorem A.3. *The stack $\mathcal{M}_{\Gamma_1(N,n)}$ is an Artin stack that is proper over $\mathbb{Z}$. It is smooth over $\mathbb{Z}[1/Nn]$, and it is Deligne–Mumford away from the open and closed substack in $\mathcal{M}^{\infty}_{\Gamma_1(N,n)}$ classifying degenerate triples (E, P, C) in positive characteristics p such that the p-part of each geometric fiber of C is nonétale and disconnected.*

The proof of [C, 3.3.4] does not use the condition $\mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ for all primes $p \mid \gcd(N, n)$ (although this condition is mentioned in the proof), so that argument gives this:

Lemma A.4. *The open substack $\mathcal{M}^{0}_{\Gamma_1(N,n)} = \mathcal{M}_{\Gamma_1(N,n)} - \mathcal{M}^{\infty}_{\Gamma_1(N,n)}$ classifying elliptic curves endowed with a $\Gamma_1(N, n)$-structure is regular and $\mathbb{Z}$-flat with pure relative dimension* 1.

We are interested in the structure of $\mathcal{M}_{\Gamma_1(N,n)}$ around its cuspidal substack, and especially in determining whether it is regular or a scheme near such points. Our analysis of $\mathcal{M}^{\infty}_{\Gamma_1(N,n)}$ rests on the following theorem.

Theorem A.5. *The map $\mathcal{M}_{\Gamma_1(N,n)} \to \mathrm{Spec}(\mathbb{Z})$ is flat and Cohen–Macaulay with pure relative dimension* 1.

Proof. By Lemma A.4, we just have to work along the cusps. Also, it suffices to check the result after localization at each prime p, and if either $p \nmid \gcd(N, n)$ or $1 \leq \mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ then [C, 3.3.1] gives the result over $\mathbb{Z}_{(p)}$. It thus remains to study the cusps in positive characteristic p when $1 \leq \mathrm{ord}_p(N) < \mathrm{ord}_p(n)$. As in the cases treated in [Conrad 2007], the key is to study the deformation theory of a related level structure on generalized elliptic curves, the so-called $\widetilde{\Gamma}_1(N, n)$-*structure*: this is a pair (P, Q), where P is a $\mathbb{Z}/N\mathbb{Z}$-structure on the smooth locus and Q is a $\mathbb{Z}/n\mathbb{Z}$-structure on the smooth locus such that $(P, \langle Q\rangle)$ is a $\Gamma_1(N, n)$-structure. The same definition is given in [C, 3.3.2] with the unnecessary restriction $\mathrm{ord}_p(n) \leq \mathrm{ord}_p(N)$ for all primes $p \mid \gcd(N, n)$, and the argument that immediately follows that definition works without such a restriction to show that the moduli stack $\mathcal{M}_{\widetilde{\Gamma}_1(N,n)}$ of $\widetilde{\Gamma}_1(N, n)$-structures is a Deligne–Mumford stack over $\mathbb{Z}$ that is a finite flat cover of the proper Artin stack $\mathcal{M}_{\Gamma_1(N,n)}$.

By the Deligne–Mumford property, any $\widetilde{\Gamma}_1(N, n)$-structure $x_0 = (E_0, P_0, Q_0)$ over an algebraically closed field k admits a universal deformation ring. Since $\mathcal{M}_{\widetilde{\Gamma}_1(N,n)}$ is a finite flat cover of $\mathcal{M}_{\Gamma_1(N,n)}$, as in the proof of [C, 3.3.1], it suffices to assume $\mathrm{char}(k) = p > 0$ and to exhibit the deformation ring at x_0 as a finite flat extension of $W(k)[\![x]\!]$ when E_0 is a standard polygon, $n = p^e$, and $N = Mp^r$ with $p \nmid M$ and $e, r \geq 1$. The case $e \leq r$ is settled in [Conrad 2007], and we will adapt that argument to handle the case $1 \leq r < e$. By the ampleness condition, at least one of MP_0 or Q_0 generates the p-part of the component group of E_0^{sm}, and moreover $\{MP_0, p^{e-r}Q_0\}$ is a Drinfeld $\mathbb{Z}/p^r\mathbb{Z}$-basis of $E_0^{\mathrm{sm}}[p^r]$. We shall break up the problem into three cases, and it is only in Case 3 that we will meet a situation essentially different from that encountered in Conrad's proof for $1 \leq e \leq r$.

CASE 1: We first assume that MP_0 generates the p-part of the component group, so by the Drinfeld $\mathbb{Z}/p^r\mathbb{Z}$-basis hypothesis, this point is a basis of $E_0^{\mathrm{sm}}(k)[p^\infty]$ over $\mathbb{Z}/p^r\mathbb{Z}$ (as we are in characteristic p and E_0 is a polygon). Hence, $Q_0 = jMP_0$ for a unique $j \in \mathbb{Z}/p^r\mathbb{Z}$ (so $p^{e-r}Q_0 = p^{e-r}jMP_0$). Since n is a p-power, it also follows that $\langle P_0\rangle$ is ample. In particular, (E_0, P_0) is a $\Gamma_1(N)$-structure. Thus, the formation of an infinitesimal deformation (E, P, Q) of (E_0, P_0, Q_0) can be given in three steps: first give an infinitesimal deformation (E, P) of (E_0, P_0) as a $\Gamma_1(N)$-structure, then give a Drinfeld $\mathbb{Z}/p^r\mathbb{Z}$-basis (MP, Q') of $E^{\mathrm{sm}}[p^r]$ with Q' deforming $p^{e-r}Q_0$, and finally specify a p^{e-r}-th root Q of Q' lifting $Q_0 = jMP_0$. The one aspect of this description that merits some explanation is to justify that such a p^{e-r}-th root Q of Q' must be a $\mathbb{Z}/p^e\mathbb{Z}$-structure on E^{sm}. The point Q is clearly killed by p^e, so the Cartier divisor $D = \sum_{j\in\mathbb{Z}/p^e\mathbb{Z}}[jQ]$ in E^{sm} makes sense, and we have to check that it is automatically a subgroup scheme.

For any $t \geq 0$, the identification $(E_0^{\mathrm{sm}})^0[p^t] = \mu_{p^t}$ uniquely lifts to an isomorphism $(E^{\mathrm{sm}})^0[p^t] \simeq \mu_{p^t}$. In particular, if p^ν is the order of the p-part of the cyclic component group of E_0^{sm} (with $\nu \geq r$), then $E^{\mathrm{sm}}[p^e]$ is an extension of $\mathbb{Z}/p^j\mathbb{Z}$ by μ_{p^e}, where $j = \min(\nu, e)$. The image of $\langle Q_0 \rangle$ in the component group can be uniquely identified with $\mathbb{Z}/p^i\mathbb{Z}$ (for some $i \leq j$) such that $Q_0 \mapsto 1$, and this $\mathbb{Z}/p^i\mathbb{Z}$ has preimage G in $E^{\mathrm{sm}}[p^e]$ that is a p^e-torsion commutative extension of $\mathbb{Z}/p^i\mathbb{Z}$ by μ_{p^e} with $0 \leq i \leq e$. Since Q is a point of G over the (artinian local) base, it follows from [C, 2.3.3] that Q is a $\mathbb{Z}/p^e\mathbb{Z}$-structure on E^{sm} if and only if the point $p^i Q$ in $\mu_{p^{e-i}}$ is a $\mathbb{Z}/p^{e-i}\mathbb{Z}$-generator of $\mu_{p^{e-i}}$. The case $i = e$ is therefore settled, so we can assume $i < e$ (that is, $\langle Q_0 \rangle$ is not étale, or equivalently $p^{e-1}Q_0 = 0$). By hypothesis, $p^{e-r}Q = Q'$ is a $\mathbb{Z}/p^r\mathbb{Z}$-structure on E^{sm} with $1 \leq r < e$, so $p^{e-1}Q = p^{r-1}Q'$ is a $\mathbb{Z}/p\mathbb{Z}$-structure on E^{sm}. This $\mathbb{Z}/p\mathbb{Z}$-structure must generate the subgroup $\mu_p \subseteq E^{\mathrm{sm}}[p^e]$ since $p^{e-1}Q$ lies in $(E^{\mathrm{sm}})^0$ (as $p^{e-1}Q_0 = 0$). Hence, $Q'' = p^i Q$ is a point of $\mu_{p^{e-i}}$ such that $p^{e-i-1}Q''$ is a $\mathbb{Z}/p\mathbb{Z}$-generator of μ_p. Since $\mathbb{Z}/m\mathbb{Z}$-generators of μ_m are simply roots of the cyclotomic polynomial Φ_m [C, 1.12.9], our problem is reduced to the assertion that if s is a positive integer (such as $e - i$), then an element ζ in a ring is a root of the cyclotomic polynomial Φ_{p^s} if $\zeta^{p^{s-1}}$ is a root of Φ_p. This assertion is obvious since $\Phi_{p^s}(T) = \Phi_p(T^{p^{s-1}})$, and so our description of the infinitesimal deformation theory of (E_0, P_0, Q_0) is justified.

The torsion subgroup $E^{\mathrm{sm}}[p^r]$ is uniquely an extension of $\mathbb{Z}/p^r\mathbb{Z}$ by μ_{p^r} deforming the canonical such description for $E_0^{\mathrm{sm}}[p^r]$, so the condition on Q' is that it has the form $\zeta + p^{e-r}jMP$ for a point ζ of the scheme of generators $\mu_{p^r}^{\times}$ of $\mu_{p^r} = (E^{\mathrm{sm}})^0[p^r]$. Thus, to give Q is to specify a p^{e-r}-th root of ζ in E^{sm} deforming the identity, which is to say a point of $\mu_{p^e}^{\times}$. It is shown in the proof of [C, 3.3.1] that the universal deformation ring A for (E_0, P_0) is finite flat over $W(k)[\![x]\!]$, and the specification of ζ amounts to giving a root of the cyclotomic polynomial Φ_{p^e}, so the case when MP_0 generates the p-part of the component group of E_0^{sm} is settled (with deformation ring $A[T]/(\Phi_{p^e}(T))$).

CASE 2: Next assume that Q_0 generates the p-part of the component group and that $\langle Q_0 \rangle$ is étale (that is, $Q_0 \in E_0^{\mathrm{sm}}(k)$ has order p^e). The point Q_0 must generate $E_0^{\mathrm{sm}}(k)[p^\infty]$ over $\mathbb{Z}/p^e\mathbb{Z}$, and the étale hypothesis ensures that Q_0 is a $\mathbb{Z}/p^e\mathbb{Z}$-basis of $E_0^{\mathrm{sm}}(k)[p^\infty]$. Thus, $MP_0 = p^{e-r}jQ_0$ for some (unique) $j \in \mathbb{Z}/p^r\mathbb{Z}$. By replacing P with $P - M^{-1}p^{e-r}jQ$ for any infinitesimal deformation (E, P, Q) of (E_0, P_0, Q_0), we can assume that the p-part of P_0 vanishes. The p-part of P must therefore be a point of $\mu_{p^r}^{\times}$. The $\mathbb{Z}/M\mathbb{Z}$-part of P together with Q constitutes a $\Gamma_1(Mp^e)$-structure on E (in particular, the ampleness condition holds), and this is an étale level structure since the cyclic subgroup $\langle Q_0 \rangle$ in E_0^{sm} is étale. Hence, the infinitesimal deformation functor of (E_0, P_0, Q_0) is pro-represented by $\mu_{p^r}^{\times}$ over the deformation ring of an étale $\Gamma_1(Mp^e)$-structure. For any $R \geq 1$, deformation rings for étale $\Gamma_1(R)$-structures on polygons over k have the form $W(k)[\![x]\!]$ (as is explained near the end of the proof of [C, 3.3.1], using [C, II, 1.17]), so not only are we done but in this case the deformation ring for (E_0, P_0, Q_0) is the ring $W(k)[\![x]\!][T]/(\Phi_{p^r}(T))$ that is visibly regular.

CASE 3: Finally, assume Q_0 generates the p-part of the component group but that $\langle Q_0 \rangle$ is not étale (that is, $Q_0 \in E_0^{\mathrm{sm}}(k)$ has order strictly less than p^e), and so $p^{e-r}Q_0 \in E_0^{\mathrm{sm}}(k)$ has order strictly dividing p^r. Since $\{MP_0, p^{e-r}Q_0\}$ is a Drinfeld $\mathbb{Z}/p^r\mathbb{Z}$-basis of $E_0^{\mathrm{sm}}[p^r]$, the point MP_0 must be a $\mathbb{Z}/p^r\mathbb{Z}$-basis for $E_0^{\mathrm{sm}}(k)[p^r]$. Hence, if we write $P_0 = P_0' + P_0''$ corresponding to the decomposition $\mathbb{Z}/N\mathbb{Z} = (\mathbb{Z}/M\mathbb{Z}) \times (\mathbb{Z}/p^r\mathbb{Z})$, then P_0'' has order exactly p^r in $E_0^{\mathrm{sm}}(k)$. We use P_0'' to identify $E_0^{\mathrm{sm}}(k)[p^r]$ with $\mathbb{Z}/p^r\mathbb{Z}$. It follows that if we make the analogous canonical decomposition $P = P' + P''$ for an infinitesimal deformation (E, P, Q) of (E_0, P_0, Q_0), then the p-part P'' deforms P_0'' and generates an étale subgroup of E^{sm} with order p^r. Thus, P' and Q together constitute a (nonétale) $\Gamma_1(Mp^e)$-structure on E (in particular, the ampleness condition holds), and the data of P'' amounts to a section over $1 \in \mathbb{Z}/p^r\mathbb{Z}$ with respect to the unique quotient map $E^{\mathrm{sm}}[p^r] \twoheadrightarrow \mathbb{Z}/p^r\mathbb{Z}$ lifting the quotient map $E_0^{\mathrm{sm}}[p^r] \twoheadrightarrow \mathbb{Z}/p^r\mathbb{Z}$ defined by P_0''. Since the specification of a $\mathbb{Z}/N\mathbb{Z}$-structure on E^{sm} is the "same" as the specification of a pair consisting of $\mathbb{Z}/M\mathbb{Z}$-structure and a $\mathbb{Z}/p^r\mathbb{Z}$-structure [C, 1.7.3], we conclude that the universal deformation ring of (E_0, P_0, Q_0) classifies the fiber over $1 \in \mathbb{Z}/p^r\mathbb{Z}$ in the connected-étale sequence for the p^r-torsion in infinitesimal deformations of the underlying $\Gamma_1(Mp^e)$-structure (E_0, P_0', Q_0). Universal deformation rings for $\Gamma_1(Mp^e)$-structures over k are finite flat over $W(k)[\![x]\!]$ (by the proof of [C, 3.3.1]), so we are therefore done. □

Corollary A.6. *The closed substack $\mathcal{M}^{\infty}_{\Gamma_1(N,n)} \hookrightarrow \mathcal{M}_{\Gamma_1(N,n)}$ is a relative effective Cartier divisor over $\mathbb{Z}$, and it has a reduced generic fiber over $\mathbb{Q}$.*

Proof. The reducedness over $\mathbb{Q}$ is shown in [C, 4.3.2], and the proof works without restriction on $\gcd(N, n)$. Likewise, the proof that $\mathcal{M}^{\infty}_{\Gamma_1(N,n)}$ is a $\mathbb{Z}$-flat Cartier divisor is part of [C, 4.1.1(1)] in case $\mathrm{ord}_p(n) \le \mathrm{ord}_p(N)$ for all primes $p \mid \gcd(N, n)$, but by using the above proof of Theorem A.5, we see that the method of proof works in general. □

Using Lemma A.4, Theorem A.5, and Corollary A.6, Serre's normality criterion can be used to prove normality for $\mathcal{M}_{\Gamma_1(N,n)}$ in general. (This is proved in [C, 4.1.4] subject to the restrictions on $\gcd(N, n)$ in the definition therein of $\Gamma_1(N, n)$-structures, but the argument works in general by using the results that are stated above without any such restriction on $\gcd(N, n)$.) However, the proof of regularity encounters complications at points of a certain locus of cusps in bad characteristics. This problematic locus is defined as follows.

Definition A.7. Let $\mathscr{Z}_{\Gamma_1(N,n)} \hookrightarrow \mathscr{M}^{\infty}_{\Gamma_1(N,n)}$ be the 0-dimensional closed substack with reduced structure that consists of geometric points (E_0, P_0, C_0) in characteristics $p \mid \gcd(N, n)$ such that $1 \le \mathrm{ord}_p(N) < \mathrm{ord}_p(n)$, C_0 is not étale, and $(N/p^{\mathrm{ord}_p(N)})P_0$ does not generate the p-part of the component group of E_0^{sm}.

Note that if $\mathrm{ord}_p(n) \le \mathrm{ord}_p(N)$ for all primes $p \mid \gcd(N, n)$ (the situation considered in [Conrad 2007]), then $\mathscr{Z}_{\Gamma_1(N,n)}$ is empty; this includes the case of $\Gamma_1(N)$-structures for any N (take $n = 1$). In all other cases, it is nonempty. The geometric points of $\mathscr{Z}_{\Gamma_1(N,n)}$ correspond to precisely the points in Case 3 in the proof of Theorem A.5. The method in [Conrad 2007] for analyzing regularity along the cusps assumes $\mathscr{Z}_{\Gamma_1(N,n)}$ is empty, and by combining it with the modified arguments in the proof of Theorem A.5 (especially the regularity observation in Case 2) we obtain the following consequence.

Theorem A.8. *Outside the closed substack $\mathscr{Z}_{\Gamma_1(N,n)} \subseteq \mathscr{M}^{\infty}_{\Gamma_1(N,n)}$, the stack $\mathscr{M}_{\Gamma_1(N,n)}$ is regular.*

A.3. *Applications.* Before we apply our results, we record a useful lemma.

Lemma A.9. *Let S be a scheme, and let $\mathscr{X}$ be an Artin stack over S. Assume $\mathscr{X}$ is S-separated. The locus of geometric points of $\mathscr{X}$ with trivial automorphism group scheme is an open substack $\mathscr{U} \subseteq \mathscr{X}$ that is an algebraic space. This algebraic space is a scheme if $\mathscr{X}$ is quasifinite over a separated S-scheme.*

Proof. The first part is [C, 2.2.5(2)], and the second part follows from the general fact that an algebraic space that is quasifinite and separated over a scheme is a scheme [C, Theorem A.2]. □

In the setting of Lemma A.9, if $\mathscr{X}$ is quasifinite over a separated S-scheme, then we call $\mathscr{U}$ the *maximal open subscheme* of $\mathscr{X}$. The case of interest to us is $\mathscr{X} = \mathscr{M}_{\Gamma_1(N,n)/S}$ over any scheme S. This is quasifinite over the S-proper stack $\mathscr{M}_{\Gamma(1)/S}$ via fibral contraction away from the identity component, and $\mathscr{M}_{\Gamma(1)/S}$ is quasifinite over $\mathbf{P}^1_S$ via the j-invariant, so $\mathscr{X}$ is quasifinite over the separated S-scheme $\mathbf{P}^1_S$.

We wish to prove results concerning when certain components of $\mathscr{M}^{\infty}_{\Gamma_1(N,n)}$ lie in the maximal open subscheme of $\mathscr{M}_{\Gamma_1(N,n)}$. So we first record a general lemma.

Lemma A.10. *Let $\mathscr{Y}$ be an irreducible Artin stack over $\mathbb{F}_p$, and let $\mathscr{C}$ be a finite locally free commutative $\mathscr{Y}$-group that is cyclic with order p^e. If $\mathscr{C}$ has a multiplicative geometric fiber over $\mathscr{Y}$, then all of its geometric fibers are connected.*

The abstract notion of cyclicity (with no ambient smooth curve group) is developed in [C, 1.5, 1.9, 1.10] over arbitrary base schemes, and the theory carries over when the base is an Artin stack. We will only need the lemma for situations that arise within torsion on generalized elliptic curves (over Artin stacks).

Proof. We can assume $e \geq 1$, and we may replace $\mathscr{C}$ with its standard subgroup $\mathscr{C}_p$ of order p because it is obvious by group theory that a *cyclic* group scheme C of p-power order over an algebraically closed field of characteristic p is étale if and only if its standard subgroup of order p is étale. Hence, we can assume that $\mathscr{C}$ has order p. Our problem is therefore to rule out the existence of étale fibers. By openness of the locus of étale fibers and irreducibility of $\mathscr{Y}$, if there is an étale fiber, then there is a Zariski-dense open $\mathscr{U} \subseteq \mathscr{Y}$ over which $\mathscr{C}$ has étale fibers. In particular, there is some geometric point u of $\mathscr{U}$ that specializes to the geometric point $y \in \mathscr{Y}$ where we assume the fiber is multiplicative, so after pullback to a suitable valuation ring, we get an étale group of order p in characteristic p specializing to a multiplicative one. Passing to Cartier duals gives a multiplicative group of order p having an étale specialization, and this is impossible since multiplicative groups of order p in characteristic p are not étale. □

Theorem A.11. *Let p be a prime, and choose a positive integer M not divisible by p such that $M > 2$. Also fix integers $e, r \geq 0$. If $e = 0$ or $r = 0$, then assume $M \neq 4$. Let $x_0 = (E_0, P_0, C_0)$ be a geometric point on the special fiber of the cuspidal substack in the proper Artin stack $\mathscr{X} = \mathscr{M}_{\Gamma_1(Mp^r, p^e)/\mathbb{Z}_{(p)}}$ over $\mathbb{Z}_{(p)}$, and assume that C_0 is étale.*

Let $\mathscr{Y}$ be the irreducible component of x_0 in $\mathscr{X}_{\mathbb{F}_p}$. For every geometric cusp $x_1 = (E_1, P_1, C_1)$ on $\mathscr{Y}$, the group C_1 is étale and x_1 lies in the maximal open subscheme of $\mathscr{X}$. Moreover, if $x \in \mathscr{X}_{\mathbb{Q}}$ is a cusp specializing into $\mathscr{Y}$, then the Zariski closure D of x in $\mathscr{X}$ lies in the maximal open subscheme and D is Cartier in $\mathscr{X}$.

The case $e = 2$ is required in the main text. It is necessary to avoid the cases $M \leq 2$ and $(M, r) = (4, 0)$ because in these cases there are cusps x_0 in characteristic p as in the theorem such that x_0 admits nontrivial automorphisms (and so x_0 cannot lie in the maximal open subscheme of $\mathscr{X}$).

Proof. We first check that the étale assumption at x_0 is inherited by all geometric cusps $x_1 \in \mathscr{Y}$. Let $(\mathscr{E}, \mathscr{P}, \mathscr{C})$ be the pullback to $\mathscr{Y}$ of the universal family over $\mathscr{X}$. The group $\mathscr{C}$ is cyclic of order p^e with $e \geq 0$, so applying Lemma A.10 to its Cartier dual gives the result (since at a cusp a connected subgroup of p-power order must be multiplicative).

Now we can rename x_1 as x_0 without loss of generality, so we have to check that x_0 lies in the maximal open subscheme of $\mathscr{X}$ and that if $x \in \mathscr{X}_{\mathbb{Q}}$ is a geometric cusp specializing to x_0, then the Zariski closure of x in $\mathscr{X}$ is Cartier. But the étale hypothesis on C_0 ensures that x_0 is not in the closed substack $\mathscr{Z}_{\Gamma_1(Mp^r, p^e)/\mathbb{Z}_{(p)}}$, so by Theorem A.8 the stack $\mathscr{X}$ is regular at x_0. Hence, since $\mathscr{X}$ is $\mathbb{Z}_{(p)}$-flat with pure relative dimension 1 (by Theorem A.5), the desired properties of D at the end of the theorem hold once we know that x_0 is in the maximal open subscheme of $\mathscr{X}$, which is to say that its automorphism group scheme G is trivial. To verify this triviality we will make essential use of the property that C_0 is étale. Let k be the algebraically closed field over which x_0 lives. Since E_0 is d-gon over k for some $d \geq 1$, G is a closed subgroup of the automorphism group $\mu_d \rtimes \langle \mathrm{inv} \rangle$ of the d-gon. Since C_0 is étale with order p^e in characteristic p, it follows that C_0 maps isomorphically into the p-part of the component group of $E_0^{\mathrm{sm}} = \mathbf{G}_m \times (\mathbb{Z}/d\mathbb{Z})$. (In particular, $p^e \mid d$.) If R is an artinian local k-algebra with residue field k, any choice of generator Q_0 of C_0 must be carried to another generator of C_0 by any $g \in G(R)$ since $C_0(R) \to C_0(k)$ is a bijection. But $\mu_d(R)$ acts on $(E_0)_R$ in a manner that preserves the components of the smooth locus, and C_0 meets each component of E_0^{sm} in at most one point. Hence, $G \cap \mu_d$ acts as automorphisms of the $\Gamma_1(Mp^e)$-structure on E_0 defined by $p^r P_0$ and Q_0. Since $Mp^e > 2$ and $Mp^e \neq 4$ (due to the cases we are avoiding), such an ample level structure on a d-gon has trivial automorphism group scheme. This shows that $G \cap \mu_d$ is trivial, so G injects into the group $\mathbb{Z}/2\mathbb{Z}$ of automorphisms of the identity component $\mathbf{G}_m$ of E_0^{sm}. Hence, the contraction operation on E_0 away from $\langle P_0 \rangle$ is faithful on G since contraction does not affect the identity component. It follows that G is a subgroup of the automorphism group of the $\Gamma_1(Mp^r)$-structure obtained by contraction away from $\langle P_0 \rangle$. But $Mp^r \notin \{1, 2, 4\}$ since we assume $M > 2$ and $(M, r) \neq (4, 0)$, so $\Gamma_1(Mp^r)$-structures on polygons have trivial automorphism functor. Thus, $G = \{1\}$ as desired. $\square$

Over the base $\mathbb{Z}_{(p)}$, the results of [C, Sections 3 and 4] concerning the properties of the stack $X_1(N, n)$ carry over if $p \nmid n$. In effect, the hypothesis on $\mathrm{ord}_p(n)$ imposed in [Conrad 2007] only intervenes in the proofs when n is not invertible on the base.

Acknowledgments

The author thanks Brian Conrad for writing the appendix, as well as for numerous helpful discussions and suggestions about the technical issues in Section 2. The author thanks the referee for several suggestions on the manuscript and for directing him to some good references to help deal with the case $p = 2$.

References

[Bosch and Lütkebohmert 1993] S. Bosch and W. Lütkebohmert, "Formal and rigid geometry, I: Rigid spaces", *Math. Ann.* **295**:2 (1993), 291–317. MR 94a:11090 Zbl 0808.14017

[Bosch et al. 1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis: A systematic approach to rigid analytic geometry*, Grundlehren der Mathematischen Wissenschaften **261**, Springer, Berlin, 1984. MR 86b:32031 Zbl 0539.14017

[Buzzard 2003] K. Buzzard, "Analytic continuation of overconvergent eigenforms", *J. Amer. Math. Soc.* **16**:1 (2003), 29–55. MR 2004c:11063 Zbl 1076.11029

[Buzzard 2007] K. Buzzard, "Eigenvarieties", pp. 59–120 in *L-functions and Galois representations* (Durham, 2004), edited by D. Burns et al., London Math. Soc. Lecture Note Ser. **320**, Cambridge Univ. Press, 2007. MR 2392353 Zbl 05282812

[Coleman 1997] R. F. Coleman, "p-adic Banach spaces and families of modular forms", *Invent. Math.* **127**:3 (1997), 417–479. MR 98b:11047 Zbl 0918.11026

[Coleman and Mazur 1998] R. Coleman and B. Mazur, "The eigencurve", pp. 1–113 in *Galois representations in arithmetic algebraic geometry* (Durham, 1996), edited by A. J. Scholl and R. L. Taylor, London Math. Soc. Lecture Note Ser. **254**, Cambridge Univ. Press, 1998. MR 2000m:11039 Zbl 0932.11030

[Conrad 1999] B. Conrad, "Irreducible components of rigid spaces", *Ann. Inst. Fourier* (*Grenoble*) **49**:2 (1999), 473–541. MR 2001c:14045 Zbl 0928.32011

[Conrad 2007] B. Conrad, "Arithmetic moduli of generalized elliptic curves", *J. Inst. Math. Jussieu* **6**:2 (2007), 209–278. MR 2008e:11073 Zbl 1140.14018

[Deligne and Rapoport 1973] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques", pp. 143–316 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **349**, Springer, Berlin, 1973. MR 49 #2762 Zbl 0281.14010

[Goren and Kassaei 2006] E. Z. Goren and P. L. Kassaei, "The canonical subgroup: A "subgroup-free" approach", *Comment. Math. Helv.* **81**:3 (2006), 617–641. MR 2007c:11066 Zbl 1122.11037

[de Jong 1995] A. J. de Jong, "Crystalline Dieudonné module theory via formal and rigid geometry", *Inst. Hautes Études Sci. Publ. Math.* 82 (1995), 5–96 (1996). MR 97f:14047 Zbl 0864.14009

[Kassaei 2005] P. L. Kassaei, "Overconvergence, analytic continuation, and classicality: The case of curves", 2005, Available at http://www.mth.kcl.ac.uk/~kassaei/research/files/OSLB.pdf. To appear in *J. Reine Angew. Math.*

[Kassaei 2006] P. L. Kassaei, "A gluing lemma and overconvergent modular forms", *Duke Math. J.* **132**:3 (2006), 509–529. MR 2007c:11056 Zbl 1112.11020

[Katz 1973] N. M. Katz, "p-adic properties of modular schemes and modular forms", pp. 69–190 in *Modular functions of one variable, III* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Mathematics **350**, Springer, Berlin, 1973. MR 56 #5434 Zbl 0271.10033

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies **108**, Princeton University Press, 1985. MR 86i:11024 Zbl 0576.14026

[Ramsey 2006] N. Ramsey, "Geometric and p-adic modular forms of half-integral weight", *Ann. Inst. Fourier* (*Grenoble*) **56**:3 (2006), 599–624. MR 2007c:11057

[Ramsey 2007] N. Ramsey, "The overconvergent Shimura lifting", preprint, 2007, Available at http://tinyurl.com/5fg4ru.

[Serre 1962] J.-P. Serre, "Endomorphismes complètement continus des espaces de Banach p-adiques", *Inst. Hautes Études Sci. Publ. Math.* 12 (1962), 69–85. MR 26 #1733

[Shimura 1973] G. Shimura, "On modular forms of half integral weight", *Ann. of Math.* (2) **97** (1973), 440–481. MR 48 #10989 Zbl 0266.10022

Communicated by Brian Conrad
Received 2007-09-25 Revised 2008-07-03 Accepted 2008-08-22

naramsey@umich.edu *Department of Mathematics, University of Michigan, 2074 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, United States*

conrad@math.stanford.edu *Department of Mathematics, Stanford University, Building 380, Sloan Hall, Stanford, CA 94305, United States*
http://math.stanford.edu/~conrad/

ALGEBRA AND NUMBER THEORY 2:7(2008)

The moduli space of curves is rigid

Paul Hacking

We prove that the moduli stack $\overline{\mathcal{M}}_{g,n}$ of stable curves of genus g with n marked points is rigid, that is, has no infinitesimal deformations. This confirms the first case of a principle proposed by Kapranov. It can also be viewed as a version of Mostow rigidity for the mapping class group.

1. Introduction

Kapranov [1997] has proposed the following informal statement: Given a smooth variety $X = X(0)$, consider the moduli space $X(1)$ of varieties obtained as deformations of $X(0)$, the moduli space $X(2)$ of deformations of $X(1)$, and so on. Then this process should stop after $n = \dim X$ steps, that is, $X(n)$ should be rigid (no infinitesimal deformations). Roughly speaking, one thinks of $X(1)$ as H^1 of a sheaf of nonabelian groups on $X(0)$. Indeed, at least the tangent space to $X(1)$ at $[X]$ is identified with $H^1(T_X)$, where T_X is the tangent sheaf, the sheaf of first order infinitesimal automorphisms of X. Then one regards $X(m)$ as a kind of nonabelian H^m, and the analogy with the usual definition of abelian H^m suggests the statement above.

In particular, the moduli space of curves should be rigid. In this paper, we verify this in the following precise form: the moduli stack of stable curves of genus g with n marked points is rigid for each g and n.

On the other hand, moduli spaces of surfaces should have nontrivial deformations in general. A simple example (for surfaces with boundary) is given in Section 6. It seems plausible that there should be a nontrivial deformation of a moduli space of surfaces whose fibres parametrise "generalised surfaces" in some sense, for example noncommutative surfaces. From this point of view the result of this paper says that the concept of a curve cannot be deformed.

Let us also note that our result can be thought of as a version of Mostow rigidity for the mapping class group. Recall that the moduli space M_g of smooth complex curves of genus g is the quotient of the Teichmüller space T_g by the mapping class

MSC2000: 14H10.
Keywords: moduli, curve, rigidity.
The author was partially supported by NSF grant DMS-0650052.

group Γ_g. The space T_g is a bounded domain in $\mathbb{C}^{3g-3}$, which is homeomorphic to a ball, and Γ_g acts discontinuously on T_g with finite stabilisers. We thus obtain M_g as a complex orbifold with orbifold fundamental group Γ_g. The space T_g admits a natural metric, the Weil–Petersson metric, which has negative holomorphic sectional curvatures. So, roughly speaking, M_g looks like a quotient of a complex ball by a discrete group Γ of isometries, with finite volume. Mostow rigidity predicts that such a quotient is uniquely determined by the group Γ up to complex conjugation. (This is certainly true if Γ acts freely with compact quotient; see [Siu 1980].) In particular, it should have no infinitesimal deformations. Unfortunately I do not know a proof along these lines.

2. Statements

We work over an algebraically closed field k of characteristic zero. Let g and n be nonnegative integers such that $2g-2+n>0$. Let $\overline{\mathcal{M}}_{g,n}$ denote the moduli stack of stable curves of genus g with n marked points. The stack $\overline{\mathcal{M}}_{g,n}$ is a smooth proper Deligne–Mumford stack of dimension $3g-3+n$.

Theorem 2.1. *The stack $\overline{\mathcal{M}}_{g,n}$ is rigid, that is, has no infinitesimal deformations.*

Let $\partial\overline{\mathcal{M}}_{g,n} \subset \overline{\mathcal{M}}_{g,n}$ denote the boundary of the moduli stack, that is, the complement of the locus of smooth curves (with its reduced structure). The locus $\partial\overline{\mathcal{M}}_{g,n}$ is a normal crossing divisor in $\overline{\mathcal{M}}_{g,n}$.

Theorem 2.2. *The pair $(\overline{\mathcal{M}}_{g,n}, \partial\overline{\mathcal{M}}_{g,n})$ has no locally trivial deformations.*

Let $\overline{M}_{g,n}$ denote the coarse moduli space of the stack $\overline{\mathcal{M}}_{g,n}$. The space $\overline{M}_{g,n}$ is a projective variety with quotient singularities.

Theorem 2.3. *The variety $\overline{M}_{g,n}$ has no locally trivial deformations if*

$$(g,n) \neq (1,2), (2,0), (2,1), (3,0).$$

Remark 2.4. In the exceptional cases, the projection $\overline{\mathcal{M}}_{g,n} \to \overline{M}_{g,n}$ is ramified in codimension one over the interior of $\overline{M}_{g,n}$, and an additional calculation is needed to relate the deformations of the stack and the deformations of the coarse moduli space (see Proposition 5.2). Presumably the result still holds.

3. Proof of Theorem 2.2

Write $\mathcal{B}$ for the boundary of $\overline{\mathcal{M}}_{g,n}$. Let $\Omega_{\overline{\mathcal{M}}_{g,n}}(\log \mathcal{B})$ denote the sheaf of 1-forms on $\overline{\mathcal{M}}_{g,n}$ with logarithmic poles along the boundary, and $T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})$ the dual of $\Omega_{\overline{\mathcal{M}}_{g,n}}(\log \mathcal{B})$. The sheaf $T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})$ is the subsheaf of the tangent sheaf $T_{\overline{\mathcal{M}}_{g,n}}$ consisting of vector fields on $\overline{\mathcal{M}}_{g,n}$ which are tangent to the boundary. In other words, it is the sheaf of first order infinitesimal automorphisms of the pair

$(\overline{\mathcal{M}}_{g,n}, \mathcal{B})$. Hence the first order locally trivial deformations of the pair $(\overline{\mathcal{M}}_{g,n}, \mathcal{B})$ are identified with the space $H^1\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big)$. To prove Theorem 2.2, we show

$$H^1\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big) = 0.$$

Let $\pi : \mathcal{U}_{g,n} \to \overline{\mathcal{M}}_{g,n}$ denote the universal family over $\overline{\mathcal{M}}_{g,n}$. That is, $\mathcal{U}_{g,n}$ is the stack of n-pointed stable curves of genus g together with an extra section (with no smoothness condition). Let Σ denote the union of the n tautological sections of π. We define the boundary $\mathcal{B}_U$ of $\mathcal{U}_{g,n}$ as the union of $\pi^*\mathcal{B}$ and Σ.

Let $\nu \colon \mathcal{B}^\nu \to \mathcal{B}$ be the normalisation of the boundary $\mathcal{B}$ of $\overline{\mathcal{M}}_{g,n}$, and $\mathcal{N}$ the normal bundle of the map $\mathcal{B}^\nu \to \overline{\mathcal{M}}_{g,n}$. Then we have an exact sequence

$$0 \to T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B}) \to T_{\overline{\mathcal{M}}_{g,n}} \to \nu_*\mathcal{N} \to 0.$$

Let ω_π denote the relative dualising sheaf of the morphism π.

Lemma 3.1. *There is a natural isomorphism*

$$\delta : T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B}) \xrightarrow{\sim} R^1\pi_*(\omega_\pi(\Sigma)^\vee).$$

Proof. For a pointed stable curve $(C, \Sigma_C = x_1 + \cdots + x_n)$, the space of first order deformations is equal to $\operatorname{Ext}^1(\Omega_C(\Sigma_C), \mathcal{O}_C)$. See [Deligne and Mumford 1969, p. 79–82]. The surjection

$$\operatorname{Ext}^1(\Omega_C(\Sigma_C), \mathcal{O}_C) \to H^0\big(\mathcal{E}xt^1(\Omega_C(\Sigma_C), \mathcal{O}_C)\big) = \bigoplus_{q \in \operatorname{Sing} C} \mathcal{E}xt^1(\Omega_C(\Sigma_C), \mathcal{O}_C)_q$$

sends a global deformation of (C, Σ_C) to the induced deformations of the nodes. Étale locally at the point $[(C, \Sigma_C)] \in \overline{\mathcal{M}}_{g,n}$, the boundary $\mathcal{B}$ is a normal crossing divisor with components B_q indexed by the nodes q of C (the divisor B_q is the locus where the node q is *not* smoothed). The Kodaira–Spencer map identifies the fibre of the normal bundle of B_q at $[(C, \Sigma_C)]$ with the stalk of $\mathcal{E}xt^1(\Omega_C(\Sigma_C), \mathcal{O}_C)$ at q.

We now work globally over $\overline{\mathcal{M}}_{g,n}$. We omit the subscripts g, n for clarity. Consider the exact sequence

$$0 \to \pi^*\Omega_{\overline{\mathcal{M}}} \to \Omega_{\mathcal{U}}(\log \Sigma) \to \Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma) \to 0. \tag{3-1}$$

For a sheaf $\mathcal{F}$ on $\mathcal{U}$, let $\mathcal{E}xt^i_\pi(\mathcal{F}, \cdot)$ denote the i-th right derived functor of

$$\pi_* \circ \mathcal{H}om(\mathcal{F}, \cdot).$$

Applying $\pi_* \circ \mathcal{H}om(\cdot, \mathcal{O}_{\mathcal{U}})$ to the exact sequence (3-1), we obtain a long exact sequence with connecting homomorphism

$$\rho \colon T_{\overline{\mathcal{M}}} \to \mathcal{E}xt^1_\pi(\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma), \mathcal{O}_{\mathcal{U}}).$$

The map ρ is the Kodaira–Spencer map for the universal family over $\overline{\mathcal{M}}$ and thus is an isomorphism. (Note that, for a point $p = [(C, \Sigma_C)] \in \overline{\mathcal{M}}$, the base change map

$$\mathcal{E}xt^1_\pi(\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma), \mathcal{O}_\mathcal{U}) \otimes k(p) \to \mathcal{E}xt^1(\Omega_C(\Sigma_C), \mathcal{O}_C)$$

is an isomorphism. Indeed, by relative duality [Kleiman 1980, Theorem 21], it suffices to show that $\pi_*(\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma) \otimes \omega_\pi)$ commutes with base change. This follows from cohomology and base change.)

Consider the exact sequences

$$0 \to T_{\overline{\mathcal{M}}}(-\log \mathcal{B}) \to T_{\overline{\mathcal{M}}} \to \nu_* \mathcal{N} \to 0$$

and

$$0 \to R^1\pi_*(\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma)^\vee) \to \mathcal{E}xt^1_\pi(\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma), \mathcal{O}_\mathcal{U}) \to \pi_* \mathcal{E}xt^1(\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma), \mathcal{O}_\mathcal{U}) \to 0.$$

The Kodaira–Spencer map ρ identifies the middle terms, and induces an identification of the right end terms determined by the deformations of the singularities of the fibres of π. We thus obtain a natural isomorphism δ of the left end terms. Finally, note that $\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma)^\vee = \omega_\pi(\Sigma)^\vee$ because $\omega_\pi(\Sigma)$ is invertible and agrees with $\Omega_{\mathcal{U}/\overline{\mathcal{M}}}(\Sigma)$ in codimension 1. This completes the proof. □

The line bundle $\omega_\pi(\Sigma)$ is ample on fibres of π. Hence $\pi_*(\omega_\pi(\Sigma)^\vee) = 0$. Also $R^i\pi_*(\omega_\pi(\Sigma)^\vee) = 0$ for $i > 1$ by dimensions. So

$$H^{i+1}(\omega_\pi(\Sigma)^\vee) = H^i\big(R^1\pi_*(\omega_\pi(\Sigma)^\vee)\big)$$

for all i by the Leray spectral sequence. Hence the isomorphism δ induces an isomorphism

$$H^i\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big) \xrightarrow{\sim} H^{i+1}(\omega_\pi(\Sigma)^\vee) \tag{3-2}$$

for each i.

Let $U_{g,n}$ denote the coarse moduli space of the stack $\mathcal{U}_{g,n}$ and $p : \mathcal{U}_{g,n} \to U_{g,n}$ the projection. The line bundle $\omega_\pi(\Sigma)$ on the stack $\mathcal{U}_{g,n}$ defines a $\mathbb{Q}$-line bundle $p_*^{\mathbb{Q}}\omega_\pi(\Sigma)$ on the coarse moduli space $U_{g,n}$ (see the Appendix). We use the following important result, which is essentially due to Arakelov [1971, Proposition 3.2, p. 1297]. We refer to [Keel 1999, Section 4] for the proof.

Theorem 3.2. *The $\mathbb{Q}$-line bundle $p_*^{\mathbb{Q}}\omega_\pi(\Sigma)$ is big and nef on $U_{g,n}$.*

It follows by Kodaira vanishing (see Theorem A.1) that $H^i(\omega_\pi(\Sigma)^\vee) = 0$ for $i < \dim \mathcal{U}_{g,n}$. Combining with (3-2), we deduce

Proposition 3.3. $H^i\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big) = 0$ *for* $i < \dim \overline{\mathcal{M}}_{g,n}$.

In particular,

$$H^1\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big) = 0$$

if $\dim \overline{\mathcal{M}}_{g,n} > 1$. The remaining cases are easy to check. This completes the proof of Theorem 2.2.

4. Proof of Theorem 2.1

We now prove that $\overline{\mathcal{M}}_{g,n}$ is rigid. Since $\overline{\mathcal{M}}_{g,n}$ is a smooth Deligne–Mumford stack, its first order infinitesimal deformations are identified with the space $H^1(T_{\overline{\mathcal{M}}_{g,n}})$, and we must show that $H^1(T_{\overline{\mathcal{M}}_{g,n}}) = 0$. Consider the exact sequence

$$0 \to T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B}) \to T_{\overline{\mathcal{M}}_{g,n}} \to \nu_* \mathcal{N} \to 0$$

and the associated long exact sequence of cohomology

$$\cdots \to H^i\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big) \to H^i(T_{\overline{\mathcal{M}}_{g,n}}) \to H^i(\mathcal{N}) \to \cdots .$$

We prove below that $H^i(\mathcal{N}) = 0$ for $i < \dim \mathcal{B}$. Now

$$H^i\big(T_{\overline{\mathcal{M}}_{g,n}}(-\log \mathcal{B})\big) = 0$$

for $i < \dim \overline{\mathcal{M}}_{g,n}$ by Proposition 3.3, so we deduce

Proposition 4.1. *$H^i(T_{\overline{\mathcal{M}}_{g,n}}) = 0$ for $i < \dim \overline{\mathcal{M}}_{g,n} - 1$.*

In particular, $H^1(T_{\overline{\mathcal{M}}_{g,n}}) = 0$ if $\dim \overline{\mathcal{M}}_{g,n} > 2$. In the remaining cases it is easy to check that $H^1(\mathcal{N}) = 0$, so again $H^1(T_{\overline{\mathcal{M}}_{g,n}}) = 0$.

The irreducible components of the normalisation $\mathcal{B}^\nu$ of the boundary $\mathcal{B}$ of $\overline{\mathcal{M}}_{g,n}$ are finite images of the following stacks [Knudsen 1983a, Definition 3.8, Corollary 3.9]:

(1) $\overline{\mathcal{M}}_{g_1, S_1 \cup \{n+1\}} \times \overline{\mathcal{M}}_{g_2, S_2 \cup \{n+2\}}$, where $g_1 + g_2 = g$ and S_1, S_2 is a partition of $\{1, \ldots, n\}$, and

(2) $\overline{\mathcal{M}}_{g-1, n+2}$.

Here $\overline{\mathcal{M}}_{h,S}$ denotes the moduli stack of stable curves of genus h with marked points labelled by a finite set S. In each case the map to $\mathcal{B}^\nu$ is given by identifying the points labelled by $n+1$ and $n+2$. The map is an isomorphism onto the component of $\mathcal{B}^\nu$ except in case (1) for $g_1 = g_2$ and $n = 0$ and case (2), when it is étale of degree 2.

For $\overline{\mathcal{M}}_{h,S}$ a moduli stack of pointed stable curves as above, let $\pi : \mathcal{U}_{h,S} \to \overline{\mathcal{M}}_{h,S}$ denote the universal family, and $x_i : \overline{\mathcal{M}}_{h,S} \to \mathcal{U}_{h,S}$ for $i \in S$, the tautological sections of π. Define $\psi_i = x_i^* \omega_\pi$, the pullback of the relative dualising sheaf of π along the section x_i. The following result is well known; see, for example, [Harris and Morrison 1998, Proposition 3.32].

Lemma 4.2. *The pullback of* $\mathcal{N}^\vee$ *to* $\overline{\mathcal{M}}_{g_1,S_1\cup\{n+1\}} \times \overline{\mathcal{M}}_{g_2,S_2\cup\{n+2\}}$ *is identified with* $\mathrm{pr}_1^*\,\psi_{n+1} \otimes \mathrm{pr}_2^*\,\psi_{n+2}$. *Similarly, the pullback of* $\mathcal{N}^\vee$ *to* $\overline{\mathcal{M}}_{g-1,n+2}$ *is identified with* $\psi_{n+1} \otimes \psi_{n+2}$.

There is an isomorphism of stacks $c : \overline{\mathcal{M}}_{g,n+1} \to \mathcal{U}_{g,n}$ which identifies the morphism $p_{n+1} : \overline{\mathcal{M}}_{g,n+1} \to \overline{\mathcal{M}}_{g,n}$ given by forgetting the last point with the projection $\pi : \mathcal{U}_{g,n} \to \overline{\mathcal{M}}_{g,n}$; see [Knudsen 1983a, Section 1–2].

Lemma 4.3 [Knudsen 1983b, Theorem 4.1(d), p. 202]. *The line bundle* ψ_{n+1} *on* $\overline{\mathcal{M}}_{g,n+1}$ *is identified with the pullback of the line bundle* $\omega_\pi(\Sigma)$ *under the isomorphism* $c\colon \overline{\mathcal{M}}_{g,n+1} \to \mathcal{U}_{g,n}$.

Corollary 4.4. *The* $\mathbb{Q}$*-line bundle on the coarse moduli space of* $\mathcal{B}^\nu$ *defined by* $\mathcal{N}^\vee$ *is big and nef on each component*

Proof. This follows immediately from Lemmas 4.2, 4.3, and Theorem 3.2. □

We deduce that $H^i(\mathcal{N}) = 0$ for $i < \dim \mathcal{B}$ by Theorem A.1. This completes the proof of Theorem 2.1.

5. Proof of Theorem 2.3

We first prove a basic result which relates the deformations of a smooth Deligne–Mumford stack and its coarse moduli space.

Let $\mathcal{X}$ be a smooth proper Deligne–Mumford stack, X the coarse moduli space of $\mathcal{X}$, and $p : \mathcal{X} \to X$ the projection. Let $T_{\mathcal{X}}$ denote the tangent sheaf of $\mathcal{X}$. Let $D \subset X$ be the union of the codimension one components of the branch locus of $p : \mathcal{X} \to X$ (with its reduced structure). Let $T_X(-\log D)$ denote the subsheaf of the tangent sheaf T_X consisting of derivations which preserve the ideal sheaf of D. It is the sheaf of first order infinitesimal automorphisms of the pair (X, D).

Lemma 5.1. $p_*T_{\mathcal{X}} = T_X(-\log D)$

Proof. The sheaves $p_*T_{\mathcal{X}}$ and $T_X(-\log D)$ satisfy Serre's S_2 condition, and are identified over the locus where p is étale. So it suffices to work in codimension 1. We reduce to the case $\mathcal{X} = [\mathbb{A}^1_x/\mu_e]$, where $\mu_e \ni \zeta : x \mapsto \zeta x$. Then $X = \mathbb{A}^1_x/\mu_e = \mathbb{A}^1_y$, where $y = x^e$, and $D = (y = 0) \subset X$. Let $\pi : \mathbb{A}^1_x \to \mathbb{A}^1_x/\mu_e$ be the quotient map. We compute

$$p_*T_{\mathcal{X}} = \left(\pi_*\mathcal{O}_{\mathbb{A}^1_x} \cdot \frac{\partial}{\partial x}\right)^{\mu_e} = \mathcal{O}_{\mathbb{A}^1_y} \cdot x\frac{\partial}{\partial x} = \mathcal{O}_{\mathbb{A}^1_y} \cdot y\frac{\partial}{\partial y} = T_X(-\log D),$$

as required. □

Proposition 5.2. *The first order deformations of the stack* $\mathcal{X}$ *are identified with the first order locally trivial deformations of the pair* (X, D).

Proof. By the Lemma, $H^1(T_{\mathcal{X}}) = H^1(p_*T_{\mathcal{X}}) = H^1(T_X(-\log D))$. □

We now apply this result to relate deformations of the stack $\overline{\mathcal{M}}_{g,n}$ and its coarse moduli space $\overline{M}_{g,n}$.

A stable n-pointed curve of genus 0 has no nontrivial automorphisms. Hence the stack $\overline{\mathcal{M}}_{0,n}$ is equal to its coarse moduli space $\overline{M}_{0,n}$, and $\overline{M}_{0,n}$ is rigid by Theorem 2.1. Also, recall that $\overline{M}_{1,1}$ is isomorphic to $\mathbb{P}^1$ and therefore rigid. So, in the following, we assume that $g \neq 0$ and $(g, n) \neq (1, 1)$.

Let $\mathcal{D} \subset \overline{\mathcal{M}}_{g,n}$ be the component of the boundary whose general point is a curve with two components of genus 1 and $g-1$ meeting in a node, with each of the n marked points on the component of genus $g-1$. Note that each point of $\mathcal{D}$ has a nontrivial automorphism given by the involution of the component of genus 1 fixing the node. Let $p : \overline{\mathcal{M}}_{g,n} \to \overline{M}_{g,n}$ be the projection, and $D \subset \overline{M}_{g,n}$ the coarse moduli space of $\mathcal{D}$.

Lemma 5.3 [Harris and Mumford 1982, § 2]. *If $g+n \geq 4$ then the automorphism group of a general point of $\overline{\mathcal{M}}_{g,n}$ is trivial, and the divisor $D \subset \overline{M}_{g,n}$ is the unique codimension* 1 *component of the branch locus of p.*

Assume $g+n \geq 4$. Let $\nu : \mathcal{D}^\nu \to \mathcal{D}$ denote the normalisation of $\mathcal{D}$, so $\mathcal{D}^\nu = \overline{\mathcal{M}}_{1,1} \times \overline{\mathcal{M}}_{g-1,n+1}$. Let $\mathcal{N}_D$ denote the normal bundle of the map $\mathcal{D}^\nu \to \overline{\mathcal{M}}_{g,n}$.

Lemma 5.4. *There is an exact sequence*

$$0 \to T_{\overline{M}_{g,n}}(-\log D) \to T_{\overline{M}_{g,n}} \to p_* \nu_* \mathcal{N}_D^{\otimes 2} \to 0.$$

Proof. This is a straightforward calculation similar to [Harris and Mumford 1982, Lemma, p. 52]. □

We have

$$H^1\big(T_{\overline{M}_{g,n}}(-\log D)\big) = H^1(T_{\overline{\mathcal{M}}_{g,n}}) = 0$$

by Proposition 5.2 and Theorem 2.1. Also $H^1(\mathcal{N}_D^{\otimes 2}) = 0$ by Theorem A.1 because the $\mathbb{Q}$-line bundle defined by $\mathcal{N}_D^\vee$ on the coarse moduli space of $\mathcal{D}^\nu$ is big and nef by Corollary 4.4. So $H^1(T_{\overline{M}_{g,n}}) = 0$ by Lemma 5.4, that is, $\overline{M}_{g,n}$ has no locally trivial deformations. This concludes the proof of Theorem 2.3.

6. Nonrigidity of moduli of surfaces

We exhibit a moduli space of surfaces with boundary that is not rigid.

Let $P_1, \ldots, P_4$ be 4 points in linear general position in $\mathbb{P}^2$. Let l_{ij} be the line through P_i and P_j. Let l be a line through the point $Q = l_{12} \cap l_{34}$ such that l does not pass through $l_{13} \cap l_{24}$ or $l_{14} \cap l_{23}$ and is not equal to l_{12} or l_{34}. Let $S \to \mathbb{P}^2$ be the blowup of the points $P_1, \ldots, P_4$, Q, and B the sum of the strict transforms of l and the l_{ij} and the exceptional curves. Then (S, B) is a smooth surface with normal crossing boundary such that $K_S + B$ is very ample. We fix an ordering $B_1, \ldots, B_{12}$ of the components of B. The moduli stack $\mathcal{M}$ of deformations of

(S, B) is isomorphic to $\mathbb{P}^1 \setminus \{q_1, \dots, q_4\}$ where the q_i are distinct points. Indeed, it suffices to observe that all deformations of (S, B) are obtained by the construction above. The moduli space $\mathcal{M}$ has a modular compactification $(\overline{M}, \partial\overline{M})$, the Kollár–Shepherd-Barron–Alexeev moduli stack of stable surfaces with boundary, which is isomorphic to $(\mathbb{P}^1, \sum q_i)$. In particular, the pair $(\overline{M}, \partial\overline{M})$ has nontrivial deformations.

Remark 6.1. The compact moduli space $\overline{M}$ is an instance of the compactifications of moduli spaces of hyperplane arrangements described in [Lafforgue 2003] (see also [Hacking et al. 2006]).

Appendix: Kodaira vanishing for stacks

Let $\mathcal{X}$ be a smooth proper Deligne–Mumford stack, X the coarse moduli space of $\mathcal{X}$, and $p : \mathcal{X} \to X$ the projection. Étale locally on X, $p : \mathcal{X} \to X$ is of the form $p : [U/G] \to U/G$, where U is a smooth affine variety and G is a finite group acting on U [Abramovich and Vistoli 2002, Lemma 2.2.3, p. 32]. A sheaf $\mathcal{F}$ on $[U/G]$ corresponds to a G-equivariant sheaf $\mathcal{F}_U$ on U, and $p_*\mathcal{F} = (\pi_*\mathcal{F}_U)^G$, where $\pi : U \to U/G$ is the quotient map.

Let $\mathcal{L}$ be a line bundle on $\mathcal{X}$. Let $n \in \mathbb{N}$ be sufficiently divisible so that for each open patch $[U/G]$ of $\mathcal{X}$ as above and point $q \in U$ the stabilizer G_q of q acts trivially on the fibre of $\mathcal{L}_U^{\otimes n}$ over q. Then the pushforward $p_*(\mathcal{L}^{\otimes n})$ is a line bundle on X. We define $p_*^{\mathbb{Q}}\mathcal{L} = \frac{1}{n} p_*(\mathcal{L}^{\otimes n}) \in \operatorname{Pic}(X) \otimes \mathbb{Q}$, and call $p_*^{\mathbb{Q}}\mathcal{L}$ the $\mathbb{Q}$-line bundle on X defined by $\mathcal{L}$.

Theorem A.1. *Assume that the coarse moduli space X is an algebraic variety. If the $\mathbb{Q}$-line bundle $p_*^{\mathbb{Q}}\mathcal{L}$ on X is big and nef then $H^i(\mathcal{L}^\vee) = 0$ for $i < \dim \mathcal{X}$.*

Remark A.2. If the coarse moduli space X is smooth then Theorem A.1 follows from [Matsuki and Olsson 2005, Theorem 2.1].

Theorem A.1 is proved by reducing to the following generalisation of the Kodaira vanishing theorem.

Theorem A.3 [Kollár and Mori 1998, Theorem 2.70, p. 73]. *Let X be a proper normal variety and Δ a $\mathbb{Q}$-divisor on X such that the pair (X, Δ) is Kawamata log terminal (klt). Let N be a $\mathbb{Q}$-Cartier Weil divisor on X such that $N \equiv M + \Delta$, where M is a big and nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor. Then $H^i(X, \mathcal{O}_X(-N)) = 0$ for $i < \dim X$.*

Proof of Theorem A.1. Observe first that X is a normal variety with quotient singularities. Consider the sheaf $p_*(\mathcal{L}^\vee)$ on X. If the automorphism group of a general point of $\mathcal{X}$ acts nontrivially on $\mathcal{L}$, then $p_*\mathcal{L}^\vee = 0$, and so $H^i(\mathcal{L}^\vee) = H^i(p_*\mathcal{L}^\vee) = 0$ for each i. Suppose now that the automorphism group of a general point acts trivially on $\mathcal{L}$. Then $p_*\mathcal{L}^\vee$ is a rank 1 reflexive sheaf on X. Write $p_*\mathcal{L}^\vee = \mathcal{O}_X(-N)$, where N is a Weil divisor on X. Let $n \in \mathbb{N}$ be sufficiently divisible so that

$$p_*^{\mathbb{Q}}(\mathscr{L}) = \frac{1}{n} p_*(\mathscr{L}^{\otimes n})$$

as above. Let M be a $\mathbb{Q}$-divisor corresponding to the $\mathbb{Q}$-line bundle $p_*^{\mathbb{Q}}\mathscr{L}$. There is a natural map $(p_*\mathscr{L}^\vee)^{\otimes n} \to p_*(\mathscr{L}^{\vee\otimes n})$, that is, a map $\mathbb{O}_X(-nN) \to \mathbb{O}_X(-nM)$, which is an isomorphism over the locus where p is étale. So $N \equiv M + \Delta$, where Δ is an effective $\mathbb{Q}$-divisor supported on the branch locus of p. Let $D_1, \dots, D_r$ be the codimension 1 components of the branch locus. Let e_i be the ramification index at D_i, and a_i the age of the line bundle $\mathscr{L}^\vee$ along D_i. That is, after removing the automorphism group of a general point of $\mathscr{X}$, a transverse slice of $\mathscr{X}$ at a general point of D_i is of the form $[\mathbb{A}^1_x/\mu_{e_i}]$, where $\mu_{e_i} \ni \zeta : x \mapsto \zeta \cdot x$, and μ_{e_i} acts on the fibre of $\mathscr{L}^\vee$ by the character $\zeta \mapsto \zeta^{-a_i}$, where $0 \le a_i \le e_i - 1$. We compute that $\Delta = \sum \frac{a_i}{e_i} D_i$.

We claim that (X, Δ) is klt. Let $\Delta' = \sum \frac{e_i - 1}{e_i} D_i$, then $K_{\mathscr{X}} = p^*(K_X + \Delta')$, and $\mathscr{X}$ is smooth, so (X, Δ') is klt by [Kollár and Mori 1998, Proposition 5.20(4), p. 160]. Now $\Delta \le \Delta'$ and X is $\mathbb{Q}$-factorial, so (X, Δ) is also klt. We deduce that $H^i(\mathscr{L}^\vee) = H^i(p_*\mathscr{L}^\vee) = H^i(\mathbb{O}_X(-N)) = 0$ for $i < \dim \mathscr{X}$ by Theorem A.3. $\square$

Acknowledgements

We thank M. Kapranov for suggesting the problem and subsequent encouragement. We also thank I. Dolgachev, G. Farkas, S. Grushevsky, S. Keel, J. Kollár, S. Kovács, R. Lazarsfeld, A. Marian, M. Olsson, S.P. Smith and T. Stafford for helpful discussions.

References

[Abramovich and Vistoli 2002] D. Abramovich and A. Vistoli, "Compactifying the space of stable maps", *J. Amer. Math. Soc.* **15**:1 (2002), 27–75. MR 2002i:14030 Zbl 0991.14007

[Arakelov 1971] S. J. Arakelov, "Familien algebraischer Kurven mit festen Entartungen", *Izv. Akad. Nauk SSSR Ser. Mat.* **35** (1971), 1269–1293. In Russian; translated in *Math. USSR-Izv.* **5** *(1971), 1277–1302*. MR 48 #298 Zbl 0238.14012

[Deligne and Mumford 1969] P. Deligne and D. Mumford, "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* **36** (1969), 75–109. MR 41 #6850 Zbl 0181.48803

[Hacking et al. 2006] P. Hacking, S. Keel, and J. Tevelev, "Compactification of the moduli space of hyperplane arrangements", *J. Algebraic Geom.* **15**:4 (2006), 657–680. MR 2007j:14016 Zbl 1117.14036

[Harris and Morrison 1998] J. Harris and I. Morrison, *Moduli of curves*, Graduate Texts in Mathematics **187**, Springer, New York, 1998. MR 99g:14031 Zbl 0913.14005

[Harris and Mumford 1982] J. Harris and D. Mumford, "On the Kodaira dimension of the moduli space of curves", *Invent. Math.* **67**:1 (1982), 23–88. MR 83i:14018 Zbl 0506.14016

[Kapranov 1997] M. Kapranov, "Deformations of moduli spaces", 1997. Unpublished manuscript.

[Keel 1999] S. Keel, "Basepoint freeness for nef and big line bundles in positive characteristic", *Ann. of Math.* (2) **149**:1 (1999), 253–286. MR 2000j:14011 Zbl 0954.14004

[Kleiman 1980] S. L. Kleiman, "Relative duality for quasicoherent sheaves", *Compositio Math.* **41**:1 (1980), 39–60. MR 81m:14017 Zbl 0403.14003

[Knudsen 1983a] F. F. Knudsen, "The projectivity of the moduli space of stable curves. II. The stacks $M_{g,n}$", *Math. Scand.* **52**:2 (1983), 161–199. MR 85d:14038a Zbl 0544.14020

[Knudsen 1983b] F. F. Knudsen, "The projectivity of the moduli space of stable curves. III. The line bundles on $M_{g,n}$, and a proof of the projectivity of $\overline{M}_{g,n}$ in characteristic 0", *Math. Scand.* **52**:2 (1983), 200–212. MR 85d:14038b Zbl 0544.14021

[Kollár and Mori 1998] J. Kollár and S. Mori, *Birational geometry of algebraic varieties*, Cambridge Tracts in Mathematics **134**, Cambridge University Press, Cambridge, 1998. Translated from the 1998 Japanese original. MR 2000b:14018 Zbl 0926.14003

[Lafforgue 2003] L. Lafforgue, *Chirurgie des grassmanniennes*, CRM Monograph Series **19**, American Mathematical Society, Providence, RI, 2003. MR 2004k:14085 Zbl 1026.14015

[Matsuki and Olsson 2005] K. Matsuki and M. Olsson, "Kawamata–Viehweg vanishing as Kodaira vanishing for stacks", *Math. Res. Lett.* **12**:2-3 (2005), 207–217. MR 2006c:14023 Zbl 1080.14023

[Siu 1980] Y. T. Siu, "The complex-analyticity of harmonic maps and the strong rigidity of compact Kähler manifolds", *Ann. of Math.* (2) **112**:1 (1980), 73–111. MR 81j:53061 Zbl 0517.53058

Communicated by Ravi Vakil
Received 2007-11-30 Revised 2008-08-06 Accepted 2008-09-17

hacking@math.washington.edu *Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350, United States*
www.math.washington.edu/~hacking/

Quasimaps, straightening laws, and quantum cohomology for the Lagrangian Grassmannian

James Ruffo

The Drinfel'd Lagrangian Grassmannian compactifies the space of algebraic maps of fixed degree from the projective line into the Lagrangian Grassmannian. It has a natural projective embedding arising from the canonical embedding of the Lagrangian Grassmannian. We show that the defining ideal of any Schubert subvariety of the Drinfel'd Lagrangian Grassmannian is generated by polynomials which give a straightening law on an ordered set. Consequentially, any such subvariety is Cohen–Macaulay and Koszul. The Hilbert function is computed from the straightening law, leading to a new derivation of certain intersection numbers in the quantum cohomology ring of the Lagrangian Grassmannian.

1. Introduction

The space of algebraic maps of degree d from $\mathbb{P}^1$ to a projective variety X has applications to mathematical physics, linear systems theory, quantum cohomology, geometric representation theory, and the geometric Langlands correspondence [Braverman 2006; Sottile 2000; 2001]. This space is (almost) never compact, so various compactifications have been introduced to help understand its geometry. Among these (at least when X is a flag variety) are Kontsevich's space of *stable maps* [Fulton and Pandharipande 1997; Kontsevich 1995], the *quot scheme* (or space of *quasiflags*) [Chen 2001; Laumon 1990; Strømme 1987], and the *Drinfel'd compactification* (or space of *quasimaps*). The latter space is defined concretely as a projective variety, and much information can be gleaned directly from its defining equations.

Inspired by the work of Hodge [1943], standard monomial theory was developed by Lakshmibai [2003], Musili [2003], Seshadri, and others (see also the references therein), to study the *flag varieties* G/P, where G is a semisimple algebraic group and $P \subseteq G$ is a parabolic subgroup. These spaces have a decomposition into

MSC2000: primary 13F50; secondary 13P10, 14N35, 14N15.

Keywords: algebra with straightening law, quasimap, Lagrangian Grassmannian, quantum cohomology.

Schubert cells, whose closures (the *Schubert varieties*) give a basis for cohomology. As consequences of standard monomial theory, Schubert varieties are normal and Cohen–Macaulay, and one has an explicit description of their singularities and defining ideals.

A key part of standard monomial theory is that any flag variety G/P (P a parabolic subgroup) has a projective embedding which presents its coordinate ring as an *algebra with straightening law* (Definition 3.3), a special case of a *Hodge algebra* [De Concini et al. 1982]. This idea originates with the work of Hodge on the Grassmannian [1943], and was extended to the Lagrangian Grassmannian by De Concini and Lakshmibai [1981]. This framework was extended to many other flag varieties by Lakshmibai [2003], Seshadri, and their coauthors [Musili 2003]. Littelmann's path model for representations of algebraic groups [1998] later provided the necessary tools to treat all flag varieties in a unified way, as carried out by Chirivì [2000; 2001]. The general case requires a more expansive notion of an algebra with straightening law (also due to Chirivì), called a *Lakshmibai–Seshadri (LS) algebra*.

Sottile and Sturmfels [2001] have extended standard monomial theory to the *Drinfel'd Grassmannian* parametrizing algebraic maps from $\mathbb{P}^1$ into the Grassmannian. They define Schubert subvarieties of this space and prove that the homogeneous coordinate ring of any Schubert variety (including the Drinfel'd Grassmannian itself) is an algebra with straightening law on a distributive lattice. Using this fact, the authors show that these Schubert varieties are normal, Cohen–Macaulay and Koszul, and have rational singularities.

We extend these results to the *Drinfel'd Lagrangian Grassmannian*, which parametrizes algebraic maps from $\mathbb{P}^1$ into the Lagrangian Grassmannian. In particular, we prove:

Theorem 1.1. *The coordinate ring of any Schubert subvariety of the Drinfel'd Lagrangian Grassmannian is an algebra with straightening law on a doset.*

See Theorems 5.9 and 5.10 in Section 5 for more details.

A doset, as introduced in [De Concini and Lakshmibai 1981], is a certain kind of ordered set (Definition 3.1). As consequences of Theorem 1.1, we show that the coordinate ring is reduced, Cohen–Macaulay, and Koszul, and obtain formulas for its degree and dimension. These formulas have an interpretation in terms of quantum cohomology, as described in Section 4.

In Section 2, we review the basic definitions and facts concerning Drinfel'd compactifications and the Lagrangian Grassmannian. Section 3 provides the necessary background on algebras with straightening law. We discuss an application to the quantum cohomology of the Lagrangian Grassmannian in Section 4. Our main result and its consequences are proved in Section 5.

2. Preliminaries

We first give a precise definition of the Drinfel'd compactification of the space of algebraic maps from $\mathbb{P}^1$ to a homogeneous variety. We then review the basic facts we will need regarding the Lagrangian Grassmannian.

2A. ***Spaces of algebraic maps.*** Let G be a semisimple linear algebraic group. Fix a Borel subgroup $B \subseteq G$ and a maximal torus $T \subseteq B$. Let R be the set of roots (determined by T), and $S := \{\rho_1, \ldots, \rho_r\}$ the simple roots (determined by B). The simple roots form an ordered basis for the Lie algebra $\mathfrak{t}$ of T; let $\{\omega_1, \ldots, \omega_r\}$ be the dual basis (the *fundamental weights*). The *Weyl group* W is the normalizer of T modulo T itself.

Let $P \subseteq G$ be the maximal parabolic subgroup associated to the fundamental weight ω, let $\mathrm{L}(\omega)$ be the irreducible representation of highest weight ω, and let $(\bullet, \bullet)$ denote the Killing form on $\mathfrak{t}$. The flag variety G/P embeds in $\mathbb{P}\mathrm{L}(\omega)$ as the orbit of a point $[v] \in \mathbb{P}\mathrm{L}(\omega)$, where $v \in \mathrm{L}(\omega)$ is a highest weight vector. Define the *degree* of a algebraic map $f : \mathbb{P}^1 \to G/P$ to be its degree as a map into $\mathbb{P}\mathrm{L}(\omega)$. For $\rho \in R$, set $\rho^\vee := 2\rho/(\rho, \rho)$. For simplicity, assume that $(\omega, \rho^\vee) \leq 2$ for all $\rho \in S$ (that is, P is of *classical type* [Lakshmibai 2003]). This condition implies that $\mathrm{L}(\omega)$ has T-fixed lines indexed by certain admissible pairs of elements of W/W_P.

Let $\mathcal{M}_d(G/P)$ be the space of algebraic maps of degree d from $\mathbb{P}^1$ into G/P. If P is of classical type then the set $\mathfrak{D}$ of *admissible pairs* indexes homogeneous coordinates on $\mathbb{P}\mathrm{L}(\omega)$ (see Definition 3.16, and [Lakshmibai 2003; Musili 2003] for a more thorough treatment). Therefore, any map $f \in \mathcal{M}_d(G/P)$ can be expressed as

$$f : [s, t] \mapsto [p_w(s, t) \mid w \in \mathfrak{D}],$$

where the $p_w(s, t)$ are homogeneous forms of degree d. This leads to an embedding of $\mathcal{M}_d(G/P)$ into $\mathbb{P}\big((S^d\mathbb{C}^2)^* \otimes \mathrm{L}(\omega)\big)$, where $(S^d\mathbb{C}^2)^*$ is the space of homogeneous forms of degree d in two variables. The coefficients of the homogeneous forms in $(S^d\mathbb{C}^2)^*$ give coordinate functions on $(S^d\mathbb{C}^2)^* \otimes \mathrm{L}(\omega)$; they are indexed by the set $\{w^{(a)} \mid w \in \mathfrak{D}, a = 0, \ldots, d\}$, a disjoint union of $d+1$ copies of $\mathfrak{D}$.

The closure of

$$\mathcal{M}_d(G/P) \subseteq \mathbb{P}\big((S^d\mathbb{C}^2)^* \otimes \mathrm{L}(\omega)\big)$$

is called the *Drinfel'd compactification* and denoted $\mathfrak{Q}_d(G/P)$. This definition is due to V. Drinfel'd, dating from the mid-1980s. Drinfel'd never published this definition himself; to the author's knowledge its first appearance in print was in [Rosenthal 1994]; see also [Kuznetsov 1997].

Let $G = \mathrm{SL}_n(\mathbb{C})$ and P be the maximal parabolic subgroup stabilizing a fixed k-dimensional subspace of $\mathbb{C}^n$, so that $G/P = \mathrm{Gr}(k, n)$. In this case we denote the *Drinfel'd Grassmannian* $\mathfrak{Q}_d(G/P)$ by $\mathfrak{Q}_d(k, n)$. In [Sottile and Sturmfels 2001] it is shown that the homogeneous coordinate ring of $\mathcal{Q}_d(k,n)$ is an algebra with straightening law on the distributive lattice

$$\binom{[n]}{k}_d := \left\{ \alpha^{(a)} \;\middle|\; \alpha \in \binom{[n]}{k}, 0 \le a \le d \right\},$$

with partial order on $\binom{[n]}{k}_d$ defined by $\alpha^{(a)} \le \beta^{(b)}$ if and only if $a \le b$ and $\alpha_i \le \beta_{b-a+i}$ for $i = 1, \ldots, k-b+a$. It follows that the homogeneous coordinate ring of $\mathcal{Q}_d(k,n)$ is normal, Cohen–Macaulay, and Koszul, and that the ideal $I^d_{k,n-k} \subseteq \mathbb{C}[\binom{[n]}{k}_d]$ has a quadratic Gröbner basis consisting of the straightening relations.

Taking $d = 0$ above, this partial order is the classical *Bruhat order* on $\binom{[n]}{k}$. In general, given a semisimple algebraic group G with parabolic subgroup P, the Bruhat order is an ordering on the set of maximal coset representatives of the quotient of the Weyl group of G by the Weyl group of P.

Suppose that $d = \ell k + q$ for nonnegative integers ℓ and q with $q < k$, and let $X = (x_{ij})_{1 \le i,j \le n}$ be a matrix with polynomial entries

$$x_{ij} = x_{ij}^{(k_i)} t^{k_i} + \cdots + x_{ij}^{(1)} t + x_{ij}^{(0)},$$

where $k_i = \ell + 1$ if $i \le q$ and $k_i = \ell$ if $i > q$. The ideal $I^d_{k,n-k}$ is the kernel of the map

$$\varphi : \mathbb{C}\left[\binom{[n]}{k}_d\right] \to \mathbb{C}[X]$$

sending the variable $p_\alpha^{(a)}$ indexed by $\alpha^{(a)} \in \binom{[n]}{k}_d$ to the coefficient of t^a in the maximal minor of X whose columns are indexed by α.

The main results of [Sottile and Sturmfels 2001] follow from the next proposition. Given any distributive lattice, we denote by $\wedge$ and $\vee$, respectively, the *meet* and *join*. The symbol $\wedge$ will also be used for exterior products of vectors, but the meaning should be clear from the context.

Proposition 2.1 [Sottile and Sturmfels 2001, Theorem 10]. *Let α, β be a pair of incomparable variables in the poset $\binom{[n]}{k}_d$. There is a quadratic polynomial $S(\alpha,\beta)$ lying in the kernel of $\varphi : \mathbb{C}[\binom{[n]}{k}_d] \to \mathbb{C}[X]$ whose first two monomials are*

$$p_\alpha p_\beta - p_{\alpha\wedge\beta} p_{\alpha\vee\beta}.$$

Moreover, if $\lambda p_\gamma p_\delta$ is any noninitial monomial in $S(\alpha,\beta)$, then α, β lies in the interval $[\gamma,\delta] = \{\theta \in \binom{[n]}{k}_d \mid \gamma \le \theta \le \delta\}$.

The quadratic polynomials $S(\alpha,\beta)$ in fact form a Gröbner basis for the ideal they generate. It is shown in [Sottile and Sturmfels 2001] that there exists a toric (SAGBI) deformation taking $S(\alpha,\beta)$ to its *initial form* $p_\alpha p_\beta - p_{\alpha\wedge\beta} p_{\alpha\vee\beta}$, deforming the Drinfel'd Grassmannian into a toric variety.

Our goal is to extend the main results of standard monomial theory to the *Lagrangian Drinfel'd Grassmannian* $L\mathfrak{Q}_d(n) := \mathfrak{Q}_d(\mathrm{LG}(n))$ of degree-d maps from $\mathbb{P}^1$ into the Lagrangian Grassmannian.

2B. ***The Lagrangian Grassmannian.*** De Concini and Lakshmibai [1981] showed that, in its natural projective embedding, the Lagrangian Grassmannian $\mathrm{LG}(n)$ is defined by quadratic relations which give a straightening law on a doset. These relations are obtained by expressing $\mathrm{LG}(n)$ as a linear section of $\mathrm{Gr}(n, 2n)$. While this is well known, the author knows of no explicit derivation of these relations which do not require the representation theory of semisimple algebraic groups. We provide a derivation which does not rely upon representation theory (although we adopt the notation and terminology). This will be useful when we consider the Drinfel'd Lagrangian Grassmannian, to which representation theory has yet to be successfully applied.

Set $[n] := \{1, 2, \dots, n\}$, $\bar{\imath} := -i$, and $\langle n\rangle := \{\bar{n}, \dots, \bar{1}, 1, \dots n\}$. If S is any set, let $\binom{S}{k}$ be the collection of subsets $\alpha = \{\alpha_1, \dots, \alpha_k\}$ of cardinality k.

The projective space $\mathbb{P}(\bigwedge^n \mathbb{C}^{2n})$ has *Plücker coordinates* indexed by the distributive lattice $\binom{\langle n\rangle}{n}$, and the Grassmannian $\mathrm{Gr}(n, 2n)$ is the subvariety, defined by the Plücker relations, of $\mathbb{P}(\bigwedge^n \mathbb{C}^{2n})$.

Proposition 2.2 [Fulton 1997; Hodge 1943]. *For $\alpha, \beta \in \binom{\langle n\rangle}{n}$ there is a Plücker relation*

$$p_\alpha p_\beta - p_{\alpha\wedge\beta} p_{\alpha\vee\beta} + \sum_{\gamma\le\alpha\wedge\beta<\alpha\vee\beta\le\delta} c^{\gamma,\delta}_{\alpha,\beta} p_\gamma p_\delta = 0.$$

The defining ideal of $\mathrm{Gr}(n, 2n) \subseteq \mathbb{P}(\bigwedge^n \mathbb{C}^{2n})$ *is generated by the Plücker relations.*

Fix an ordered basis $\{e_{\bar{n}}, \dots, e_{\bar{1}}, e_1, \dots, e_n\}$ of the vector space $\mathbb{C}^{2n}$, and let $\Omega := \sum_{i=1}^n e_{\bar{\imath}} \wedge e_i$ be a nondegenerate alternating bilinear form. The *Lagrangian Grassmannian* $\mathrm{LG}(n)$ is the set of maximal isotropic subspaces of $\mathbb{C}^{2n}$ (relative to Ω).

Let $\{h_i := E_{ii} - E_{\bar{\imath}\bar{\imath}} \mid i \in [n]\}$ be the usual basis for the Lie algebra $\mathfrak{t}$ of T [Fulton and Harris 1991], and let $\{h_i^* \mid i \in [n]\} \subseteq \mathfrak{t}^*$ be the dual basis. Observe that $h_{\bar{\imath}}^* = -h_i^*$. The weights of any representation of $\mathrm{Sp}_{2n}(\mathbb{C})$ are $\mathbb{Z}$-linear combinations of the fundamental weights $\omega_i = h_{n-i+1}^* + \cdots + h_n^*$.

The weights of the representation $\bigwedge^n \mathbb{C}^{2n}$, and hence those of the subrepresentation $\mathrm{L}(\omega_n)$, are of the form $\omega = \sum_{i=1}^n h_{\alpha_i}^*$ for some $\alpha \in \binom{\langle n\rangle}{n}$. If $\alpha_j = \bar{\alpha}_{j'}$ for some $j, j' \in [n]$, then $h_{\alpha_j}^* = -h_{\alpha_{j'}}^*$, and thus the support of ω does not contain $h_{\alpha_j}^*$. Hence the set of all such weights ω are indexed by elements $\alpha \in \binom{\langle n\rangle}{k}$ $(k = 1, \dots, n)$ which do not involve both i and $\bar{\imath}$ for any $i = 1, \dots, n$.

Let V be a vector space. For simple alternating tensors $v := v_1 \wedge \cdots \wedge v_l \in \bigwedge^l V$ and $\varphi := \varphi_1 \wedge \cdots \wedge \varphi_k \in \bigwedge^k V^*$, there is a *contraction* defined by setting

$$\varphi \lrcorner v := \begin{cases} \sum_{I \in \binom{[l]}{k}} \pm v_1 \wedge \cdots \wedge \varphi_1(v_{i_1}) \wedge \cdots \wedge \varphi_k(v_{i_k}) \wedge \cdots \wedge v_l, & k \leq l, \\ 0, & k > l, \end{cases}$$

and extending bilinearly to a map $\bigwedge^k V^* \otimes \bigwedge^l V \to \bigwedge^{l-k} V$. In particular, for a fixed element $\Phi \in \bigwedge^k V^*$, we obtain a linear map $\Phi \lrcorner \bullet : \bigwedge^l V \to \bigwedge^{l-k} V$.

The Lagrangian Grassmannian embeds in $\mathbb{P}\mathrm{L}(\omega_n)$, where $\mathrm{L}(\omega_n)$ is the irreducible $\mathrm{Sp}_{2n}(\mathbb{C})$-representation of highest weight $\omega_n = h_1^* + \cdots + h_n^*$. This representation is isomorphic to the kernel of the contraction $\Omega \lrcorner \bullet : \bigwedge^n \mathbb{C}^{2n} \to \bigwedge^{n-2} \mathbb{C}^{2n}$ by Proposition 2.3. We thus have a commutative diagram of injective maps:

$$\begin{array}{ccc} \mathrm{LG}(n) & \longrightarrow & \mathrm{Gr}(n, 2n) \\ \big\downarrow & & \big\downarrow \\ \mathbb{P}\mathrm{L}(\omega_n) & \longrightarrow & \mathbb{P}(\bigwedge^n \mathbb{C}^{2n}). \end{array}$$

The next proposition implies that $\mathrm{LG}(n) = \mathrm{Gr}(n, 2n) \cap \mathbb{P}\mathrm{L}(\omega_n)$.

Proposition 2.3. *The dual of the contraction map*

$$\Omega \lrcorner \bullet : \bigwedge^n \mathbb{C}^{2n} \to \bigwedge^{n-2} \mathbb{C}^{2n}$$

is the multiplication map

$$\Omega \wedge \bullet : \bigwedge^{n-2} \mathbb{C}^{2n*} \to \bigwedge^{n} \mathbb{C}^{2n*}.$$

Furthermore, the irreducible representation $\mathrm{L}(\omega_n)$ *is defined by the ideal generated by the linear forms*

$$L_n := \mathrm{span}\{\Omega \wedge e^*_{\alpha_1} \wedge \cdots \wedge e^*_{\alpha_{n-2}} \mid \alpha \in \tbinom{\langle n \rangle}{n-2}\}.$$

These linear forms cut out $\mathrm{LG}(n)$ *scheme-theoretically in* $\mathrm{Gr}(n, 2n)$*. Dually,*

$$\mathrm{L}(\omega_n) = \ker(\Omega \lrcorner \bullet).$$

Proof. The proof of first statement is straightforward, and the second can be found in [Weyman 2003, Chapter 3, Exercise 1; Chapter 6, Exercise 24]. □

Since the linear forms spanning L_n are supported on variables indexed by $\alpha \in \binom{\langle n \rangle}{n}$ such that $\{\bar{\imath}, i\} \in \alpha$ for some $i \in [n]$, the set of complementary variables is linearly independent. These are indexed by the set $\mathcal{P}_n$ of *admissible* elements of $\binom{\langle n \rangle}{n}$:

$$\mathcal{P}_n := \left\{\alpha \in \tbinom{\langle n \rangle}{n} \;\middle|\; i \in \alpha \Leftrightarrow \bar{\imath} \notin \alpha\right\},$$

and have a simple description in terms of partitions (see Proposition 2.4).

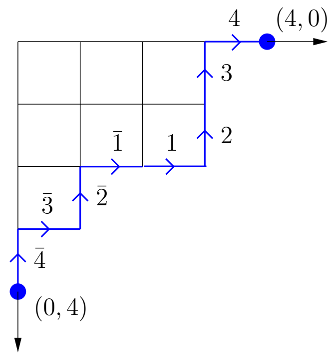


Figure 1. The partition (3, 3, 1) associated to $\bar{4}\bar{2}23$.

Consider the lattice $\mathbb{Z}^2$ with coordinates (a, b) corresponding to the point a units to the right of the origin and b units below the origin. Given an increasing sequence $\alpha \in \binom{\langle n\rangle}{n}$, let $[\alpha]$ be the lattice path beginning at $(0, n)$, ending at $(n, 0)$, and whose i-th step is vertical if $i \in \alpha$ and horizontal if $i \notin \alpha$. We can associate a partition to α by taking the boxes lying in the region bounded by the coordinate axes and $[\alpha]$. For instance, the sequence $\alpha = \bar{4}\bar{2}23 \in \binom{\langle 4\rangle}{4}$ is associated to the partition shown in Figure 1.

Proposition 2.4. *The bijection between increasing sequences and partitions induces a bijection between sequences α which do not contain both i and $\bar{\imath}$ for any $i \in [n]$, and partitions which lie inside the $n \times n$ square (n^n) and are symmetric with respect to reflection about the diagonal* $\{(a, a) \mid a \in \mathbb{Z}\} \subseteq \mathbb{Z}^2$.

Proof. The poset $\mathcal{P}_n$ consists of those $\alpha \in \binom{\langle n\rangle}{n}$ which are fixed upon negating each element of α and taking the complement in $\langle n\rangle$. On the other hand, the composition of these two operations (in either order) corresponds to reflecting the associated diagram about the diagonal. $\square$

Remark 2.5. We will use an element of $\binom{\langle n\rangle}{n}$ and its associated partition interchangeably. We denote by α^t the *transpose* partition obtained by reflecting α about the diagonal in $\mathbb{Z}^2$. As a sequence, α^t is the complement of $\{\bar{\alpha}_1, \dots, \bar{\alpha}_n\} \subseteq \langle n\rangle$. We denote by α_+ (respectively, α_-) the subsequence of positive (negative) elements of α.

Definition 2.6. The *Lagrangian involution* is the map $\tau : p_\alpha \mapsto \sigma_\alpha p_{\alpha^t}$, where $\sigma_\alpha := \operatorname{sgn}(\alpha_+^c, \alpha_+) \cdot \operatorname{sgn}(\alpha_-, \alpha_-^c) = \pm 1$, and $\operatorname{sgn}(a_1, \dots, a_s)$ denotes the sign of the permutation sorting the sequence $(a_1, \dots, a_s)$.

For example, if $\alpha = \bar{4}\bar{1}23$, then $\alpha_+ = 23$, $\alpha_- = \bar{4}\bar{1}$, and $\sigma_\alpha = 1$.

The Grassmannian $\mathrm{Gr}(n, 2n)$ has a natural geometric involution

$$\bullet^\perp : \mathrm{Gr}(n, 2n) \to \mathrm{Gr}(n, 2n)$$

sending an n-plane U to its orthogonal complement

$$U^{\perp} := \{u \in \mathbb{C}^{2n} \mid \Omega(u, u') = 0, \text{ for all } u' \in U\}$$

with respect to Ω. The next proposition relates $\bullet^{\perp}$ to the Lagrangian involution.

Proposition 2.7. *The map* $\bullet^{\perp} : \mathrm{Gr}(n, 2n) \to \mathrm{Gr}(n, 2n)$ *expressed in Plücker coordinates coincides with the Lagrangian involution*:

$$\left[p_\alpha \;\middle|\; \alpha \in \binom{\langle n\rangle}{n}\right] \mapsto \left[\sigma_\alpha p_{\alpha^t} \;\middle|\; \alpha \in \binom{\langle n\rangle}{n}\right].$$

In particular, the relation $p_\alpha - \sigma_\alpha p_{\alpha^t} = 0$ *holds on* $\mathrm{LG}(n)$.

Proof. The set of n-planes in $\mathbb{C}^{2n}$ which do not meet the span of the first n standard basis vectors is open and dense in $\mathrm{Gr}(n, 2n)$. Any such n-plane is the row space of an $n \times 2n$ matrix

$$Y := (I \mid X)$$

where I is the $n \times n$ identity matrix and X is a generic $n \times n$ matrix. We work in the affine coordinates given by the entries in X. For $\alpha \in \binom{\langle n\rangle}{n}$, denote the α-th minor of Y by $p_\alpha(Y)$. For a set of indices $\alpha = \{\alpha_1, \dots, \alpha_k\} \subseteq [n]$, let $\alpha^c := [n] \setminus \alpha$ be the complement, $\alpha' := \{n-\alpha_k+1, \dots, n-\alpha_1+1\}$, and $\bar{\alpha} := \{\bar{\alpha}_1, \dots, \bar{\alpha}_k\}$. Via the correspondence between partitions and sequences (Proposition 2.4), $\alpha^t = \bar{\alpha}^c$.

We claim that $\bullet^{\perp}$ reflects X along the antidiagonal. To see this, we simply observe how the rows of Y pair under Ω. For vectors $u, v \in \mathbb{C}^n$, let $(u, v) \in \mathbb{C}^{2n}$ be the concatenation. Let $r_i := (e_i, v_i) \in \mathbb{C}^{2n}$ be the i-th row of Y. For $k \in \langle n\rangle$, we let $r_{ik} \in \mathbb{C}$ be the k-th entry of r_i. Then, for $i, j \in \langle n\rangle$,

$$\Omega(r_i, r_j) = (e_i, v_i) \cdot (-v_j, e_j)^t = r_{i,n-j+1} - r_{j,n-i+1}.$$

It follows that the effect of $\bullet^{\perp}$ on the minor $X_{\rho,\gamma}$ of X given by row indices ρ and column indices γ is

$$(X^{\perp})_{\rho,\gamma} = X_{\gamma',\rho'}.$$

Let $\alpha = \bar{\epsilon} \cup \phi \in \binom{\langle n\rangle}{n}$, where ϵ and ϕ are subsets of $[n]$ whose cardinalities sum to n. Combining the above description of $\bullet^{\perp}$ with the identity

$$p_\alpha(Y) = \mathrm{sgn}(\epsilon^c, \epsilon) X_{(\epsilon^c)',\phi},$$

we have

$$\begin{aligned} p_\alpha(Y^{\perp}) &= \mathrm{sgn}(\epsilon^c, \epsilon)(X^{\perp})_{(\epsilon^c)',\phi} = \mathrm{sgn}(\epsilon^c, \epsilon) X_{\phi',\epsilon^c} \\ &= \mathrm{sgn}(\epsilon^c, \epsilon)\mathrm{sgn}(\phi, \phi^c) p_{(\bar{\phi}^c,\epsilon^c)}(Y) = \sigma_\alpha p_{\alpha^t}(Y). \end{aligned}$$

It follows that the relation $p_\alpha - \sigma_\alpha p_{\alpha^t} = 0$ holds on a dense Zariski-open subset and hence identically on all of $\mathrm{LG}(n)$. $\square$

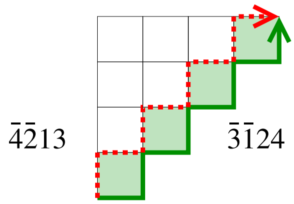


Figure 2. The paths associated to $\bar{4}\bar{2}13$ and $\bar{3}\bar{1}24$ in $\mathcal{P}_4$.

By Proposition 2.7, the system of linear forms

$$L'_n := \operatorname{span}\{p_\alpha - \sigma_\alpha p_{\alpha^t} \mid \alpha \in \tbinom{\langle n\rangle}{n}\} \tag{2-1}$$

defines $\mathrm{LG}(n) \subseteq \mathrm{Gr}(n, 2n)$ *set*-theoretically. Since $\mathrm{LG}(n)$ lies in no hyperplane of $\mathbb{P}\mathrm{L}(\omega_n)$, L'_n is a linear subspace of the span L_n of the defining equations of $\mathrm{L}(\omega_n) \subseteq \bigwedge^n \mathbb{C}^{2n}$. The generators of L'_n given in (2-1) suggest that homogeneous coordinates for the Lagrangian Grassmannian should be indexed by some sort of quotient (which we will call $\mathcal{D}_n$) of the poset $\binom{\langle n\rangle}{n}$. The correct notion is that of a doset (Definition 3.1). An important set of representatives for $\mathcal{D}_n$ in $\binom{\langle n\rangle}{n}$ is the set of *Northeast* partitions (Proposition 2.10).

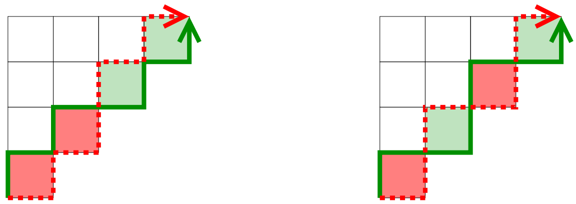
Remark 2.8. The set of *strict* partitions with at most n rows and columns is commonly used to index Plücker coordinates for the Lagrangian Grassmannian. Given a symmetric partition $\alpha \in \mathcal{P}_n$, we can obtain a strict partition by first removing the boxes of α which lie below the diagonal, and then left-justifying the remaining boxes. This gives a bijection between the two sets of partitions.

By Proposition 2.4, we may identify elements of $\binom{\langle n\rangle}{n}$ with partitions lying in the $n \times n$ square (n^n), and $\mathcal{P}_n$ with the set of symmetric partitions. Define a map

$$\pi_n : \tbinom{\langle n\rangle}{n} \to \mathcal{P}_n \times \mathcal{P}_n, \qquad \pi_n(\alpha) := (\alpha \wedge \alpha^t, \alpha \vee \alpha^t).$$

Let $\mathcal{D}_n$ be the image of π_n. It is called the set of *admissible pairs*, and is a subset of $\mathcal{O}_{\mathcal{P}_n} := \{(\alpha, \beta) \in \mathcal{P}_n \times \mathcal{P}_n \mid \alpha \leq \beta\}$. The image of $\mathcal{P}_n \subseteq \binom{\langle n\rangle}{n}$ under π_n is the diagonal $\Delta_{\mathcal{P}_n} \subseteq \mathcal{P}_n \times \mathcal{P}_n$.

To show that $\mathcal{D}_n$ indexes coordinates on $\mathrm{LG}(n)$, we will work with a convenient set of representatives of the fibers of π_n. The fiber over $(\alpha, \beta) \in \mathcal{D}_n$ can be described as follows. The lattice paths $[\alpha]$ and $[\beta]$ must meet at the diagonal. Since α and β are symmetric, they are determined by the segments of their associated paths to the right and above the diagonal. Let $\Pi(\alpha, \beta)$ be the set of boxes bounded by these segments. Taking $n = 4$ for example, $\Pi(\bar{4}\bar{2}13, \bar{3}\bar{1}24)$ consists of the two shaded boxes above the diagonal in Figure 2. The lattice path $[\bar{4}\bar{2}13]$ is above and to the left of the path $[\bar{3}\bar{1}24]$.



$\bar{4}\bar{2}24$ and $\bar{3}\bar{1}13$ $\bar{4}\bar{1}14$ and $\bar{3}\bar{2}23$

Figure 3. Elements of $\pi_n^{-1}(\bar{4}\bar{2}13, \bar{3}\bar{1}24)$.

For any partition $\alpha \subseteq (n^n)$, the set $\alpha_+ \subseteq \alpha$ consists of the boxes of α on or above the main diagonal, and $\alpha_- \subseteq \alpha$ consists of the boxes of α on or below the main diagonal (compare Remark 2.5). Similarly, let $\Pi_+(\alpha, \beta) \subseteq \Pi(\alpha, \beta)$ be the set of boxes above the diagonal and let $\Pi_-(\alpha, \beta) \subseteq \Pi(\alpha, \beta)$ be the set of boxes below the diagonal.

A subset $S \subseteq (n^n)$ of boxes is *disconnected* if $S = S' \sqcup S''$ and no box of S' shares an edge with a box of S''. A subset S is connected if it is not disconnected. Let $\bigsqcup_{i=1}^k S_i$ be the decomposition of $\Pi_+(\alpha, \beta)$ into its connected components (so that $\bigsqcup_{i=1}^k S_i^t$ is the decomposition of $\Pi_-(\alpha, \beta)$).

Any element γ of the fiber $\pi_n^{-1}(\alpha, \beta)$ is obtained by choosing a subset $I \subseteq [k]$ and setting

$$\gamma = \alpha \cup \Big(\bigcup_{i \notin I} S_i^t\Big) \cup \Big(\bigcup_{i \in I} S_i\Big).$$

The elements of $\pi_n^{-1}(\bar{4}\bar{2}13, \bar{3}\bar{1}24)$ are shown in Figure 3.

Definition 2.9. A partition α is *Northeast* if $\alpha_-^t \subseteq \alpha_+$ and is *Southwest* if its transpose is Northeast.

For example, $\bar{4}\bar{2}24 \in \mathcal{P}_4$ is Northeast while $\bar{4}\bar{1}14 \in \mathcal{P}_4$ is neither Northeast nor Southwest. We summarize these ideas as:

Proposition 2.10. *Let*

$$(\alpha, \beta) \in \mathcal{D}_n$$

be an element of the image of π_n. Then $\pi_n^{-1}(\alpha, \beta)$ is in bijection with the set of subsets of connected components of $\Pi_+(\alpha, \beta)$. There exists a unique Northeast element of $\pi_n^{-1}(\alpha, \beta)$, namely, the element corresponding to all the connected components. Similarly, there is a unique Southwest element corresponding to the empty set of components.

Example 2.11. The Lagrangian Grassmannian $\mathrm{LG}(4) \subseteq \mathrm{Gr}(4, 8)$ is defined by the ideal L_4. From the explicit linear generators given in Proposition 2.3, it is evident that $L_4 \subseteq \bigwedge^4 \mathbb{C}^{8*}$ is spanned by weight vectors. For example, the generators of $L_4 \cap (\bigwedge^4 \mathbb{C}^{8*})_0$ are the vectors of weight zero:

$$\begin{aligned}\Omega \wedge p_{\bar{1}1} &= p_{\bar{4}\bar{1}14} + p_{\bar{3}\bar{1}13} + p_{\bar{2}\bar{1}12},\\ \Omega \wedge p_{\bar{2}2} &= p_{\bar{4}\bar{2}24} + p_{\bar{3}\bar{2}23} + p_{\bar{2}\bar{1}12},\\ \Omega \wedge p_{\bar{3}3} &= p_{\bar{4}\bar{3}34} + p_{\bar{3}\bar{2}23} + p_{\bar{3}\bar{1}13},\\ \Omega \wedge p_{\bar{4}4} &= p_{\bar{4}\bar{3}34} + p_{\bar{4}\bar{2}24} + p_{\bar{4}\bar{1}14}.\end{aligned} \tag{2-2}$$

The following linear forms lie in the span of the right-hand side of (2-2):

$$p_{\bar{2}\bar{1}12} + p_{\bar{4}\bar{3}34},\ \ p_{\bar{3}\bar{1}13} + p_{\bar{4}\bar{2}24},\ \ p_{\bar{3}\bar{2}23} + p_{\bar{4}\bar{1}14},\ \text{ and } p_{\bar{4}\bar{1}14} + p_{\bar{4}\bar{2}24} + p_{\bar{4}\bar{3}34}. \tag{2-3}$$

Three of the linear forms in (2-3) are supported on a pair $\{p_\alpha, p_{\alpha^t}\}$, and the remaining linear form expresses the Plücker coordinate $p_{\bar{4}\bar{1}14}$ as a linear combination of coordinates indexed by Northeast partitions (this follows from Lemma 5.5 in general).

Since each pair $\{p_\alpha, p_{\alpha^t}\}$ is incomparable, there is a Plücker relation which, after reduction by the linear forms (2-2), takes the form

$$\pm p_\alpha^2 - p_\beta p_\gamma + \text{lower order terms}$$

where $\beta := \alpha \wedge \alpha^t$ and $\gamma := \alpha \vee \alpha^t$ are respectively the meet and join of α and α^t. Defining $p_{(\beta,\gamma)} := p_\alpha = \sigma_\alpha p_{\alpha^t}$ we can regard such an equation as giving a rule for rewriting $p^2_{(\beta,\gamma)}$ as a linear combination of monomials supported on a chain. This general case is treated in Section 5.

3. Algebras with straightening law

3A. *Generalities.* The following definitions are due to De Concini and Lakshmibai [1981]. Let $\mathcal{P}$ be a poset, $\Delta_{\mathcal{P}}$ the diagonal in $\mathcal{P} \times \mathcal{P}$, and

$$\mathbb{O}_{\mathcal{P}} := \{(\alpha, \beta) \in \mathcal{P} \times \mathcal{P} \mid \alpha \le \beta\}$$

the subset of $\mathcal{P} \times \mathcal{P}$ defining the order relation on $\mathcal{P}$.

Definition 3.1. [De Concini and Lakshmibai 1981] A *doset* on $\mathcal{P}$ is a set $\mathcal{D}$ such that $\Delta_{\mathcal{P}} \subseteq \mathcal{D} \subseteq \mathbb{O}_{\mathcal{P}}$, and if $\alpha \le \beta \le \gamma$, then $(\alpha, \gamma) \in \mathcal{D}$ if and only if $(\alpha, \beta) \in \mathcal{D}$ and $(\beta, \gamma) \in \mathcal{D}$. The ordering on $\mathcal{D}$ is given by $(\alpha, \beta) \le (\gamma, \delta)$ if and only if $\beta \le \gamma$ in $\mathcal{P}$. We call $\mathcal{P}$ the *underlying poset*.

Remark 3.2. The doset ordering just defined does not, in general, satisfy the reflexive property. That is, for $(\alpha, \beta) \in \mathcal{D}$, it is not generally true that $(\alpha, \beta) \le (\alpha, \beta)$. Indeed, this is the case if and only if $(\alpha, \beta) \in \Delta_{\mathcal{P}} \cong \mathcal{P}$, that is, if and only if $\alpha = \beta$.

The *Hasse diagram* of a doset $\mathcal{D}$ on $\mathcal{P}$ is obtained from the Hasse diagram of $\mathcal{P} \subseteq \mathcal{D}$ by drawing a double line for each cover $\alpha \lessdot \beta$ such that (α, β) is in $\mathcal{D}$. The defining property of a doset implies that we can recover all the information in the doset from its Hasse diagram. See Figure 6 in Section 3C for an example.

An algebra with straightening law (Definition 3.3 below) is an algebra generated by indeterminates $\{p_\alpha \mid \alpha \in \mathfrak{D}\}$ indexed by a (finite) doset $\mathfrak{D}$ with a basis consisting of *standard* monomials supported on a chain. That is, a monomial $p_{(\alpha_1,\beta_1)} \cdots p_{(\alpha_k,\beta_k)}$ is standard if $\alpha_1 \le \beta_1 \le \alpha_2 \le \beta_2 \le \cdots \le \alpha_k \le \beta_k$. Furthermore, monomials which are not standard are subject to certain *straightening relations*, as described in the following definition.

Definition 3.3. [De Concini and Lakshmibai 1981] Let $\mathfrak{D}$ be a doset. A graded $\mathbb{C}$-algebra

$$A = \bigoplus_{q \ge 0} A_q$$

is an *algebra with straightening law* on $\mathfrak{D}$ if there is an injection $\mathfrak{D} \ni (\alpha, \beta) \mapsto p_{(\alpha,\beta)} \in A_1$ such that:

(1) The set $\{p_{(\alpha,\beta)} \mid (\alpha, \beta) \in \mathfrak{D}\}$ generates A.

(2) The set of standard monomials are a $\mathbb{C}$-basis of A.

(3) For any monomial $m = p_{(\alpha_1,\beta_1)} \cdots p_{(\alpha_k,\beta_k)}$, $(\alpha_i, \beta_i) \in \mathfrak{D}$ and $i = 1, \ldots, k$, if

$$m = \sum_{j=1}^{N} c_j \, p_{(\alpha_{j1},\beta_{j1})} \cdots p_{(\alpha_{jk},\beta_{jk})},$$

is the unique expression of m as a linear combination of distinct standard monomials, then the sequence $(\alpha_{j1} \le \beta_{j1} \le \cdots \le \alpha_{jk} \le \beta_{jk})$ is lexicographically smaller than $(\alpha_1 \le \beta_1 \le \cdots \le \alpha_k \le \beta_k)$. That is, if $\ell \in [2k]$ is minimal such that $\alpha_{j\ell} \ne \alpha_\ell$, then $\alpha_{j\ell} < \alpha_\ell$.

(4) If $\alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4$ are such that for some permutation $\sigma \in S_4$ we have $(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}) \in \mathfrak{D}$ and $(\alpha_{\sigma(3)}, \alpha_{\sigma(4)}) \in \mathfrak{D}$, then

$$p_{(\alpha_{\sigma(1)},\alpha_{\sigma(2)})} p_{(\alpha_{\sigma(3)},\alpha_{\sigma(4)})} = \pm p_{(\alpha_1,\alpha_2)} p_{(\alpha_3,\alpha_4)} + \sum_{i=1}^{N} r_i m_i$$

where the m_i are quadratic standard monomials distinct from $p_{(\alpha_1,\alpha_2)} p_{(\alpha_3,\alpha_4)}$.

The ideal of straightening relations is generated by homogeneous quadratic forms in the p_α ($\alpha \in \mathfrak{D}$), so we may consider the projective variety $X := \operatorname{Proj} A$ they define. For each $\alpha \in \mathcal{P}$, we have the *Schubert variety*

$$X_\alpha := \{x \in X \mid p_{(\beta,\gamma)}(x) = 0 \text{ for } \gamma \not\le \alpha\}$$

and the *dual Schubert variety*

$$X^\alpha := \{x \in X \mid p_{(\beta,\gamma)}(x) = 0 \text{ for } \beta \not\ge \alpha\}.$$

Remark 3.4. If $\alpha \lessdot \beta$ and $(\alpha, \beta) \in \mathcal{D}$, then the multiplicity of X_α in X_β is 2, and likewise for the multiplicity of X^α in X^β. This fact will arise in Section 4 when we consider enumerative questions.

We recall the case when X is the Grassmannian of k-planes in $\mathbb{C}^n$, whose coordinate ring is an algebra with straightening law on the poset (with trivial doset structure) $\binom{[n]}{k}$.

For each $i \in [n]$, set $F_i := \langle e_1, \ldots, e_i \rangle$ and $F_i' := \langle e_n, \ldots, e_{n-i+1} \rangle$, where $\langle \cdots \rangle$ denotes linear span and $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{C}^n$. We call $F_\bullet := \{F_1 \subseteq \cdots \subseteq F_n\}$ the *standard coordinate* flag, and $F_\bullet' := \{F_1' \subseteq \cdots \subseteq F_n'\}$ the *opposite* flag.

We represent any k-plane $E \in \mathrm{Gr}(k, n)$ as the row space of a $k \times n$ matrix. Furthermore, any such k-plane E is the row space of a unique reduced row echelon matrix. The Schubert variety X_α consists of precisely the k-planes E such that the pivot in row i is weakly to the left of column α_i. Since the Plücker coordinate $p_\beta(E)$ is just the β-th maximal minor of this matrix, we see that $E \in X_\alpha$ if and only if $p_\beta(E) = 0$ for all $\beta \not\leq \alpha$; hence the general definition of the Schubert variety X_α (and by a similar argument, the dual Schubert variety X^α) agrees with the well-known geometric definition in the case of the Grassmannian. Namely, for $\alpha \in \binom{[n]}{k}$, the Schubert variety X_α is

$$X_\alpha = \{E \in \mathrm{Gr}(k, n) \mid \dim(E \cap F_{\alpha_i}) \geq i, \text{ for } i = 1, \ldots, k\},$$

and the dual Schubert variety X^α is

$$X^\alpha = \{E \in \mathrm{Gr}(k, n) \mid \dim(E \cap F'_{n-\alpha_i+1}) \geq k-i+1, \text{ for } i = 1, \ldots, k\}.$$

For a fixed projective variety $X \subseteq \mathbb{P}^n$, there are many homogeneous ideals which cut out X set-theoretically. However, there exists a unique such ideal which is saturated and radical. Under mild hypotheses, any ideal generated by straightening relations on a doset is saturated and radical. The proofs of Theorems 3.8 and 3.9 illustrate the usefulness of Schubert varieties in the study of an algebra with straightening law.

Definition 3.5. An ideal $I \subseteq \mathbb{C}[x_0, \ldots, x_n]$ is *saturated* if, given a polynomial $f \in \mathbb{C}[x_0, \ldots, x_n]$ and an integer $N \in \mathbb{N}$,

$$x_i^N f = 0 \mod I$$

for all $i = 0, \ldots, n$ implies that

$$f = 0 \mod I.$$

Definition 3.6. A ring A is *reduced* if it has no nilpotent elements; that is, if $f \in A$ satisfies $f^N = 0$ for some $N \in \mathbb{N}$, then $f = 0$.

Definition 3.7. An ideal $I \subseteq \mathbb{C}[x_0, \dots, x_n]$ is *radical* if, given a polynomial $f \in \mathbb{C}[x_0, \dots, x_n]$ and an integer $N \in \mathbb{N}$,

$$f^N = 0 \mod I$$

implies that

$$f = 0 \mod I.$$

That is, I is radical if the quotient $\mathbb{C}[x_0, \dots, x_n]/I$ is reduced.

The next two results concern an algebra with straightening law on a doset $\mathcal{D}$ with underlying poset $\mathcal{P}$. We write $A = \mathbb{C}[\mathcal{D}]/J$ for this algebra, where J is the ideal generated by the straightening relations. Proposition 3.9 is a special case of [Chirivì 2000, Proposition 27].

Theorem 3.8. *Let $\mathcal{D}$ be a doset whose underlying poset has a unique minimal element α_0. Then any ideal J of straightening relations on $\mathcal{D}$ is saturated.*

Proof. Let $f \notin J$. Modulo J, we may write $f = \sum_{i=1}^{k} a_i m_i$ where the m_i are (distinct) standard monomials, and $a_i \in \mathbb{C}$. For each $N \in \mathbb{N}$,

$$p_{\alpha_0}^N f = \sum_{i=1}^{k} a_i p_{\alpha_0}^N m_i$$

is a linear combination of standard monomials, since $\operatorname{supp} m_i \cup \{\alpha_0\}$ is a chain for each $i \in [k]$. It is nontrivial since $p_{\alpha_0}^N m_i = p_{\alpha_0}^N m_j$ implies $i = j$. Thus $p_{\alpha_0}^N f \notin J$ for any $N \in \mathbb{N}$. ☐

Proposition 3.9. *An algebra with straightening law on a doset is reduced.*

Proof. Let A be an algebra with straightening law. For $f \in A$ and $\alpha \in \mathcal{P}$, denote by f_α the restriction of f to the dual Schubert variety X^α.

We will show by induction on the poset $\mathcal{P}$ that $f_\alpha^n = 0$ implies $f_\alpha = 0$. Note that by induction on n it suffices to do this for $n = 2$. Indeed, assume that we have shown that $f^2 = 0$ implies $f = 0$ for any f in some ring A, and suppose $f^n = 0$. Then $(f^{\lceil \frac{n}{2} \rceil})^2 = 0$, so that $f^{\lceil \frac{n}{2} \rceil} = 0$ by our assumption, and thus $f = 0$ by induction.

Let $f \in A$ be such that $f_\alpha^2 = 0$. In particular, $f_\beta^2 = 0$ for all $\beta \geq \alpha$ (since $X^\beta \subseteq X^\alpha$), so that $f_\beta = 0$ by induction. It follows that f_α is supported on monomials on X^α which vanish on X^β for all $\beta \geq \alpha$. That is,

$$f_\alpha = \sum_{i=1}^{m} c_i p_\alpha^{e_i} p_{(\alpha, \beta_{1,i})} \cdots p_{(\alpha, \beta_{\ell_i, i})}. \tag{3-1}$$

For the right hand side of (3-1) to be standard, we must have $\ell_i = 1$ for all $i = 1, \dots m$. Also, homogeneity implies that $e := e_1 = \cdots = e_m$ for $i = 1, \dots m$. Thus, if we set $\beta_i := \beta_{1,i}$, then f_α has the form

$$f_\alpha = p_\alpha^e \sum_{i=1}^{m} c_i\, p_{(\alpha,\beta_i)}. \tag{3-2}$$

Choose a linear extension of $\mathcal{D}$ as follows. Begin with a linear extension of $\mathcal{P} \subseteq \mathcal{D}$. For incomparable elements (α,β), (γ,δ) of $\mathcal{D}$, set $(\alpha,\beta) \le (\gamma,\delta)$ if $\beta < \delta$ or $\beta = \delta$ and $\alpha \le \gamma$. With respect to the resulting linear ordering of the variables, take the lexicographic term order on monomials in A.

For an element $g \in A$, denote by $\mathrm{lt}(g)$ (respectively, $\mathrm{lm}(g)$) the lead term (respectively, lead monomial) of g. Reordering the terms in (3-2) if necessary, we may assume that

$$\mathrm{lt}(f_\alpha) = c_1 p_\alpha^e p_{(\alpha,\beta_1)}.$$

Writing f_α^2 as a linear combination of standard monomials (by first expanding the square of the right hand side of (3-2) and then applying the straightening relations), we see that

$$\mathrm{lt}(f_\alpha^2) = \pm c_1^2 p_\alpha^{2e+1} p_{\beta_1}.$$

This follows from our choice of term order and the condition in Definition 3.3 (4).

We claim that $\mathrm{lt}(f_\alpha^2)$ cannot be cancelled in the expression for f_α^2 as a sum of standard monomials. Indeed, suppose there are $i, j \in [m]$ such that

$$\mathrm{lm}\big((p_\alpha^e p_{(\alpha,\beta_i)}) \cdot (p_\alpha^e p_{(\alpha,\beta_j)})\big) = p_\alpha^{2e+1} p_{\beta_1}.$$

Then by the straightening relations, $\beta_1 \le \beta_i, \beta_j$. But $\beta_1 \not< \beta_i$ since

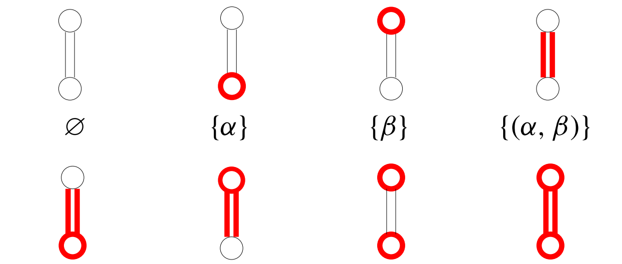
$$\mathrm{lm}(f_\alpha) = p_\alpha^e p_{(\alpha,\beta_1)}.$$

For the same reasons, $\beta_1 \not< \beta_j$. Therefore $\beta_i = \beta_j = \beta_1$, so $c_1 p_\alpha^e p_{(\alpha,\beta_1)}$ is the only term contributing to the monomial $p_\alpha^{2e+1} p_{\beta_1}$ in f_α^2. □

Remark 3.10. Proposition 3.9 was first proved for an algebra with straightening law on a *poset* in [Eisenbud 1980], but the methods used (deformation to the initial ideal) are not well-suited for a doset. The proof given here is essentially an extension of the proof of Bruns and Vetter [1988, Theorem 5.7] to the doset case.

3B. ***Hilbert series of an algebra with straightening law.*** We compute the Hilbert series of an algebra with straightening law A on a doset, and thus obtain formulas for the dimension and degree of Proj A. Let $\mathcal{P}$ be a poset and $\mathcal{D}$ a doset on $\mathcal{P}$. Assume that $\mathcal{P}$ and $\mathcal{D}$ are *ranked*; that is, any two maximal chains in $\mathcal{D}$ (respectively, $\mathcal{P}$) have the same length. Define rank $\mathcal{D}$ (respectively, rank $\mathcal{P}$) to be the length of any maximal chain in $\mathcal{D}$ (respectively, $\mathcal{P}$).

First, we compute the Hilbert series of A with respect to a suitably chosen fine grading of A by the elements of a semigroup, as follows.



$\varnothing$ $\{\alpha\}$ $\{\beta\}$ $\{(\alpha, \beta)\}$

$\{\alpha, (\alpha, \beta)\}$ $\{(\alpha, \beta), \beta\}$ $\{\alpha, \beta\}$ $\{\alpha, (\alpha, \beta), \beta\}$

Figure 4. The set $\mathrm{Ch}(\mathscr{D})$ of chains in $\mathscr{D}$.

Monomials in $\mathbb{C}[\mathscr{D}]$ are determined by their exponent vectors. We can therefore identify the set of such monomials with the semigroup $\mathbb{N}^{\mathscr{D}}$. Define the *weight map* $\mathbf{w} : \mathbb{N}^{\mathscr{D}} \to \mathbb{Q}^{\mathscr{P}}$ by setting $\mathbf{w}(\alpha, \beta) := \frac{\epsilon_\alpha + \epsilon_\beta}{2}$, where $\epsilon_\alpha \in \mathbb{Q}^{\mathscr{P}}$ ($\alpha \in \mathscr{P}$) is the vector with α-coordinate equal to 1 and all other coordinates equal to 0. This gives a grading of A by the semigroup $\mathrm{im}(\mathbf{w})$. Let $\mathrm{Ch}(\mathscr{D})$ be the set of all chains in $\mathscr{D}$. Since the standard monomials (those supported on a chain) form a $\mathbb{C}$-basis for A, the Hilbert series with respect to this fine grading is

$$H_A(r) = \sum_{c \in \mathrm{Ch}(\mathscr{D})} \sum_{\substack{a \in \mathrm{im}(\mathbf{w}) \\ \mathrm{supp}(a) = c}} r^a,$$

where $r := (r_\alpha \mid \alpha \in \mathscr{P})$, $a = (a_\alpha \mid \alpha \in \mathscr{P})$, and $r^a = \prod_{\alpha \in \mathscr{P}} r_\alpha^{a_\alpha}$. Note that elements of $\mathrm{im}(\mathbf{w})$ correspond to certain monomials with *rational* exponents (supported on $\mathscr{P}$). For example, $(\alpha, \beta) \in \mathscr{D}$ corresponds to $\sqrt{r_\alpha r_\beta}$. Setting all $r_\alpha = r$, we obtain the usual (coarse) Hilbert series, defined with respect to the usual $\mathbb{Z}$-grading on A by degree.

Example 3.11. Consider the doset $\mathscr{D} := \{\alpha, (\alpha, \beta), \beta\}$ on the two element poset $\{\alpha < \beta\}$. The elements of $\mathrm{Ch}(\mathscr{D})$ are shown in Figure 4.

$$\mathrm{Ch}(\mathscr{D}) = \big\{\varnothing, \{\alpha\}, \{\beta\}, \{(\alpha, \beta)\}, \{\alpha, (\alpha, \beta)\}, \{(\alpha, \beta), \beta\}, \{\alpha, \beta\}, \{\alpha, (\alpha, \beta), \beta\}\big\}.$$

We have

$$H_A(r) = 1 + \frac{r_\alpha}{1 - r_\alpha} + \frac{r_\beta}{1 - r_\beta} + \sqrt{r_\alpha r_\beta} + \frac{\sqrt{r_\alpha^3 r_\beta}}{1 - r_\alpha} + \frac{\sqrt{r_\alpha r_\beta^3}}{(1 - r_\beta)} + \frac{r_\alpha r_\beta}{(1 - r_\alpha)(1 - r_\beta)} + \frac{\sqrt{r_\alpha^3 r_\beta^3}}{(1 - r_\alpha)(1 - r_\beta)}.$$

Setting $r = r_\alpha = r_\beta$, we obtain the Hilbert series with respect to the usual $\mathbb{Z}$-grading of $\mathbb{C}[\mathcal{D}]$:

$$h_A(r) = 1 + \frac{2r}{1-r} + r + \frac{2r^2}{1-r} + \frac{r^3+r^2}{(1-r)^2} = \frac{r+1}{(1-r)^2} = 1 + \sum_{i=1}^{\infty}(2i+1)r^i.$$

We see that the Hilbert polynomial is $p(i) = 2i+1$, so $\dim(\mathrm{Proj}\ A) = 1$, and $\deg(\mathrm{Proj}\ A) = 2$.

Remark 3.12. The coordinate ring of the Lagrangian Grassmannian LG(2) is an algebra with straightening law on the five-element doset obtained by adding two elements $\hat{0} < \alpha$ and $\hat{1} > \beta$ to the doset of Example 3.11. The addition of these elements does not affect the degree, which is also 2. Theorem 5.9 allows us to carry out such degree computations for the Drinfel'd Lagrangian Grassmannian, giving a new derivation of the intersection numbers computed in quantum cohomology.

Fix a poset $\mathcal{P}$, and a doset $\mathcal{D}$ on $\mathcal{P}$. For the remainder of this section, set $P :=$ rank $\mathcal{P}$ and $D :=$ rank $\mathcal{D}$. Given a chain

$$\{\alpha_1, \dots, \alpha_u, (\beta_{11}, \beta_{12}), \dots, (\beta_{v1}, \beta_{v2})\} \subseteq \mathcal{D}$$

(not necessarily written in order), let r_i be the formal variable corresponding to α_i ($i = 1, \cdots, u$), and let s_{jk} correspond to β_{jk} ($j = 1, \dots, v$, $k = 1, 2$). The variables r and s are not necessarily disjoint; in the example above, the chain $\{\alpha, (\alpha, \beta)\}$ has $r_1 = s_{11}$. We have

$$\sum_{\substack{a\in \mathrm{im}(\mathbf{w}) \\ \mathrm{supp}(a)=c}} r^a = \prod_{i=1}^{u} \frac{r_i}{1-r_i} \cdot \prod_{j=1}^{v} \sqrt{s_{j1}s_{j2}}.$$

Recall that we may identify $\mathcal{P}$ with the diagonal $\Delta_{\mathcal{P}} \subseteq \mathcal{D} \subseteq \mathcal{P}\times\mathcal{P}$. Letting c_u^v denote the number of chains consisting of u elements of $\mathcal{P}$ and v elements of $\mathcal{D} \setminus \mathcal{P}$, we have

$$\begin{aligned}\mathrm{HS}_A(r) &= \sum_{u=0}^{P+1}\sum_{v=0}^{D-P} c_u^v \frac{r^{u+v}}{(1-r)^u} = \sum_{u=0}^{P+1}\sum_{v=0}^{D-P} c_u^v r^{u+v}\left(\sum_{k=0}^{\infty} r^k\right)^u \\ &= \sum_{v=0}^{D-P} c_0^v r^v + \sum_{\ell=0}^{\infty}\sum_{u=1}^{P+1}\sum_{v=0}^{D-P} c_u^v \binom{u+\ell-1}{u-1} r^{u+v+\ell}.\end{aligned}$$

When $w > D-P$, the coefficient of r^w agrees with the Hilbert polynomial:

$$\mathrm{HP}_A(w) = \sum_{u=1}^{P+1}\sum_{v=0}^{D-P} c_u^v \binom{w-v-1}{u-1}. \tag{3-3}$$

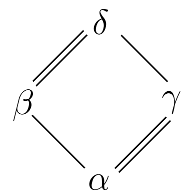


Figure 5. A doset on a four-element poset.

In particular, the dimension of Proj A is P, since this is the largest value of

$$u-1=\deg_w\binom{w-v-1}{u-1}.$$

The leading monomial of $HP_A(w)$ is

$$\sum_{v=0}^{D-P} c_{P+1}^{v}\binom{w-v-1}{P-1}.$$

By our assumption that the maximal chains in $\mathcal{P}$ (respectively, $\mathcal{D}$) have the same length, we have $c_{P+1}^{v}=\binom{D-P}{v}c_{P+1}^{0}$, so that the leading coefficient of $HP_A(w)$ is

$$\frac{c_{P+1}^{0}}{(P-1)!}\sum_{v=0}^{D-P}\binom{D-P}{v}=\frac{2^{D-P}c_{P+1}^{0}}{(P-1)!},$$

from which we deduce the degree and dimension of Proj A.

Theorem 3.13. *The degree of* Proj A *is* $2^{D-P}c_{P+1}^{0}$. *The dimension of* Proj A *is* P.

Example 3.14. Let A be an algebra with straightening law on the doset $\mathcal{D}$ shown in Figure 5.

We have rank $\mathcal{P}=2$, rank $\mathcal{D}=3$, and

$$\begin{aligned}\mathrm{Ch}(\mathcal{D})=\big\{&\varnothing,\{\alpha\},\{\beta\},\{\gamma\},\{\delta\},\{(\alpha,\gamma)\},\{(\beta,\delta)\},\\&\{\alpha,\gamma\},\{\alpha,\beta\},\{\alpha,\delta\},\{\beta,\delta\},\{\gamma,\delta\},\{\alpha,(\alpha,\gamma)\},\\&\{\alpha,(\beta,\delta)\},\{(\alpha,\gamma),\gamma\},\{(\alpha,\gamma),\delta\},\{\beta,(\beta,\delta)\},\\&\{(\beta,\delta),\delta\},\{\alpha,\beta,\delta\},\{\alpha,\gamma,\delta\},\{\alpha,(\alpha,\gamma),\gamma\},\\&\{\alpha,(\alpha,\gamma),\delta\},\{\alpha,\beta,(\beta,\delta)\},\{\alpha,(\beta,\delta),\delta\},\\&\{\beta,(\beta,\delta),\delta\},\{\alpha,(\alpha,\gamma),\gamma,\delta\},\{\alpha,\beta,(\beta,\delta),\delta\}\big\},\end{aligned}$$

and the values of c_u^v are given by the matrix

$$\begin{pmatrix}1&4&5&2\\2&6&5&2\end{pmatrix},$$

whose entry in row i and column j is c_{j-1}^{i-1}.

In view of (3-3), the Hilbert polynomial is therefore

$$
\begin{aligned}
\mathrm{HP}_A(w)& \\
&= 4\binom{w-1}{0}+5\binom{w-1}{1}+2\binom{w-1}{2}+6\binom{w-2}{0}+5\binom{w-2}{1}+2\binom{w-2}{2} \\
&= 2w^2+2w+3 = 4\frac{w^2}{2!}+2w+3.
\end{aligned}
$$

In particular, $\dim(\operatorname{Proj} A) = 2$ and $\deg(\operatorname{Proj} A) = 4$.

Theorems 5.9 and 3.13 will allow us to compute intersection numbers in quantum cohomology in the same manner as Example 3.14. The essential step is to show that the Drinfel'd Lagrangian Grassmannian is a algebra with straightening law on the doset of admissible pairs $\mathcal{D}_{d,n}$.

3C. ***The doset of admissible pairs.*** We define the doset of admissible pairs on the poset $\mathcal{P}_{d,n}$. Let us first consider an example.

Example 3.15. Consider the poset

$$\mathcal{P}_{2,4} := \{\alpha^{(a)} \in \tbinom{\langle 4\rangle}{4}_2 \mid i \in \alpha \Longleftrightarrow \bar{\imath} \notin \alpha\}$$

of admissible elements of $\binom{\langle 4\rangle}{4}_2$. Let $\mathcal{D}_{2,4}$ be the set of elements $(\alpha,\beta)^{(a)} \in \mathcal{O}_{\mathcal{P}_{2,4}}$ such that α and β have the same number of negative elements. It is a doset on $\mathcal{P}_{2,4}$. The Hasse diagram (drawn so that going up in the doset corresponds to moving to the right) for $\mathcal{D}_{2,4}$ is shown in Figure 6.

To each $(\alpha,\beta)^{(a)} \in \mathcal{P}_{2,4}$, we have the Plücker coordinate

$$p_{(\alpha,\beta)}^{(a)} := u^a v^{d-a} \otimes p_{(\alpha,\beta)} \in S^d\mathbb{C}^2 \otimes \mathrm{L}(\omega_n)^*,$$

where $\{u, v\} \subseteq \mathbb{C}^2$ is a basis dual to $\{s, t\} \subseteq (\mathbb{C}^2)^*$.

Let $\binom{\langle n\rangle}{n}_d \cong \binom{[2n]}{n}_d$ be the poset associated to the (ordinary) Drinfel'd Grassmannian $\mathcal{Q}_d(n, 2n)$, and recall that $\mathcal{P}_{d,n} \subseteq \binom{\langle n\rangle}{n}_d$ is the subposet consisting of the elements $\alpha^{(a)}$ such that $\alpha^t = \alpha$. There are three types of covers in $\mathcal{P}_{d,n}$.

(1) $\alpha^{(a)} \lessdot \beta^{(a)}$, where α and β have the same number of negative elements. For example, $\bar{4}\bar{2}13^{(a)} \lessdot \bar{4}\bar{1}23^{(a)} \in \mathcal{P}_{d,4}$ for any nonnegative integers $a \le d$.

(2) $\alpha^{(a)} \lessdot \beta^{(a)}$, where the number of negative elements in β is one less than the number of negative elements of α. For example, $\bar{4}\bar{1}23^{(a)} \lessdot \bar{4}123^{(a)} \in \mathcal{P}_{d,4}$ for any nonnegative integers $a \le d$.

(3) $\alpha^{(a)} \lessdot \beta^{(a+1)}$, where the number of negative elements of β is one more than the number of negative elements of α, $\bar{n} \in \beta$, and $n \in \alpha$. For example, $\bar{3}\bar{2}14^{(a)} \lessdot \bar{4}\bar{3}\bar{2}1^{(a+1)}$ for any nonnegative integers $a \le d$.

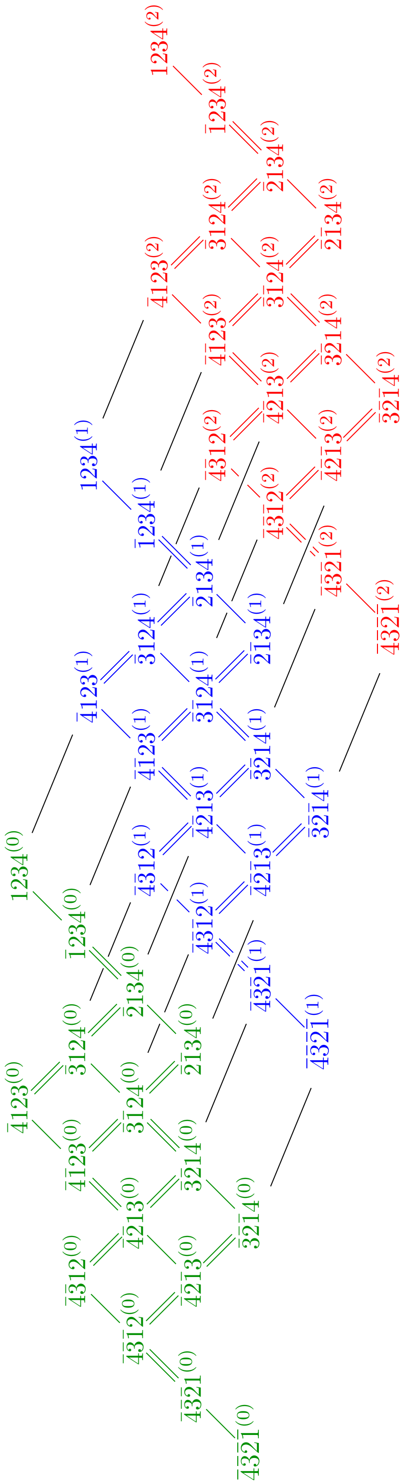


Figure 6. The doset $\mathscr{D}_{2,4}$. Elements increase as one moves to the right.

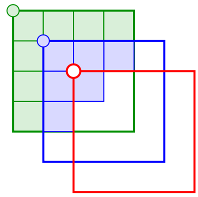


Figure 8. The subset of $S_{2,4}$ associated to $\bar{4}\bar{2}13^{(1)} \in \mathcal{P}_{2,4}$.

The first two types are those appearing in the classical Bruhat order on $\mathcal{P}_{0,n}$. It follows that $\mathcal{P}_{d,n}$ is a union of levels $\mathcal{P}_{d,n}^{(a)}$, each isomorphic to the Bruhat order, with order relations between levels imposed by covers of the type (3) above. We define the doset $\mathcal{D}_{d,n}$ of admissible pairs in $\mathcal{P}_{d,n}$.

Definition 3.16. A pair $(\alpha^{(a)} < \beta^{(a)})$ is *admissible* if there exists a saturated chain $\alpha = \alpha_0 \lessdot \alpha_1 \lessdot \cdots \lessdot \alpha_s = \beta$, where each $\alpha_i \lessdot \alpha_{i+1}$ is a cover of type (1).

We denote the set of admissible pairs by $\mathcal{D}_{d,n}$. Observe that the pair $(\alpha^{(a)} < \beta^{(b)})$ is never admissible if $a < b$.

Proposition 3.17. *The set $\mathcal{D}_{d,n} \subseteq \mathcal{P}_{d,n} \times \mathcal{P}_{d,n}$ is a doset on $\mathcal{P}_{d,n}$. The poset $\mathcal{P}_{d,n}$ is a distributive lattice.*

Proof. In view of our description of the covers in $\mathcal{D}_{d,n}$, it is clear that, for all $d \geq 0$, $\mathcal{D}_{d,n}$ is a doset if and only if $\mathcal{D}_n = \mathcal{D}_{0,n}$ is a doset. The latter is proved in [De Concini and Lakshmibai 1981]. To prove that $\mathcal{P}_{d,n}$ is a distributive lattice, we give an isomorphism with a certain lattice of subsets of the union of $d+1$ shifted $n \times n$ squares in $\mathbb{Z}^2$ which generalize the usual notion of a partition.

Let

$$S_{d,n} := \bigcup_{a=0}^{d} \{(i+a,\, j+a) \mid 0 \leq i, j \leq n\}.$$

To $\alpha^{(a)} \in \mathcal{P}_{d,n}$, we associate the subset of $S_{d,n}$ obtained by shifting the (open) squares in α by (a, a), and adding the boxes obtained by translating a box of α by a vector (v_1, v_2) with $v_1, v_2 \leq 0$ and the points (i, i) for $i = 0, \ldots, a$. See Figure 8 for an example. It is straightforward to check that the (symmetric) subsets obtained in this way form a distributive lattice (ordered by inclusion) isomorphic to $\mathcal{P}_{d,n}$. $\square$

4. Schubert varieties and Gromov–Witten invariants

Gromov–Witten invariants are solutions to enumerative questions involving algebraic maps from $\mathbb{P}^1$ to a projective variety X. When X is the Lagrangian Grassmannian (or the ordinary Grassmannian [Sottile and Sturmfels 2001]), these questions can be studied geometrically via the Drinfel'd compactification, as advocated by A. Braverman [2006]. We do this in Section 4A, and relate our findings to the quantum cohomology of the Lagrangian Grassmannian in Section 4B. See [Braverman 2006; Sottile 2000; 2001] for further reading on applications of Drinfel'd compactifications to quantum cohomology. The study of Gromov–Witten invariants (in various special cases) has also been approached via the quot scheme [Bertram 1997; Chen 2003; Ciocan-Fontanine 1999; Fulton and Pandharipande 1997] and the space of stable maps [Bertram et al. 2005; Givental 1996; Oprea 2006].

4A. *Intersection problems on the Drinfel'd compactification.* Given an isotropic flag $F_\bullet$ and a symmetric partition $\alpha \in \mathcal{P}_n$, we have the Schubert variety

$$X^\alpha(F_\bullet) := \{E \in \mathrm{LG}(n) \mid \dim(E \cap F_{n-\alpha_i+i}) \geq i\}.$$

The enumerative problems we consider involve conditions that the image of a map $M \in L\mathcal{M}_d(X)$ pass through Schubert varieties at prescribed points of $\mathbb{P}^1$.

Question 4.1. Let $F_\bullet^1, \dots, F_\bullet^N$ be general Lagrangian flags, $\alpha_1, \dots, \alpha_N \in \mathcal{P}_n$, and let $s_1, \dots, s_N \in \mathbb{P}^1$ be distinct points. Assume

$$\sum_{i=1}^{N} |\alpha_i| = \dim \mathrm{LG}(n) + d(n+1).$$

How many degree-d algebraic maps $M : \mathbb{P}^1 \to \mathrm{LG}(n)$ satisfy

$$M(s_i) \in X^{\alpha_i}(F_\bullet^i)$$

for all $i = 1, \dots, N$?

Our answer to Question 4.1 is given in Theorem 4.3. In order to prove this result, we must first establish some results on the geometry of certain subvarieties of $L\mathcal{Q}_d(n)$ defined in terms of the *universal evaluation map*

$$\mathrm{ev} : \mathbb{P}^1 \times L\mathcal{M}_d(n) \to \mathrm{LG}(n), \quad \mathrm{ev}(s, M) := M(s)$$

for $s \in \mathbb{P}^1$ and $M \in L\mathcal{M}_d(n)$.

Fix a point $s \in \mathbb{P}^1$ and define

$$\mathrm{ev}_s := \mathrm{ev}(s, \bullet) : L\mathcal{M}_d(n) \to \mathrm{LG}(n).$$

Given a Schubert variety $X^\alpha(F_\bullet) \subseteq \mathrm{LG}(n)$, the set of maps $M \in L\mathcal{M}_d(n)$ such that $M(s)$ lies in $X^\alpha(F_\bullet)$ is the preimage $\mathrm{ev}_s^{-1}(X^\alpha(F_\bullet))$. This is a general translate of the locally closed subset $X^{\alpha^{(0)}} \cap L\mathcal{M}_d(n)$ under the action of the group $\mathrm{SL}_2\mathbb{C} \times \mathrm{Sp}_{2n}\mathbb{C}$. By a Schubert variety, we will mean the closure of $\mathrm{ev}_s^{-1}(X^\alpha(F_\bullet))$ in $L\mathcal{Q}_d(n)$, and denote it by $X^{\alpha^{(0)}}(s; F_\bullet)$. In order to understand these subvarieties, we extend the evaluation map to a globally defined map $\mathbb{P}^1 \times L\mathcal{Q}_d(n) \to \mathrm{LG}(n)$. To do this, we must first study the boundary $L\mathcal{Q}_d(n) \setminus L\mathcal{M}_d(n)$.

The embedding

$$L\mathcal{M}_d(n) \hookrightarrow \mathbb{P}\big((S^d\mathbb{C}^2)^* \otimes \mathrm{L}(\omega_n)\big)$$

is defined by regarding a map $M \in L\mathcal{M}_d(n)$ as a $\binom{2n}{n}$-tuple of degree-d homogeneous forms. We identify the space $L\mathcal{M}_d(n)$ of maps with its image, which is a locally closed subset of $\mathbb{P}\big((S^d\mathbb{C}^2)^* \otimes \mathrm{L}(\omega_n)\big)$. The Drinfel'd compactification $L\mathcal{Q}_d(n)$ is by definition the closure of the image.

On the other hand, $L\mathcal{M}_d(n) \subseteq L\mathcal{Q}_d(n)$ is the set of points corresponding to a $\binom{2n}{n}$-tuple of homogeneous forms satisfying the Zariski open condition that they have no common factor. Therefore, the boundary $L\mathcal{Q}_d(n) \setminus L\mathcal{M}_d(n)$ consists of $\binom{2n}{n}$-tuples of homogeneous forms which do have a common factor. Such a list of forms gives a regular map of degree $a < d$ together with an effective Weil divisor of degree $d - a$ on $\mathbb{P}^1$ defined by the base points of the map. We thus have a stratification

$$L\mathcal{Q}_d(n) = \bigsqcup_{a=0}^{d} \mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n)$$

where $\mathbb{P}(S^a\mathbb{C}^{2^*})$ is the space of degree-a forms in two variables, or alternatively, the space of effective Weil divisors on $\mathbb{P}^1$ of degree a. In particular, the boundary of $L\mathcal{Q}_d(n)$ is simply $\bigsqcup_{a=1}^{d} \mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n)$. We may regard any point of $L\mathcal{Q}_d(n)$ as a pair (D, M), where $M \in L\mathcal{M}_{d-a}(n)$ and D is a divisor on $\mathbb{P}^1$.

Fixing a point $s \in \mathbb{P}^1$, the evaluation map $\mathrm{ev}_s := \mathrm{ev}(s, \bullet)$ is undefined at each point $(D, M) \in \mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n)$ such that $s \in D$. Thus, restricting to the stratum $\mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n)$, the map ev_s is defined on $U_s^a \times L\mathcal{M}_{d-a}(n)$, where $U_s^a \subseteq \mathbb{P}(S_a\mathbb{C}^2)$ is the set of forms which do not vanish at $s \in \mathbb{P}^1$.

For each $a = 0, \ldots, d$, define a map

$$\epsilon_s^a : \mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n) \to \mathrm{LG}(n)$$

by the formula $\epsilon_s^a(D, M) := M(s)$, and let

$$\epsilon_s : L\mathcal{Q}_d(n) = \bigsqcup_{a=0}^{d} \mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n) \to \mathrm{LG}(n)$$

be the (globally-defined) map which restricts to ϵ_s^a on $\mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n)$. The evaluation map ev_s agrees with ϵ_s wherever it is defined. Hence ϵ_s extends ev_s to a globally defined map, which is a morphism on each stratum $\mathbb{P}(S^a\mathbb{C}^{2^*}) \times L\mathcal{M}_{d-a}(n)$. The Schubert variety $X^{\alpha^{(0)}}(s; F_\bullet)$ is the preimage of $X^\alpha(F_\bullet)$ under this globally defined map; hence we have the following fact.

Lemma 4.2. *Given a point $s \in \mathbb{P}^1$ and a isotropic flag $F_\bullet$, the Schubert variety $X^{\alpha^{(0)}}(s, F_\bullet)$ is the disjoint union of the strata*

$$\mathbb{P}(S^a\mathbb{C}^{2*}) \times \big(X^{\alpha^{(a)}}(s; F_\bullet) \cap L\mathcal{M}_{d-a}(n)\big).$$

Proof. For each $a \in \{0, \dots d\}$, we have

$$\big(\mathbb{P}(S^a\mathbb{C}^{2*}) \times L\mathcal{M}_{d-a}(n)\big) \cap \epsilon_s^{-1}(X^\alpha(F_\bullet)) = \mathbb{P}(S^a\mathbb{C}^{2*}) \times X^{\alpha^{(a)}}(s; F_\bullet). \qquad \square$$

We now state and prove the main theorem of this section.

Theorem 4.3. *Given partitions $\alpha_1, \dots \alpha_N \in \mathcal{P}_n$ such that*

$$\sum_{i=1}^{N} |\alpha_i| = \binom{n+1}{2} + d(n+1),$$

general isotropic flags $F_\bullet^1, \dots, F_\bullet^N$, and distinct points $s_1, \dots, s_N \in \mathbb{P}^1$, the intersection

$$X^{\alpha_1^{(0)}}(s_1; F_\bullet^1) \cap \cdots \cap X^{\alpha_N^{(0)}}(s_N; F_\bullet^N) \tag{4-1}$$

is transverse, and hence consists only of reduced points. Each point of the intersection (4-1) *lies in $L\mathcal{M}_d(n)$, that is, corresponds to a degree-d map whose image $M(s_i)$ lies in $X^{\alpha_i}(F_\bullet^i)$ for $i = 1, \dots, N$.*

Proof. For each $a = 0, \dots, d$, the Schubert variety $X^{\alpha^{(a)}}(s, F_\bullet)$ is the preimage of the Schubert variety $X^\alpha(F_\bullet) \subseteq \mathrm{LG}(n)$ under the evaluation map ϵ_s, which is regular on the stratum $\mathbb{P}(S^a\mathbb{C}^{2*}) \times L\mathcal{M}_{d-a}(n)$. By Lemma 4.2, it suffices to consider the intersection (4-1) on each of these strata. Fix $a \in \{0, \dots, d\}$, and consider the product of evaluation maps

$$\prod_{i=1}^{N} \epsilon_{s_i} : \big(\mathbb{P}(S^a\mathbb{C}^{2*}) \times L\mathcal{M}_{d-a}(n)\big)^N \to \mathrm{LG}(n)^N$$

and the injection

$$X^{\alpha_1}(F_\bullet^1) \times \cdots \times X^{\alpha_n}(F_\bullet^N) \hookrightarrow \mathrm{LG}(n)^N.$$

The intersection $\big(X^{\alpha_1^{(0)}} \cap \cdots \cap X^{\alpha_N^{(0)}}\big) \cap \mathbb{P}(S^a\mathbb{C}^{2*}) \times L\mathcal{M}_{d-a}(n)$ is isomorphic to the fiber product

$$\big(\mathbb{P}(S^a\mathbb{C}^{2*}) \times L\mathcal{M}_{d-a}(n)\big)^N \times_{\mathrm{LG}(n)^N} \big(X^{\alpha_1}(F_\bullet^1) \times \cdots \times X^{\alpha_N}(F_\bullet^N)\big).$$

For each $a = 0, \dots, d$, Kleiman's theorem [1974, Corollary 2] implies that this intersection is proper and transverse. Considering the dimensions of these subvarieties, we see that this intersection is therefore zero-dimensional when $a = 0$ and empty when $a > 0$. $\square$

4B. ***Gromov–Witten invariants and quantum cohomology.*** A common approach to Question 4.1 is through the quantum cohomology ring of the Lagrangian Grassmannian $QH^*(\mathrm{LG}(n))$, defined as follows. The cohomology ring $H^*(\mathrm{LG}(n);\mathbb{Z})$ has a $\mathbb{Z}$-basis consisting of the classes of Schubert varieties (the *Schubert classes*) $\sigma_\alpha := [X^\alpha]$, where $\alpha \in \mathcal{P}_n$. We will denote by α^* the dual partition, defined so that $\sigma_\alpha \cdot \sigma_{\alpha^*} = [pt] \in H^*(\mathrm{LG}(n);\mathbb{Z})$ [Hiller and Boe 1986]. The correspondence $\alpha \leftrightarrow \alpha^*$ is bijective and order reversing.

The (small) *quantum cohomology ring* is the $\mathbb{Z}[q]$-algebra isomorphic to

$$H^*(\mathrm{LG}(n);\mathbb{Z}) \otimes \mathbb{Z}[q]$$

as a $\mathbb{Z}[q]$-module, and with multiplication defined by the formula

$$\sigma_\alpha \cdot \sigma_\beta = \sum \langle \alpha, \beta, \gamma^* \rangle_d \, \sigma_\gamma q^d,$$

where the sum is over all $d \geq 0$ and γ such that $|\gamma| = |\alpha| + |\beta| - d\binom{n+1}{2}$. For partitions α, β, and γ in $\mathcal{P}_n$, the coefficients $\langle \alpha, \beta, \gamma^* \rangle_d$ are the *Gromov–Witten invariants*, defined as the number of algebraic maps $M : \mathbb{P}^1 \to \mathrm{LG}(n)$ of degree d such that

$$M(0) \in X^\alpha(F_\bullet), \quad M(1) \in X^\beta(G_\bullet), \quad \text{and} \quad M(\infty) \in X^{\gamma^*}(H_\bullet),$$

where $F_\bullet$, $G_\bullet$, and $H_\bullet$ are general isotropic flags (that is, general translates of the standard flag under the action of the group $\mathrm{Sp}_{2n}\mathbb{C}$).

A special case of Pieri's rule gives a formula for the product of a Schubert class $\sigma_\alpha \in H^*(\mathrm{LG}(n);\mathbb{Z})$ with the *simple* Schubert class $\sigma_\square$ [Hiller and Boe 1986]:

$$\sigma_\alpha \cdot \sigma_\square = \sum_{\alpha \lessdot \beta} 2^{N(\alpha,\beta)} \sigma_\beta,$$

the sum over all partitions β obtained from α by adding a box above the diagonal, along with its image under reflection about the diagonal. The exponent $N(\alpha,\beta) = 1$ if $(\alpha,\beta) \in \mathcal{D}_n$ and $N(\alpha,\beta) = 0$ otherwise (compare Proposition 2.10). Kresch and Tamvakis [2003] give a quantum analogue of Pieri's rule. We state the relevant special case of this rule:

Proposition 4.4 [Kresch and Tamvakis 2003]. *For any $\alpha \in \mathcal{P}_n$, we have*

$$\sigma_\alpha \cdot \sigma_\square = \sum_{\alpha \lessdot \beta} 2^{N(\alpha,\beta)} \sigma_\beta + \sigma_\gamma q$$

in $QH^*(\mathrm{LG}(n))$, where the first sum is from the classical Pieri rule, and $\sigma_\gamma = 0$ unless α contains the hook-shaped partition $(n, 1^{n-1})$, in which case γ is the partition obtained from α by removing this hook.

For $\alpha^{(a)} \lessdot \beta^{(b)}$ in $\mathfrak{D}_{d,n}$, let $N'(\alpha^{(a)}, \beta^{(b)}) = N(\alpha, \beta)$ if $b = a$, and let $N'(\alpha^{(a)}, \beta^{(b)}) = 0$ if $b = a+1$. Let $\alpha^{(0)} \in \mathcal{P}_{d,n}$, and let $\pi \in \mathbb{N}$ be its *corank* in $\mathcal{P}_{d,n}$; that is, π is the length of any saturated chain of elements $\alpha^{(d)} = x_0 \lessdot \cdots \lessdot x_\pi = (n^n)^{(d)}$, where $x_i \in \mathcal{P}_{d,n}$ for all $i = 0, \ldots, \pi$ and $(n^n)^{(d)}$ is the maximal element of $\mathcal{P}_{d,n}$. By Theorem 3.13, π is the dimension of $X^{\alpha^{(0)}}$. The quantum Pieri rule of Proposition 4.4 has a simple formulation in terms of the distributive lattice $\mathcal{P}_{d,n}$:

Theorem 4.5. *The quantum Pieri rule in Proposition 4.4 has the formulation in terms of the poset* $\mathcal{P}_{d,n}$:

$$(\sigma_\alpha q^a) \cdot \sigma_\square = \sum_{\alpha^{(a)} \lessdot \beta^{(b)}} 2^{N'(\alpha^{(a)}, \beta^{(b)})} \sigma_\beta q^b.$$

As a consequence, we have

$$\sigma_\alpha \cdot (\sigma_\square)^\pi = \deg(X^{\alpha^{(0)}}) \cdot \sigma_{(n^n)} q^d \pmod{d+1}. \tag{4-2}$$

Proof. The element $\gamma^{(a+1)} \in \mathcal{P}_{d,n}$, where γ is as the partition obtained by removing a maximal hook from α in Proposition 4.4, is the unique cover of $\alpha^{(a)} \in \mathcal{P}_{d,n}$ with superscript $a+1$. The remaining covers (with superscript a) index the sum in Proposition 4.4.

The second formula follows by induction from the first. □

The appearance of the number $\deg(X^{\alpha^{(0)}})$ in (4-2) is for purely combinatorial reasons: it is the number of saturated chains $\alpha^{(0)} \lessdot \cdots \lessdot (n^n)^{(d)}$ in $\mathfrak{D}_{d,n}$, counted with multiplicity. Since $X^{\square^{(0)}}$ is a hyperplane section of $L\mathfrak{Q}_d(n)$, this is also the number of points in the intersection

$$X^{\alpha^{(0)}}(s; F_\bullet) \cap \left(\bigcap_{i=1}^{\pi} X^{\square^{(0)}}(s_i; F^i_\bullet) \right), \tag{4-3}$$

the intersection of $X^{\alpha^{(0)}}(s; F_\bullet)$ with $\pi = \operatorname{codim}(X^{\alpha^{(0)}}(s; F_\bullet))$ general translates of the hyperplane section $X^{\square^{(0)}}$. On the other hand, multiplication in $QH^*(\mathrm{LG}(n))$ represents the conjunction of conditions that a map takes values in Schubert varieties at generic points of $\mathbb{P}^1$. In this way, the quantum cohomology identity of Theorem 4.5 has an interpretation as the number of points in the intersection (4-3) of Schubert varieties in $L\mathfrak{Q}_d(n)$.

5. The straightening law

5A. ***A basis for $S^d\mathbb{C}^2 \otimes \mathrm{L}(\omega_n)^*$.*** The Drinfel'd Lagrangian Grassmannian embeds in the projective space $\mathbb{P}\big((S^d\mathbb{C}^2)^* \otimes \mathrm{L}(\omega_n)\big)$. We begin by describing convenient bases for the representation $\mathrm{L}(\omega_n)$ and its dual $\mathrm{L}(\omega_n)^*$.

For $\alpha \in \binom{\langle n\rangle}{n}$ and positive integers, set

$$v_\alpha^{(a)} := s^a t^{d-a} \otimes e_{\alpha_1} \wedge \cdots \wedge e_{\alpha_n} \in (S^d\mathbb{C}^2)^* \otimes \bigwedge^n \mathbb{C}^{2n},$$

and let

$$p_\alpha^{(a)} := u^a v^{d-a} \otimes e_{\alpha_1}^* \wedge \cdots \wedge e_{\alpha_n}^* \in S^d\mathbb{C}^2 \otimes \bigwedge^n \mathbb{C}^{2n*}$$

be the *Plücker coordinate* indexed by $\alpha^{(a)} \in \mathfrak{D}_{d,n}$, where $\{u, v\} \in \mathbb{C}^2$ and $\{s, t\} \in (\mathbb{C}^2)^*$ are dual bases.

The representation $\mathrm{L}(\omega_n)^*$ is the quotient of $\bigwedge^n \mathbb{C}^{2n*}$ by the linear subspace

$$L_n = \Omega \wedge \bigwedge^{n-2} \mathbb{C}^{2n*}$$

described in Proposition 2.3. Thus $S^d\mathbb{C}^2 \otimes \mathrm{L}(\omega_n)^*$ is the quotient of $S^d\mathbb{C}^2 \otimes \bigwedge^n \mathbb{C}^{2n*}$ by the linear subspace

$$L_{d,n} := S^d\mathbb{C}^2 \otimes L_n.$$

Note that $L_{d,n}$ is spanned by the linear forms

$$\ell_\alpha^{(a)} := u^a v^{d-a} \otimes \sum_{i \mid \{\bar{\imath}, i\} \cap \alpha = \varnothing} e_{\bar{\imath}}^* \wedge e_i^* \wedge e_{\alpha_1}^* \wedge \cdots \wedge e_{\alpha_{n-2}}^* \tag{5-1}$$

for $\alpha \in \binom{\langle n\rangle}{n-2}$ and $a = 0, \ldots, d$. The linear form (5-1) is simply $u^a v^{d-a}$ tensored with a linear form generating L_n. Each term in the linear form (5-1) is a Plücker coordinate indexed by a sequence involving both i and $\bar{\imath}$, for some $i \in [n]$.

Let $S \subseteq \mathrm{SL}_2(\mathbb{C})$ and $T \subseteq \mathrm{Sp}_{2n}(\mathbb{C})$ be maximal tori. The torus S is one-dimensional, so that its Lie algebra $\mathfrak{s}$ has basis consisting of a single element $H \in \mathfrak{s}$. For $i \in \langle n\rangle$, let $h_i := E_{ii} - E_{\bar{\imath}\bar{\imath}}$. The set $\{h_i \mid i \in [n]\}$ is a basis for the Lie algebra $\mathfrak{t}$ of $T \subseteq \mathrm{Sp}_{2n}(\mathbb{C})$. The weights of the maximal torus $S \times T \subseteq \mathrm{SL}_2(\mathbb{C}) \times \mathrm{Sp}_{2n}(\mathbb{C})$ are elements of $\mathfrak{s}^* \oplus \mathfrak{t}^*$. The Plücker coordinate $p_\alpha^{(a)} \in S^d\mathbb{C}^2 \otimes (\bigwedge^n \mathbb{C}^{2n})^*$ is a weight vector of weight

$$(d-2a)H^* + \sum_{i \mid \bar{\alpha}_i \notin \alpha} h_{\alpha_i}^*. \tag{5-2}$$

Each linear form (5-1) lies in a unique weight space. Thus, to find a basis for $S^d\mathbb{C}^2 \otimes \mathrm{L}(\omega_n)^*$, it suffices to find a basis for each weight space. We therefore fix the weight (5-2) and its corresponding weight space in the following discussion. We reduce to the case that the weight (5-2) is in fact 0, as follows.

For each $\alpha \in \binom{\langle n\rangle}{n-2}$, we have an element $\ell_\alpha = \Omega \wedge p_\alpha \in L_n$. This is a weight vector of weight

$$\omega_\alpha := h_{\alpha_1}^* + \cdots + h_{\alpha_k}^* \in \mathfrak{t}^*.$$

Set $\widetilde{\alpha} := \{i \in \alpha \mid \bar{\imath} \notin \alpha\}$ and observe that $\omega_{\widetilde{\alpha}} = \omega_\alpha$. The elements $\alpha \in \binom{\langle n\rangle}{n-2}$ such that $\ell_\alpha \in (L_n)_\omega$ are those satisfying $\omega_\alpha = \omega$. That is,

$$(L_n)_\omega = \langle \Omega \wedge p_\alpha \mid \omega_\alpha = \omega\rangle.$$

The shape of the linear form ℓ_α is determined by the number of pairs $\{\bar{\imath}, i\} \subseteq \alpha$; it is the same, up to multiplication of some variables by -1, as the linear form $\ell_{\alpha\setminus\widetilde{\alpha}} = \Omega \wedge p_{\alpha\setminus\widetilde{\alpha}} \in L_{n-|\widetilde{\alpha}|}$, of weight $\omega_{\alpha\setminus\widetilde{\alpha}} = 0$. It follows that the generators of $(L_n)_{\omega_\alpha}$ have the same form as those of $(L_{n-|\widetilde{\alpha}|})_{\omega_{\alpha\setminus\widetilde{\alpha}}}$, up to some signs arising from sorting the indices. Since these signs do not affect linear independence, it suffices to find a basis for $(L_n)_0$, from which it is then straightforward to deduce a basis for $(L_n)_{\omega_\alpha}$. We thus assume that the weight space in question is $(L_n)_0$. This implies that n is even; set $m := \frac{n}{2}$.

Example 5.1. We consider linear forms which span $(L_6)_{h_1^*+h_3^*}$. Let $m = 3$ (so $n = 6$) and $\omega = h_1^* + h_3^*$. If $\alpha = \bar{6}136$, then $\widetilde{\alpha} = 13$ and $\omega_\alpha = \omega$. We have

$$\ell_\alpha = p_{\bar{6}\bar{5}1356} + p_{\bar{6}\bar{4}1346} - p_{\bar{6}\bar{2}1236}.$$

The equations for the weight space $(L_n)_\omega$ are

$$\begin{aligned}
\ell_{\bar{6}136} &= p_{\bar{6}\bar{5}1356} + p_{\bar{6}\bar{4}1346} - p_{\bar{6}\bar{2}1236},\\
\ell_{\bar{5}135} &= p_{\bar{6}\bar{5}1356} + p_{\bar{5}\bar{4}1345} - p_{\bar{5}\bar{2}1235},\\
\ell_{\bar{4}134} &= p_{\bar{6}\bar{4}1346} + p_{\bar{5}\bar{4}1345} - p_{\bar{4}\bar{2}1234},\\
\ell_{\bar{2}123} &= p_{\bar{6}\bar{2}1236} + p_{\bar{5}\bar{2}1235} + p_{\bar{4}\bar{2}1234}.
\end{aligned}$$

We can obtain the linear forms which span $(L_4)_0$ (see Example 2.11) by first removing every occurrence of 1 and 3 in the subscripts above and then flattening the remaining indices. That is, we apply the following replacement (and similarly for the negative indices): $6 \mapsto 4$, $5 \mapsto 3$, $4 \mapsto 2$, and $2 \mapsto 1$. We then replace a variable by its negative if 2 appears in its index; this is to keep track of the sign of the permutation sorting the sequence $(\bar{\imath}, i, \alpha_1, \dots, \alpha_{n-2})$ in each term of ℓ_α (see (5-1)).

By Proposition 2.3, the map

$$\Bigl(\bigwedge^{2m} \mathbb{C}^{4m}\Bigr)_0 \to \Bigl(\bigwedge^{2m-2} \mathbb{C}^{4m}\Bigr)_0$$

given by contraction with the form

$$\Omega \in \bigwedge^2 (\mathbb{C}^{4m})^*$$

is surjective, with kernel $(\mathrm{L}(\omega_{2m}))_0$. Since the set $\{(\bar{\alpha},\alpha) \mid \alpha \in \binom{[2m]}{k}\}$ is a basis of $(\bigwedge^{2k}\mathbb{C}^{4m})_0$ (for any $k \le m$), we have

$$\begin{aligned}\dim(\mathrm{L}(\omega_{2m}))_0 &= \dim\Bigl(\bigwedge^{2m}\mathbb{C}^{4m}\Bigr)_0 - \dim\Bigl(\bigwedge^{2m-2}\mathbb{C}^{4m}\Bigr)_0\\ &= \binom{2m}{m} - \binom{2m}{m-2} = \frac{1}{m+1}\binom{2m}{m}.\end{aligned}$$

This (Catalan) number is equal to the number of admissible pairs of weight 0.

Lemma 5.2. $\dim(\mathrm{L}(\omega_n))_0$ *is equal to the number of admissible pairs* $(\alpha,\beta) \in \mathfrak{D}_n$ *of weight* $\frac{\omega_\alpha+\omega_\beta}{2} = 0$.

Proof. Recall that each trivial admissible pair (α,α), where

$$\alpha = (\bar{a}_1, \dots, \bar{a}_s, b_1, \dots, b_{n-s}) \in \mathfrak{D}_n,$$

indexes a weight vector of weight $\sum_{i=1}^{n-s} h^*_{b_i} - \sum_{i=1}^{s} h^*_{a_i}$. Also, the nontrivial admissible pairs are those (α,β) for which $\alpha < \beta$ have the same number of negative elements. Therefore, the admissible pairs of weight zero are the $(\alpha,\beta) \in \mathfrak{D}_n$ such that $\beta = (\bar{a}_m, \dots, \bar{a}_1, b_1, \dots, b_m)$, $\alpha = (\bar{b}_m, \dots, \bar{b}_1, a_1, \dots, a_m)$, and the sets $\{a_1, \dots, a_m\}$ and $\{b_1, \dots, b_m\}$ are disjoint. This last condition is equivalent to $a_i > b_i$ for all $i \in [m]$. The number of such pairs is equal to the number of standard tableaux of shape (m^2) (that is, a rectangular box with 2 rows and m columns) with entries in $[2m]$. By the hook length formula [Fulton 1997] this number is $\frac{1}{m+1}\binom{2m}{m}$. □

The weight vectors $p_\alpha \in (\bigwedge^n \mathbb{C}^{4m*})_0$ are indexed by sequences of the form

$$\alpha = (\bar{\alpha}_m, \dots, \bar{\alpha}_1, \alpha_1, \dots, \alpha_m)$$

which can be abbreviated by the positive subsequence $\alpha_+ := (\alpha_1, \dots, \alpha_m) \in \binom{[2m]}{m}$ without ambiguity. We take these as an indexing set for the variables appearing in the linear forms (5-1).

With this notation, the positive parts of Northeast sequences are characterized in Proposition 5.4. The proof requires the following definition.

Definition 5.3. A *tableau* is a partition whose boxes are filled with integers from the set $[n]$, for some $n \in \mathbb{N}$. A tableau is *standard* if the entries strictly increase from left to right and top to bottom.

Proposition 5.4. *Let* $\alpha \in \binom{\langle 2m\rangle}{2m}$ *be a Northeast sequence. Then the positive part of* α *satisfies* $\alpha_+ \ge 24 \cdots (2m) \in \binom{[2m]}{m}$. *In particular, no Northeast sequence contains* $1 \in [2m]$ *and every Northeast sequence contains* $2m \in [2m]$.

Proof. $\alpha_+ \geq 24\cdots(2m)$ if and only if the tableau of shape (m^2) whose first row is filled with the sequence $(\alpha^t)_+ = [n] \setminus \alpha_+$ and whose second row is filled with the α_+ is standard. This is equivalent to α being Northeast. $\square$

It follows from Proposition 2.10 that the set $\mathcal{NE}$ of Northeast sequences indexing vectors of weight zero has cardinality equal to the dimension of the zero-weight space of the representation $\mathrm{L}(\omega_{2m})^*$. This weight space is the cokernel of the map

$$\Omega \wedge \bullet : \Big(\bigwedge^{2m-2} \mathbb{C}^{4m}\Big)^*_0 \to \Big(\bigwedge^{2m} \mathbb{C}^{4m}\Big)^*_0.$$

Similarly, the weight space $\mathrm{L}(\omega_{2m})_0$ is the kernel of the dual map

$$\Omega \lrcorner \bullet : \Big(\bigwedge^{2m} \mathbb{C}^{4m}\Big)_0 \to \Big(\bigwedge^{2m-2} \mathbb{C}^{4m}\Big)_0.$$

We fix the positive integer m, and consider only the positive subsequence α_+ of the sequence $\alpha \in \binom{\langle 2m\rangle}{2m}$. When the weight of p_α is 0, α_+ is an element of $\binom{[2m]}{m}$. For $\alpha \in \binom{[2m]}{m}$, we call a bijection $M : \alpha \to \alpha^c$ a *matching* of α. Fixing a matching $M : \alpha \to \alpha^c$, we have an element of the kernel $\mathrm{L}(\omega_{2m})$, as follows. Let H_α be the set of all sequences in $\binom{[2m]}{m}$ obtained by interchanging $M(\alpha_i)$ and α_i, for $i \in I$, $I \subseteq [m]$.

Elements of the set H_α are the vertices of an m-dimensional hypercube, whose edges connect pairs of sequences which are related by the interchange of a single element. Equivalently, a pair of sequences are connected by an edge if they share a subsequence of size $m-1$. For any such subsequence $\beta \subseteq \alpha$ there exists a unique edge of H_α connecting the two vertices which share the subsequence β. Let $I \cdot \alpha$ denote the element of H_α obtained from α by the interchange of $M(\alpha_i)$ and α_i for $i \in I$. The element

$$K_\alpha := \sum_{I \subseteq [m]} (-1)^{|I|} v_{I\cdot\alpha}$$

lies in the kernel $\mathrm{L}(\omega_{2m})$. Indeed, for each $I \subseteq [m]$, we have

$$\Omega \lrcorner v_{I\cdot\alpha} = \sum_{i=1}^{m} v_{(I\cdot\alpha)\setminus\{(I\cdot\alpha)_i\}}.$$

For each term $v_{(I\cdot\alpha)\setminus\{(I\cdot\alpha)_i\}}$ on the right hand side, let $j \in [m]$ be such that either $(I\cdot\alpha)_i = \alpha_j$ or $(I\cdot\alpha)_i = \alpha_j^c$. Set

$$J = \begin{cases} I \cup \{j\}, & \text{if } (I\cdot\alpha)_i = \alpha_j, \\ I \setminus \{j\}, & \text{if } (I\cdot\alpha)_i = \alpha_j^c. \end{cases}$$

The set J is the unique subset of $[m]$ such that $(I\cdot\alpha) \setminus \{(I\cdot\alpha)_i\}$ is in the support of $\Omega \lrcorner v_{J\cdot\alpha}$, with coefficient $(-1)^{|J|} = (-1)^{|I|+1}$. Hence these terms cancel in K_α, and we see that the coefficient of each v_β for $\beta \in \binom{[2m]}{m-1}$ in the support of $\Omega \lrcorner K_\alpha$ is zero. Therefore $\Omega \lrcorner K_\alpha = 0$. See Example 5.8 for the case $m = 2$.

If $\alpha \in \mathcal{NE}$ then there exists a *descending matching*, that is, $M(\alpha_i) < \alpha_i$ for all $i \in [m]$. For example, the condition that the matching $M(\alpha_i) := \alpha_i^c$ be descending is equivalent to the condition that α be Northeast. If we choose a descending matching for each $\alpha \in \mathcal{NE}$, the element $K_\alpha \in \mathrm{L}(\omega_{2m})$ is supported on sequences which precede α in the poset $\binom{[2m]}{m}$. It follows that the set $\mathcal{B} := \{K_\alpha \in \mathrm{L}(\omega_{2m}) \mid \alpha \in \mathcal{NE}\}$ is a basis for $\mathrm{L}(\omega_{2m})$.

Lemma 5.5. *The Plücker coordinates p_α with $\alpha \in \mathcal{NE}$ are a basis for* $\mathrm{L}(\omega_{2m})^*$.

Proof. Fix a basis $\mathcal{B}$ of $\mathrm{L}(\omega_{2m})$ obtained from descending matchings of each Northeast sequence. We use this basis to show that the set of Plücker coordinates p_α such that α is Northeast is a basis for the dual $\mathrm{L}(\omega_{2m})^*$.

Suppose not. Then there exists a linear form

$$\ell = \sum_{\alpha \in \mathcal{NE}} c_\alpha p_\alpha$$

vanishing on each element of the basis $\mathcal{B}$. We show by induction on the poset $\mathcal{NE}$ that all of the coefficients c_α appearing in this form vanish.

Fix a Northeast sequence $\alpha \in \mathcal{NE}$, and assume that $c_\beta = 0$ for all Northeast $\beta < \alpha$. Since K_α involves only the basis vectors v_β with $\beta \le \alpha$, we have $\ell(K_\alpha) = c_\alpha$, hence $c_\alpha = 0$. This completes the inductive step of the proof.

The initial step of the induction is simply the inductive step applied to the unique minimal Northeast sequence $\alpha = 24 \cdots (2m)$. □

It follows that every Plücker coordinate p_α indexed by a non-Northeast sequence α can be written uniquely as a linear combination of Plücker coordinates indexed by Northeast sequences. We can be more precise about the form of these linear combinations. Recall that each fiber of the map π_{2m} contains a unique Northeast sequence. For a sequence α_0, let α be the Northeast sequence in the same fiber as α_0.

Lemma 5.6. *For each non-Northeast sequence α_0, let ℓ'_{α_0} be the linear relation among the Plücker coordinates expressing p_{α_0} as a linear combination of the p_β with β Northeast. Then p_α appears in ℓ'_{α_0} with coefficient $(-1)^{|I|}$, where $\alpha = I \cdot \alpha_0$, and every other Northeast β with p_β in the support of ℓ'_{α_0} satisfies $\beta > \alpha$.*

Proof. Let M be the descending matching of α with α^c defined by $M(\alpha_i) := \alpha_i^c$. Let K_α be the kernel element obtained by the process described above. Any linear form

$$\ell = p_{\alpha_0} + (-1)^{|I|+1} p_\alpha + \sum_{\alpha < \beta \in \mathcal{NE}} c_\beta p_\beta$$

vanishes on K_α.

We extend this relation to one which vanishes on all of $\mathrm{L}(\omega_n)_0$, proceeding inductively on the poset of Northeast sequences greater than or equal to α. Suppose that $\beta > \alpha$ is Northeast. By induction, suppose that for each Northeast sequence γ in the interval $[\alpha, \beta]$ the coefficient c_γ of ℓ has been determined in such a way that $\ell(K_\gamma) = 0$.

Let S be the set of Northeast sequences γ in the open interval (α, β) such that v_γ appears in K_β. Then

$$\ell(K_\beta) = \left(\sum_{\gamma \in S} c_\gamma\right) + c_\beta,$$

so setting $c_\beta := -\sum_{\gamma \in S} c_\gamma$ implies that $\ell(K_\beta) = 0$.

This completes the inductive part of the proof. We now have a linear form ℓ vanishing on $\mathrm{L}(\omega_n)_0$ which expresses p_{α_0} as a linear combination of Plücker coordinates indexed by Northeast sequences. Since such a linear form is unique, $\ell = \ell'_{\alpha_0}$. □

By Lemmas 5.5 and 5.6 and the argument preceding them, we deduce:

Theorem 5.7. *The system of linear relations*

$$\{\ell_\alpha^{(a)} = u^a v^{d-a} \otimes \Omega \wedge p_\alpha \mid a = 0, \ldots, d,\ \alpha \in \tbinom{\langle n \rangle}{n-2}\}$$

has a reduced normal form consisting of linear forms expressing each Plücker coordinate $p_\beta^{(b)}$ with $\beta \notin \mathcal{NE} \subseteq \binom{\langle n \rangle}{n}$ as a linear combination of Plücker coordinates indexed by Northeast elements of $\binom{\langle n \rangle}{n}$.

Proof. We have seen that the linear relations preserve weight spaces, and Lemmas 5.5 and 5.6 provide the required normal form on each of these. The union of the relations constitute a normal form for the linear relations generating the entire linear subspace $L_{d,n}$. □

Example 5.8. Consider the zero weight space $(\bigwedge^4 \mathbb{C}^8)_0$ (so that $m = 2$). This is spanned by the vectors

$$v_\alpha := e_{\alpha_1} \wedge e_{\alpha_2} \wedge e_{\alpha_3} \wedge e_{\alpha_4}$$

(with dual basis the Plücker coordinates $p_\alpha = v_\alpha^*$), where

$$\alpha \in \{\bar{4}\bar{3}34, \bar{4}\bar{2}24, \bar{4}\bar{1}14, \bar{3}\bar{2}23, \bar{3}\bar{1}13, \bar{2}\bar{1}12\}.$$

The Northeast sequences are $\bar{4}\bar{3}34$ and $\bar{4}\bar{2}24$. The kernel of

$$\Omega \lrcorner \bullet : \left(\bigwedge^4 \mathbb{C}^8\right)_0 \to \left(\bigwedge^2 \mathbb{C}^8\right)_0$$

is spanned by the vectors

$$K_{\bar{4}\bar{2}24} = v_{\bar{4}\bar{2}24} - v_{\bar{4}\bar{1}14} - v_{\bar{3}\bar{2}23} + v_{\bar{3}\bar{1}13}, \text{ and}$$
$$K_{\bar{4}\bar{3}34} = v_{\bar{4}\bar{3}34} - v_{\bar{4}\bar{1}14} - v_{\bar{3}\bar{2}23} + v_{\bar{2}\bar{1}12}.$$

To see this concretely, we compute:

$$\Omega \lrcorner K_{\bar{4}\bar{2}24} = v_{\bar{4}4} + v_{\bar{2}2} - v_{\bar{4}4} - v_{\bar{1}1} - v_{\bar{3}3} - v_{\bar{2}2} + v_{\bar{3}3} + v_{\bar{1}1} = 0,$$

and similarly $\Omega \lrcorner K_{\bar{4}\bar{3}34} = 0$. The fibers of the map $\pi_4 : \binom{\langle 4\rangle}{4} \to \mathscr{D}_4$ are

$$\pi_4^{-1}(\bar{4}\bar{3}12, \bar{2}\bar{1}34) = \{\bar{4}\bar{3}34, \bar{2}\bar{1}12\}, \text{ and}$$
$$\pi_4^{-1}(\bar{4}\bar{2}13, \bar{3}\bar{1}24) = \{\bar{4}\bar{2}24, \bar{4}\bar{1}14, \bar{3}\bar{2}23, \bar{3}\bar{1}13\}.$$

The expression for $p_{\bar{4}\bar{1}14}$ as a linear combination of Plücker coordinates indexed by Northeast sequences is

$$\ell_{\bar{4}\bar{1}14} = p_{\bar{4}\bar{1}14} + c_{\bar{4}\bar{2}24} p_{\bar{4}\bar{2}24} + c_{\bar{4}\bar{3}34} p_{\bar{4}\bar{3}34},$$

for some $c_{\bar{4}\bar{2}24}, c_{\bar{4}\bar{3}34} \in \mathbb{C}$, which we can compute as follows. Since

$$0 = \ell_{\bar{4}\bar{1}14}(K_{\bar{4}\bar{2}24}) = c_{\bar{4}\bar{2}24} - 1,$$

we have $c_{\bar{4}\bar{2}24} = 1$. Similarly,

$$0 = \ell_{\bar{4}\bar{1}14}(K_{\bar{4}\bar{3}34}) = c_{\bar{4}\bar{3}34} - 1,$$

so $c_{\bar{4}\bar{3}34} = 1$. Hence $\ell_{\bar{4}\bar{1}14} = p_{\bar{4}\bar{1}14} + p_{\bar{4}\bar{2}24} + p_{\bar{4}\bar{3}34}$, which agrees with (2-3).

5B. ***Proof of the straightening law.*** We find generators of $(I_{d,n}+L_{d,n}) \cap \mathbb{C}[\mathscr{D}_{d,n}]$ which express the quotient as an algebra with straightening law on $\mathscr{D}_{d,n}$. Such a generating set is automatically a Gröbner basis with respect to the degree reverse lexicographic term order where variables are ordered by a refinement of the doset order. We begin with a Gröbner basis $G_{I_{d,n}+L_{d,n}}$ for $I_{d,n}+L_{d,n}$ with respect to a similar term order. For $\alpha^{(a)} \in \binom{\langle n\rangle}{n}_d$, write

$$\check{\alpha}^{(a)} := \alpha^{(a)} \vee (\alpha^t)^{(a)} \quad \text{and} \quad \hat{\alpha}^{(a)} := \alpha^{(a)} \wedge (\alpha^t)^{(a)},$$

so that

$$\pi_n(\alpha^{(a)}) = (\hat{\alpha}^{(a)}, \check{\alpha}^{(a)}).$$

We call an element $\alpha^{(a)} \in \binom{\langle n\rangle}{n}_d$ *Northeast* if $\alpha \in \binom{\langle n\rangle}{n}$ is Northeast.

Let $<$ be a linear refinement of the partial order on $\mathscr{P}_{d,n}$ satisfying the following conditions. First, the Northeast sequence is minimal among those in a given fiber of π_n. This is possible since every weight space is an antichain (that is, no two elements are comparable). Second, $\alpha^{(a)} < \beta^{(b)}$ if $(\hat{\alpha}^{(a)}, \check{\alpha}^{(a)})$ is lexicographically smaller than $(\hat{\beta}^{(b)}, \check{\beta}^{(b)})$.

With respect to any such refinement, consider the degree reverse lexicographic term order. A reduced Gröbner basis $G_{d,n}$ for $I_{d,n}+L_{d,n}$ with respect to this term order will have standard monomials indexed by chains (in $\mathcal{P}_{d,n}$) of Northeast partitions. While every monomial supported on a chain of Northeast partitions is standard modulo $I_{d,n}$, this is not always the case modulo $I_{d,n}+L_{d,n}$. In other words, upon identifying each Northeast partition appearing in a given monomial with an element of $\mathcal{D}_{d,n}$, we do not necessarily obtain a monomial supported on a chain in $\mathcal{D}_{d,n}$. It is thus necessary to identify precisely which Northeast chains in $\binom{\langle n\rangle}{n}$ correspond to chains in $\mathcal{D}_{d,n}$ via the map π_n.

A monomial $p_\alpha^{(a)} p_\beta^{(b)}$ such that $\alpha^{(a)} < \beta^{(b)}, (\beta^t)^{(b)}$, and $\alpha^{(a)}, \beta^{(b)} \in \mathcal{NE}$ cannot be reduced modulo $G_{I_{d,n}}$ or $G_{L_{d,n}}$. On the other hand, if $\alpha^{(a)} < \beta^{(b)}$ (say), but $\alpha^{(a)}$ and $(\beta^t)^{(b)}$ are incomparable (written $\alpha^{(a)} \not\sim (\beta^t)^{(b)}$) then there is a relation in $G_{I_{d,n}}$ with leading term $p_\alpha^{(a)} p_{\beta^t}^{(b)}$. It follows that the degree-two standard monomials are indexed by Northeast partitions $p_\alpha^{(a)} p_\beta^{(b)}$ with $\alpha^{(a)} < \beta^{(b)}, (\beta^t)^{(b)}$.

Conversely, any monomial $p_\alpha^{(a)} p_\beta^{(b)}$ with $\alpha^{(a)} < \beta^{(b)}, (\beta^t)^{(b)}$, and $\alpha^{(a)}, \beta^{(b)} \in \mathcal{NE}$ cannot be the leading term of any element of $G_{I_{d,n}+L_{d,n}}$. To see this, observe that $G_{I_{d,n}+L_{d,n}}$ is obtained by Buchberger's algorithm [1965] applied to $G_{I_{d,n}} \cup G_{L_{d,n}}$, and we may consider only the S-polynomials $S(f, g)$ with $f \in G_{I_{d,n}}$ and $g \in G_{L_{d,n}}$. In this case we may assume $\mathrm{in}_< g$ divides $\mathrm{in}_< f$.

Let α_0 be the partition such that $\mathrm{in}_< g = p_{\alpha_0}^{(a)}$ (that is, g is the unique expression of $p_{\alpha_0}^{(a)}$ as a linear combination of Plücker coordinates indexed by Northeast partitions), and let α be the unique Northeast partition such that $\pi_n(\alpha_0) = \pi_n(\alpha)$. By the reduced normal form given in Theorem 5.7, $S(f, g)$ is the obtained by replacing $p_{\alpha_0}^{(a)}$ with $\pm p_\alpha^{(a)} + \ell$, where ℓ is a linear combination of Plücker coordinates $p_\gamma^{(a)}$ with γ Northeast and $\alpha_+ < \gamma_+$. This latter condition implies that $\hat{\alpha} < \hat{\gamma}$ (also, $\check{\alpha} > \check{\gamma}$), and therefore $(\hat{\alpha}, \check{\alpha})$ is lexicographically smaller than $(\hat{\gamma}, \check{\gamma})$.

Hence, with respect to the reduced Gröbner basis $G_{I_{d,n}+L_{d,n}}$, the standard monomials are precisely the monomials $p_\alpha^{(a)} p_\beta^{(b)}$ with $\alpha^{(a)} \le \beta^{(b)}, (\beta^t)^{(b)}$, and $\alpha^{(a)}, \beta^{(b)} \in \mathcal{NE}$.

Recall that elements of the doset $\mathcal{D}_{d,n}$ are pairs (α, β) of admissible elements (Definition 2.6) of $\binom{\langle n\rangle}{n}_d$ such that (regarded as sequences):

- $\alpha \le \beta$;
- α and β have the same number of negative (or positive) elements.

Equivalently, regarding α and β as partitions, the elements of $\mathcal{D}_{d,n}$ are pairs (α, β) of symmetric partitions such that:

- $\alpha \subseteq \beta$;
- α and β have the same Durfee square, where the *Durfee square* of a partition α is the largest square subpartition $(p^p) \subseteq \alpha$ (for some $p \le n$).

Theorem 5.9. *$\mathbb{C}[\binom{\langle n\rangle}{n}_d]/\langle I_{d,n}+L_{d,n}\rangle$ is an algebra with straightening law on $\mathfrak{D}_{d,n}$.*

Proof. Since standard monomials with respect to a Gröbner basis are linearly independent, the arguments above establish the conditions in Definition 3.3 (1) and (2).

To establish the condition (3), note that it suffices to consider the expression for a degree-2 monomial as a sum of standard monomials. For simplicity, we absorb the superscripts into our notation and write $\alpha \in \binom{\langle n\rangle}{n}_d$ and similarly for the corresponding Plücker coordinate. Let

$$p_{(\hat{\alpha},\check{\alpha})}p_{(\hat{\beta},\check{\beta})} = \sum_{j=1}^{k} c_j\, p_{(\hat{\alpha}_j,\check{\alpha}_j)}p_{(\hat{\beta}_j,\check{\beta}_j)} \tag{5-3}$$

be a reduced expression in $G_{I_{d,n}+L_{d,n}}$ for $p_{(\hat{\alpha},\check{\alpha})}p_{(\hat{\beta},\check{\beta})}$ as a sum of standard monomials. That is, $p_{(\hat{\alpha},\check{\alpha})}p_{(\hat{\beta},\check{\beta})}$ is nonstandard and $p_{(\hat{\alpha}_j,\check{\alpha}_j)}p_{(\hat{\beta}_j,\check{\beta}_j)}$ is standard for $j = 1, \ldots, k$. We assume that α (respectively, β) be the unique Northeast partition such that $\pi_n(\alpha) = (\hat{\alpha}, \check{\alpha})$ (respectively, $\pi_n(\beta) = (\hat{\beta}, \check{\beta})$), and similarly for each α_j and β_j appearing in (5-3).

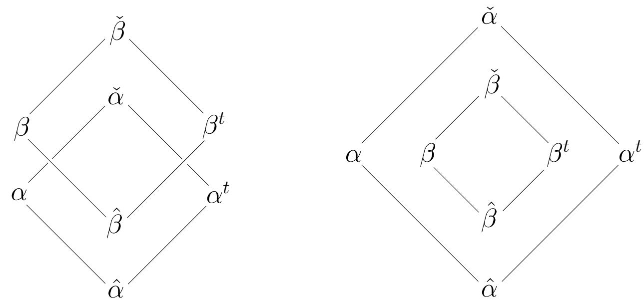
Fix $j = 1, \ldots, k$. The standard monomial $p_{(\hat{\alpha}_j,\check{\alpha}_j)}p_{(\hat{\beta}_j,\check{\beta}_j)}$ is obtained by the reduction modulo $G_{L_{d,n}}$ of a standard monomial $p_\gamma p_\delta$ appearing in the straightening relation for $p_\alpha p_\beta$, which is an element of the Gröbner basis $G_{I_{d,n}}$. If γ and δ are both Northeast, then nothing happens, that is, $\gamma = \alpha_j$ and $\delta = \beta_j$. If γ is not Northeast, then we rewrite p_γ as a linear combination of Plücker coordinates indexed by Northeast sequences. Lemma 5.6 ensures that the leading term of the new expression is $p_{(\hat{\gamma},\check{\gamma})}$, and the lower order terms $p_{(\hat{\epsilon},\check{\epsilon})}$ satisfy $\hat{\epsilon} < \hat{\gamma}$.

It follows that the lexicographic comparison in the condition of Definition 3.3 (3) terminates with the first Plücker coordinate. That is, if $(\hat{\alpha}_j \le \check{\alpha}_j \le \hat{\beta}_j \le \check{\beta}_j)$ is lexicographically smaller than $(\hat{\alpha} \le \check{\alpha} \le \hat{\beta} \le \check{\beta})$, then either $\hat{\alpha}_j < \hat{\alpha}$ or $\hat{\alpha}_j = \hat{\alpha}$ and $\hat{\alpha}_j < \hat{\alpha}$. Therefore the reduction process applied to p_δ does not affect the result, and the condition (3) is proven.

It remains to prove the condition (4). Suppose that $(\hat{\alpha}, \check{\alpha})$ and $(\hat{\beta}, \check{\beta})$ are incomparable elements of $\mathfrak{D}_{d,n}$ (α and β Northeast). This means that α is incomparable to either β or β^t (possibly both). Without loss of generality, we will deal only with the more complicated case that α and β^t are incomparable. The hypothesis of the condition (4) is that the set $\{\hat{\alpha}, \check{\alpha}, \hat{\beta}, \check{\beta}\}$ forms a chain in $\binom{\langle n\rangle}{n}_d$. Up to interchanging the roles of α and β, there are two possible cases (see Figure 9):

$$\hat{\alpha} < \hat{\beta} < \check{\alpha} < \check{\beta}, \quad \text{or} \quad \hat{\alpha} < \hat{\beta} < \check{\beta} < \check{\alpha}.$$

First, suppose $\hat{\alpha} < \hat{\beta} < \check{\alpha} < \check{\beta}$. Recall that for any $\gamma_0 \in \binom{\langle n\rangle}{n}_d$, with Northeast sequence γ in the same fiber of π_n, the expression for the Plücker coordinate p_{γ_0} as a linear combination of Plücker coordinates indexed by Northeast sequences is



$\hat{\alpha} < \hat{\beta} < \check{\alpha} < \check{\beta}$ $\qquad\qquad$ $\check{\alpha} < \hat{\beta} < \check{\beta} < \check{\alpha}$

Figure 9. The two cases in the proof of the condition of Definition 3.3 (4).

supported on Plücker coordinates p_δ such that $\delta_+ \geq \gamma_+$ with equality if and only if $\delta = \gamma$, and the Plücker coordinate p_γ appears with coefficient ± 1 (Lemma 5.6).

Upon replacing each Northeast (or Southwest) partition with its associated doset element using the map π_n from Section 2B, the first two terms of straightening relation for $p_\alpha p_{\beta^t}$ are

$$\begin{aligned} p_\alpha p_{\beta^t} - p_{\alpha\wedge\beta^t} p_{\alpha\vee\beta^t} \\ &= p_\alpha p_{\beta^t} - \sigma p_{((\alpha\wedge\beta^t)^\wedge,(\alpha\wedge\beta^t)^\vee)} p_{((\alpha\vee\beta^t)^\wedge,(\alpha\vee\beta^t)^\vee)} + \text{lower order terms} \\ &= \sigma_\beta p_{(\hat{\alpha},\check{\alpha})} p_{(\hat{\beta},\check{\beta})} - \sigma p_{(\hat{\alpha},\hat{\beta})} p_{(\check{\alpha},\check{\beta})} + \text{lower order terms,} \end{aligned}$$

where $\sigma = \pm 1$. The second equation is justified as follows. For any element $\alpha \in \binom{\langle n\rangle}{n}_d$, recall that α_+ (respectively, α_-) denotes the subsequence of positive (negative) elements of α. This was previous defined for elements of $\binom{\langle n\rangle}{n}$, but extends to elements of $\binom{\langle n\rangle}{n}_d$ in the obvious way, that is, by ignoring the superscript. The condition

$$\hat{\alpha} < \hat{\beta} < \check{\alpha} < \check{\beta}$$

is equivalent to

$$\bar{\alpha}_-^c < \bar{\beta}_-^c < \alpha_+ < \beta_+.$$

Note that this implies that

$$\alpha \wedge \beta^t = \alpha_- \cup \bar{\beta}_-^c, \quad \text{and} \quad \alpha \vee \beta^t = \bar{\alpha}_+^c \cup \beta_+.$$

We compute in the distributive lattice $\binom{\langle n\rangle}{n}_d$.

$$\begin{aligned} (\alpha\wedge\beta^t)\wedge(\alpha^t\wedge\beta) &= (\alpha_-\cup\bar{\beta}_-^c)\wedge(\beta_-\cup\bar{\alpha}_-^c) = \hat{\alpha} \\ (\alpha\wedge\beta^t)\vee(\alpha^t\wedge\beta) &= (\alpha_-\cup\bar{\beta}_-^c)\vee(\beta_-\cup\bar{\alpha}_-^c) = \hat{\beta} \end{aligned}$$

Similarly,

$$(\alpha \vee \beta^t) \wedge (\alpha^t \vee \beta) = \check{\alpha}, \quad \text{and} \quad (\alpha \vee \beta^t) \vee (\alpha^t \vee \beta) = \check{\beta}.$$

In the remaining case, we have $\check{\alpha} < \hat{\beta} < \check{\beta} < \check{\alpha}$, and it follows that $\alpha \not\sim \beta$ and $\alpha \not\sim \beta^t$ both hold. We use the relation for the incomparable pair $\alpha \not\sim \beta^t$:

$$\begin{aligned} p_\alpha p_{\beta^t} - p_{\alpha\wedge\beta^t} p_{\alpha\vee\beta^t} \\ &= p_\alpha p_{\beta^t} - \sigma p_{((\alpha\wedge\beta^t)^\wedge,(\alpha\wedge\beta^t)^\vee)} p_{((\alpha\vee\beta^t)^\wedge,(\alpha\vee\beta^t)^\vee)} + \text{lower order terms} \\ &= \sigma_\beta p_{(\hat{\alpha},\check{\alpha})} p_{(\hat{\beta},\check{\beta})} - \sigma p_{(\hat{\alpha},\hat{\beta})} p_{(\check{\beta},\check{\alpha})} + \text{lower order terms,} \end{aligned}$$

where the second equality holds by a similar computation in $\binom{\langle n\rangle}{n}_d$. □

The next result shows that the algebra with straightening law just constructed is indeed the coordinate ring of $L\mathcal{Q}_d(n)$.

Theorem 5.10. $\mathbb{C}[\binom{\langle n\rangle}{n}_d]/\langle I_{d,n}+L_{d,n}\rangle \cong \mathbb{C}[L\mathcal{Q}_d(n)]$.

Proof. Let $I' := I(L\mathcal{Q}_d(n))$. By definition, we have $I_{d,n}+L_{d,n} \subseteq I'$. Since the degree and codimension of these ideals are equal, I' is nilpotent modulo $I_{d,n}+L_{d,n}$. On the other hand $I_{d,n}+L_{d,n}$ is radical, so $I_{d,n}+L_{d,n} = I'$. □

The arguments of De Concini and Lakshmibai [1981, Theorem 4.5] extend to the case of Schubert subvarieties of $L\mathcal{Q}_d(n)$.

Corollary 5.11. *The coordinate ring of any Schubert subvariety of $L\mathcal{Q}_d(n)$ is an algebra with straightening law on a doset, hence Cohen–Macaulay and Koszul.*

Proof. For $\alpha^{(a)} \in \mathcal{D}_{d,n}$, the Schubert variety $X_{\alpha^{(a)}}$ is defined by the vanishing of the Plücker coordinates $p^{(b)}_{(\beta,\gamma)}$ for $\gamma^{(b)} \not\leq \alpha^{(a)}$. The conditions in Definition 3.3 (4) are stable upon setting these variables to zero, so we obtain an algebra with straightening law on the doset

$$\{(\beta,\gamma)^{(b)} \in \mathcal{D}_{d,n} \mid \gamma^{(b)} \leq \alpha^{(a)}\}.$$

Let $\mathcal{D} \subseteq \mathcal{P}\times\mathcal{P}$ be a doset on the poset $\mathcal{P}$, A any algebra with straightening law on $\mathcal{D}$, and $\mathbb{C}\{\mathcal{P}\}$ the unique *discrete* algebra with straightening law on $\mathcal{P}$. That is, $\mathbb{C}\{\mathcal{P}\}$ has algebra generators corresponding to the elements of $\mathcal{P}$, and the straightening relations are $\alpha\beta = 0$ if α and β are incomparable elements of $\mathcal{P}$. Then A is Cohen–Macaulay if and only if $\mathbb{C}\{\mathcal{P}\}$ is Cohen–Macaulay [De Concini and Lakshmibai 1981].

On the other hand, $\mathbb{C}\{\mathcal{P}\}$ is the face ring of the order complex of $\mathcal{P}$. The order complex of a locally upper semimodular poset is shellable. The face ring of a shellable simplicial complex is Cohen–Macaulay [Bruns and Herzog 1993]. By Proposition 3.17, any interval in the poset $\mathcal{P}_{d,n}$ is a distributive lattice, hence locally upper semimodular. This proves that $\mathbb{C}[L\mathcal{Q}_d(n)]$ is Cohen–Macaulay. The

Koszul property is a consequence of the quadratic Gröbner basis consisting of the straightening relations. □

The main results of this paper suggest that the space of quasimaps is an adequate setting for the study of the enumerative geometry of curves into a general flag variety. They also give a new and interesting example of a family of varieties whose coordinate rings are Hodge algebras.

After the ordinary Grassmannian, the Lagrangian Grassmannian was the first space to be well understood in terms of (classical) standard monomial theory. Our results thus lend credence to the expectation that further study of the space of quasimaps into a flag variety of general type, possibly incorporating the ideas of Chirivì [2000; 2001], will yield new results in parallel (to some extent) with the classical theory.

Acknowledgments

We thank Frank Sottile for suggesting this problem, and for valuable feedback during the preparation of this paper.

References

[Bertram 1997] A. Bertram, "Quantum Schubert calculus", *Adv. Math.* **128**:2 (1997), 289–305. MR 98j:14067 Zbl 0945.14031

[Bertram et al. 2005] A. Bertram, I. Ciocan-Fontanine, and B. Kim, "Two proofs of a conjecture of Hori and Vafa", *Duke Math. J.* **126**:1 (2005), 101–136. MR 2006e:14077 Zbl 1082.14055

[Braverman 2006] A. Braverman, "Spaces of quasi-maps into the flag varieties and their applications", pp. 1145–1170 in *International Congress of Mathematicians* (Madrid, 2006), vol. II, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2006. MR 2008i:14019 Zbl 1118.22007

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

[Bruns and Vetter 1988] W. Bruns and U. Vetter, *Determinantal rings*, Lecture Notes in Mathematics **1327**, Springer, Berlin, 1988. MR 89i:13001 Zbl 0673.13006

[Buchberger 1965] B. Buchberger, *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*, PhD thesis, University of Innsbruck, Austria, 1965.

[Chen 2001] L. Chen, "Poincaré polynomials of hyperquot schemes", *Math. Ann.* **321**:2 (2001), 235–251. MR 2002g:14075 Zbl 1076.14536

[Chen 2003] L. Chen, "Quantum cohomology of flag manifolds", *Adv. Math.* **174**:1 (2003), 1–34. MR 2004a:14055 Zbl 1061.14061

[Chirivì 2000] R. Chirivì, "LS algebras and application to Schubert varieties", *Transform. Groups* **5**:3 (2000), 245–264. MR 2001h:14060 Zbl 1019.14019

[Chirivì 2001] R. Chirivì, "Deformation and Cohen–Macaulayness of the multicone over the flag variety", *Comment. Math. Helv.* **76**:3 (2001), 436–466. MR 2002f:14064 Zbl 1074.14529

[Ciocan-Fontanine 1999] I. Ciocan-Fontanine, "On quantum cohomology rings of partial flag varieties", *Duke Math. J.* **98**:3 (1999), 485–524. MR 2000d:14058 Zbl 0969.14039

[De Concini and Lakshmibai 1981] C. De Concini and V. Lakshmibai, "Arithmetic Cohen–Macaulayness and arithmetic normality for Schubert varieties", *Amer. J. Math.* **103**:5 (1981), 835–850. MR 83e:14035 Zbl 0475.14045

[De Concini et al. 1982] C. De Concini, D. Eisenbud, and C. Procesi, *Hodge algebras*, Astérisque **91**, Société Mathématique de France, Paris, 1982. MR 85d:13009 Zbl 0509.13026

[Eisenbud 1980] D. Eisenbud, "Introduction to algebras with straightening laws", pp. 243–268 in *Ring theory and algebra, III* (Norman, OK, 1979), edited by B. R. McDonald, Lecture Notes in Pure and Appl. Math. **55**, Marcel Dekker, New York, 1980. MR 82i:14030 Zbl 0448.13010

[Fulton 1997] W. Fulton, *Young tableaux: With applications to representation theory and geometry*, London Mathematical Society Student Texts **35**, Cambridge University Press, Cambridge, 1997. MR 99f:05119 Zbl 0878.14034

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory: A first course*, Graduate Texts in Mathematics **129**, Springer, New York, 1991. MR 93a:20069 Zbl 0744.22001

[Fulton and Pandharipande 1997] W. Fulton and R. Pandharipande, "Notes on stable maps and quantum cohomology", pp. 45–96 in *Algebraic geometry* (Santa Cruz, 1995), edited by J. Kollár et al., Proc. Sympos. Pure Math. **62**, Amer. Math. Soc., Providence, RI, 1997. MR 98m:14025 Zbl 0898.14018

[Givental 1996] A. B. Givental, "Equivariant Gromov–Witten invariants", *Internat. Math. Res. Notices* 13 (1996), 613–663. MR 97e:14015 Zbl 0881.55006

[Hiller and Boe 1986] H. Hiller and B. Boe, "Pieri formula for $\mathrm{SO}_{2n+1}/\mathrm{U}_n$ and $\mathrm{Sp}_n/\mathrm{U}_n$", *Adv. in Math.* **62**:1 (1986), 49–67. MR 87k:14058 Zbl 0611.14036

[Hodge 1943] W. V. D. Hodge, "Some enumerative results in the theory of forms", *Proc. Cambridge Philos. Soc.* **39** (1943), 22–30. MR 4,184e Zbl 0060.04107

[Kleiman 1974] S. L. Kleiman, "The transversality of a general translate", *Compositio Math.* **28** (1974), 287–297. MR 50 #13063 Zbl 0288.14014

[Kontsevich 1995] M. Kontsevich, "Enumeration of rational curves via torus actions", pp. 335–368 in *The moduli space of curves* (Texel, Netherland, 1994), edited by R. Dijkgraaf et al., Progr. Math. **129**, Birkhäuser, Boston, 1995. MR 97d:14077 Zbl 0885.14028

[Kresch and Tamvakis 2003] A. Kresch and H. Tamvakis, "Quantum cohomology of the Lagrangian Grassmannian", *J. Algebraic Geom.* **12**:4 (2003), 777–810. MR 2004h:14063 Zbl 1051.53070

[Kuznetsov 1997] A. Kuznetsov, "Laumon's resolution of Drinfel′d's compactification is small", *Math. Res. Lett.* **4**:2-3 (1997), 349–364. MR 98d:14022 Zbl 0910.14026

[Lakshmibai 2003] V. Lakshmibai, "The development of standard monomial theory, II", pp. 283–309 in *A tribute to C. S. Seshadri* (Chennai, 2002), edited by V. Lakshmibai et al., Birkhäuser, Basel, 2003. MR 2004m:14107b Zbl 1056.14066

[Laumon 1990] G. Laumon, "Faisceaux automorphes liés aux séries d'Eisenstein", pp. 227–281 in *Automorphic forms, Shimura varieties, and L-functions* (Ann Arbor, MI, 1988), vol. I, edited by L. Clozel and J. S. Milne, Perspect. Math. **10**, Academic Press, Boston, 1990. MR 91k:11106 Zbl 0773.11032

[Littelmann 1998] P. Littelmann, "Contracting modules and standard monomial theory for symmetrizable Kac–Moody algebras", *J. Amer. Math. Soc.* **11**:3 (1998), 551–567. MR 99d:17027 Zbl 0915.20022

[Musili 2003] C. Musili, "The development of standard monomial theory, I", pp. 385–420 in *A tribute to C. S. Seshadri* (Chennai, 2002), edited by V. Lakshmibai et al., Birkhäuser, Basel, 2003. MR 2004m:14107a Zbl 1056.14067

[Oprea 2006] D. Oprea, "The tautological rings of the moduli spaces of stable maps to flag varieties", *J. Algebraic Geom.* **15**:4 (2006), 623–655. MR 2007i:14015 Zbl 1117.14056

[Rosenthal 1994] J. Rosenthal, "On dynamic feedback compensation and compactification of systems", *SIAM J. Control Optim.* **32**:1 (1994), 279–296. MR 95b:93084 Zbl 0797.93018

[Sottile 2000] F. Sottile, "Real rational curves in Grassmannians", *J. Amer. Math. Soc.* **13**:2 (2000), 333–341. MR 2000j:14080 Zbl 0946.14035

[Sottile 2001] F. Sottile, "Rational curves on Grassmannians: systems theory, reality, and transversality", pp. 9–42 in *Advances in algebraic geometry motivated by physics* (Lowell, MA, 2000), edited by E. Previato, Contemp. Math. **276**, Amer. Math. Soc., Providence, RI, 2001. MR 2002d:14080 Zbl 1095.14503

[Sottile and Sturmfels 2001] F. Sottile and B. Sturmfels, "A sagbi basis for the quantum Grassmannian", *J. Pure Appl. Algebra* **158**:2-3 (2001), 347–366. MR 2002a:13027 Zbl 0992.14023

[Strømme 1987] S. A. Strømme, "On parametrized rational curves in Grassmann varieties", pp. 251–272 in *Space curves* (Rocca di Papa, 1985), edited by F. Ghione et al., Lecture Notes in Math. **1266**, Springer, Berlin, 1987. MR 88i:14020 Zbl 0625.14027

[Weyman 2003] J. Weyman, *Cohomology of vector bundles and syzygies*, Cambridge Tracts in Mathematics **149**, Cambridge University Press, Cambridge, 2003. MR 2004d:13020 Zbl 1075.13007

Communicated by Bernd Sturmfels
Received 2008-06-04 Revised 2008-07-31 Accepted 2008-09-12

ruffojv@oneonta.edu *Department of Mathematics, Science, and Statistics, State University of New York - College at Oneonta, Oneonta, NY 13820, United States*
http://employees.oneonta.edu/ruffojv

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@mathscipub.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 2 No. 7 2008

1937-0652(2008)2:7;1-9