Algebra & Number Theory

Volume 8

2014

No. 7

msp

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2014 is US $225/year for the electronic version, and $400/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscribers address should be sent to MSP.

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

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1539

Double Dirichlet series and quantum unique ergodicity of weight one-half Eisenstein series

Yiannis N. Petridis, Nicole Raulf and Morten S. Risager

The problem of quantum unique ergodicity (QUE) of weight $\frac{1}{2}$ Eisenstein series for $\Gamma_0(4)$ leads to the study of certain double Dirichlet series involving GL_2 automorphic forms and Dirichlet characters. We study the analytic properties of this family of double Dirichlet series (analytic continuation, convexity estimate) and prove that a subconvex estimate implies the QUE result.

1. Introduction

An important problem of quantum chaos is to describe the behavior of eigenfunctions of Laplacians ϕ_λ with eigenvalue λ, as $\lambda \to \infty$. This problem has a rich and interesting history; see [Shnirelman 1974; Zelditch 1987; Colin de Verdière 1985; Zelditch 1992; Lindenstrauss 2006; Soundararajan 2010a], for example. For the weight 0 Eisenstein series $E(z, s)$ on the surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, Luo and Sarnak [1995] determined the asymptotic behavior of the measures

$$d\mu_t(z) = |E(z, \tfrac{1}{2} + it)|^2 \, d\mu(z)$$

on compact sets. Here $d\mu(z) = dx\, dy/y^2$ denotes the volume element corresponding to the hyperbolic metric on the upper half-plane $\mathbb{H}$. The main input in doing so was subconvex bounds on certain standard GL_1 and GL_2 L-functions, namely the Riemann zeta function and the L-function of a Maaß cusp form. Their work was later generalized to the corresponding micro-local lifts [Jakobson 1994] and other arithmetic symmetric spaces [Koyama 2000; Truelsen 2011]. Also for these generalizations, subconvex bounds were at the heart of the proofs. In [Petridis et al. 2013] we studied similar questions for scattering states.

In this paper we study the analogous problem for Eisenstein series of weight $\frac{1}{2}$. To be precise: Let $E(z, s, \frac{1}{2})$ be the weight $\frac{1}{2}$ Eisenstein series at the cusp infinity for

Petridis was supported by NSF grant DMS-0401318. Raulf was partly supported by the Labex CEMPI (ANR-11-LABX-0007-01). Risager was supported by a Sapere Aude grant from the Danish Natural Science Research Council.
MSC2010: primary 58J51; secondary 11F68, 11F72.
Keywords: quantum limits, Eisenstein series, multiple Dirichlet series.

the group $\Gamma = \Gamma_0(4)$ (see Section 3). We study the limiting behavior as $|t| \to \infty$ of

$$d\mu_t(z) = |E(z, \tfrac{1}{2} + it, \tfrac{1}{2})|^2 \, d\mu(z). \tag{1-1}$$

Since the Fourier coefficients $\phi_n(s, \frac{1}{2})$ of $E(z, \frac{1}{2} + it, \frac{1}{2})$ are essentially values of Dirichlet L-functions on the critical line — see (3-3) — and, therefore, are not multiplicative, the problem is much harder. The Rankin–Selberg convolutions that appear are not factored into standard L-functions. Instead, we find that certain double Dirichlet series play a crucial role. The relevant double Dirichlet series are the following.

Let χ, χ' be characters mod 8, and let t_n be either the eigenvalue of the Hecke operator T_n for a weight 0 Maaß form ψ on $\Gamma_0(4)\backslash\mathbb{H}$ or $t_n = \tau(n)$ be the divisor function. Let $s_0(1 - s_0)$ be the corresponding Laplace eigenvalue of ψ, with $\Re(s_0) \geq \frac{1}{2}$, and if $t_n = \tau(n)$ let $s_0 = \frac{1}{2}$.

We then define

$$Z(s, w, \chi, \chi') = \zeta_2(4s - 1) \sum_{\substack{n=1 \\ (n,2)=1}}^{\infty} \frac{\chi(n) t_n L^*(2w - \frac{1}{2}, n, \chi')}{n^{s-w+\frac{1}{2}}}, \tag{1-2}$$

where $L^*(w, n, \chi) = q(w, n, \chi) L_2(w, \chi_{n_0}\chi)$. Here n_0 is the squarefree part of n, $\chi_{n_0}(c) = \left(\frac{n_0}{c}\right)$ and $L_2(w, \chi_{n_0}\chi)$ is the standard L-function with the 2-factor removed. The functions $q(w, n, \chi)$ are explicitly given so-called "correction polynomials"; see (2-7) below. The function $L^*(w, n, \chi)$ may seem strange at first, but it occurs naturally as the n-th Fourier coefficient of the Eisenstein series of weight $\frac{1}{2}$, and it has many nice properties. See, for example, [Shimura 1973] or Section 3 below.

Friedberg and Hoffstein [1995] have studied a Rankin–Selberg integral (see (3-13) below) which turns out to be a linear combination of $Z(s, w, \chi, \chi')$ and $Z(s, w, \chi, \chi_4\chi')$, where χ_4 is the primitive character mod 4. They observed that this admits meromorphic continuation and that certain linear combinations have a pole at $(s, w) = (\frac{3}{4}, \frac{3}{4})$ (in our normalization). They did this in order to prove nonvanishing of quadratic twists of GL_2-L-functions at the central point.

Furthermore similar series with higher-order twists instead of the quadratic characters χ_{n_0} were studied by Brubaker, Bucur, Chinta, Frechette and Hoffstein [Brubaker et al. 2004] in order to prove nonvanishing of higher-order twists. To understand the new series $Z(s, w, \chi, \chi')$ we follow essentially the program introduced in [Bump et al. 1996] to prove the following.

The series defining $Z(s, w, \chi, \chi')$ converges absolutely and uniformly in certain regions in $\mathbb{C}^2$, and hence defines an analytic function there. The functions $Z(s, w, \chi, \chi')$ admit meromorphic continuation to $\mathbb{C}^2$ and they satisfy a group of functional equations generated by

$$\alpha : (s, w) \mapsto (s, 1 - w), \quad \beta : (s, w) \mapsto (w, s).$$

The functions $Z(s, w, \chi, \chi')$ grow at most polynomially for $(\Re(s), \Re(w))$ in compact sets. For the precise form of the functional equations we refer to Theorems 2.11 and 2.13. The group of functional equations is isomorphic to the dihedral group of order 8. A similar result for higher-order twists may be found in [Brubaker et al. 2004].

We want to investigate the growth of $Z(s, w, \chi, \chi')$ in s and w. The notions of analytic conductor and subconvexity are not completely well established for general multiple Dirichlet series. Certain cases are dealt with in [Blomer 2011; Blomer et al. 2014] but a general theory is missing.

To define these notions in the present case we note that when $\Re(s), \Re(w) > \frac{3}{4}$ the function $Z(s, w, \chi, \chi')$ has a representation

$$Z(s, w, \chi, \chi') = \sum_{\substack{c=1 \\ (c,2)=1}}^{\infty} \frac{\chi'(c) L^{**}(s - w + \frac{1}{2}, \psi, c, \chi)}{c^{2w-1/2}}, \tag{1-3}$$

where $L^{**}(s, \psi, c, \chi) = Q^*(s, c, \chi) L_2(s, \psi \otimes \tilde{\chi}_{c_0} \chi)$ (see (2-19) and Theorem 2.13). Here c_0 is the squarefree part of c, $\tilde{\chi}_{c_0}(n) = \left(\frac{n}{c_0}\right)$ and $L_2(s, \psi \otimes \tilde{\chi}_{c_0} \chi)$ is the standard L-function with the 2-factor removed. The functions $Q^*(s, c, \chi)$ are explicitly given so-called "correction polynomials"; see (2-20) below.

When proving bounds on standard L-functions one usually normalizes the coefficients to be essentially bounded, at least on average. In our case it is not so clear how to do that since the true size of $L^{**}(s, \psi, c, \chi)$ is known only conjecturally. If the generalized Lindelöf hypothesis is true the coefficients of the series (1-3) are essentially bounded. We investigate what happens when this is true on average (over c). To be precise: we want to know what bound on $Z(s, w, \chi, \chi')$ can be proved if we assume that the coefficients are essentially bounded, i.e., if

$$\sum_{\substack{c \le X \\ (c,2)=1}} |L^{**}(s, \psi, c, \chi)| = O(X^{1+\epsilon}(1 + |s|)^{\epsilon}). \tag{1-4}$$

Using the properties of $Q^*(s, c, \chi)$ we will see that this follows from assuming

$$\sum_{\substack{1 \le c_0 \le X \\ c_0 \text{ odd,} \\ \text{squarefree}}} |L_2(s, \psi \otimes \tilde{\chi}_{c_0} \chi)|^2 = O(X^{1+\epsilon}(1 + |s|)^{\epsilon}) \quad \text{when } \Re(s) = \tfrac{1}{2}. \tag{1-5}$$

Also, it is easy to see that (1-4) implies (1-5) with the exponent 2 replaced by a 1. In particular it implies the generalized Lindelöf hypothesis in the t parameter.

We now define the analytic conductor of $Z(\frac{1}{2} + it, \frac{1}{2} + iu, \chi, \chi')$ to be

$$\mathfrak{q}(t, u) = (1 + |t|)(1 + |t + u|)^2 (1 + |u|). \tag{1-6}$$

Using an approximate functional equation argument for $Z(s, w, \chi, \chi')$ we can prove the following bound on the critical line.

Theorem 1.1. *Assume* (1-4). *Then*

$$Z(\tfrac{1}{2}+it, \tfrac{1}{2}+iu, \chi, \chi') = O_\psi(\mathfrak{q}(t, u)^{\frac{1}{4}+\varepsilon}). \tag{1-7}$$

Unconditionally,

$$Z(\tfrac{1}{2}+it, \tfrac{1}{2}+iu, \chi, \chi') = O_\psi\big((\mathfrak{q}(t, u)(1+|t-u|)^2)^{\frac{1}{4}+\varepsilon}\big). \tag{1-8}$$

Remark 1.2. We call the unconditional bound (1-8) the *trivial* bound. The conditional bound (1-7) is called the *convexity* bound. Any bound $O(\mathfrak{q}(t, u)^{\frac{1}{4}-\delta})$ with $\delta > 0$ is called a *subconvex* bound with saving δ. If $\delta = \frac{1}{4} - \varepsilon$ is permitted, we say that $Z(s, w, \chi, \chi')$ admits a Lindelöf-type bound. In the theory of L-functions, the notion of convexity and subconvexity is standard and has numerous applications; see, e.g., [Iwaniec and Kowalski 2004].

Remark 1.3. We note that even proving the *trivial* bound requires strong input. In particular, in order to prove Theorem 1.1 (1-8), we need the Lindelöf hypothesis on average in the conductor aspect for $L(s, \chi_n)$, and the convexity estimate in the s aspect. This bound is available, as follows from Heath-Brown's famous large sieve inequality for quadratic characters (2-28); see (2-29) below.

Also, we note that we can prove unconditionally (see Lemma 3.2 below) that, if $\{t_n\}$ comes from a cusp form,

$$Z(\tfrac{1}{2}+it, \tfrac{1}{2}-it, \chi, \chi') + bZ(\tfrac{1}{2}+it, \tfrac{1}{2}-it, \chi, \chi') = O_\psi(\mathfrak{q}(t, -t)^{\frac{1}{4}+\varepsilon}).$$

Here b is the product of the sign of χ and the sign of the cusp form. We note that this is of the same order as the convexity estimate above *without* assuming (1-4).

Remark 1.4. For special configurations of s, w (in our case $s-w$ constant) the trivial bound and the convexity bound coincide. This is because, in this case, (1-5) follows from Heath-Brown's estimate (2-30).

We emphasize that our notion of convexity is *different* from that of Blomer, Goldmakher and Louvel [Blomer 2011; Blomer et al. 2014]. What we call the trivial bound corresponds to what they call the convexity bound.

Remark 1.5. Even though we cannot prove it, it is not unreasonable to expect subconvexity for $Z(s, w, \chi, \chi')$! Double Dirichlet series similar to $Z(s, w, \chi, \chi')$ — with degree-one L-functions as coefficients — *are* known to satisfy subconvex bounds [Blomer 2011; Blomer et al. 2014]. (Blomer et al. [2014] consider a configuration such that the bound they prove would be considered a subconvex bound also by our definition. Likewise the bound proved in [Blomer 2011, Theorem 1] is a subconvex bound by our definition if one restricts to $s = \frac{1}{2}$ or $w = \frac{1}{2}$.)

Furthermore, it is known that on average the double Dirichlet series considered by Blomer admits Lindelöf-type bounds [Blomer 2011, Theorem 2] in the (s, w) aspect. In the conductor aspect (which is here the conductor related to the form with eigenvalues $\{t_n\}$), Hoffstein and Kontorovich [2010, (1.23)] conjecture Lindelöf-type bounds to hold.

Theorem 1.6. *Assume that, for all χ, χ', $\{t_n\}$ the function $Z(s, w, \chi, \chi')$ admits a subconvex bound. Then, for any compact Jordan measurable subsets A and B of $\Gamma\backslash\mathbb{H}$, we have*

$$\frac{\int_A |E(z, \frac{1}{2}+it, \frac{1}{2})|^2 \, d\mu(z)}{\int_B |E(z, \frac{1}{2}+it, \frac{1}{2})|^2 \, d\mu(z)} \to \frac{\mathrm{vol}(A)}{\mathrm{vol}(B)} \quad \text{as } |t| \to \infty. \tag{1-9}$$

Remark 1.7. Theorem 1.6 is the analogue of the Luo–Sarnak theorem [1995] for the weight 0 Eisenstein series. Their theorem, however, is unconditional, as in their case subconvex bounds for standard GL_1 and GL_2-L-functions are readily available. As in that paper, we really prove — conditionally on any subconvex bound — the asymptotic result

$$\int_A |E(z, \tfrac{1}{2}+it, \tfrac{1}{2})|^2 \, d\mu(z) \sim \frac{4}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} \mathrm{vol}(A) \log|t| \quad \text{as } |t| \to \infty. \tag{1-10}$$

In contrast to the case of quantum unique ergodicity of Maaß cusp forms, the rate of convergence in (1-10) is very slow. As in [Luo and Sarnak 1995] one can prove $O(\log t / \log\log t)$.

It is understood in many arithmetic cases that the equidistribution of masses is implied by subconvexity bounds for appropriate L-functions of degree 8; see, e.g., [Sarnak 2011; Soundararajan 2010b; Nelson et al. 2014].

Remark 1.8. The structure of the paper is as follows. In Section 2 we study the double Dirichlet series $Z(s, w, \chi, \chi')$ which arise when we address QUE of the weight $\frac{1}{2}$ Eisenstein series $E(z, s, \frac{1}{2})$. In Section 3 we review the theory for $E(z, s, \frac{1}{2})$ with explicit computations. In Section 4, which is the main section of the paper, we analyze (1-10) by splitting it into a cuspidal contribution and incomplete Eisenstein series contributions. For example, in the cuspidal space we find that, for a cusp form ψ with eigenvalue $s_0(1-s_0)$, the integral

$$\int_{\Gamma\backslash\mathbb{H}} \psi(z) |E(z, \tfrac{1}{2}+it, \tfrac{1}{2})|^2 \, d\mu(z) \tag{1-11}$$

equals a linear combination of terms of the form

$$c_{\chi,\chi'_\pm}(s, w) Z(s, w, \chi, \chi') \frac{1}{\Gamma(w \pm \frac{1}{4})} \int_0^\infty W_{0,s_0-\frac{1}{2}}(2y) W_{\pm\frac{1}{4},w-\frac{1}{2}}(2y) y^{s-1} \frac{dy}{y} \tag{1-12}$$

evaluated at $(s, w) = (\frac{1}{2} + it, \frac{1}{2} - it)$. Here $c_{\chi,\chi'_\pm}(s, w)$ are functions which can easily be understood when $\Re(w) = \Re(s) = \frac{1}{2}$, and $W_{\mu,\nu}$ are Whittaker functions. In the Appendix we analyze the Mellin transform of the product of Whittaker functions.

We can then deal with (1-12) using bounds on $Z(s, w, \chi, \chi')$. To deal with the cuspidal space we need subconvexity for $Z(\frac{1}{2} + it, \frac{1}{2} - it, \chi, \chi')$, with t_n corresponding to Hecke eigenvalues for Maaß forms. For the incomplete Eisenstein series a similar analysis shows that we need the same type of bound for $t_n = \tau(n)$, the divisor function, for all configurations of s and w. We also use Zagier's theory of Rankin–Selberg integrals for functions not of rapid decay.

Remark 1.9. Although the analytic continuation of

$$I(s, w) = \int_{\Gamma\backslash\mathbb{H}} \psi(z) E(z, w, \tfrac{1}{2}) \overline{E(z, \bar{s}, \tfrac{1}{2})}\, d\mu(z)$$

(of which (1-11) is a special case) follows from the well-known analytic properties of $E(z, w, \frac{1}{2})$, its growth/decay properties jointly in (s, w) are less clear. This is why we have to unfold and eventually analyze $Z(s, w, \chi, \chi')$ to see that the above integral is $O(|t|^{-\delta})$ for $s = 1 - w = \frac{1}{2} + it$ when $|t| \to \infty$, assuming subconvexity with saving δ. The Maaß–Selberg relation gives an upper bound (see, e.g., (3-15) below), but this is not good enough to prove Theorem 1.6.

Remark 1.10. One could speculate whether the implication in Theorem 1.6 could be reversed, i.e., to what extent bounds on integrals like (1-11) would imply bounds on $Z(s, w, \chi, \chi')$ via the expression (1-12). Such speculation is problematic at least for the following reason. We have good control over the asymptotics of the Mellin transform (see, e.g., Lemma A.1) but since integrals like (1-11) are *linear combinations* of terms of the form (1-12), we cannot conclude from bounds on integrals like (1-11) the same bounds on the individual summands. We elaborate on this in Lemma 3.2 and Remark 3.3 below.

2. A double Dirichlet series

In this section we define and prove various properties of the double Dirichlet series. To derive its meromorphic continuation and functional equation we proceed as in [Brubaker et al. 2004], but with some simplifications and refinements. We show, for instance, that knowing optimal bounds towards the Ramanujan–Petersson conjecture is not necessary to get optimal regions of convergence. To prove the convexity bounds we use a combination of techniques from [Blomer 2011; Blomer et al. 2014]. Although the techniques we use are certainly known to the experts in the field, we were not able to find precise enough statements in the existing literature for the double Dirichlet series (1-2).

We start by introducing some notation and deriving some basic results about Gauss sums and Dirichlet series involving Gauss sums.

Let $\{t_n\}_{n\in\mathbb{N}}$ be the coefficients of the normalized L-function of a self-dual GL_2 automorphic form ψ. For good primes — and we assume that only $p=2$ could potentially be a bad prime — the Satake parameters α_p, β_p satisfy $\alpha_p+\beta_p=t_p$, $\alpha_p\cdot\beta_p=1$ and

$$t_{p^\lambda}=\sum_{j=0}^{\lambda}\alpha_p^j\beta_p^{\lambda-j}=\frac{\alpha_p^{\lambda+1}-\beta_p^{\lambda+1}}{\alpha_p-\beta_p}. \tag{2-1}$$

The Fourier coefficients satisfy the Ramanujan–Petersson conjecture on average, since the Rankin–Selberg method gives

$$\sum_{|n|\le X}|t_n|^2\sim CX \tag{2-2}$$

as $X\to\infty$. Here C is an explicit constant; see, e.g., [Iwaniec 2002, (8.15)]. The corresponding p-factor, i.e., the local L-function, is given by

$$L^{(p)}(s,\psi)=\sum_{\lambda=0}^{\infty}\frac{t_{p^\lambda}}{p^{\lambda s}}=(1-t_pp^{-s}+p^{-2s})^{-1}=(1-\alpha_pp^{-s})^{-1}(1-\beta_pp^{-s})^{-1}.$$

Similar but easier identities and estimates are true for the divisor function $t_n=\tau(n)$, where $\alpha_p=\beta_p=1$.

For any L-function we will write $L^{(p)}(s)$ for its corresponding p-factor and $L_2(s)$ for the L-function with the 2-factor removed.

2A. ***Gauss sums and some related series.*** We now recall a few basic relevant results about Gauss sums for real characters. Let n, d be integers with d odd and positive and let $\left(\frac{n}{d}\right)$ be the Jacobi–Legendre symbol

$$\left(\frac{n}{d}\right)=\prod_{p^v\|d}\left(\frac{n}{p}\right)^v,$$

where for an odd prime p we denote by $\left(\frac{n}{p}\right)$ the usual Legendre symbol. The symbol $\left(\frac{n}{d}\right)$ is then extended to all odd $d\in\mathbb{Z}$ as in [Shimura 1973, p. 442]; see also [Koblitz 1984, p. 147, 187–188].

For an integer n and a positive odd integer d we define Gauss sums

$$G_n(d):=\sum_{m \bmod d}\left(\frac{m}{d}\right)e\left(\frac{nm}{d}\right). \tag{2-3}$$

Here $e(x)=e^{2\pi ix}$. Gauss ingeniously proved that for odd squarefree d we have $G_1(d)=\varepsilon_d\sqrt{d}$, where $\varepsilon_d=1$ if $d\equiv 1 \pmod 4$ and $\varepsilon_d=i$ if $d\equiv -1 \pmod 4$.

Quadratic reciprocity states that for relatively prime odd positive integers n, d,

$$\left(\frac{n}{d}\right)\left(\frac{d}{n}\right)=(-1)^{\frac{n-1}{2}\frac{d-1}{2}}. \tag{2-4}$$

It is elementary to verify that the right-hand side equals $\varepsilon_n\varepsilon_d/\varepsilon_{nd}$. For odd d it turns out to be convenient to consider

$$H_n(d):=\varepsilon_d^{-1}G_n(d).$$

Proposition 2.1. *The function $H_n(d)$ has the following properties*:

(1) *For fixed n, $H_n(d)$ is multiplicative*, *i.e.*, *if d_1, d_2 are coprime odd positive integers*, *then*

$$H_n(d_1d_2)=H_n(d_1)H_n(d_2).$$

(2) *If* $(n_1,d)=1$, *then*

$$H_{n_1n_2}(d)=\left(\frac{n_1}{d}\right)H_{n_2}(d).$$

(3) *Let α, β be nonnegative integers and let p be an odd prime. Then*

$$H_{p^\alpha}(p^\beta)=\begin{cases}\phi(p^\beta) & \text{if } \alpha\ge\beta,\ \beta\equiv 0 \pmod 2,\\ p^{\beta-\frac12}(\delta_{\beta\equiv 1 \,(\mathrm{mod}\, 2)}-p^{-\frac12}\delta_{\beta\equiv 0\,(\mathrm{mod}\,2)}) & \text{if } \alpha=\beta-1,\\ 0 & \text{otherwise.}\end{cases}$$

Proof. (1) follows from the Chinese remainder theorem and quadratic reciprocity; (2) from the fact that if $(n_1,d)=1$ then n_1m runs through a set of representatives mod d; and (3) from elementary considerations. □

We now compute

$$\sum_{\substack{c=1\\(c,2)=1}}^{\infty}\frac{\chi(c)H_n(c)}{c^{2s}} \quad\text{and}\quad \sum_{\substack{n=1\\(n,2)=1}}^{\infty}\frac{t_n\chi(n)H_n(c)}{n^s}, \tag{2-5}$$

where χ is a character mod q with $q|8$. As we shall see later these sums occur naturally in the Fourier coefficients of the weight $\frac12$ Eisenstein series of $\Gamma_0(4)$, and in Rankin–Selberg-type integrals formed from these Eisenstein series.

For n odd and positive we denote

$$\tilde{\chi}_n(c)=\left(\frac{c}{n}\right),$$

which is a character mod n. When n is squarefree its conductor is n.

For c odd we denote

$$\chi_n(c)=\left(\frac{n}{c}\right),$$

which for n odd and squarefree has an extension to all c which is a character of conductor $|n|$ if $n \equiv 1 \pmod 4$ and $4|n|$ if $n \equiv 3 \pmod 4$. See [Koblitz 1984, p. 147, 187–188].

By quadratic reciprocity (2-4) we have, for odd positive m, n,

$$\chi_n(m) = \tilde{\chi}_n(m) \begin{cases} 1 & \text{if } n \equiv 1 \pmod 4, \\ \chi_4(m) & \text{if } n \equiv 3 \pmod 4, \end{cases} \tag{2-6}$$

where χ_4 is the primitive character mod 4. We can write any nonzero integer n uniquely as $n = n_0 n_1^2$, where n_0 is squarefree and $n_1 > 0$. We define correction polynomials as

$$q(s, n, \chi) = \prod_{2 \neq p | n_1} \sum_{\beta=0}^{v_p(n_1)} \frac{1 - \delta_{\beta < v_p(n_1)} \chi_{n_0}(p) \chi(p) p^{-s}}{p^{2\beta(s-\frac{1}{2})}}, \tag{2-7}$$

where v_p is the p-adic valuation. For $\chi = 1$, we sometimes write $q(s, n) = q(s, n, \chi)$.

We define

$$L^*(s, n, \chi) = q(s, n, \chi) L_2(s, \chi_{n_0} \chi). \tag{2-8}$$

Lemma 2.2. *We have*

$$\sum_{\substack{c=1 \\ (c,2)=1}}^{\infty} \frac{\chi(c) H_n(c)}{c^{2s}} = \frac{L^*(2s - \frac{1}{2}, n, \chi)}{\zeta_2(4s-1)}.$$

Proof. Using multiplicativity of $H_n(d)$ (Proposition 2.1) we see that the sum factors into local factors. For a prime $p \neq 2$ we compute the corresponding factor

$$R_p(s) = \sum_{\beta=0}^{\infty} \frac{\chi(p^\beta) H_n(p^\beta)}{p^{\beta 2s}}.$$

Write $n = n' p^\alpha$, where $(n', p) = 1$. Then using Proposition 2.1 (2), (3) we have

$$R_p(s) = \sum_{\beta=0}^{\infty} \frac{\left(\frac{n'}{p^\beta}\right) \chi(p^\beta) H_{p^\alpha}(p^\beta)}{p^{\beta 2s}} = \sum_{\substack{\beta=0 \\ \beta \text{ even}}}^{\alpha} \frac{\phi(p^\beta)}{p^{\beta 2s}} + \frac{\left(\frac{n'}{p^{\alpha+1}}\right) \chi(p^{\alpha+1}) H_{p^\alpha}(p^{\alpha+1})}{p^{(\alpha+1)2s}} \tag{2-9}$$

Consider first α even, in which case $\alpha = 2v_p(n_1)$. Then we find

$$R_p(s) = 1 + \sum_{\substack{\beta=1 \\ \beta \text{ even}}}^{\alpha} \frac{p^{\beta-1}(p-1)}{p^{\beta 2s}} + \frac{\chi_{n_0}(p) \chi(p) p^{\alpha+\frac{1}{2}}}{p^{(\alpha+1)2s}},$$

noting that $\chi_{n_0}(p) = \left(\frac{n'}{p}\right)$. By induction we find

$$R_p(s) = \frac{L^{(p)}(2s-\frac{1}{2}, \chi_{n_0}\chi)}{\zeta^{(p)}(4s-1)} \Bigg(\sum_{\substack{\beta=0 \\ \beta \text{ even}}}^{\alpha} p^{\beta(1-2s)} - \sum_{\substack{\beta=0 \\ \beta \text{ even}}}^{\alpha-2} \chi_{n_0}(p)\chi(p) p^{-(2s-\frac{1}{2})} p^{\beta(1-2s)} \Bigg).$$

Here we have used $\chi_{n_0}^2(p) = 1$.

Returning to (2-9), we assume instead that α is odd, in which case $\alpha - 1 = 2v_p(n_1)$. We find that in this case

$$\begin{aligned} R_p(s) &= 1 + \sum_{\substack{\beta=1 \\ \beta \text{ even}}}^{\alpha} \frac{p^{\beta-1}(p-1)}{p^{\beta 2s}} + \frac{-p^{\alpha}}{p^{(\alpha+1)2s}} \\ &= (1 - p^{-(4s-1)}) \sum_{\substack{\beta=0 \\ \beta \text{ even}}}^{\alpha-1} p^{\beta(1-2s)}, \end{aligned}$$

where again we have used induction. Using that, for α odd, $\chi_{n_0}(p) = 0$, we may write this as

$$R_p(s) = \frac{L^{(p)}(2s-\frac{1}{2}, \chi_{n_0}\chi)}{\zeta^{(p)}(4s-1)} \sum_{\substack{\beta=0 \\ \beta \text{ even}}}^{\alpha-1} p^{\beta(1-2s)}.$$

Since $\chi_{n_0}(p) = 0$, we arrive at the desired result. □

Proposition 2.3. *The function $q(s, n, \chi)$ has the following properties*:

(1) *If n is squarefree*, *then $q(s, n, \chi) = 1$.*

(2) *If $n = n_0 n_1^2$ with n_0 squarefree and n_0, n_1 odd*, *then*

$$q(s, n, \chi) = (n_1^2)^{\frac{1}{2}-s} q(1-s, n, \chi).$$

(3) *If $\Re(s) \geq \frac{1}{2}$*, *then $q(s, n, \chi) = O(n^{\epsilon})$ uniformly in $\Re(s)$.*

Proof. These statements are all straightforward to verify from the definition. (1) is clear and (2) is easily verified by considering factors. Trivial estimates for $\Re(s) \geq \frac{1}{2}$ lead to $|q(s, n, \chi)| \leq 2^{\#\{p|n\}}\tau(n)$, which gives (3). □

Write $c = c_0 c_1^2$ with c_0 squarefree and set $v = v_p(c_1)$. We then define, for odd c,

$$Q_{\psi}(s, c, \chi) = \prod_{p|c_1} \frac{t_{p^{2v}} - t_{p^{2v-1}} \tilde{\chi}_{c_0}(p)\chi(p)(p^{1-s} + p^s)/p + t_{p^{2v-2}} \tilde{\chi}_{c_0}(p)^2/p}{p^{2v(s-\frac{1}{2})}}. \tag{2-10}$$

Since ψ is fixed, we shall often omit it from the notation and simply write $Q(s, c, \chi)$. We define

$$L^*(s, c, \psi, \chi) := Q_{\psi}(s, c, \chi) L_2(s, \psi \otimes \tilde{\chi}_{c_0}\chi). \tag{2-11}$$

Lemma 2.4. *Let c be an odd natural number. Then*

$$\sum_{\substack{n=1\\(n,2)=1}}^{\infty} \frac{t_n \chi(n) H_n(c)}{n^s} = \sqrt{c} L^*(s, c, \psi, \chi).$$

Proof. A similar computation can be found in [Brubaker et al. 2004, Section 3]. We first show that the Dirichlet series factors into local factors. For p an odd prime, write $c = c'p^l$ with $(c', p) = 1$, and $m = p^{v_p(m)} m/p^{v_p(m)}$. Then using Proposition 2.1 (1) and (2) we find

$$H_m(c) = \left(\frac{m/p^{v_p(m)}}{p^l}\right)\left(\frac{p^{v_p(m)}}{c'}\right) H_{p^{v_p(m)}}(p^l) H_{m/p^{v_p(m)}}(c').$$

Writing $m = np^\lambda$, we can write the Dirichlet series as

$$\sum_{\substack{n=1\\(n,2p)=1}}^{\infty} \sum_{\lambda=0}^{\infty} \frac{t_{np^\lambda} \chi(np^\lambda)}{(np^\lambda)^s} H_{np^\lambda}(c'p^l) = \sum_{\substack{n=1\\(n,2p)=1}}^{\infty} \frac{t_n \chi(n) H_n(c')\left(\frac{n}{p^l}\right)}{n^s} \left(\sum_{\lambda=0}^{\infty} \frac{t_{p^\lambda} \chi(p^\lambda)}{p^{\lambda s}} H_{p^\lambda}(p^l)\left(\frac{p^\lambda}{c'}\right)\right).$$

Repeating this argument for every prime p, it follows that the series factors as

$$\prod_{p \neq 2} \left(\sum_{\lambda=0}^{\infty} \frac{t_{p^\lambda}}{p^{\lambda s}} H_{p^\lambda}(p^{v_p(c)}) \left(\frac{p^\lambda}{c/p^{v_p(c)}}\right) \chi(p^\lambda)\right). \tag{2-12}$$

We now compute the local factors of (2-12), i.e., we compute, for $p \neq 2$,

$$\sum_{\lambda=0}^{\infty} \frac{t_{p^\lambda} \chi(p^\lambda)}{p^{\lambda s}} H_{p^\lambda}(p^l) \left(\frac{p^\lambda}{c'}\right), \tag{2-13}$$

where $l = v_p(c)$ and $c' = c/p^{v_p(c)}$. If $l = 0$ the sum reduces to

$$\sum_{\lambda=0}^{\infty} \frac{t_{p^\lambda} \chi(p^\lambda)}{p^{\lambda s}} \left(\frac{p^\lambda}{c}\right) = L^{(p)}(s, \psi \otimes \tilde{\chi}_{c_0} \chi),$$

where we have used that $\tilde{\chi}_c(p) = \tilde{\chi}_{c_0}(p)$ if $(p, c) = 1$. Here c_0 denotes the squarefree part of c.

If $l > 0$ is even we use Proposition 2.1 (3) to see that in this case (2-13) is equal to

$$\left(-\frac{t_{p^{l-1}} p^{l-1} \chi(p^{l-1})}{p^{(l-1)s}} \left(\frac{p^{l-1}}{c'}\right) + \sum_{\lambda=l}^{\infty} \frac{t_{p^\lambda} p^{l-1}(p-1) \chi(p^\lambda)}{p^{\lambda s}} \left(\frac{p^\lambda}{c'}\right)\right). \tag{2-14}$$

For t_n being a Hecke eigenvalue we can use the Satake parameters and evaluate the resulting geometric sums to see that

$$\begin{aligned}
&\sum_{\lambda=l}^{\infty}\frac{t_{p^\lambda}\chi(p^\lambda)}{p^{\lambda s}}\left(\frac{p^\lambda}{c'}\right)\\
&\quad=\frac{1}{\alpha_p-\beta_p}\sum_{\lambda=l}^{\infty}\frac{\alpha_p^{\lambda+1}-\beta_p^{\lambda+1}}{p^{\lambda s}}\chi(p^\lambda)\left(\frac{p^\lambda}{c'}\right)\\
&\quad=\frac{1}{\alpha_p-\beta_p}\left(\frac{\alpha_p^{l+1}}{p^{ls}}\left(1-\alpha_p\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right)^{-1}-\frac{\beta_p^{l+1}}{p^{ls}}\left(1-\beta_p\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right)^{-1}\right)\\
&\qquad\qquad\text{(where we have used } \left(\left(\tfrac{p}{c'}\right)\chi(p)\right)^l=1\text{)},\\
&\quad=\frac{L^{(p)}(s,\psi\otimes\tilde{\chi}_{c_0}\chi)}{p^{ls}}\\
&\qquad\cdot\frac{1}{\alpha_p-\beta_p}\left(\alpha_p^{l+1}\left(1-\beta_p\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right)-\beta_p^{l+1}\left(1-\alpha_p\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right)\right)\\
&\quad=\frac{L^{(p)}(s,\psi\otimes\tilde{\chi}_{c_0}\chi)}{p^{ls}}\left(t_{p^l}-t_{p^{l-1}}\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right).
\end{aligned}\tag{2-15}$$

This is also true when $t_n=\tau(n)$ by a similar computation, which we omit.

It follows that (2-14) can be written as

$$\begin{aligned}
&p^{l-1}\left[\frac{-t_{p^{l-1}}}{p^{(l-1)s}}\tilde{\chi}_{c_0}(p^{l-1})\chi(p^{l-1})\right.\\
&\qquad\qquad\left.+\frac{L^{(p)}(s,\psi\otimes\tilde{\chi}_{c_0}\chi)}{p^{ls}}(p-1)\left(t_{p^l}-t_{p^{l-1}}\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right)\right]\\
&=p^{l-1}\frac{L^{(p)}(s,\psi\otimes\tilde{\chi}_{c_0}\chi)}{p^{ls}}\left[\frac{-t_{p^{l-1}}}{p^{-s}}\tilde{\chi}_{c_0}(p^{l-1})\chi(p^{l-1})\left(1-t_p\tilde{\chi}_{c_0}(p)\chi(p)p^{-s}+p^{-2s}\right)\right.\\
&\qquad\qquad\left.+(p-1)\left(t_{p^l}-t_{p^{l-1}}\left(\frac{p}{c'}\right)\chi(p)p^{-s}\right)\right]\\
&=p^{l/2}\frac{L^{(p)}(s,\psi\otimes\tilde{\chi}_{c_0}\chi)}{p^{l(s-\frac{1}{2})+1}}\left[pt_{p^l}-t_{p^{l-1}}\tilde{\chi}_{c_0}(p)\chi(p)(p^{1-s}+p^s)+t_{p^{l-2}}\right],
\end{aligned}$$

using that the Hecke-eigenvalues satisfy $t_{p^{l-1}}t_p=t_{p^l}+t_{p^{l-2}}$.

If instead $l>0$ is odd we can again use Proposition 2.3 (3) and we find that in this case (2-13) is equal to

$$\frac{t_{p^{l-1}}}{p^{(l-1)s}}p^{l-\frac{1}{2}}\left(\frac{p^{l-1}}{c'}\right)\chi(p^{l-1})=\frac{t_{p^{l-1}}}{p^{(l-1)(s-1)-\frac{1}{2}}}.$$

We note also that $\tilde{\chi}_{c_0}(p) = \left(\frac{p}{c_0}\right) = 0$ since by l being odd we may conclude that c_0 is divisible by p. It follows that, in this case, $L^{(p)}(s, \psi \otimes \tilde{\chi}_{c_0}\chi) = 1$, and we conclude that (2-13) can be written as

$$\frac{p^{l/2} t_{p^{l-1}}}{p^{(l-1)(s-\frac{1}{2})}} L^{(p)}(s, \psi \otimes \tilde{\chi}_{c_0}\chi),$$

which gives the desired result in this case. □

Proposition 2.5. *The function $Q(s, c, \chi)$ has the following properties*:

(1) *If c is squarefree*, *then $Q(s, c, \chi) = 1$.*

(2) *If $c = c_0 c_1^2$ with c_0 squarefree and c_0, c_1 odd*, *then*

$$(c_1^2)^{1-2s} Q(1-s, c, \chi) = Q(s, c, \chi).$$

Proof. Statement (1) is clear and (2) is easily verified by considering factors. □

We would like to have bounds analogous to Proposition 2.3 (3). Any bound of the form $|t_{p^l}| \le \tau(p^l) p^{\theta l}$ implies that, when $\Re(s) \ge \frac{1}{2}$,

$$|Q(s, c, \chi)| \le \tau(c) 4^{\#\{p|c\}} c^{\theta} = O(c^{\theta+\epsilon}). \tag{2-16}$$

The Ramanujan–Petersson conjecture will give the strongest bound with $\theta = 0$. Since the Ramanujan–Petersson conjecture is true on average by (2-2), we can prove that $Q(s, c, \chi)$ is bounded on average:

Lemma 2.6. *For $\Re(s) \ge \frac{1}{2}$ we have*

$$\sum_{\substack{c \le X, \\ c \text{ odd}}} |Q(s, c, \chi)|^2 = O(X^{1+\varepsilon})$$

uniformly in s.

Proof. Write $c = c_0 c_1^2$ with c_0 squarefree and c odd. It is easy to see that

$$\begin{aligned} |Q(s, c, \chi)| &\le \prod_{p|c_1} \left(\left| t_{p^{2v_p(c_1)}} \right| + 2\left| t_{p^{2v_p(c_1)-1}} \right| + \left| t_{p^{2v_p(c_1)-2}} \right| \right) \\ &\le \prod_{p|c_1} 4 \max_{i=0,1,2} \left| t_{p^{2v_p(c_1)-i}} \right| \\ &= 4^{\#\{p|c_1\}} |t_{d_0}|, \quad \text{where } d_0 \text{ is some divisor of } c_1^2. \end{aligned}$$

It follows that

$$|Q(s, c, \chi)|^2 \le 16^{\#\{p|c_1\}} |t_{d_0}|^2 \le 16^{\#\{p|c\}} \sum_{d|c} |t_d|^2.$$

Using the Ramanujan–Petersson conjecture on average, (2-2), and $16^{\#\{p|c\}} = O(c^{\varepsilon})$, we find

$$\begin{aligned}\sum_{c\le X}|Q(s,c,\chi)|^2 &= O\biggl(X^{\varepsilon}\sum_{c\le X}\sum_{d|c}|t_d|^2\biggr)\\ &= O\biggl(X^{\varepsilon}\sum_{d\le X}|t_d|^2\#\{c\le X\mid d \text{ divides } c\}\biggr)\\ &= O\biggl(X^{1+\varepsilon}\sum_{d\le X}\frac{|t_d|^2}{d}\biggr) = O(X^{1+\varepsilon}). \qquad \square\end{aligned}$$

We are now ready to define the double Dirichlet series. Let χ_4 be the primitive character mod 4, $\chi_4(n) = \left(\frac{-1}{n}\right) = (-1)^{(n-1)/2}$ for $(n,2)=1$, and let χ_8 be the primitive character mod 8 given by $\chi_8(n) = \left(\frac{2}{n}\right) = (-1)^{\frac{1}{8}(n-1)(n+1)}$ for $(n,2)=1$. Let χ, χ' be characters mod 8, i.e., χ, χ' are induced from 1, χ_4, χ_8, or $\chi_4\chi_8$. We then define

$$Z(s,w,\chi,\chi') = \zeta_2(4s-1)\sum_{\substack{n=1\\(n,2)=1}}^{\infty}\frac{\chi(n)t_nL^*(2w-\frac12,n,\chi')}{n^{s-w+\frac12}}. \tag{2-17}$$

It is easy to see — using Proposition 2.3 (3) and (2-8) — that for $\Re(2w-\frac12)$, $\Re(s-w+\frac12)$ large enough the series is absolutely and locally uniformly convergent.

By Lemma 2.2 we see that

$$Z(s,w,\chi,\chi') = \zeta_2(4s-1)\zeta_2(4w-1)\sum_{\substack{n=1\\(n,2)=1}}^{\infty}\frac{t_n\chi(n)}{n^{s-w+\frac12}}\sum_{\substack{c=1\\(c,2)=1}}^{\infty}\frac{\chi'(c)H_n(c)}{c^{2w}}.$$

Interchanging summations and using Lemma 2.4 we see that this equals

$$Z(s,w,\chi,\chi') = \zeta_2(4s-1)\zeta_2(4w-1)\sum_{\substack{c=1\\(c,2)=1}}^{\infty}\frac{\chi'(c)L^*(s-w+\frac12,c,\psi,\chi)}{c^{2w-\frac12}}. \tag{2-18}$$

Note that, since

$$\zeta_2(4s-1)\zeta_2(4w-1) = \sum_{\substack{n=1\\(n,2)=1}}^{\infty}\frac{\sigma_{2-4(s-w+\frac12)}(n)}{n^{2(2w-\frac12)}},$$

we also have the series representation

$$Z(s,w,\chi,\chi') = \sum_{\substack{c=1\\(c,2)=1}}^{\infty}\frac{\chi'(c)L^{**}(s-w+\frac12,\psi,c,\chi)}{c^{2w-\frac12}}, \tag{2-19}$$

where

$$L^{**}(s, \psi, c, \chi) = Q^*(s, c, \chi) L_2(s, \psi \otimes \tilde{\chi}_{c_0} \chi)$$

with

$$Q^*(s, c, \chi) = \sum_{l^2 | c} \sigma_{2-4s}(l) Q(s, c/l^2, \chi). \tag{2-20}$$

Remark 2.7. The two representations (2-17), (2-18) will be instrumental in proving meromorphic continuation of $Z(s, w, \chi, \chi')$ to $\mathbb{C}^2$. The proof follows the strategy outlined in [Bump et al. 1996; Diaconu et al. 2003]. The choice of arguments in the definition of (2-17), $2w - \frac{1}{2}$ and $s - w + \frac{1}{2}$, might seem a bit strange, but for the purpose we have in mind it is the most natural one. We shall see that with this choice the functional equations are especially simple.

2B. ***Functional equations of the standard L-functions.*** We now recall the functional equations for the two L-functions $L(s, \chi_{n_0} \chi)$ and $L(s, \psi \otimes \tilde{\chi}_{c_0} \chi)$.

2B1. GL_1. We will use the functional equation for $L_2(s, \chi_{n_0} \chi)$ for n_0 a squarefree odd natural number, and χ mod 8: Let χ_0^8 be the trivial character mod 8. We have that $\chi_{n_0} \chi$ is odd precisely if $\chi = \chi_4 \chi_0^8$ or $\chi = \chi_4 \chi_8$. Also it is known (see [Davenport 2000, Chapter 5], for example) that $\chi_{n_0} \chi$ is induced from the primitive character

$$(\chi_{n_0} \chi)^* = \begin{cases} \chi_{n_0} & \text{if } n_0 \equiv 1 \pmod 4,\ \chi = \chi_0^8, \\ \chi_4 \chi_{-n_0} & \text{if } n_0 \not\equiv 1 \pmod 4,\ \chi = \chi_0^8, \\ \chi_4 \chi_{n_0} & \text{if } n_0 \equiv 1 \pmod 4,\ \chi = \chi_4 \chi_0^8, \\ \chi_{-n_0} & \text{if } n_0 \not\equiv 1 \pmod 4,\ \chi = \chi_4 \chi_0^8, \\ \chi_8 \chi_{n_0} & \text{if } n_0 \equiv 1 \pmod 4,\ \chi = \chi_8 \chi_0^8, \\ \chi_4 \chi_8 \chi_{-n_0} & \text{if } n_0 \not\equiv 1 \pmod 4,\ \chi = \chi_8 \chi_0^8, \\ \chi_4 \chi_8 \chi_{n_0} & \text{if } n_0 \equiv 1 \pmod 4,\ \chi = \chi_4 \chi_8 \chi_0^8, \\ \chi_8 \chi_{-n_0} & \text{if } n_0 \not\equiv 1 \pmod 4,\ \chi = \chi_4 \chi_8 \chi_0^8. \end{cases}$$

It follows that

$$L(s, (\chi_{n_0} \chi)^*) = \left(\frac{\delta_{n_0, \chi}}{\pi} \right)^{\frac{1}{2} - s} \frac{\Gamma\left(\frac{1}{2}(1 - s + \kappa_\chi)\right)}{\Gamma\left(\frac{1}{2}(s + \kappa_\chi)\right)} L(1 - s, (\chi_{n_0} \chi)^*), \tag{2-21}$$

where

$$\kappa_\chi = \begin{cases} 0 & \text{if } \chi = \chi_0^8, \chi_8, \\ 1 & \text{if } \chi = \chi_4 \chi_0^8, \chi_4 \chi_8, \end{cases} \tag{2-22}$$

$$\delta_{n_0, \chi} = \begin{cases} n_0 & \text{if } \chi = \chi_0^8, n_0 \equiv 1 \pmod 4 \text{ or } \chi = \chi_4 \chi_0^8, n_0 \not\equiv 1 \pmod 4, \\ 4n_0 & \text{if } \chi = \chi_0^8, n_0 \not\equiv 1 \pmod 4 \text{ or } \chi = \chi_4 \chi_0^8, n_0 \equiv 1 \pmod 4, \\ 8n_0 & \text{if } \chi = \chi_8, \chi_4 \chi_8. \end{cases}$$

Note that all the functional equations are even, i.e., $\dfrac{G_1((\chi_{n_0} \chi)^*)}{i^{\kappa_\chi} \sqrt{\delta_{n_0, \chi}}} = 1$.

We have

$$L(s,\chi_{n_0}\chi)=\prod_{p\,\mid\, 8n_0/\delta_{n_0,\chi}}(1-(\chi_{n_0}\chi)^*(p)p^{-s})L(s,(\chi_{n_0}\chi)^*)$$

and also

$$L_2(s,\chi_{n_0}\chi)=L_2(s,(\chi_{n_0}\chi)^*)=L(s,(\chi_{n_0}\chi)^*)h_2(s,n_0,\chi), \tag{2-23}$$

where $h_2(s,n_0,\chi)$ is either 1, $1-2^{-s}$, or $1+2^{-s}$. Since $(\chi_{n_0}\chi)^*(2)$ depends only on χ and n_0 mod 8, h_2 has the same dependence.

2B2. GL_2. We now turn to $L_2(s,\psi\otimes\tilde{\chi}_{c_0}\chi)$ for c_0 a squarefree odd natural number, and χ mod 8. The character $\tilde{\chi}_{c_0}$ is primitive of conductor c_0, and is even precisely when $\tilde{\chi}_{c_0}(-1)=\chi_4(c_0)=1$, i.e., when $c_0\equiv 1 \pmod 4$. A reference on twisting of automorphic forms (at least for modular forms) is [Iwaniec and Kowalski 2004, Section 14.8].

We need to take special care of 2-factors. For any primitive automorphic form f for GL_2 we define a polynomial $p_{2,f}(z)$ of degree 1 or 2, depending on whether 2 is ramified or not, by

$$\frac{1}{p_{2,f}(z)}=\sum_{j=0}^{\infty}t_{2^j}(f)z^j, \tag{2-24}$$

where $t_n(f)$ are the coefficients of $L(s,f)$. In particular the 2-factor of $L(s,f)$ equals $p_{2,f}^{-1}(2^{-s})$. If $p_{2,f}$ is of degree 2, $p_{2,f}(z)=(1-\alpha_2 z)(1-\beta_2 z)$, the estimate $|\alpha_2|,|\beta_2|<2^{1/5}$ [Shahidi 1988, p. 549] shows that $p_{2,f}(\pm 2^{-s})$ is uniformly bounded away from 0 at $\Re(s)\geq\frac{1}{2}$. If $p_{2,f}(z)$ is of degree 1, the explicit value of t_2 ($=0$ or $\pm 1/\sqrt{2}$) shows that $p_{2,f}(\pm 2^{-s})$ does not vanish on $\Re(s)\geq\frac{1}{2}$ and, as a result,

$$\frac{1}{p_{2,f}(\pm 2^{-s})}=O(1) \tag{2-25}$$

uniformly in f when $\Re(s)\geq\frac{1}{2}$.

We assume now that ψ is primitive Maaß Hecke form for $\Gamma_0(4)$ with real Fourier coefficients. The twisted function $\psi\otimes\chi$ is still a Hecke form with trivial character χ^2 but not necessarily primitive. Let $g=(\psi\otimes\chi)^*$ be the primitive form whose Fourier coefficients agree with those of $\psi\otimes\chi$ except possibly at the 2-factor. This is a cusp form of level $N=N_{\psi,\chi}=2^j$, a divisor of 64. For fixed ψ there are 4 such forms g, as there are 4 characters mod 8. We have that $L_2(s,\psi\otimes\chi)=L_2(s,g)$ since the Fourier coefficients of g and $\psi\otimes\chi$ agree on odd numbers.

We now twist g by $\tilde{\chi}_{c_0}$. Since the conductor of $\tilde{\chi}_{c_0}$ is relatively prime to the level of g, the result is a primitive cusp form of level $N\cdot c_0^2$. The twisted L-function

$L(s, \psi \otimes \tilde{\chi}_{c_0}\chi)$ agrees with $L(s, g \otimes \tilde{\chi}_{c_0})$ outside the prime 2, so that

$$L_2(s, g \otimes \tilde{\chi}_{c_0}) = L_2(s, \psi \otimes \tilde{\chi}_{c_0}\chi).$$

We have the functional equation of $g \otimes \tilde{\chi}_{c_0}$:

$$L(s, g\otimes\tilde{\chi}_{c_0}) = \epsilon(g, \tilde{\chi}_{c_0})\left(\frac{Nc_0^2}{\pi^2}\right)^{\frac{1}{2}-s} \prod_{\epsilon\in\{\pm1\}} \frac{\Gamma\left(\frac{1}{2}(1-s+\kappa_{\chi,\psi,c_0}+\epsilon(s_0-\frac{1}{2}))\right)}{\Gamma\left(\frac{1}{2}(s+\kappa_{\chi,\psi,c_0}+\epsilon(s_0-\frac{1}{2}))\right)} L(1-s, g\otimes\tilde{\chi}_{c_0}). \tag{2-26}$$

This functional equation involves the root number $\epsilon(g, \tilde{\chi}_{c_0})$ that depends on c_0 mod 8, as it is given by

$$\epsilon(g)\chi^2(c_0)\tilde{\chi}_{c_0}(2^j)G(\tilde{\chi}_{c_0})^2/c_0,$$

where $\epsilon(g)$ is the root number of g. We have

$$L_2(s, \psi \otimes \tilde{\chi}_{c_0}\chi) = H_2(s, g, c_0)L(s, g \otimes \tilde{\chi}_{c_0}),$$

where

$$H_2(s, g, c_0) = p_{2,g\otimes\tilde{\chi}_{c_0}}(2^{-s}) = p_{2,g}(\tilde{\chi}_{c_0}(2)2^{-s}). \tag{2-27}$$

The dependence of $H_2(s, g, c_0)$ on c_0 is only mod 8, as it involves $\tilde{\chi}_{c_0}(2)$. We note also that $\kappa_{\chi,\psi,c_0} = \kappa_{\chi,\psi}\tilde{\chi}_{c_0}(-1)$ depends only on c_0 mod 4 since $\tilde{\chi}_{c_0}(-1) = \chi_4(c_0)$.

Remark 2.8. In the $GL_1 \times GL_1$ case, i.e., if $\psi = \psi_\tau$ and $t_n = \tau(n)$, we have

$$L(s, \psi_\tau \otimes \tilde{\chi}_{c_0}\chi) := \sum_{n=1}^{\infty} \frac{\tau(n)\tilde{\chi}_{c_0}\chi(n)}{n^s} = L(s, \tilde{\chi}_{c_0}\chi)^2.$$

We see (after using quadratic reciprocity) that the analogues of the results of this section follow from Section 2B1.

2C. ***Average bounds on twisted L-functions.*** Before we can give the proof of the meromorphic continuation we recall a few facts concerning the involved L-series. We first recall an average bound on L-functions twisted with quadratic characters. The main ingredient in proving such a bound is Heath-Brown's large sieve estimate for quadratic characters. He proves [1995, Theorem 1] that for any positive $\varepsilon > 0$ there exists a constant $C > 0$ such that for any positive integers M, N and for arbitrary complex numbers $a_1, \ldots, a_N$ we have

$$\sum_{m\le M}^{*} \left| \sum_{n\le N}^{*} a_n \left(\frac{n}{m}\right) \right|^2 \le C(MN)^{\varepsilon}(M+N) \sum_{n\le N}^{*} |a_n|^2. \tag{2-28}$$

Here a $*$ means summation over positive odd squarefree integers. From this one can prove the following.

Theorem 2.9. *For* $\Re(s) \geq \frac{1}{2}$,

$$\sum_{\substack{1<d_0\leq X\\ d_0 \text{ odd,}\\ \text{squarefree}}} |L(s, \chi_{d_0}\chi)|^4 = O((X|s|)^{1+\epsilon}), \tag{2-29}$$

$$\sum_{\substack{1<d_0\leq X\\ d_0 \text{ odd,}\\ \text{squarefree}}} |L(s, \psi \otimes \tilde{\chi}_{d_0}\chi)|^2 = O((X|s|)^{1+\epsilon}). \tag{2-30}$$

The bound (2-29) is already in [Heath-Brown 1995, Theorem 2] and (2-30) is essentially proved in the same way. See also [Soundararajan and Young 2010, Section 2.3; Chinta and Diaconu 2005, Lemma 3.2]. These bounds give the Lindelöf hypothesis on average in the character aspect, while keeping the convexity bound in the s aspect when $\Re(s) = \frac{1}{2}$.

Remark 2.10. By considering 2-factors it is straightforward to see that the above bounds (2-29) and (2-30) are true also if we remove 2-factors, i.e., replace L by L_2.

2D. ***Meromorphic continuation and functional equations of $Z(s, w, \chi, \chi')$.*** We first analyze $Z(s, w, \chi, \chi')$ from the representation (2-17).

Theorem 2.11. *The function* $(w - \frac{3}{4})Z(s, w, \chi, \chi')$ *is analytic in*

$$D_1 = \{(s, w) : \Re(s - w) > \tfrac{1}{2}, \Re(s + w) > \tfrac{3}{2}\},$$

and satisfies a functional equation $\alpha : (s, w) \mapsto (s, 1 - w)$ *given by*

$$(1-2^{-(3-4w)})Z(s, w, \chi, \chi') = \frac{\Gamma\left(\frac{1}{2}(\frac{3}{2}-2w+\kappa_{\chi'})\right)}{\Gamma\left(\frac{1}{2}(2w-\frac{1}{2}+\kappa_{\chi'})\right)} \sum_{\chi'' \bmod 8} p_{\chi''}(w) Z(s, 1-w, \chi'', \chi').$$

Here the $p_{\chi''}(w)$ *are polynomials in* 2^{-w}. *In particular they are bounded in vertical strips. Furthermore, away from* $w = \frac{3}{4}$,

$$Z(s, w, \chi, \chi') = \begin{cases} O((|w|+1)^{\frac{1}{4}+\varepsilon}) & \text{if } \frac{1}{2} \leq \Re w \leq K, \Re(s-w) \geq \frac{1}{2}+\delta, \\ O((|w|+1)^{\frac{1}{4}+1-2\Re(w)+\varepsilon}) & \text{if } -K \leq \Re w \leq \frac{1}{2}, \Re(s+w) \geq \frac{3}{2}+\delta, \end{cases}$$

for any fixed $K > \frac{1}{2}$ *and* $\delta > 0$.

Remark 2.12. We shall see in the proof that the factor $(w - \frac{3}{4})$ is only necessary when χ' is trivial. We note also that the implied constant may depend on ψ. Moreover, the bounds given above are not necessarily optimal. All we need for Theorem 2.15 and Lemma 2.18 below is polynomial control.

Proof. We remark that the factor $\zeta_2(4s - 1)$ appearing in (2-17) does not have a pole in the region D_1. Thus we only have to study the series from (2-17) to prove the analytic properties of $(w - \frac{3}{4})Z(s, w, \chi, \chi')$.

We consider the regions where the series representation (2-17) is absolutely convergent. We consider first the sum over all *non*perfect squares ($n \neq m^2$).

If $\Re(w) \geq \frac{1}{2}$ (which corresponds to $\Re(2w - \frac{1}{2}) \geq \frac{1}{2}$) we use (2-2), Theorem 2.9, Proposition 2.3 (3), and Cauchy–Schwarz to see that, away from $w = \frac{3}{4}$,

$$\sum_{\substack{n \leq X \\ n \neq m^2}} \left| t_n \chi(n) q(2w - \tfrac{1}{2}, n, \chi') L_2(2w - \tfrac{1}{2}, \chi_{n_0} \chi') \right| = O(X^{1+\varepsilon} |w|^{\frac{1}{4}+\varepsilon}). \qquad (2\text{-}31)$$

It follows that the nonperfect square contribution is convergent for $\Re(s-w) \geq \frac{1}{2}+\delta$ and $\Re(w) \geq \frac{1}{2}$ and that, in the region $\Re(s-w) \geq \frac{1}{2}+\delta$, $\Re(w) \geq \frac{1}{2}$, it is analytic and bounded by $O(|w|^{\frac{1}{4}+\varepsilon})$.

For $\Re(w) \leq \frac{1}{2}$, we use Proposition 2.3 (2) and the functional equation for $L_2(2w - \frac{1}{2}, \chi_{n_0}\chi')$ to see that the product $q(2w - \frac{1}{2}, n, \chi') L_2(2w - \frac{1}{2}, \chi_{n_0}\chi')$ equals

$$n^{1-2w} \left(\frac{\delta_{n_0,\chi'}}{n_0 \pi} \right)^{1-2w} \frac{\Gamma\left(\frac{1}{2}(\frac{3}{2} - 2w + \kappa_{\chi'})\right)}{\Gamma\left(\frac{1}{2}(2w - \frac{1}{2} + \kappa_{\chi'})\right)} q(2(1-w) - \tfrac{1}{2}, n, \chi') L_2(2(1-w) - \tfrac{1}{2}, \chi_{n_0}\chi')$$

times a factor $h_2(2w - \frac{1}{2}, n_0, \chi) \big/ h_2(2(1-w) - \frac{1}{2}, n_0, \chi)$, which is bounded when $\Re(w) \leq \frac{1}{2}$ (recall (2-23) for the definition of h_2). We notice that $\delta_{n_0,\chi'}/n_0$ is 1, 4, or 8, and that in bounded w-strips the quotient of Γ-factors is $O(|w|^{1-2\Re(w)})$. It follows that in bounded w-strips and for $\Re(w) \leq \frac{1}{2}$ we have

$$\begin{aligned}
&\sum_{\substack{n \leq X \\ n \neq m^2}} \left| n^{2w-1} t_n \chi(n) q(2w - \tfrac{1}{2}, n, \chi') L_2(2w - \tfrac{1}{2}, \chi_{n_0}\chi') \right| \\
&\quad = O(|w|^{1-2\Re(w)}) \sum_{n \leq X} \left| t_n \chi(n) q(2(1-w) - \tfrac{1}{2}, n, \chi') L_2(2(1-w) - \tfrac{1}{2}, \chi_{n_0}\chi') \right| \\
&\quad = O(|w|^{\frac{1}{4}+1-2\Re(w)+\epsilon} X^{1+\varepsilon}),
\end{aligned}$$

where in the last line we have used the same argument as used to bound (2-31). It follows that when $\Re(s+w) \geq \frac{3}{2}+\delta$, $\Re(w) \leq \frac{1}{2}$ the nonsquare contribution from the series in (2-17) converges absolutely and that in this region this contribution is analytic and bounded by $O(|w|^{\frac{1}{4}+1-2\Re(w)+\varepsilon})$.

We next consider the sum over all perfect squares $n = m^2$,

$$L_2(2w - \tfrac{1}{2}, \chi') \sum_{\substack{m=1 \\ (m,2)=1}}^{\infty} \frac{t_n \chi(n) q(2w - \frac{1}{2}, m^2, \chi')}{m^{2(s-w+\frac{1}{2})}}.$$

Using Proposition 2.3 and (2-2) we easily see that the sum is convergent in

$$\{(s, w) : \Re(s-w) > 0, \Re(s+w) > 1\},$$

and that the factor in front has a simple pole at $w=\frac{3}{4}$ if χ' is trivial. That this contribution has the desired growth properties follows from the convexity estimate on $L_2(2w-\frac{1}{2},\chi')$.

Having established that $(w-\frac{3}{4})Z(s,w,\chi,\chi')$ is analytic in D_1, we now show that it satisfies a functional equation here. For (s,w) in this region we use the functional equation (2-21) and Proposition 2.3 and the subsequent discussion to see that $Z(s,w,\chi,\chi')/\zeta_2(4s-1)$ equals

$$\begin{aligned}
&\sum_{\substack{n=1\\(n,2)=1}}^{\infty}\frac{t_n\chi(n)q(2w-\frac{1}{2},n,\chi')L_2(2w-\frac{1}{2},\chi_{n_0}\chi')}{n^{s-w+\frac{1}{2}}}\\
&=\sum_{\substack{n=1\\(n,2)=1}}^{\infty} n^{1-2w}\left(\frac{\delta_{n_0,\chi'}}{n_0\pi}\right)^{1-2w}\frac{h_2(2w-\frac{1}{2},n_0,\chi')}{h_2(2(1-w)-\frac{1}{2},n_0,\chi')}\frac{\Gamma\big(\frac{1}{2}(\frac{3}{2}-2w+\kappa_{\chi'})\big)}{\Gamma\big(\frac{1}{2}(2w-\frac{1}{2}+\kappa_{\chi'})\big)}\\
&\qquad\cdot\frac{t_n\chi(n)q(2(1-w)-\frac{1}{2},n,\chi')L_2(2(1-w)-\frac{1}{2},\chi_{n_0}\chi')}{n^{s-w+\frac{1}{2}}}\\
&=\pi^{2w-1}\frac{\Gamma\big(\frac{1}{2}(\frac{3}{2}-2w+\kappa_{\chi'})\big)}{\Gamma\big(\frac{1}{2}(2w-\frac{1}{2}+\kappa_{\chi'})\big)}\sum_{\substack{n=1\\(n,2)=1}}^{\infty}\frac{h_2(2w-\frac{1}{2},n_0,\chi')}{h_2(2(1-w)-\frac{1}{2},n_0,\chi')}\left(\frac{\delta_{n_0,\chi'}}{n_0}\right)^{1-2w}\\
&\qquad\cdot\frac{t_n\chi(n)q(2(1-w)-\frac{1}{2},n,\chi')L_2(2(1-w)-\frac{1}{2},\chi_{n_0}\chi')}{n^{s-(1-w)+\frac{1}{2}}}.
\end{aligned}$$

We split the sum according to $n \bmod 8$, and notice that for fixed χ' the function

$$\frac{h_2(2w-\frac{1}{2},n_0,\chi')}{h_2(2(1-w)-\frac{1}{2},n_0,\chi')}\left(\frac{\delta_{n_0,\chi'}}{n_0}\right)^{1-2w}$$

is the same fraction of Dirichlet polynomials in 2^{-w} throughout each of these sums, so that we can put them outside the sums. Using again that the indicator function of residue class mod 8 can be written as a linear combination of characters mod 8 (at least on the odd numbers), we arrive at the functional equation for $Z(s,w,\chi,\chi')$. We note that the factor $1-2^{-(3-4w)}$ is the product of all possible $h_2(2(1-w)-\frac{1}{2},n_0,\chi')$. This shows that the $p_{\chi''}(w)$ are in fact polynomials in 2^{-w}. □

We now apply the same type of analysis to the second series representation of $Z(s,w,\chi,\chi')$ given in (2-18). Recall from Section 2B2 that g denotes $(\psi\otimes\chi)^*$, where χ is one of the 4 characters mod 8. Let

$$V(s,w)=\prod_g p_{2,g}(2^{-(w-s+\frac{1}{2})})p_{2,g}(-2^{-(w-s+\frac{1}{2})}),$$

where $p_{2,g}(z)$ is as in (2-24).

Theorem 2.13. *The function* $(s-w-\frac{1}{2})^2 Z(s,w,\chi,\chi')$ *is analytic in*

$$D_2 = \{(s,w) : \Re(s) > \tfrac{3}{4}, \Re(w) > \tfrac{3}{4}\},$$

and satisfies a functional equation $\beta : (s,w) \mapsto (w,s)$ *given by*

$$\begin{aligned} &V(s,w)Z(s,w,\chi,\chi') \\ &= \sum_{\substack{k=0,1 \\ \chi'' \bmod 8}} \prod_{\epsilon\in\{\pm 1\}} \frac{\Gamma\big(\frac{1}{2}(1-(s-w+\frac{1}{2})+k+\epsilon(s_0-\frac{1}{2}))\big)}{\Gamma\big(\frac{1}{2}((s-w+\frac{1}{2})+k+\epsilon(s_0-\frac{1}{2}))\big)} P_{\psi,\chi,\chi''}(s,w)Z(w,s,\chi,\chi''). \end{aligned}$$

Here the $P_{\psi,\chi,\chi''}(s,w)$ *are polynomials in* $2^{-(s-w)}$. *In particular they are functions bounded in vertical strips. Furthermore, away from* $s-w-\frac{1}{2}=0$,

$$Z(s,w,\chi,\chi') = \begin{cases} O((|s-w|+1)^{\frac{1}{2}+\varepsilon}) & \textit{for } \frac{3}{4}+\delta \le \Re w \le \Re(s) \le K, \\ O((|s-w|+1)^{\frac{3}{2}-2\Re(s-w+\frac{1}{2}+\varepsilon)}) & \textit{for } \frac{3}{4}+\delta \le \Re s \le \Re(w) \le K, \end{cases}$$

where K, δ *are any constants with* $K > \frac{3}{4}$ *and* $\delta > 0$.

Remark 2.14. We shall see in the proof that the factor $(s-w-\frac{1}{2})^2$ is only necessary when ψ is $\mathrm{GL}_1 \times \mathrm{GL}_1$ and χ is trivial. We note also that the implied constant may depend on ψ. Moreover, as before the bounds given above are not necessarily optimal, as all we need for Theorem 2.15 and Lemma 2.18 below is polynomial control.

Proof. We now want to find the region of absolute convergence of (2-18). Consider first the region $\Re(s-w+\frac{1}{2}) \ge \frac{1}{2}$. We can use Cauchy–Schwarz, Theorem 2.9, and Lemma 2.6 to see that the sum over nonperfect squares satisfies

$$\sum_{\substack{c\le X \\ c\ne r^2, c \text{ odd}}} \big|\chi'(c)Q(s-w+\tfrac{1}{2},c,\chi)L_2(s-w+\tfrac{1}{2},\psi\otimes\tilde{\chi}_{c_0}\chi)\big| = O\big(X^{1+\varepsilon}(1+|s-w|)^{\frac{1}{2}+\varepsilon}\big).$$

Hence the sum over these terms is absolutely convergent when $\Re(2w-\frac{1}{2}) \ge 1+\delta$. The sum over the perfect squares potentially has a double pole at $s-w+\frac{1}{2}=1$: For $t_n = \tau(n)$ we have $L_2(s,\psi\otimes\chi_0^8) = \zeta_2^2(s)$. The sum over perfect squares is

$$L_2(s-w+\tfrac{1}{2},\psi\otimes\chi)\sum_{\substack{c=1 \\ c=r^2}}^{\infty} \frac{\chi'(c)Q(s-w+\frac{1}{2},c,\chi)}{c^{2w-\frac{1}{2}}},$$

where the sum is again absolutely convergent for $\Re(2w-\frac{1}{2}) \ge 1+\delta$, using Cauchy–Schwarz and Lemma 2.6. It follows that, when $\Re(s-w+\frac{1}{2}) \ge \frac{1}{2}$, the sums are convergent for $\Re(w) \ge \frac{3}{4}+\delta$, and hence $Z(s,w,\chi,\chi')$ is analytic in this region except for a potential double polar line at $s-w+\frac{1}{2}=1$. We also find that in this region we have the bound $Z(s,w,\chi,\chi') = O((1+|s-w|)^{\frac{1}{2}+\varepsilon})$.

Turning now to $\Re(s-w+\frac{1}{2}) \le \frac{1}{2}$, we use the functional equation (2-26) and Proposition 2.5 (2) to move to a region where we can use the same bounds as for $\Re(s-w+\frac{1}{2}) \ge \frac{1}{2}$:

$$
\begin{aligned}
&\frac{Z(s,w,\chi,\chi')}{\zeta_2(4s-1)\zeta_2(4w-1)}\\
&=\sum_{\substack{c=1\\(c,2)=1}}^{\infty}\frac{1}{c^{2w-\frac{1}{2}}}\chi'(c)Q(s-w+\tfrac{1}{2},c,\chi)L_2(s-w+\tfrac{1}{2},\psi\otimes\tilde{\chi}_{c_0}\chi)\\
&=\sum_{\substack{c=1\\(c,2)=1}}^{\infty}\frac{1}{c^{2w-\frac{1}{2}}}\chi'(c)c^{-2(s-w)}Q(1-(s-w+\frac{1}{2}),c,\chi)\epsilon(\psi,\tilde{\chi}_{c_0}\chi)\left(\frac{N_1}{\pi^2}\right)^{-(s-w)}\\
&\quad\times\frac{\Gamma\big(\frac{1}{2}(1-(s-w+\frac{1}{2})+\kappa_{\chi,\psi,c_0}+(s_0-\frac{1}{2}))\big)\Gamma\big(\frac{1}{2}(1-(s-w+\frac{1}{2})+\kappa_{\chi,\psi,c_0}-(s_0-\frac{1}{2}))\big)}{\Gamma\big(\frac{1}{2}((s-w+\frac{1}{2})+\kappa_{\chi,\psi,c_0}+(s_0-\frac{1}{2}))\big)\Gamma\big(\frac{1}{2}((s-w+\frac{1}{2})+\kappa_{\chi,\psi,c_0}-(s_0-\frac{1}{2}))\big)}\\
&\qquad\times\frac{H_2(s-w+\frac{1}{2},g_1,c_0)}{H_2(1-(s-w+\frac{1}{2}),g_1,c_0)}L_2(1-(s-w+\tfrac{1}{2}),\psi\otimes\tilde{\chi}_{c_0}\chi),
\end{aligned}
\tag{2-32}
$$

where $g_1=(\psi\otimes\tilde{\chi}_{c_0}\chi)^*$ with level $N_1c_0^2$ for N_1 a divisor of 64 depending on χ, ψ (recall (2-27) for the definition of H_2). Using the same trick as before of splitting the sum into perfect squares and nonperfect squares, and using the bounds from Lemma 2.6 and Theorem 2.9 as well as the Stirling bound on the Gamma factors and a trivial bound on the 2-factors, we find that $Z(s,w,\chi,\chi')$ is analytic in

$$\{(s,w):\Re(s-w+\tfrac{1}{2})\le\tfrac{1}{2},\ \Re s\ge\tfrac{3}{4}+\delta\}$$

and bounded, as $Z(s,w,\chi,\chi')=O\big(1+|s-w|^{\frac{1}{2}+\varepsilon}|s-w|^{1-2\Re(s-w+\frac{1}{2})}\big)$ for $\Re(s)$, $\Re(w)$ bounded in this region.

We have established that $Z(s,w,\chi,\chi')$ is analytic in D_2. We now show that it also satisfies a functional equation in this region. Consider (2-32). We noticed that $\epsilon(\psi,\tilde{\chi}_0\chi)$, κ_{χ,ψ,c_0}, and $H_2(s,g_1,c_0)$ depend only on c_0 modulo 8 (see Section 2B2). We split the sum into residue classes modulo 8 and we can put these data outside the sum. Since $H_2(1-(s-w+\frac{1}{2}),g_1,c_0)$ can have zeros in the region we multiply the left-hand side with all possible expressions of it, which is $V(s,w)$, and arrive at the desired functional equation. □

Using the two previous theorems we can now show that $Z(s,w,\chi,\chi')$ admits a meromorphic continuation to all of $\mathbb{C}^2$.

Theorem 2.15. *The function*

$$Z^*(s,w,\chi,\chi')=(s-w-\tfrac{1}{2})^2(s+w-\tfrac{3}{2})^2(w-\tfrac{3}{4})(s-\tfrac{3}{4})Z(s,w,\chi,\chi')\tag{2-33}$$

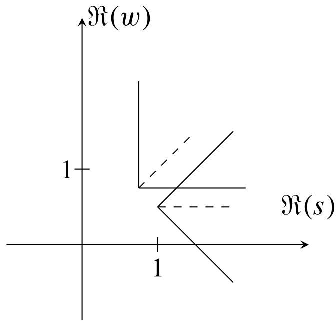


Figure 1. $D_1 \cup D_2$.

admits an analytic continuation to $(s, w) \in \mathbb{C}^2$ *with at most polynomial growth for* $\Re(s), \Re(w)$ *in bounded regions.*

Proof. We use repeatedly the functional equations in Theorems 2.11 and 2.13. We notice that these two theorems show that $Z^*(s, w, \chi, \chi')$ is analytic in the union of the two overlapping sets

$$D_1 = \{(s, w) : \Re(s - w) > \tfrac{1}{2}, \Re(s + w) > \tfrac{3}{2}\},$$
$$D_2 = \{(s, w) : \Re(s) > \tfrac{3}{4}, \Re(w) > \tfrac{3}{4}\},$$

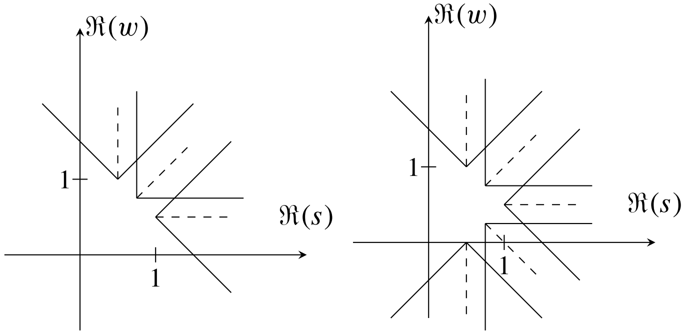
since $(w - \frac{3}{4})Z(s, w, \chi, \chi')$ is analytic in D_1 and $(s - w - \frac{1}{2})^2 Z(s, w, \chi, \chi')$ is analytic in D_2. (See Figure 1.)

We now use the group of functional equations generated by the two functional equations

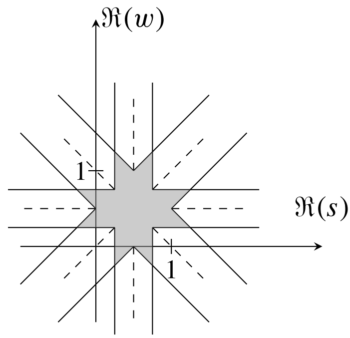
$$\alpha : (s, w) \mapsto (s, 1 - w), \quad \beta : (s, w) \mapsto (w, s).$$

They generate a group of order 8 isomorphic to the dihedral group D_4 of order 8. We note that $\alpha^2 = \beta^2 = \mathrm{Id}$. Using β, we see that $(s - \frac{3}{4})Z(s, w, \chi, \chi')$ is a holomorphic function of at most bounded polynomial growth (bounding the ratio of Gamma functions using Stirling asymptotics) in $D_3 = \beta D_1$, which then extends $Z^*(s, w, \chi, \chi')$ to $D_1 \cup D_2 \cup D_3$. We notice that the Gamma factor on the right-hand side of the functional equation in Theorem 2.13 and $V(s, w)^{-1}$ does not have poles when $\Re(w - s) > 0$ (by (2-25) and properties of the Gamma function).

Using α, we extend $Z^*(s, w, \chi, \chi')$ analytically to $D_1 \cup D_2 \cup \beta D_1 \cup \alpha D_2 \cup \alpha\beta D_1$. We notice that the 2 factor $(1 - 2^{-(3-4w)})^{-1}$ and the Gamma factor in Theorem 2.11 are analytic when $\Re(w) < \frac{3}{4}$. The reflection α of the double polar line $s - w = \frac{1}{2}$ in D_2 produces the double polar line $s + w = \frac{3}{2}$ in αD_2.



The regions $D_4 = \beta\alpha D_2$, $D_5 = \beta\alpha\beta D_1$, and $D_6 = \alpha\beta\alpha D_2 = \alpha D_4$ can be dealt with using Theorems 2.11 and 2.13 in the same way and no new polar lines are introduced, neither due to the 2 factors, nor the Gamma factors.



The function in (2-33) is now extended to a holomorphic function on the complement of the domain with tube given by the shaded region. It is bounded polynomially for $\Re(w)$, $\Re(s)$ bounded. We can therefore use Bochner's tube theorem (see [Diaconu et al. 2003, Propositions 4.6 and 4.7 and the argument on p. 341]) to extend the holomorphic function to the convex hull of this region (which is $\mathbb{C}^2$) with at most polynomial bounds for $(\Re(s), \Re(w))$ in compact sets. Therefore, $Z(s, w, \chi, \chi')$ has the same properties, apart from being meromorphic with the specified polar lines in (2-33). $\square$

Remark 2.16. Combining Theorems 2.11 and 2.13 we note that $\alpha\circ\beta\circ\alpha\circ\beta(s, w) = (1-s, 1-w)$, and it follows that there exist functions $\alpha_{\rho,\rho',\chi,\chi'}(s, w)$ bounded in vertical strips such that

$$F(s, w)Z(s, w, \chi, \chi') = \sum_{\bar{k}\in\{0,1\}^4} \frac{G(1-s, 1-w, \bar{k})}{G(s, w, \bar{k})} \sum_{\rho,\rho' \bmod 8} \alpha_{\bar{k},\rho,\rho',\chi,\chi'}(s, w)Z(1-s, 1-w, \rho, \rho'), \quad (2\text{-}34)$$

where

$$G(s,w,\bar{k}) := \Gamma\big(\tfrac{1}{2}(2w-\tfrac{1}{2}+k_1)\big)\prod_{\epsilon_1\in\{\pm1\}}\Gamma\big(\tfrac{1}{2}(s+w-\tfrac{1}{2}+k_2+\epsilon_1(s_0-\tfrac{1}{2}))\big)$$
$$\cdot\Gamma\big(\tfrac{1}{2}(2s-\tfrac{1}{2}+k_3)\big)\prod_{\epsilon_2\in\{\pm1\}}\Gamma\big(\tfrac{1}{2}(s-w+\tfrac{1}{2}+k_4+\epsilon_2(s_0-\tfrac{1}{2}))\big) \tag{2-35}$$

and

$$F(s,w) := (1-2^{-(3-4w)})(1-2^{-(3-4s)})V(s,w)V(w,1-s).$$

Using $\dfrac{\Gamma(\frac{1}{2}(1-z+1))}{\Gamma(\frac{1}{2}(z+1))} = \dfrac{\Gamma(\frac{1}{2}(1-z))}{\Gamma(\frac{1}{2}z)}\cot(\pi z/2)$ we see that

$$\frac{G(1-s,1-w,\bar{k})}{G(s,w,\bar{k})} = \frac{G(1-s,1-w,0)}{G(s,w,0)}\cot_{\bar{k}}(s,w),$$

where

$$\cot_{\bar{k}}(s,w) = \cot^{k_1}\Big(\frac{\pi(2w-\frac{1}{2})}{2}\Big)\prod_{\epsilon_1\in\{\pm1\}}\cot^{k_2}\Big(\frac{\pi(s+w-\frac{1}{2}+\epsilon_1(s_0-\frac{1}{2}))}{2}\Big)$$
$$\cdot\cot^{k_3}\Big(\frac{\pi(2s-\frac{1}{2})}{2}\Big)\prod_{\epsilon_2\in\{\pm1\}}\cot^{k_4}\Big(\frac{\pi(s-w+\frac{1}{2}+\epsilon_2(s_0-\frac{1}{2}))}{2}\Big).$$

Away from poles of cot we have uniform bounds $\cot(\frac{1}{2}\pi z) = i\,\mathrm{sign}(y)+O(e^{-\pi y})$, so we see that $\cot_{\bar{k}}(s,w)$ is bounded in vertical strips (for the arguments away from the poles of $\cot_{\bar{k}}$). It follows that the functional equation (2-34) can be written simply as

$$F(s,w)Z(s,w,\chi,\chi') = \frac{G(1-s,1-w,0)}{G(s,w,\bar{0})}\sum_{\substack{\rho,\rho' \bmod 8\\ \bar{k}\in\{0,1\}^4}}\beta_{\bar{k},\rho,\rho',\chi,\chi'}(s,w)Z(1-s,1-w,\rho,\rho'), \tag{2-36}$$

where the functions $\beta_{\bar{k},\rho,\rho',\chi,\chi'}(s,w)$ are bounded in vertical strips (away from any poles).

2E. ***Bounds on*** $\boldsymbol{Z(s,w,\chi,\chi')}$**.** In this section we bound $Z(s,w,\chi,\chi')$ when $\Re(s)=\Re(w)=\frac{1}{2}$. Recall that in (1-6) we defined the analytic conductor to be

$$\mathfrak{q}(t,u) := (1+|t|)(1+|t+u|)^2(1+|u|). \tag{2-37}$$

Theorem 2.17. *Assume (1-4). Then*

$$Z(\tfrac{1}{2}+it,\tfrac{1}{2}+iu,\chi,\chi') = O(\mathfrak{q}(t,u)^{\frac{1}{4}+\varepsilon}).$$

Unconditionally,

$$Z(\tfrac{1}{2}+it,\tfrac{1}{2}+iu,\chi,\chi')=O\big((q(t,u)(1+|t-u|)^2)^{\frac{1}{4}+\varepsilon}\big).$$

We call the bound obtained in Theorem 2.17 the *convexity* bound. Any bound of the form $O(\mathfrak{q}(t,u)^{\frac{1}{4}-\delta})$ is called a *subconvex* bound.

To prove Theorem 2.17 we first prove an approximate functional equation similar to the one in [Blomer et al. 2014, Lemma 4.2].

Lemma 2.18. *Let $t, u \in \mathbb{R}$ and χ, χ' mod 8. There exist smooth functions $W_\pm:\mathbb{R}_+\to\mathbb{C}$ depending on u, t, and the characters satisfying*

$$y^j\frac{d^j}{dy^j}W_\pm(y)=O(1+y)^{-A}$$

for all $j, A\in\mathbb{N}_0$, uniformly in u, t, such that

$$Z(\tfrac{1}{2}+it,\tfrac{1}{2}+iu,\chi,\chi') \\ =\sum_{\rho,\rho' \bmod 8}\sum_{\pm}\sum_{c=1}^{\infty}\frac{\rho'(c)L^{**}(\frac{1}{2}\pm i(t-u),\psi,c,\rho)}{c^{\frac{1}{2}\pm 2iu}}W_\pm\bigg(\frac{c}{\sqrt{\mathfrak{q}(t,u)}}\bigg).$$

Proof. Recall $1/\cos(z)$ is holomorphic in the strip $|\Re(z)|<\pi/2$ and satisfies $1/\cos(z)=O_{\epsilon_0}(e^{-|z|})$ for $|\Re(z)|\le\pi/2-\epsilon_0$. For $\eta(\log 2)/(\pi i)$ bounded away from $\mathbb{Z}$, the function $P_\eta(z)=(1-2^{\eta-z})(1-2^{\eta+z})/(1-2^\eta)^2$ is uniformly bounded in vertical strips, holomorphic in $\mathbb{C}$, even in z, with a simple zero at η, and satisfies $P_\eta(0)=1$. For a given multiset B let

$$H_B(z)=\Big(\cos\Big(\frac{\pi z}{3A}\Big)\Big)^{-12A}\prod_{\eta\in B}P_\eta(z),$$

which is $O(e^{-4\pi|z|})$ for, say, $|\Re(z)|\le(\frac{3}{2}-\delta)A$ with $\delta>0$ sufficiently small. For an appropriate choice of multiset $B=B_{t,u}$ we set $H_{t,u}(z)=H_{B_{t,u}}(z)$ so that the integrand of

$$\frac{1}{2\pi i}\int_{(1)}\frac{F(s+z,w+z)}{F(s,w)}Z(s+z,w+z,\chi,\chi')\frac{G(s+z,w+z,0)}{G(s,w,0)}H_{t,u}(z)\,\frac{dz}{z} \tag{2-38}$$

is holomorphic in the entire z-plane except for a simple pole at $z=0$. (The function $H_{t,u}$ has been used to remove the poles of $Z(s+z,w+z,\chi,\chi')$.) Also it has rapid decay in z on vertical lines due to Theorem 2.15. Moving the line of integration to $\Re(s)=-1$, we see that (2-38) equals

$$Z(s,w,\chi,\chi') \\ +\frac{1}{2\pi i}\int_{(-1)}\frac{F(s+z,w+z)}{F(s,w)}Z(s+z,w+z,\chi,\chi')\frac{G(s+z,w+z,0)}{G(s,w,0)}H_{t,u}(z)\,\frac{dz}{z}.$$

Using the functional equation (2-36) and the change of variable $z \mapsto -z$, the last integral equals

$$\sum_{\substack{\rho,\rho' \bmod 8 \\ \bar{k}\in\{0,1\}^4}} \frac{1}{2\pi i}\int_{(1)} \frac{\beta_{\bar{k},\rho,\rho',\chi,\chi'}(s-z,w-z)}{F(s,w)} Z(1-s+z,1-w+z,\rho,\rho') \times \frac{G(1-s+z,1-w+z,0)}{G(s,w,0)} H_{t,u}(z)\,\frac{dz}{z}.$$

Thus there exist functions $\gamma_{\bar{k},\rho,\rho',\chi,\chi',\pm}(x,x')$, bounded if $\Re(x)=\Re(x')=-\frac{1}{2}$ (note that using (2-25) we see that $F(s,w)^{-1}$ is uniformly bounded), such that $Z(\frac{1}{2}+it,\frac{1}{2}+iu,\chi,\chi')$ equals

$$\sum_{\substack{\pm \\ \rho,\rho' \bmod 8 \\ \bar{k}\in\{0,1\}^4}} \frac{1}{2\pi i}\int_{(1)} \gamma_{\bar{k},\rho,\rho',\chi,\chi',\pm}(\tfrac{1}{2}\pm it-z,\tfrac{1}{2}\pm iu-z) Z(\tfrac{1}{2}\pm it+z,\tfrac{1}{2}\pm iu+z,\rho,\rho') \times \frac{G(\frac{1}{2}\pm it+z,\frac{1}{2}\pm iu+z,0)}{G(\frac{1}{2}+it,\frac{1}{2}+iu,0)} H_{t,u}(z)\,\frac{dz}{z}.$$

Using the series representation (2-19) we arrive at the result with $W_{\pm}(y)$ equal to

$$\sum_{\bar{k}\in\{0,1\}^4} \frac{1}{2\pi i}\int_{(1)} \gamma_{\bar{k},\rho,\rho',\chi,\chi',\pm}(\tfrac{1}{2}\pm it-z,\tfrac{1}{2}\pm iu-z)\big(y\sqrt{C'(t,u)}\big)^{-2z} \times \frac{G(\frac{1}{2}\pm it+z,\frac{1}{2}\pm iu+z,0)}{G(\frac{1}{2}+it,\frac{1}{2}+iu,0)} H_{t,u}(z)\,\frac{dz}{z}.$$

From Stirling's formula we find that $\Gamma(s+z)/\Gamma(s) = O((1+|s|)^{\Re(z)}e^{\pi|z|/2})$ uniformly for s, z in bounded strips away from poles. It follows that we have

$$\frac{G(\frac{1}{2}\pm it+z,\frac{1}{2}\pm iu+z,0)}{G(\frac{1}{2}+it,\frac{1}{2}+iu,0)} = O(\mathfrak{q}(u,t)^{\Re(z)}e^{2\pi|z|}).$$

By shifting the contour to σ and differentiating under the integral sign we see that

$$y^j\frac{\partial^j W_{\pm}}{\partial^j y} = O\bigg(y^{-2\sigma}\int_{(\sigma)} e^{(-4\pi+2\pi)|z|}\frac{(1+|z|)^j}{|z|}\,dz+\delta_{j=0,\sigma<0}\bigg)$$

for $-\delta \le \sigma < (\frac{3}{2}-\delta)A$. The last term comes from the pole at $z=0$. For $y \le 1$ we can choose $\sigma = -\delta/2$, and for $y>1$ we choose $\sigma = A$ and find the desired bound. □

Proof of Theorem 2.17. Let $\varepsilon > 0$. For $\Re(z)=\frac{1}{2}$ we have, assuming (1-4),

$$\sum_{\substack{c\le Y \\ c \text{ odd}}} |L^{**}(z,\psi,c,\rho)| = O\big(Y^{1+\epsilon}(1+|z|)^{a+\epsilon}\big) \tag{2-39}$$

with $a=0$. Unconditionally, (2-39) holds with $a=\frac{1}{2}$, as is straightforward to verify from Lemma 2.6, Theorem 2.9 (2-30), and Cauchy–Schwartz.

It follows that for an appropriate choice of A in Lemma 2.18 we have

$$\sum_{c>\mathfrak{q}(u,t)^{\frac{1}{2}+\varepsilon}} \frac{|L^{**}(\frac{1}{2}\pm i(t-u),\psi,c,\rho)|}{c^{\frac{1}{2}}}\left|W_{\pm}\left(\frac{c}{\sqrt{\mathfrak{q}(t,u)}}\right)\right| \leq C_{\varepsilon}((1+|t-u|)^{a}\mathfrak{q}(u,t)^{\epsilon}).$$

It follows also that

$$\sum_{c\leq\mathfrak{q}(u,t)^{\frac{1}{2}+\varepsilon}} \frac{|L^{**}(\frac{1}{2}\pm i(t-u),\psi,c,\rho)|}{c^{\frac{1}{2}}} = O\big(\mathfrak{q}(u,t)^{\frac{1}{4}+\epsilon}(1+|t-u|)^{a+\epsilon}\big).$$

Theorem 2.17 now follows from the approximate functional equation. □

Remark 2.19. We notice that for the special configuration $w=1-s$ the conductor drops to essentially

$$(1+|t|)(1+|u|).$$

This configuration will be the relevant one in Theorem 4.3 below.

Remark 2.20. One could speculate whether using another functional equation could lead to a smaller conductor. During the proof of Theorem 2.17, or more precisely in the proof of the approximate functional equation Lemma 2.18, we have made certain choices: we have chosen a particular functional equation $(s,w)\to(1-s,1-w)$ and a particular series representation (2-19). In principle, there is nothing that prohibits running the same type of argument with the other series representation (2-17) and/or another functional equation.

Let us consider what happens if we make other choices. If we use (2-17) and if $\Re(z)=1$ and $\Re(s)=\Re(w)=\frac{1}{2}$, then the function $Z(s+z,w,\chi,\chi')$ in (2-38) is evaluated in D_1, where the series representation (2-17) is convergent. Similarly, if we consider (2-19) and if $\Re(z)=1$ and $\Re(s)=\Re(w)=\frac{1}{2}$, then the function $Z(s+z,w+z,\chi,\chi')$ is evaluated in D_2, where the series representation (2-19) is convergent. In order for the argument in Lemma 2.18 to work we need to use a functional equation $\gamma:\mathbb{C}^2\to\mathbb{C}^2$ with the property that, when $\Re(z)=1$ and $\Re(s)=\Re(w)=\frac{1}{2}$, the numbers $\gamma(s-z,w-z)/\gamma(s-z,w)$ lie in D_1 or D_2. Only in this case is the integrand evaluated where the double Dirichlet series has a series representation (after moving the line of integration to $\Re(z)=-1$, using the functional equation and making a change of variable $z\to -z$).

When we are using (2-19) we assume (1-4). When we are using (2-17) we make the similar assumption for this series, namely that

$$\sum_{\substack{n\leq X\\(n,2)=1}} |t_n L^{*}(w,n,\chi')| = O\big(X^{1+\epsilon}(1+|w|)^{\epsilon}\big) \quad \text{for } \Re(w)=\tfrac{1}{2}.$$

With these restrictions we list the possible “analytic conductors” in Table 1.

Functional equation	Series repn.	Analytic conductor
$\beta\alpha\beta : (s, w) \to (1-s, w)$	(2-17)	$(1+\|t+u\|)^2(1+\|t-u\|)^2(1+\|t\|)^2$
$\alpha\beta\alpha : (s, w) \to (1-w, 1-s)$	(2-19)	$(1+\|t\|)(1+\|t+u\|)^2(1+\|u\|)$
$\alpha\beta\alpha\beta : (s, w) \to (1-s, 1-w)$	(2-17)	$(1+\|t+u\|)^2(1+\|t-u\|)^2(1+\|t\|)^2$
$\alpha\beta\alpha\beta : (s, w) \to (1-s, 1-w)$	(2-19)	$(1+\|t\|)(1+\|t+u\|)^2(1+\|u\|)$

Table 1. Different choices of analytic conductors.

Since, for all $t, u \in \mathbb{R}$,

$$(1+|t|)(1+|t+u|)^2(1+|u|) \leq (1+|t+u|)^2(1+|t-u|)^2(1+|t|)^2,$$

the conductor defined in (2-37) is the smallest among these.

2F. ***Another double Dirichlet series.*** It turns out that there is another double Dirichlet series which is relevant in the applications to QUE. We now define it and then immediately show that it can be understood in terms of the series $Z(s, w, \chi, \chi')$ which was analyzed in the previous sections. Let

$$\widehat{Z}(s, w, \chi, \chi') = \sum_{\substack{c=1 \\ (c,2)=1}} \frac{\chi'(c) L^*(s-w+\frac{1}{2}, c, \chi)^2}{c^{2w-\frac{1}{2}}}. \tag{2-40}$$

In order to understand $\widehat{Z}(s, w, \chi, \chi')$ we exhibit an interesting nontrivial relation between the q-polynomials and the Q-polynomials in the case of the Eisenstein series when $t_n = \tau(n)$. Let $\widehat{Q}$ be defined as Q but with the one exception that we use χ_{c_0} instead of $\tilde{\chi}_{c_0}$, i.e., with $v = v_p(c_1)$,

$$\widehat{Q}(s, c_0 c_1^2, \chi) = \prod_{p \mid c_1} \frac{t_{p^{2v}} - t_{p^{2v-1}} \chi_{c_0}(p)\chi(p)(p^{1-s}+p^s)/p + t_{p^{2v-2}} \chi_{c_0}(p)^2/p}{p^{2v(s-\frac{1}{2})}},$$

defined for c_0, c_1 odd. By (2-6) we see that

$$\widehat{Q}(s, c, \chi) = \begin{cases} Q(s, c, \chi) & \text{if } c_0 \equiv 1 \pmod 4, \\ Q(s, c, \chi\chi_4) & \text{if } c_0 \equiv 3 \pmod 4. \end{cases} \tag{2-41}$$

Lemma 2.21. *Let d_0 be an odd squarefree positive integer, d_1 odd, and $t_n = \tau(n)$. Then*

$$\sum_{d \mid d_1} d^{1-2s} \left(q\left(s, d_0 \frac{d_1^2}{d^2}, \chi \right) \right)^2 = \sum_{d \mid d_1} \sigma_{2-4s}(d) \widehat{Q}\left(s, d_0 \frac{d_1^2}{d^2}, \chi \right).$$

Proof. Since the arithmetical functions involved are multiplicative, it is enough to verify the claim on prime powers $d_1 = p^n$, i.e., we need to verify

$$\sum_{i=0}^{n} p^{2i(\frac{1}{2}-s)} q^2(s, d_0 p^{2(n-i)}) = \sum_{i=0}^{n}\sum_{j=0}^{i} p^{4j(\frac{1}{2}-s)} Q(s, d_0 p^{2(n-i)}, \chi).$$

Using the definitions of $q(s, d, \chi)$ and $Q(s, d, \chi)$, it is a straightforward but tedious algebraic computation with sums and products of geometric sums. The details are omitted. □

Using the above lemma we can now show that many properties of $\widehat{Z}(s, w, \chi, \chi')$ can be understood on the basis of the properties of $Z(s, w, \chi, \chi')$. The following lemma implies in particular that $\widehat{Z}(s, w, \chi, \chi')$ admits a meromorphic continuation, and that any bound we have on $Z_{\psi_\tau}(s, w, \chi, \chi')$ translates into a bound for $\widehat{Z}(s, w, \chi, \chi')$.

Lemma 2.22. *Assume that $\psi = \psi_\tau$, i.e., $t_n = \tau(n)$. Then*

$$\begin{aligned}\widehat{Z}(s, w, \chi, \chi') = \frac{1}{2\zeta_2(2s+2w-1)}\big(&Z_{\psi_\tau}(s, w, \chi, \chi') + Z_{\psi_\tau}(s, w, \chi\chi_4, \chi') \\ &+ Z_{\psi_\tau}(s, w, \chi, \chi'\chi_4) - Z_{\psi_\tau}(s, w, \chi\chi_4, \chi'\chi_4)\big).\end{aligned}$$

Proof. We start by noticing that $L_2(s, \chi_{c_0}\chi)^2 = L_2(s, \psi_\tau \otimes \chi_{c_0}\chi)$. Now let d_0 be an odd squarefree natural number. Then

$$\begin{aligned}\zeta_2(2s+2w-1)\sum_{\substack{d_1=1\\ d_1 \text{ odd}}}^{\infty} \frac{q^2(s, d_0 d_1^2, \chi)}{d_1^{2w}} &= \sum_{\substack{d,d_1=1\\ d_1,\, d \text{ odd}}}^{\infty} \frac{d^{1-2s} q^2(s, d_0 d_1^2, \chi)}{(dd_1)^{2w}} \\ &= \sum_{\substack{d_1=1\\ d_1 \text{ odd}}}^{\infty} \frac{1}{d_1^{2w}} \sum_{d|d_1} d^{1-2s} q^2(s, d_0 d_1^2/d^2, \chi).\end{aligned}$$

We then use Lemma 2.21 and arrive at

$$\begin{aligned}&\sum_{\substack{d_1=1\\ d_1 \text{ odd}}}^{\infty} \frac{1}{d_1^{2w}} \sum_{d|d_1} \sigma_{2-4s}(d)\widehat{Q}(s, d_0 d_1^2/d^2, \chi) \\ &\qquad = \sum_{\substack{l=1\\ l \text{ odd}}}^{\infty} \frac{\sigma_{2-4s}(l)}{l^{2w}} \sum_{\substack{d_1=1\\ d_1 \text{ odd}}}^{\infty} \frac{\widehat{Q}(s, d_0 d_1^2, \chi)}{d_1^{2w}} \\ &\qquad = \zeta_2(4s+2w-2)\zeta_2(2w) \sum_{\substack{d_1=1\\ d_1 \text{ odd}}}^{\infty} \frac{\widehat{Q}(s, d_0 d_1^2, \chi)}{d_1^{2w}}.\end{aligned}$$

Multiply the first and last expression by $\chi'(d_0)L_2(s, \chi_{d_0}\chi)^2/d_0^w$, then summing over all odd squarefree natural numbers d_0 we get

$$\zeta_2(2s+2w-1)\sum_{\substack{d=1\\(d,2)=1}}^{\infty}\frac{\chi'(d)q^2(s,d,\chi)L_2(s,\chi_{d_0}\chi)^2}{d^w}\\ =\zeta_2(4s+2w-2)\zeta_2(2w)\sum_{\substack{d=1\\(d,2)=1}}^{\infty}\frac{\chi'(d)\widehat{Q}(s,d,\chi)L_2(s,\psi_\tau\otimes\chi_{d_0}\chi)}{d^w}.$$

By (2-6) we see that

$$\widehat{Q}(s,d,\chi)L_2(s,\psi_\tau\otimes\chi_{d_0}\chi)\\ =\begin{cases}Q_{\psi_\tau}(s,d,\chi)L_2(s,\psi_\tau\otimes\chi_{d_0}\chi) & \text{if } d\equiv 1 \pmod 4,\\ Q_{\psi_\tau}(s,d,\chi\chi_4)L_2(s,\psi_\tau\otimes\chi_{d_0}\chi\chi_4) & \text{if } d\equiv 3 \pmod 4.\end{cases}$$

Substituting $(s-w+\frac{1}{2}, 2w-\frac{1}{2})$ for (s,w) and comparing with (2-18), we obtain the desired result. $\square$

3. Eisenstein series

We briefly recall a few facts about Eisenstein series with weights. For $\gamma\in\mathrm{SL}_2(\mathbb{R})$ and $z\in\mathbb{H}$ we define $j(\gamma,z)=cz+d$ and $j_\gamma(z)=(cz+d)/|cz+d|$. We let arg denote the principal argument and define $j_\gamma(z)^k=e^{ik\arg(cz+d)}$. Since $j(\gamma_1\gamma_2,z)=j(\gamma_1,\gamma_2 z)j(\gamma_2,z)$,

$$\widetilde{\omega}(\gamma_1,\gamma_2)=\frac{1}{2\pi}(\arg j(\gamma_1,\gamma_2 z)+\arg j(\gamma_2,z)-\arg j(\gamma_1\gamma_2,z))$$

is an integer independent of z. The factor system of weight $k\in\mathbb{R}$ is then defined as

$$\omega(\gamma_1,\gamma_2)=e(k\widetilde{\omega}(\gamma_1,\gamma_2)).$$

Then we have $\omega(\gamma_1,\gamma_2)j_{\gamma_1\gamma_2}(z)^k=j_{\gamma_1}(\gamma_2 z)^k j_{\gamma_2}(z)^k$. We refer to [Iwaniec 1997, Chapters 2.6, 3] for the basic properties of multiplier systems as well as for further explanations of the generalities of Fourier expansions.

Let ν be a weight k multiplier system, and let Γ be a cofinite subgroup of $\mathrm{SL}_2(\mathbb{R})$. For an open cusp $\mathfrak{a}$, i.e., $\nu(\mathfrak{a})=1$, we define the weight k Eisenstein series for Γ by

$$E_{\mathfrak{a}}(z,s,k):=\sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma}\overline{\nu(\gamma)\omega(\sigma_{\mathfrak{a}}^{-1},\gamma)}j_{\sigma_{\mathfrak{a}}^{-1}\gamma}(z)^{-k}\Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)^s \quad \text{for } \Re(s)>1,$$

where $\sigma_{\mathfrak{a}}$ is a scaling matrix of the cusp $\mathfrak{a}$, i.e., $\sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\sigma_{\mathfrak{a}}=\Gamma_\infty$, with Γ_∞ being generated by $\gamma_\infty=\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $-\gamma_\infty$ if $-I\in\Gamma$. The function satisfies, for $\gamma\in\Gamma$, $E_{\mathfrak{a}}(\gamma z,s,k)=\nu(\gamma)j_\gamma^k(z)E_{\mathfrak{a}}(z,s,k)$, it is an eigenfunction of the weight k Laplacian with eigenvalue $s(1-s)$, and admits a meromorphic continuation to $s\in\mathbb{C}$. We now briefly recall how to find the Fourier coefficients of $E_{\mathfrak{a}}(z, s, k)$ at an open cusp $\mathfrak{b}$. We have

$$j_{\sigma_{\mathfrak{b}}}(z)^{-k} E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, k) = \sum_{\gamma \in \Gamma_\infty \backslash \sigma_{\mathfrak{a}}^{-1} \Gamma \sigma_{\mathfrak{b}}} \overline{\nu_{\mathfrak{ab}}(\gamma)}\, j_\gamma(z)^{-k} \Im(\gamma z)^s,$$

where $\nu_{\mathfrak{ab}}(\gamma) = \nu(\sigma_{\mathfrak{a}} \gamma \sigma_{\mathfrak{b}}^{-1}) \omega(\sigma_{\mathfrak{a}}^{-1}, \sigma_{\mathfrak{a}} \gamma \sigma_{\mathfrak{b}}^{-1}) \omega(\gamma \sigma_{\mathfrak{b}}^{-1}, \sigma_{\mathfrak{b}})$. For the rest of the paper we can assume that $-I \in \Gamma$. Summing over a set of representatives of $\Gamma_\infty \backslash \sigma_{\mathfrak{a}}^{-1} \Gamma \sigma_{\mathfrak{b}} / \Gamma_\infty$, which we can assume have $c_\gamma > 0$ for $\gamma \notin \Gamma_\infty$, we see that

$$j_{\sigma_{\mathfrak{b}}}(z)^{-k} E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, k) = \delta_{\mathfrak{a}=\mathfrak{b}} y^s + \sum_{I \neq \gamma \in \Gamma_\infty \backslash \sigma_{\mathfrak{a}}^{-1} \Gamma \sigma_{\mathfrak{b}} / \Gamma_\infty} \overline{\nu_{\mathfrak{ab}}(\gamma)} \sum_{l \in \mathbb{Z}} j_{\gamma \gamma_\infty^l}(z)^{-k} \Im(\gamma \gamma_\infty^l z)^s.$$

Therefore, by a familiar computation, we have

$$\begin{aligned}\int_0^1 \big(j_{\sigma_{\mathfrak{b}}}(z)^{-k} & E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, k) - \delta_{\mathfrak{a}=\mathfrak{b}} y^s\big) e(-nx)\, dx \\ &= \sum_{I \neq \gamma \in \Gamma_\infty \backslash \sigma_{\mathfrak{a}}^{-1} \Gamma \sigma_{\mathfrak{b}} / \Gamma_\infty} \frac{\overline{\nu_{\mathfrak{ab}}(\gamma)}}{c^{2s}} e\Big(n \frac{d}{c}\Big) y^s \int_{-\infty}^{\infty} \Big(\frac{z}{|z|}\Big)^{-k} \frac{1}{|z|^{2s}} e(-nx)\, dx.\end{aligned}$$

Substituting $t = x/y$ in the last integral we see that

$$\begin{aligned} y^s \int_{-\infty}^{\infty} \Big(\frac{z}{|z|}\Big)^{-k} \frac{e(-nx)}{|z|^{2s}}\, dx &= y^{1-s} \int_{-\infty}^{\infty} \Big(\frac{t+i}{|t+i|}\Big)^{-k} \frac{e(-nty)}{|t+i|^{2s}}\, dt \\ &= e^{-ik\pi/2} y^{1-s} \int_{-\infty}^{\infty} \Big(\frac{1-it}{|1-it|}\Big)^{-k} \frac{e(-nty)}{|1+it|^{2s}}\, dt \\ &= \begin{cases} \pi^s e^{-ik\pi/2} \dfrac{|n|^{s-1}}{\Gamma\big(s + \frac{kn}{2|n|}\big)} W_{\frac{kn}{2|n|}, s-\frac{1}{2}}(4\pi |n| y) & \text{if } n \neq 0, \\ \pi 4^{1-s} e^{-ik\pi/2} \dfrac{\Gamma(2s-1) y^{1-s}}{\Gamma\big(s+\frac{k}{2}\big)\Gamma\big(s-\frac{k}{2}\big)} & \text{if } n = 0, \end{cases}\end{aligned}$$

where $W_{\mu,\nu}(y)$ is the Whittaker function and where we have used [Gradshteyn and Ryzhik 2007, 3.384 (9), p. 349] for $n \neq 0$ and [Shimura 1975, p. 84–85] for $n = 0$.

3A. ***Eisenstein series of level 4.*** We now specialize to $\Gamma = \Gamma_0(4)$. In this case the Fourier coefficients of half-integral weight Eisenstein series were originally studied by Shimura [1975]. We consider the weight $\frac{1}{2}$ multiplier system ν related to the theta series

$$\theta(z) := y^{\frac{1}{4}} \sum_{m \in \mathbb{Z}} e(m^2 z),$$

i.e., $\theta(\gamma z) = \nu(\gamma) j_\gamma(z)^{\frac{1}{2}} \theta(z)$ for $\gamma \in \Gamma$. It is well known that

$$\nu(\gamma) = \left(\frac{c}{d}\right)\varepsilon_d^{-1} \quad \text{for} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \gamma \in \Gamma_0(4).$$

Here the Jacobi–Legendre symbol is extended as in [Shimura 1973, p. 442]. The group $\Gamma_0(4)$ has 3 cusps, $\mathfrak{a}_1 = \infty$, $\mathfrak{a}_2 = 0$, $\mathfrak{a}_3 = \frac{1}{2}$, with corresponding stabilizers $\Gamma_{\mathfrak{a}_i}$ generated by $\pm\gamma_{\mathfrak{a}_i}$ where

$$\gamma_{\mathfrak{a}_1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \gamma_{\mathfrak{a}_2} = \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}, \quad \gamma_{\mathfrak{a}_3} = \begin{pmatrix} -1 & 1 \\ -4 & 3 \end{pmatrix}$$

and we define scaling matrices

$$\sigma_{\mathfrak{a}_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_{\mathfrak{a}_2} = \begin{pmatrix} 0 & -\frac{1}{2} \\ 2 & 0 \end{pmatrix}, \quad \sigma_{\mathfrak{a}_3} = \begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 0 \end{pmatrix}.$$

Only the cusps ∞ and 0 are open with respect to ν, as

$$\nu(\gamma_{\mathfrak{a}_1}) = \nu(\gamma_{\mathfrak{a}_2}) = 1, \quad \nu(\gamma_{\mathfrak{a}_3}) = \left(\frac{-4}{3}\right)\epsilon_3^{-1} = i.$$

We now compute the Fourier expansion for the weight $\frac{1}{2}$ Eisenstein series. We focus on the cusp at infinity but the analysis for the other cusps is similar, although slightly more technical. The main extra complication at the other cusps comes from the factor system. This can be dealt with as follows: For $k = \frac{1}{2}$ we can use $z = \gamma_2^{-1} i$ in the definition of the factor system to see that

$$\omega(\gamma_1, \gamma_2) = \begin{cases} 1 & \text{if } -\pi < \arg(c_{\gamma_1} i + d_{\gamma_1}) + \arg(c_{\gamma_2} i + a_{\gamma_2}) \le \pi, \\ -1 & \text{otherwise.} \end{cases}$$

Using the properties of a multiplier system one finds (see [Iwaniec 1997, (3.5)]) that

$$\nu_{\mathfrak{a}\mathfrak{b}}(\gamma) = \nu(\sigma_{\mathfrak{a}} \gamma \sigma_{\mathfrak{b}}^{-1}) \frac{\omega(\sigma_{\mathfrak{a}} \gamma \sigma_{\mathfrak{b}}^{-1}, \sigma_{\mathfrak{b}})}{\omega(\sigma_{\mathfrak{a}}, \gamma)}.$$

This is explicit enough that one can do the computations also for the other cusps.

We now focus on $(\mathfrak{a}_1, \mathfrak{a}_1) = (\infty, \infty)$, and omit the corresponding subscripts. Using that all the nonidentity elements of $\Gamma_\infty \backslash \Gamma / \Gamma_\infty$ are parametrized by $\left(\begin{smallmatrix} * & * \\ 4c & d \end{smallmatrix}\right)$ with $c > 0$, $d \bmod 4c$, $(d, 4c) = 1$, we find that

$$E(z, s, \tfrac{1}{2}) = y^s + \phi(s, \tfrac{1}{2}) y^{1-s} + \sum_{n \neq 0} \phi_n(s, \tfrac{1}{2}) W_{n/(4|n|), s-\frac{1}{2}}(4\pi |n| y) e(nx)$$

with

$$\phi_n(s,\tfrac{1}{2}) = \frac{\pi^s e^{-i\pi/4}|n|^{s-1}}{\Gamma\big(s+\frac{n}{4|n|}\big)} \sum_{c=1}^{\infty} \frac{1}{(4c)^{2s}} \sum_{\substack{d \bmod 4c\\ (d,4c)=1}} \overline{\nu\begin{pmatrix} * & * \\ 4c & d\end{pmatrix}} e(nd/4c)$$

$$= \frac{\pi^s e^{-i\pi/4}|n|^{s-1}}{\Gamma\big(s+\frac{n}{4|n|}\big)} \sum_{c=1}^{\infty} \frac{1}{(4c)^{2s}} \sum_{d \bmod 4c} \varepsilon_d \Big(\frac{4c}{d}\Big) e(nd/4c), \tag{3-1}$$

and

$$\phi(s,\tfrac{1}{2}) = \frac{\pi 4^{1-s} e^{-i\pi/4}\Gamma(2s-1)}{\Gamma(s+\frac{1}{4})\Gamma(s-\frac{1}{4})} \sum_{c=1}^{\infty} \frac{1}{(4c)^{2s}} \sum_{d \bmod 4c} \varepsilon_d \Big(\frac{4c}{d}\Big).$$

If we write $4c = 2^k c'$ with c' odd then Sturm proved [1980, Lemma 1] — using quadratic reciprocity and the Chinese remainder theorem — that

$$\sum_{d \bmod 4c} \varepsilon_d \Big(\frac{4c}{d}\Big) e(nd/4c) = H_n(c') \sum_{r \bmod 2^k} \Big(\frac{2^k}{r}\Big) \varepsilon_r e(nr/2^k). \tag{3-2}$$

It follows that, for $n \neq 0$,

$$\phi_n\big(s,\tfrac{1}{2}\big) = \frac{\pi^s e^{-i\pi/4}|n|^{s-1}}{\Gamma\big(s+\frac{n}{4|n|}\big)} \sum_{\substack{c'=1\\ (c',2)=1}}^{\infty} \frac{H_n(c')}{c'^{2s}} \sum_{k=2}^{\infty} \frac{\sum_{r \bmod 2^k} \big(\frac{2^k}{r}\big)\varepsilon_r e(nr/2^k)}{2^{2ks}},$$

which by Lemma 2.2 equals

$$\frac{\pi^s e^{-i\pi/4}|n|^{s-1}}{\Gamma\big(s+\frac{n}{4|n|}\big)} \frac{L^*(2s-\frac{1}{2},n,1)}{\zeta_2(4s-1)} r_2(s,n), \tag{3-3}$$

where we have written

$$r_2(s,n) := \sum_{k=2}^{\infty} \frac{\sum_{r \bmod 2^k} \big(\frac{2^k}{r}\big)\varepsilon_r e(nr/2^k)}{2^{k2s}}. \tag{3-4}$$

The function $r_2(s,n)$ can also be computed. One uses that ε_d can be expressed as a sum of characters mod 4 as

$$\varepsilon_d = \tfrac{1}{2}(1+i)\chi_4^0(d) + \tfrac{1}{2}(1-i)\chi_4(d).$$

Inserting this in (3-4) the numerator becomes

$$\tfrac{1}{2}(1+i)G_n(\chi_8^k \chi_{2^k}^0) + \tfrac{1}{2}(1-i)G_n(\chi_8^k \chi_4 \chi_{2^k}^0), \tag{3-5}$$

where χ_8 is the primitive character mod 8 given by $\chi_8(n) = (-1)^{\frac{1}{8}(n-1)(n+1)}$ for $(n,2)=1$, and the G_n denote the usual Gauss sums. Using [Shimura 1975, Lemma 3] as well as explicit computations of $G_1(\chi_1)$, $G_1(\chi_8)$, $G_1(\chi_4)$, $G_1(\chi_4\chi_8)$, these can all be computed and using the result one can compute $r_2(s,n)$. We omit the details but state the result. Assume first $n \not\equiv 0 \pmod 4$. Then

$$r_2(s,n) = \tfrac{1}{4}(1+i)\begin{cases} -\dfrac{1}{2^{2(2s-1)}} & n \not\equiv 1 \pmod 4, \\ \dfrac{1}{2^{2(2s-1)}} + \dfrac{\chi_8(n)\sqrt{2}}{2^{3(2s-1)}} & n \equiv 1 \pmod 4. \end{cases} \tag{3-6}$$

More generally we find that, if $n = 4^r n_0$ with $n_0 \not\equiv 0 \pmod 4$, then

$$r_2(s,n) = \tfrac{1}{4}(1+i)u_r(2^{-(2s-1)}) + 4^{-r(2s-1)} r_2(s,n_0), \tag{3-7}$$

where

$$u_r(x) = \frac{(x^2)^{r+1} - x^2}{x^2-1}. \tag{3-8}$$

We remark that $r_2(s,n)$ is entire.

3A1. *Scattering term.* We now compute the scattering term $\phi(s,\frac{1}{2})$, which by (3-2) equals

$$\frac{\pi 4^{1-s} e^{-i\pi/4}\Gamma(2s-1)}{\Gamma(s+\frac{1}{4})\Gamma(s-\frac{1}{4})} \sum_{\substack{c'=1 \\ (c',2)=1}}^{\infty} \frac{H_0(c')}{c'^{2s}} \sum_{k=2}^{\infty} \sum_{r \bmod 2^k} \left(\frac{2^k}{r}\right) \frac{\varepsilon_r}{2^{2ks}}.$$

The sum $\displaystyle\sum_{\substack{c'=1 \\ (c',2)=1}}^{\infty} \frac{H_0(c')}{c'^{2s}}$ factors, and for an odd prime p we observe that

$$H_0(p^\beta) = \begin{cases} \varphi(p^\beta) & \text{if } \beta \equiv 0 \pmod 2, \\ 0 & \text{otherwise.} \end{cases}$$

Here φ is Euler's φ-function. Therefore

$$\sum_{\beta=0}^{\infty} \frac{H_0(p^\beta)}{p^{\beta 2s}} = \sum_{\beta=0}^{\infty} \frac{\varphi(p^{2\beta})}{p^{2\beta 2s}} = \frac{\zeta^{(p)}(4s-2)}{\zeta^{(p)}(4s-1)}.$$

For the prime 2 we note that for $k \ge 2$ we have

$$G_0(\chi_4\chi^0_{2^k}) = G_0(\chi_8\chi^0_{2^{k+1}}) = G_0(\chi_4\chi_8\chi^0_{2^{k+1}}) = 0$$

Using this, we find

$$\begin{aligned} &\sum_{k=2}^{\infty} \sum_{r \bmod 2^k} \left(\frac{2^k}{r}\right) \frac{\varepsilon_r}{2^{k2s}} \\ &\quad = \sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{\frac{1+i}{2} G_0(\chi_4^0 \chi^0_{2^k}) + \frac{1-i}{2} G_0(\chi_4\chi^0_{2^k})}{2^{k2s}} + \sum_{\substack{k=2 \\ k \text{ odd}}}^{\infty} \frac{\frac{1+i}{2} G_0(\chi_8\chi^0_{2^k}) + \frac{1-i}{2} G_0(\chi_8\chi_4\chi^0_{2^k})}{2^{k2s}} \\ &\quad = \sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{\frac{1+i}{2}\varphi(2^k)}{2^{k2s}} = (1+i)\frac{2^{-4s}}{1-2^{-(4s-2)}}. \end{aligned}$$

It follows that

$$\phi(s,\tfrac{1}{2})=\pi 4^{1-s}e^{-i\pi/4}\frac{\Gamma(2s-1)}{\Gamma(s+\frac{1}{4})\Gamma(s-\frac{1}{4})}\frac{(1+i)}{2^{4s}}\frac{\zeta(4s-2)}{\zeta_2(4s-1)}.$$

Using that $\Gamma(s+\frac{1}{4})\Gamma(s-\frac{1}{4})=\sqrt{\pi}2^{3/2-2s}\Gamma(2s-\frac{1}{2})$, this simplifies to

$$\frac{1}{2^{4s-1}-1}\frac{\xi(4s-2)}{\xi(4s-1)}, \tag{3-9}$$

where $\xi(s)=\pi^{-s/2}\Gamma(s/2)\zeta(s)$ (compare [Iwaniec 1997, p. 247–248]). The other entries in the scattering matrix $\Phi(s,\frac{1}{2})$ can be computed in a similar way and we find

$$\Phi(s,\tfrac{1}{2})=\begin{pmatrix}\frac{2^{-(4s-1)}}{1-2^{-(4s-2)}} & \frac{1-i}{2^{2s}}\\ \frac{1+i}{2^{2s}} & \frac{2^{-(4s-1)}}{1-2^{-(4s-2)}}\end{pmatrix}\frac{1-2^{-(4s-2)}}{1-2^{-(4s-1)}}\frac{\xi(4s-2)}{\xi(4s-1)}. \tag{3-10}$$

As a consistency check we note that a direct computation and the functional equation for ξ show that the scattering matrix verifies $\Phi(s,\frac{1}{2})\Phi(1-s,\frac{1}{2})=I$, as predicted by the general theory.

3B. ***Eisenstein series of level 2^n.*** We now consider the group $\Gamma_0(N)$, where $N=2^n$ with $n\geq 2$. Let χ be a Dirichlet character modulo N, and consider the weight $\frac{1}{2}$ multiplier system

$$\nu(\gamma)=\chi(d)\Big(\frac{c}{d}\Big)\varepsilon_d^{-1}\quad\text{for}\quad\begin{pmatrix}a&b\\c&d\end{pmatrix}=\gamma\in\Gamma_0(N).$$

We consider the corresponding Eisenstein series of weight $\frac{1}{2}$ at the cusp at 0, denoted by

$$E_{0,\chi}(z,s,\tfrac{1}{2}).$$

Similarly one denotes $E_{\infty,\chi}(z,s,\frac{1}{2})$ the corresponding Eisenstein series at the cusp ∞. The Fourier coefficients at infinity of the Eisenstein series at zero has a simpler 2-factor than the Eisenstein series at infinity. The stabilizer at 0 is generated by $\pm\gamma_0$ and has corresponding scaling matrix σ_0, where

$$\gamma_0=\begin{pmatrix}1&0\\-2^n&1\end{pmatrix},\quad \sigma_0=\begin{pmatrix}0&-1/\sqrt{2^n}\\ \sqrt{2^n}&0\end{pmatrix}.$$

From the general considerations in the beginning of Section 3 we find that the nonzero Fourier coefficients at infinity equal

$$\sum_{I\neq\gamma\in\Gamma_\infty\backslash\sigma_0^{-1}\Gamma_0(N)/\Gamma_\infty}\frac{\overline{\nu_{0\infty}(\gamma)}}{c^{2s}}e\Big(n\frac{d}{c}\Big)\pi^s e^{-i\pi/4}\frac{|n|^{s-1}}{\Gamma\big(s+\frac{n}{4|n|}\big)}W_{\frac{n}{4|n|},s-\frac{1}{2}}(4\pi|n|y).$$

After some computations one finds

$$\sum_{I\neq\gamma\in\Gamma_\infty\backslash\sigma_0^{-1}\Gamma_0(N)/\Gamma_\infty} \frac{\overline{\nu_{0\infty}(\gamma)}}{c^{2s}} e\left(n\frac{d}{c}\right) = \frac{i\chi(-1)}{N^s} \sum_{\substack{a=1\\(a,2)=1}}^{\infty} \frac{\chi(a)H_n(a)}{a^{2s}}$$

$$= \frac{i\chi(-1)}{N^s} \frac{L^*(2s-\frac{1}{2}, n, \chi)}{\zeta_2(4s-1)}, \tag{3-11}$$

where in the last equality we have used Lemma 2.2. Using this it is straightforward to see how $Z(s, w, \chi, \chi')$ relates directly to a Rankin–Selberg integral in the case where $\{t_n\}$ comes from a cusp form. Let ψ be a cuspidal Hecke newform of weight zero, and trivial multiplier for $\Gamma_0(2^k)$ with eigenvalue $s_0(1-s_0)$ and Fourier expansion

$$\psi(z) = \sum_{n\neq 0} b_n W_{0,s_0-\frac{1}{2}}(4\pi|n|y)e(nx). \tag{3-12}$$

Let χ be a Dirichlet character mod 8. Consider the twisted Maaß form

$$\psi\otimes\chi(z) = \sum_{n\neq 0} \chi(n) b_n W_{0,s_0-\frac{1}{2}}(4\pi|n|y)e(nx),$$

which is a weight zero cusp form for some $\Gamma_0(M)$ and character χ_0^M for some $M \mid \operatorname{lcm}(64, 2^k)$ and $8|M$. Let χ' be another Dirichlet character mod 8. Consider now the Rankin–Selberg integral

$$I(\psi, \chi, \chi', s, w) = \int_{\Gamma_0(M)\backslash\mathbb{H}} \psi\otimes\chi(z) E_{0,\chi_0^M\chi'}(z, w, \tfrac{1}{2}) \overline{E_{\infty,\chi_0^M\chi'}\big(z, \bar{s}, \tfrac{1}{2}\big)}\, d\mu(z).$$

This is the integral studied by Friedberg and Hoffstein [1995, (1.2) p. 388].

Unfolding, using $b_n = b_{n/|n|}|n|^{-\frac{1}{2}} t_{|n|}$, (3-11), and $L^*(s, -n, \chi) = L^*(s, n, \chi_4\chi)$ we arrive at

$$I(\psi, \chi, \chi', s, w)$$
$$= \frac{\pi^w e^{-i\pi/4} i\chi'(-1)}{(2\pi)^{s-1} M^w \zeta_2(4w-1)} \sum_{\substack{n\neq 0\\(n,2)=1}} \frac{\chi(n) b_{n/|n|} t_{|n|} L^*(2w-\frac{1}{2}, n, \chi')}{|n|^{s-w+\frac{1}{2}}} G_{\frac{n}{|n|}}(w)$$

$$= \frac{\pi^w e^{-i\pi/4} i\chi'(-1)\big[Z(s, w, \chi, \chi')G_+(w) + \chi(-1) b_{-1} Z(s, w, \chi, \chi_4\chi')G_-(w)\big]}{(2\pi)^{s-1} M^w \zeta_2(4w-1)\zeta_2(4s-1)}, \tag{3-13}$$

where

$$G_\pm(w) = \frac{1}{\Gamma(w\pm\frac{1}{4})} \int_0^\infty W_{\pm\frac{1}{4},w-\frac{1}{2}}(2y) W_{0,s_0-\frac{1}{2}}(2y) y^{w-1} \frac{dy}{y}.$$

Lemma 3.1. $I(\psi, \chi, \chi', \frac{1}{2}+it, \frac{1}{2}+iu) = O\big(\log((2+|t|)(2+|u|))\big)$.

Proof. This follows from the Maaß–Selberg relation, and known properties of the relevant scattering matrix. □

It is tempting to speculate whether the above bound on $I(\psi,\chi,\chi',\frac{1}{2}+it,\frac{1}{2}+iu)$ can be used to bound $Z(s,w,\chi,\chi')$ through (3-13). What we can prove is the following:

Denote the expression in the square brackets of (3-13) by $\tilde{I}(\psi,\chi,\chi',s,w)$. We then find that

$$\begin{aligned}\tilde{I}(\psi,\chi,\chi',s,w)&\pm\tilde{I}(\psi,\chi,\chi_4\chi',s,w)\\ &=(Z(s,w,\chi,\chi')\pm Z(s,w,\chi,\chi_4\chi'))(G_+(w)\pm\chi(-1)b_{-1}G_-(w)).\end{aligned}\tag{3-14}$$

Lemma 3.2. *Assume that ψ is a cusp form. Then, for $s=1-w=\frac{1}{2}+it$,*

$$Z(s,w,\chi,\chi')+\chi(-1)b_{-1}Z(s,w,\chi,\chi_4\chi')=O((1+|t|)^{\frac{1}{2}+\epsilon}).\tag{3-15}$$

Proof. From (3-14) we see that

$$\begin{aligned}Z(s,w,\chi,\chi')+\chi(-1)b_{-1}Z(s,w,\chi,\chi_4\chi')(G_+(w)+G_-(w))&\\ =\tilde{I}(\psi,\chi,\chi',s,w)+\chi(-1)b_{-1}\tilde{I}(\psi,\chi,\chi_4\chi',s,w).&\end{aligned}$$

The claim now follows from Lemmas 3.1 and A.1, combined with Remark A.2. □

Remark 3.3. We notice that with the restriction above on s, w the conductor $\mathfrak{q}(t,-t)$ is of order $(1+|t|)^2$. So the right-hand side in (3-15) is of order $\mathfrak{q}(t,-t)^{\frac{1}{4}+\epsilon}$, i.e., for the linear combination $Z(s,w,\chi,\chi')+\chi(-1)b_{-1}Z(s,w,\chi,\chi_4\chi')$ we have proved the convexity estimate unconditionally. Surprisingly this “soft” method of using the Maaß–Selberg relations gives much stronger bounds than the harder method using Heath-Brown’s equation (2-29) and approximate functional equations. Unfortunately we do not know how to prove this unconditionally for $Z(s,w,\chi,\chi')$ and $Z(s,w,\chi,\chi_4\chi')$ separately. The main reason for this is that $G_+(w)-G_-(w)$ decays much faster than $G_+(w)+G_-(w)$, so using a similar argument on

$$Z(s,w,\chi,\chi')-\chi(-1)b_{-1}Z(s,w,\chi,\chi_4\chi')$$

gives very poor bounds.

If we use (2-19) (i.e., interchange sums) we find, like [Friedberg and Hoffstein 1995, (1.2) p. 389], that $\tilde{I}(\psi,\chi,\chi',s,w)$ equals

$$\sum_{\substack{c=1\\(c,2)=1}}\frac{\chi'(c)L^{**}(s-w+\frac{1}{2},\psi,c,\chi)}{c^{2w-\frac{1}{2}}}(G_+(w)+\chi(-1)b_{-1}\chi_4(c)G_-(w)).$$

By taking linear combinations over different χ' we can restrict to c in a specific residue class, as in the work of [Friedberg and Hoffstein 1995].

4. Limits of weight $\frac{1}{2}$ Eisenstein series

We consider separately Maaß cusp forms and incomplete Eisenstein series, i.e., we analyze

$$\int_{\Gamma\backslash\mathbb{H}} \psi(z)|E(z, \tfrac{1}{2}+it, \tfrac{1}{2})|^2 \, d\mu(z),$$

where ψ is either a Maaß cusp form or an incomplete Eisenstein series. Then a standard approximation argument — see [Luo and Sarnak 1995, p. 217] — implies the result (1-10).

4A. ***The cuspidal contribution.*** Let ψ be a cuspidal element of a weight zero Hecke basis for $\Gamma_0(4)$ with eigenvalue $s_0(1-s_0)$ and Fourier expansion

$$\psi(z) = \sum_{n\neq 0} b_n W_{0,s_0-\frac{1}{2}}(4\pi|n|y)e(nx).$$

We will freely use that we can assume that the Fourier coefficients are real.

We want to study

$$\int_{\Gamma\backslash\mathbb{H}} \psi(z)|E(z, s, \tfrac{1}{2})|^2 \, d\mu(z)$$

when $\Re(s) = \frac{1}{2}$. It turns out to be convenient to consider the slightly more general integral

$$I(s, w) = \int_{\Gamma\backslash\mathbb{H}} \psi(z)E(z, w, \tfrac{1}{2})\overline{E(z, \bar{s}, \tfrac{1}{2})} \, d\mu(z).$$

For sufficiently large $\Re(s)$, we can unfold to get

$$I(s, w) = \int_{\Gamma_\infty\backslash\mathbb{H}} \psi(z)E(z, w, \tfrac{1}{2})y^s \, d\mu(z). \tag{4-1}$$

Using the Fourier expansions of ψ and $E_\infty(z, w, \frac{1}{2})$, and computing the x-integral, we find

$$\begin{aligned} I(s, w) &= \int_0^\infty \sum_{n\neq 0} b_n\phi_{-n}(w, \tfrac{1}{2}) W_{0,s_0-\frac{1}{2}}(4\pi|n|y) W_{-\frac{1}{4}n/|n|, w-\frac{1}{2}}(4\pi|n|y) y^{s-1} \frac{dy}{y} \\ &= \sum_{n\neq 0} \frac{b_n\phi_{-n}(w, \frac{1}{2})}{(2\pi|n|)^{s-1}} \int_0^\infty W_{0,s_0-\frac{1}{2}}(2y) W_{-\frac{1}{4}n/|n|, w-\frac{1}{2}}(2y) y^{s-1} \frac{dy}{y}. \end{aligned} \tag{4-2}$$

We consider the series

$$Z_\pm(s, w) := \frac{\Gamma(w \mp \frac{1}{4})}{\pi^w e^{-i\pi/4}} \zeta_2(4s-1)\zeta_2(4w-1) \sum_{\pm n=1}^{\infty} \frac{b_n\phi_{-n}(w, \frac{1}{2})}{|n|^{s-1}}.$$

By (3-3) we see that

$$Z_{\pm}(s,w)=\zeta_2(4s-1)\sum_{\pm n=1}^{\infty}\frac{b_n r_2(w,-n)L^*(2w-\frac{1}{2},-n,1)}{|n|^{s-w}}. \tag{4-3}$$

The next proposition reduces many questions about $Z_{\pm}(s,w)$ to questions about $Z(s,w,\chi,\chi')$. Consider the Dirichlet polynomial

$$T(s,w):=\prod_{\epsilon\in\{\pm 1\}} p_2(\epsilon 2^{-(s+w-\frac{1}{2})})p_2(\epsilon 2^{-(s-w+\frac{1}{2})}),$$

where $p_2(z)$ is defined in (2-24).

Proposition 4.1. *There exist functions $f_{\pm}(s,w,\chi,\chi')$ bounded in vertical strips such that*

$$T(s,w)Z_{\pm}(s,w)=\sum_{\chi,\chi'} f_{\pm}(s,w,\chi,\chi')Z(s,w,\chi,\chi'),$$

where the sum is over all pairs of characters mod 8.

Proof. We first assume that ψ is a newform. Then we have

$$b_n=b_{n/|n|}|n|^{-\frac{1}{2}}t_{|n|},$$

where $\{t_n\}_{n\in\mathbb{N}}$ are the coefficients of $L(s,\psi)$. We note that if $m\geq 1$ is odd then $\chi_{(\pm 2^l m)_0}=\chi_{m_0}\chi$ where m_0 denotes the squarefree part of m for some character χ whose conductor divides 8, namely

$$\chi(d)=\begin{cases}\left(\dfrac{\pm 2}{d}\right) & \text{if } l \text{ odd},\\[2mm] \left(\dfrac{\pm 1}{d}\right) & \text{if } l \text{ even}.\end{cases} \tag{4-4}$$

Notice that χ depends only on l mod 2 and the sign $\pm$. For the same χ we have $q(w,m,\chi)=q(w,\pm 2^l m)$. It follows that $L^*(s,m,\chi)=L^*(s,\pm 2^l m,1)$. We write the summation index n in (4-3) as $n=2^l m$, where m is odd, and split the sum as

$$\sum_{\substack{l=0\\ l \text{ odd}}}^{\infty}\sum_{\substack{\pm m=1\\ (m,2)=1}}^{\infty}\cdots+\sum_{\substack{l=0\\ l \text{ even}}}^{\infty}\sum_{\substack{\pm m=1\\ (m,2)=1}}^{\infty}\cdots.$$

We split the m sum further according to $m\equiv 1,3,5,7 \pmod 8$, which can be done by using a linear combination of characters. We then use the explicit formulae for $r_2(w,-n)$ in (3-6), (3-7) and that the Fourier coefficients satisfy the Hecke relations to see that $Z_{\pm}(s,w)$ can be written as a linear combination of $Z(s,w,\chi,\chi')$ with coefficients being functions bounded on vertical strips multiplied by one of the following series:

$$\sum_{j=0}^{\infty} \frac{t_{2^{2j}}}{2^{2j(s+w-\frac{1}{2})}}, \quad \sum_{j=0}^{\infty} \frac{t_{2^{2j+1}}}{2^{(2j+1)(s+w-\frac{1}{2})}},$$
$$\sum_{j=0}^{\infty} \frac{t_{2^{2j}} u_j(2^{-(2w-1)})}{2^{2j(s-w+\frac{1}{2})}}, \quad \sum_{j=0}^{\infty} \frac{t_{2^{2j+1}} u_j(2^{-(2w-1)})}{2^{(2j+1)(s-w+\frac{1}{2})}}. \tag{4-5}$$

We easily see that

$$2\sum_{j=0}^{\infty} \frac{t_{2^{2j}}}{2^{2js}} = \frac{1}{p_2(2^{-s})} + \frac{1}{p_2(-2^{-s})}, \quad 2\sum_{j=0}^{\infty} \frac{t_{2^{2j+1}}}{2^{(2j+1)s}} = \frac{1}{p_2(2^{-s})} - \frac{1}{p_2(-2^{-s})}.$$

We see also that, using (3-8),

$$\sum_{j=0}^{\infty} \frac{t_{2^{2j}} u_j(x)}{2^{2js}} = \frac{x^2}{2(1-x^2)}\left(\frac{1}{p_2(2^{-s})} + \frac{1}{p_2(-2^{-s})} - \frac{1}{p_2(x2^{-s})} - \frac{1}{p_2(-x2^{-s})}\right),$$

which has no poles coming out of x^2-1 in the denominator. Similarly, we see that

$$\sum_{j=0}^{\infty} \frac{t_{2^{2j+1}} u_j(x)}{2^{(2j+1)s}} = \frac{x^2}{2(1-x^2)}\left(\frac{1}{p_2(2^{-s})} - \frac{1}{p_2(-2^{-s})} - \frac{1}{x}\left(\frac{1}{p_2(x2^{-s})} - \frac{1}{p_2(-x2^{-s})}\right)\right).$$

We substitute in the last four equations $s+w-\frac{1}{2}$ or $s-w+\frac{1}{2}$ for s as required and $x = 2^{-(2w-1)}$ to identify the possible polynomials that appear in the denominators. These have product $T(s,w)$. We now notice that multiplying any of the 4 functions in (4-5) by $T(s,w)$ we get holomorphic functions bounded on vertical strips, which proves the claim.

If ψ is an oldform with, say, $\psi = \psi_1(2^j z)$ with ψ_1 a primitive form, and $j = 1, 2$, then the series in (4-2) becomes

$$\sum_{n\neq 0} \frac{b_n(\psi_1)\phi_{-2^j n}(w, \frac{1}{2})}{(2\pi|2^j n|)^{s-1}},$$

which by the explicit expression for $\phi_n(w, \frac{1}{2})$ can be analyzed similarly to the newform case. □

Remark 4.2. In Theorem 4.3 below, we need to study $Z_{\pm}(\frac{1}{2}+it, \frac{1}{2}-it)$. For $\Re(s) = \Re(w) = \frac{1}{2}$ we notice that by (2-25) we have $1/T(s,w) = O(1)$.

Theorem 4.3. *Assume that for any χ, χ' mod 8 the function $Z(s, 1-s, \chi, \chi')$ satisfies a subconvex bound. Then*

$$\int_{\Gamma\backslash\mathbb{H}} \psi(z)|E(z, \tfrac{1}{2}+it, \tfrac{1}{2})|^2\, d\mu(z) \to 0 \quad \text{as } |t| \to \infty.$$

Proof. By Proposition 4.1, a subconvex bound with saving δ translates into a bound $Z_{\pm}(s, 1-s) = O(|t|^{2(\frac{1}{4}-\delta)})$ when $\Re(s) = \frac{1}{2}$. Combining this with the bound in Lemma A.1, the estimate $1/\zeta(1+it) = O(\log|t|)$ [Titchmarsh 1986, Equation 3.11.8], and the identity (4-2) we see that $I(s, 1-s) = O(|t|^{2(\frac{1}{4}-\delta)-\frac{1}{2}+\varepsilon})$ for any $\varepsilon > 0$ when $\Re(s) = \frac{1}{2}$. Since

$$I(\tfrac{1}{2}+it, \tfrac{1}{2}-it) = \int_{\Gamma\backslash\mathbb{H}} \psi(z)|E(z, \tfrac{1}{2}-it, \tfrac{1}{2})|^2\, d\mu(z),$$

we find that, when $\delta > 0$, $I(\frac{1}{2}+it, \frac{1}{2}-it) \to 0$ as $|t| \to \infty$. □

Remark 4.4. In the proof above we see that the trivial bound from Theorem 2.17 only gives $O(|t|^{\frac{1}{2}+\varepsilon})$.

4B. ***The incomplete Eisenstein series contribution.*** In the following we choose a fundamental domain of Γ such that

$$\mathcal{D} = \mathcal{D}_0 \cup \bigcup_{j=1}^{3} \sigma_{\mathfrak{a}_j}\mathcal{D}^Y,$$

where $\mathcal{D}^Y := \{x+iy : 0 < x < 1, y > Y\}$, Y sufficiently large, $\mathcal{D}_0$ is a suitable compact set and, as before, $\sigma_{\mathfrak{a}_j}$ denotes the scaling matrix of the cusp $\mathfrak{a}_j$.

In order to introduce the incomplete Eisenstein series, let $h(y) \in C^\infty(\mathbb{R}^+)$ be a function which decreases rapidly at 0 and ∞, and whose derivatives are also of rapid decay. Its Mellin transform evaluated at $-s$ is

$$H(s) = \int_0^\infty h(y)y^{-s}\,\frac{dy}{y} \tag{4-6}$$

and thus by the Mellin inversion formula we have

$$h(y) = \frac{1}{2\pi i}\int_{\Re s = a} H(s)y^s\, ds \tag{4-7}$$

for any $a \in \mathbb{R}$. The function $H(s)$ is entire and $H(a+it)$ is in the Schwartz space in the t variable for any $a \in \mathbb{R}$. The incomplete Eisenstein series corresponding to the cusp $\mathfrak{a}$ is then given by

$$F_h(z, \mathfrak{a}) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma} h(\Im\sigma_{\mathfrak{a}}^{-1}\gamma z) = \frac{1}{2\pi i}\int_{\Re s = a > 1} H(s)E_{\mathfrak{a}}(z, s, 0)\, ds. \tag{4-8}$$

For $i = 1, 2, 3$ we are interested in the behavior of

$$J(t, \mathfrak{a}_i) = \int_{\Gamma\backslash\mathbb{H}} F_h(z, \mathfrak{a}_i)|E(z, \tfrac{1}{2}+it, \tfrac{1}{2})|^2\, d\mu(z) \quad \text{as } |t| \to \infty.$$

In the following we only treat the contribution from the cusp at infinity, but the other contributions can be dealt with similarly. Unfolding the incomplete Eisenstein series we find

$$J(t,\infty)=\int_{\Gamma\backslash\mathbb{H}} F_h(z,\infty)\,d\mu_t(z)=J_1(t,\infty)+J_2(t,\infty)$$

with

$$J_1(t,\infty):=\int_0^\infty h(y)\bigl|y^{\frac12+it}+\phi(\tfrac12+it,\tfrac12)y^{\frac12-it}\bigr|^2\frac{dy}{y^2}, \tag{4-9}$$

$$J_2(t,\infty):=\int_0^\infty h(y)\sum_{n\neq 0}\bigl|\phi_n(\tfrac12+it,\tfrac12)\bigr|^2\bigl|W_{n/(4|n|),it}(4\pi|n|y)\bigr|^2\frac{dy}{y^2}. \tag{4-10}$$

The integral $J_1(t,\infty)$ is easily dealt with. Namely, we obtain

$$\begin{aligned}
&J_1(t,\infty)\\
&\quad=\bigl(1+|\phi(\tfrac12+it,\tfrac12)|^2\bigr)\int_0^\infty h(y)\frac{dy}{y}+\phi(\tfrac12+it,\tfrac12)\int_0^\infty h(y)y^{-2it}\frac{dy}{y}\\
&\qquad\qquad\qquad\qquad\qquad\qquad+\overline{\phi(\tfrac12+it,\tfrac12)}\int_0^\infty h(y)y^{2it}\frac{dy}{y}\\
&\quad=\bigl(1+|\phi(\tfrac12+it,\tfrac12)|^2\bigr)H(0)+\phi(\tfrac12+it,\tfrac12)H(2it)+\overline{\phi(\tfrac12+it,\tfrac12)}H(-2it)\\
&\quad=O(1).
\end{aligned} \tag{4-11}$$

For the integral $J_2(t,\infty)$ we find, using the rapid decay of the Whittaker function and the Mellin inversion formula,

$$J_2(t,\infty)=\frac{1}{2\pi i}\int_{\Re s=a>1} H(s)R_1\bigl(|E(z,\tfrac12+it,\tfrac12)|^2,s\bigr)\,ds, \tag{4-12}$$

where

$$\begin{aligned}
R_1\bigl(|E(z,w,\tfrac12)|^2,s\bigr)&=\sum_{n\neq 0}|\phi_n(w,\tfrac12)|^2\int_0^\infty\bigl|W_{n/(4|n|),w-\frac12}(4\pi|n|y)\bigr|^2y^{s-1}\frac{dy}{y}\\
&=\sum_{n\neq 0}\frac{|\phi_n(w,\tfrac12)|^2}{(2\pi|n|)^{s-1}}\int_0^\infty\bigl|W_{n/(4|n|),w-\frac12}(2y)\bigr|^2y^{s-1}\frac{dy}{y}.
\end{aligned} \tag{4-13}$$

In order to analyze the asymptotic behavior of $J_2(t,\infty)$ we need to understand the function $R_1\bigl(|E(z,w,\tfrac12)|^2,s\bigr)$. There are (at least) two ways to do this: to use properties of the double Dirichlet series we defined in Section 2, or to use Zagier's theory of the Rankin–Selberg method for functions that are not of rapid decay but satisfy a certain mild growth condition. We will actually use a combination of these two techniques. We want to shift the line of integration in (4-12) to $\Re(s)=\frac12$. For this we need to identify the poles, estimate them, see what the contribution of $\int_{\Re(s)=\frac12} H(s)R_1\bigl(|E(z,w,\tfrac12)|^2,s\bigr)$ is to the asymptotics. For the first and third aspect we use the Rankin–Selberg approach and for the second aspect the multiple Dirichlet approach works best.

We first describe why double Dirichlet series techniques apply. The growth of the Mellin transform of the absolute value of the Whittaker function is analyzed in Lemma A.5. By combining (3-3), (3-6), and (3-7) we see that $\overline{\phi_n(\bar{w},\frac{1}{2})}=\phi_n(w,\frac{1}{2})$. This shows that when $\Re(w)=\frac{1}{2}$ we have

$$|\phi_n(w,\tfrac{1}{2})|^2=\phi_n(w,\tfrac{1}{2})\phi_n(1-w,\tfrac{1}{2}).$$

The right-hand side has the advantage of being meromorphic in w. We define

$$\widehat{Z}_\pm(s,w)=\frac{\Gamma(w\pm\frac{1}{4})\Gamma(1-w\pm\frac{1}{4})}{\pi i}\sum_{\pm n=1}^{\infty}\frac{\phi_n(w,\frac{1}{2})\phi_n(1-w,\frac{1}{2})}{|n|^{s-1}},$$

which by (3-3) equals

$$\frac{1}{\zeta_2(4w-1)\zeta_2(4(1-w)-1)}\times\sum_{\pm n=1}^{\infty}\frac{L^*(2w-\frac{1}{2},n,1)L^*(2(1-w)-\frac{1}{2},n,1)}{|n|^s}r_2(w,n)r_2(1-w,n).$$

We now show that $\widehat{Z}_\pm(s,w)$ is directly related to the function $\widehat{Z}(s,w,\chi,\chi')$ defined in (2-40). Let

$$U(s,w)=(1-2^{-(4w-1)})(1-2^{-2s})(1-2^{-(4w-2+2s)})(1-2^{-(-4w+2+2s)}).$$

Proposition 4.5. *There exist functions $\hat{f}_{\pm,\kappa}(s,w,\chi,\chi')$ bounded in vertical strips such that*

$$U(s,w)\widehat{Z}_\pm(s,w)=\frac{1}{\zeta_2(4w-1)\zeta_2(4(1-w)-1)}\sum_{\kappa\in\{0,1\}}\frac{\Gamma(\frac{1}{2}(2w-\frac{1}{2}+\kappa))}{\Gamma\big(\frac{1}{2}(2(1-w)-\frac{1}{2}+\kappa)\big)}\times\sum_{\chi,\chi'}\hat{f}_{\pm,\kappa}(s,w,\chi,\chi')\widehat{Z}\bigg(\frac{s+2w-\frac{1}{2}}{2},\frac{s-2w+\frac{3}{2}}{2},\chi,\chi'\bigg).$$

Proof. As in the proof of Proposition 4.1 we write $n=2^l m$ and split into sums over l even, odd respectively. We then split the m sum according to the residue class mod 8 which is a linear combination over characters mod 8. Inserting the explicit formulae for $r_2(w,n)$, (3-6), (3-7) we are led to consider the series

$$\sum_{j=0}^{\infty}u_j(x)u_j(y)z^j,\quad\sum_{j=0}^{\infty}u_j(x)z^j,\quad\sum_{j=0}^{\infty}z^j$$

with x, y, z being appropriate powers of 2. Since these are all sums of geometric series — see (3-8) — they are explicitly computable and after multiplying by

$(1-2^{-2s})(1-2^{-(4w-2+2s)})(1-2^{-(-4w+2+2s)})$ they become Dirichlet polynomials in powers of 2, hence holomorphic and bounded in vertical strips. Therefore

$$(1-2^{-2s})(1-2^{-(4w-2+2s)})(1-2^{-(-4w+2+2s)})\widehat{Z}_{\pm}(s,w) \\ =\sum_{\chi,\chi'}\tilde{f}_{\pm}(s,w,\chi,\chi')\widetilde{Z}(s,w,\chi,\chi'),$$

where

$$\widetilde{Z}(s,w,\chi,\chi') \\ =\frac{1}{\zeta_2(4w-1)\zeta_2(4(1-w)-1)}\sum_{n=1}^{\infty}\frac{\chi'(n)L^*(2w-\frac{1}{2},n,\chi)L^*(2(1-w)-\frac{1}{2},n,\chi)}{n^s}$$

and $\tilde{f}_{\pm}(s,w,\chi,\chi')$ are bounded in vertical strips. Using the functional equation on $L^*(2(1-w)-\frac{1}{2},n,\chi)$ we see — as in the proof of Theorem 2.11 — that

$$(1-2^{-(4w-1)})\sum_{n=1}^{\infty}\frac{\chi'(n)L^*(2w-\frac{1}{2},n,\chi)L^*(2(1-w)-\frac{1}{2},n,\chi)}{n^s} \\ =\sum_{\kappa\in\{0,1\}}\frac{\Gamma(\frac{1}{2}(2w-\frac{1}{2}+\kappa))}{\Gamma(\frac{1}{2}(2(1-w)-\frac{1}{2}+\kappa))}\sum_{\chi,\chi'}\tilde{\tilde{f}}_{\kappa}(x,y,\chi,\chi')\widetilde{\widetilde{Z}}(s,w,\chi,\chi'),$$

where $\tilde{\tilde{f}}_{\kappa}(x,y,\chi,\chi')$ is another set of functions bounded in vertical strips and

$$\widetilde{\widetilde{Z}}(s,w,\chi,\chi')=\sum_{n=1}^{\infty}\frac{\chi'(n)L^*(2w-\frac{1}{2},n,\chi)^2}{n^{s-2w+1}}.$$

Combining the above equations and comparing with (2-40) finishes the proof. □

The above lemma implies that many questions about $R_1\big(|E(z,w,\frac{1}{2})|^2,s\big)$ can be dealt with using $Z(s,w,\chi,\chi')$. We now describe a different method for understanding $R_1\big(|E(z,w,\frac{1}{2})|^2,s\big)$, namely Zagier's Rankin–Selberg method for functions not of rapid decay. This method was introduced by Zagier [1981] for the group $\mathrm{SL}_2(\mathbb{Z})$ and generalized by Kudla (unpublished), Dutta Gupta [1997], and Mizuno [2005]. Its usefulness for determining the contribution of the incomplete Eisenstein series to the asymptotics can already be seen in [Zelditch 1991]. We introduce the generalized Rankin–Selberg transform, following [Zagier 1981] and [Mizuno 2005]. We write $e_{ij}(y,s,k)=\delta_{ij}y^s+\phi_{ij}(s,k)y^{1-s}$ for the zero Fourier coefficient of $E_{\mathfrak{a}_i}(z,s,k)$ at $\mathfrak{a}_j$ and we denote the scattering matrix by $\Phi(s,k)=(\phi_{ij}(s,k))$. We note that for $\Gamma_0(4)$ the matrix $\Phi(s,0)$ is 3×3 whereas $\Phi(s,\frac{1}{2})$ is 2×2. For the weight 0 Eisenstein series we use the notation $E_i(z,s,0)=E_{\mathfrak{a}_i}(z,s,0)$.

Theorem 4.6 [Mizuno 2005, Theorem 2]. *Let F be a continuous functions on $\mathbb{H}$ that is Γ-invariant and satisfies, for $i = 1, 2, 3$,*

$$F(\sigma_{\mathfrak{a}_i} z) = \psi_i(y) + O(y^{-N}) \quad \text{for all } N \text{ as } y \to \infty,$$

where

$$\psi_i(y) = \sum_{j=1}^{l} \frac{c_{ij}}{n_{ij}!}\, y^{\alpha_{ij}} \log^{n_{ij}} y, \quad n_{ij} \in \mathbb{N} \cup \{0\}, \quad i = 1, 2, 3.$$

For such a function F the Rankin–Selberg transform $R_i(F, s)$ corresponding to the cusp $\mathfrak{a}_i$, $i = 1, 2, 3$, is defined by

$$R_i(F, s) := \int_0^\infty \int_0^1 (F(\sigma_{\mathfrak{a}_i} z) - \psi_i(y)) y^s \, d\mu(z),$$

for $\Re s$ sufficiently large. Then we have

$$\begin{aligned}
R_i(F, s) = {} & \int_{\mathcal{D}_0} F(z) E_i(z, s, 0)\, d\mu(z) \\
& + \sum_{j=1}^{3} \int_{\mathcal{D}^Y} \bigl(F(\sigma_{\mathfrak{a}_j} z) E_i(\sigma_{\mathfrak{a}_j} z, s, 0) - \psi_j(y) e_{ij}(y, s, 0)\bigr)\, d\mu(z) \\
& \qquad + \sum_{j=1}^{3} \phi_{ij}(s, 0) \int_Y^\infty \psi_j(y) y^{-s-1}\, dy - \int_0^Y \psi_i(y) y^{s-2}\, dy \\
= {} & \int_{\mathcal{D}_0} F(z) E_i(z, s, 0)\, d\mu(z) \\
& + \sum_{j=1}^{3} \int_{\mathcal{D}^Y} \bigl(F(\sigma_{\mathfrak{a}_j} z) E_i(\sigma_{\mathfrak{a}_j} z, s, 0) - \psi_j(y) e_{ij}(y, s, 0)\bigr)\, d\mu(z) \\
& \qquad - \sum_{j=1}^{3} \phi_{ij}(s, 0) \widehat{\psi_j}(1 - s, Y) - \widehat{\psi_i}(s, Y),
\end{aligned} \tag{4-14}$$

where

$$\widehat{\psi_i}(s, Y) = \sum_{j=1}^{l} c_{ij} \sum_{m=0}^{n_{ij}} \frac{(-1)^{n_{ij}-m}}{m!} \frac{Y^{s+\alpha_{ij}-1} \log^m Y}{(s + \alpha_{ij} - 1)^{n_{ij}-m+1}}.$$

Furthermore, for each $i = 1, 2, 3$, the function $R_i(F, s)$ can be meromorphically continued to $\mathbb{C}$ and we have the functional equation

$$\mathcal{R}(F, s) := {}^t(R_1(F, s), R_2(F, s), R_3(F, s)) = \Phi(s, 0)\mathcal{R}(F, 1 - s).$$

We want to move the line of integration in (4-12) to $\Re(s) = \frac{1}{2}$ and Theorem 4.6 plays a major role, as it allows to identify the relevant poles and to calculate the corresponding residues. By the above theorem, in particular by (4-14), we infer

$$\begin{aligned}
&R_1\big(|E(z,\tfrac12+it,\tfrac12)|^2,s\big)\\
&=\int_{\mathfrak{D}_0}|E(z,\tfrac12+it,\tfrac12)|^2E_1(z,s,0)\,d\mu(z)-\widehat{\psi(s,Y)}\\
&+\sum_{j=1}^{3}\int_{\mathfrak{D}^Y}\big(|E(\sigma_{\mathfrak{a}_j}z,\tfrac12+it,\tfrac12)|^2E_1(\sigma_{\mathfrak{a}_j}z,s,0)-\psi_j(y)e_{1j}(y,s,0)\big)\,d\mu(z), \qquad (4\text{-}15)
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\psi(s,Y)}&=\widehat{\psi_1}(s,Y)+\phi_{11}(s,0)\widehat{\psi_1}(1-s,Y)+\frac{Y^{1-s}}{1-s}\phi_{12}(s,0)|\phi_{12}(\tfrac12+it,\tfrac12)|^2,\\
\widehat{\psi_1}(s,Y)&=\frac{Y^s}{s}\big(1+|\phi_{11}(\tfrac12+it,\tfrac12)|^2\big)+\frac{Y^{s-2it}}{s-2it}\phi_{11}(\tfrac12+it,\tfrac12)+\frac{Y^{s+2it}}{s+2it}\overline{\phi_{11}(\tfrac12+it,\tfrac12)},\\
\psi_j(y)&=\big|\delta_{1j}y^{\frac12+it}+\phi_{1j}(\tfrac12+it,\tfrac12)y^{\frac12-it}\big|^2, \qquad j=1,2,\\
\psi_3(y)&=0.
\end{aligned}$$

Thus we easily see that we pick up residues at $s=1$ and $s=1\pm 2it$ when we shift the line of integration. The pole at $s=1$ is responsible for the contribution of the $\log|t|$ term in (1-10), as we will see. We therefore examine $H(s)R_1\big(|E(z,\frac12+it,\frac12)|^2,s\big)$ at $s=1$. In order to determine the order of the pole at $s=1$ and its residue we use the Laurent expansion of $H(s)$ and $R_1\big(|E(z,\frac12+it,\frac12)|^2,s\big)$. The first two terms of (4-15) are easily understood because of the Eisenstein series, which has simple poles at $s=1$ and no other poles in $\Re(s)\geq\frac12$. In order to treat the last term of (4-15) we write

$$\begin{aligned}
\frac{Y^{1-s}}{1-s}&=-\frac{1}{s-1}+\log Y+O(|s-1|),\\
\phi_{1j}(s,0)&=\frac{1}{\operatorname{vol}(\Gamma\backslash\mathbb{H})}\frac{1}{s-1}+b_0^{1j}+O(|s-1|).
\end{aligned}$$

These expansions and the fact that the scattering matrix $\Phi(s,\frac12)=(\phi_{ij}(s,\frac12))_{1\leq i,j\leq 2}$ is unitary for $\Re(s)=\frac12$ (see [Roelcke 1966, Lemma 10.5]) yield

$$\begin{aligned}
&\widehat{\psi(s,Y)}\\
&=-\frac{1}{\operatorname{vol}(\Gamma\backslash\mathbb{H})}\Big(1+\sum_{j=1}^{2}|\phi_{1j}(\tfrac12+it,\tfrac12)|^2\Big)\frac{1}{(s-1)^2}\\
&\quad+\Bigg(\Big(1+\sum_{j=1}^{2}|\phi_{1j}(\tfrac12+it,\tfrac12)|^2\Big)\frac{\log Y}{\operatorname{vol}(\Gamma\backslash\mathbb{H})}\\
&\qquad-\big(1+|\phi_{11}(\tfrac12+it,\tfrac12)|^2\big)b_0^{11}-|\phi_{12}(\tfrac12+it,\tfrac12)|^2b_0^{12}\\
&\qquad+\frac{1}{\operatorname{vol}(\Gamma\backslash\mathbb{H})}\frac{\overline{\phi_{11}(\frac12+it,\frac12)}Y^{2it}-\phi_{11}(\frac12+it,\frac12)Y^{-2it}}{2it}\Bigg)\frac{1}{s-1}+O(1)
\end{aligned}$$

$$= -\frac{2}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\frac{1}{(s-1)^2} + \biggl(\frac{2\log Y}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} - \bigl(1+|\phi_{11}(\tfrac12+it,\tfrac12)|^2\bigr)b_0^{11} - |\phi_{12}(\tfrac12+it,\tfrac12)|^2 b_0^{12} + \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\frac{\overline{\phi_{11}(\tfrac12+it,\tfrac12)}Y^{2it} - \phi_{11}(\tfrac12+it,\tfrac12)Y^{-2it}}{2it}\biggr)\frac{1}{s-1} + O(1).$$

Consequently we see that $R_1(|E(z,\tfrac12+ir,\tfrac12)|^2,s)$ has a pole of order 2 in $s=1$. Furthermore,

$$\begin{aligned}
&\operatorname*{res}_{s=1} H(s)R_1\bigl(|E(z,\tfrac12+it,\tfrac12)|^2,s\bigr)\\
&= \Biggl(\frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\biggl(-2\log Y + \int_{\mathcal{D}_0}|E(z,\tfrac12+it,\tfrac12)|^2\,d\mu(z)\\
&\qquad + \sum_{j=1}^{3}\int_{\mathcal{D}^Y}\bigl(|E(\sigma_{\mathfrak{a}_j}z,\tfrac12+it,\tfrac12)|^2-\psi_j(y)\bigr)\,d\mu(z)\\
&\qquad - \frac{\overline{\phi_{11}(\tfrac12+it,\tfrac12)}Y^{2it} - \phi_{11}(\tfrac12+it,\tfrac12)Y^{-2it}}{2it}\biggr)\\
&\qquad + b_0^{11} + \sum_{j=1}^{2}|\phi_{1j}(\tfrac12+it,\tfrac12)|^2 b_0^{1j}\Biggr)H(1) + \frac{2H'(1)}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\\
&= \Biggl(-\frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\sum_{j=1}^{2}{\phi_{1j}}'(\tfrac12+it,\tfrac12)\overline{\phi_{1j}(\tfrac12+it,\tfrac12)}\\
&\qquad + b_0^{11} + \sum_{j=1}^{2}|\phi_{1j}(\tfrac12+it,\tfrac12)|^2 b_0^{1j}\Biggr)H(1) + \frac{2H'(1)}{\mathrm{vol}(\Gamma\backslash\mathbb{H})},
\end{aligned}\tag{4-16}$$

where we used the Maaß–Selberg relations (see [Roelcke 1966, Lemma 11.2], for example). For the remaining poles at $s=1\pm 2it$ we obtain

$$\operatorname*{res}_{s=1+2it} H(s)R_1\bigl(|E(z,\tfrac12+it,\tfrac12)|^2,s\bigr) = H(1+2it)\phi_{11}(1+2it,0)\overline{\phi_{11}(\tfrac12+it,\tfrac12)},$$

and this expression is of rapid decay as $|t|\to\infty$. This follows from the following general facts: the entries of the scattering matrix of weight zero are uniformly bounded for $\Re(s)\ge\frac12$, $|\Im(s)|\ge 1$ (see [Selberg 1989, p. 655], for example), $\phi_{11}(\tfrac12\pm it,\tfrac12)$ is bounded since $\Phi(\tfrac12+it,k)$ is unitary, and we have the rapid decay of $H(1\pm 2it)$. The same bound holds for the residue of $H(s)R_1\bigl(|E(z,\tfrac12+it,\tfrac12)|^2,s\bigr)$ at $s=1-2it$. We now want to shift the line of integration in (4-12). To do this we need to control the growth of the $R_1\bigl(|E(z,\tfrac12+it,\tfrac12)|^2,s\bigr)$ as well as knowing the residues.

Lemma 4.7. *Let $F(z) = |E_1(z, \frac{1}{2}+it, \frac{1}{2})|^2$. The function $R_1(F(z), \sigma+iv)$ is of at most polynomial growth as $|v| \to \infty$ for $\sigma \geq \frac{1}{2}$.*

Proof. In order to avoid the poles of the Eisenstein series coming from the zeros of the zeta function in the critical strip we work with $R_i^*(F, s) := \zeta(2s) R_i(F, s)$, $i = 1, 2, 3$. Then the function $R_i^*(F, s)$ has only finitely many poles in the strip $0 \leq \Re(s) \leq 1$. The estimates for the Eisenstein series and the scattering matrix imply that

$$R_i^*(F, s) = O(1)$$

as $|\Im(s)| \to \infty$ for $\Re(s) > 1$, $i = 1, 2, 3$. Using the functional equation as well as explicit expressions for $\phi_{1j}(s, 0)$ we then get

$$R_1^*(F, s) = \frac{\zeta(2s)}{\zeta(2(1-s))} \sum_{j=1}^{3} \phi_{1j}(s, 0) R_j^*(F, 1-s) = O(|\Im(s)|^{1-2\sigma})$$

as $|\Im(s)| \to \infty$ for $\sigma = \Re(s) < 0$, $i = 1, 2, 3$. Thus by the Phragmén–Lindelöf principle we finally obtain that

$$R_1(F, \sigma+iv) = O(|v|^k) \quad \text{as } |v| \to \infty,\ \sigma \geq \tfrac{1}{2}, \text{ for some } k \in \mathbb{N}. \qquad \square$$

Now that polynomial growth has been established it follows from (4-16) that

$$\begin{aligned} J_2(t, \infty) = \Biggl(&-\frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} \sum_{j=1}^{2} \frac{{\phi_{1j}}'}{\phi_{1j}}(\tfrac{1}{2}+it, \tfrac{1}{2}) |\phi_{1j}(\tfrac{1}{2}+it, \tfrac{1}{2})|^2 + b_0^{11} \\ &+ \sum_{j=1}^{2} |\phi_{1j}(\tfrac{1}{2}+it, \tfrac{1}{2})|^2 b_0^{1j} \Biggr) H(1) + \frac{H'(1)}{\pi} \\ &+ \frac{1}{2\pi i} \int_{\Re s = \frac{1}{2}} H(s) R_1\bigl(|E(z, \tfrac{1}{2}+it, \tfrac{1}{2})|^2, s\bigr)\, ds + O(1). \end{aligned} \tag{4-17}$$

In Section 3 we saw that, up to constants and fractions of polynomials in powers of 2, the entries of the scattering matrix are equal to $\xi(3-4s)/\xi(4s-1)$; see (3-10). Hence, in order to determine the asymptotic behavior of the first term in (4-17) with respect to the t-variable, we need to understand the logarithmic derivative of $\xi(3-4s)/\xi(4s-1)$ at $s = \frac{1}{2}+it$. The contribution from the remaining terms is $O(1)$. We have

$$\begin{aligned} \left(\log \frac{\xi(3-4s)}{\xi(4s-1)} \right)' \Bigg|_{s=\frac{1}{2}+it} &= 4 \log \pi - 2 \frac{\Gamma'}{\Gamma}\bigl(\tfrac{1}{2} - 2it\bigr) - 2 \frac{\Gamma'}{\Gamma}\bigl(\tfrac{1}{2} + 2it\bigr) \\ &\qquad - 4 \left(\frac{\zeta'}{\zeta}(1-4it) - \frac{1}{4it} + \frac{\zeta'}{\zeta}(1+4it) + \frac{1}{4it} \right) \\ &= -4 \log |t| + o(\log |t|) \end{aligned}$$

by Stirling's formula and [Titchmarsh 1986, Theorem 5.17]. Since $\Phi(s, \frac{1}{2})$ is unitary for $\Re s = \frac{1}{2}$, we finally arrive at

$$J_2(t,\infty) = \frac{4H(1)}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\log|t| + \frac{1}{2\pi i}\int_{\Re s=\frac{1}{2}} H(s)R_1\big(|E(z,\tfrac{1}{2}+it,\tfrac{1}{2})|^2, s\big)\,ds + o(\log|t|) \tag{4-18}$$

as $|t|\to\infty$. To treat the last integral we use again the connection to double Dirichlet series.

Lemma 4.8. *Assume that for any χ, χ' mod 8 the function $Z_{\psi_\tau}(s, 1-s, \chi, \chi')$ satisfies a subconvex bound with saving $\delta > 0$. Then, as $|t|\to\infty$,*

$$\frac{1}{2\pi i}\int_{\Re s=\frac{1}{2}} H(s)R_1\big(|E(z,\tfrac{1}{2}+it,\tfrac{1}{2})|^2, s\big)\,ds = o(1).$$

Proof. We find, by (4-13), Proposition 4.5 combined with $U(s,w)^{-1} = O(1)$ when $\Re(s) = \Re(w) = \frac{1}{2}$, Lemma 2.22, Stirling's formula, Lemma A.5 and, finally, $1/\zeta(1+it) = O(\log|t|)$, that

$$\begin{aligned} R_1\big(|E(z,\tfrac{1}{2}+it,\tfrac{1}{2})|^2, \tfrac{1}{2}+iu\big) \\ &= O\Big(|t|^{-\frac{1}{2}+\varepsilon}\max_{\chi,\chi'}\big|Z_{\psi_\tau}(\tfrac{1}{2}+i(u+2t), \tfrac{1}{2}+i(u-2t), \chi, \chi')\big|\Big). \end{aligned}$$

Subconvexity implies that the max is

$$O\Big(\big((1+|u+2t|)(1+|u-2t|)(1+2|u|)^2\big)^{\frac{1}{4}-\delta}\Big).$$

Using the rapid decay of $H(s)$ we finally obtain that

$$J_3(t,\infty) = O\Big(|t|^{-\frac{1}{2}+\varepsilon}|t|^{2(\frac{1}{4}-\delta)}\Big) = o(1). \qquad \square$$

Remark 4.9. In the above proof we see that, as in the cuspidal case, the trivial bound from Theorem 2.17 only gives $O(|t|^{\frac{1}{2}+\varepsilon})$. However, for a compact set A the Maaß-Selberg relations easily yield

$$\int_A |E(z,\tfrac{1}{2}+it,\tfrac{1}{2})|^2\,d\mu(z) = O(\log t).$$

To summarize, we have proved:

Theorem 4.10. *Assume that for any χ, χ' mod 8 the function $Z(s, 1-s, \chi, \chi')$ satisfies a subconvex bound. Then, as $|t|\to\infty$,*

$$\int_{\Gamma\backslash\mathbb{H}} F_h(z)|E_\infty(z,\tfrac{1}{2}+it,\tfrac{1}{2})|^2\,d\mu(z) = \frac{4}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}H(1)\log|t| + o(\log|t|).$$

The asymptotics (1-10) and hence Theorem 1.6 now follow from Theorems 4.3 and 4.10 by an approximation argument as in [Luo and Sarnak 1995, p. 217].

Appendix: Mellin transforms of products of Whittaker functions

In this appendix we prove various bounds on Mellin transforms of products of Whittaker functions that we have not been able to find in the literature in the generality needed.

Lemma A.1. *Let $p \in \{\pm 1\}$. For $s = \frac{1}{2} + it$, $w = 1 - s$, and s_0 fixed, we have the bound*

$$\frac{1}{\Gamma(w+p/4)} \int_0^\infty W_{0,s_0-\frac{1}{2}}(y) W_{p/4,w-\frac{1}{2}}(y) y^{s-1} \frac{dy}{y} = O\big((1+|t|)^{-\frac{1}{2}}\big)$$

as $|t| \to \infty$.

Remark A.2. The estimate in Lemma A.1 cannot be improved, as the proof below shows that the estimate can be turned into an asymptotic rate of decay of the same order.

Proof. Using [Gradshteyn and Ryzhik 2007, 7.611 7., p. 821] we obtain

$$\begin{aligned}
&\int_0^\infty W_{0,s_0-\frac{1}{2}}(y) W_{p/4,w-\frac{1}{2}}(y) y^{s-1} \frac{dy}{y} \\
&= \frac{\Gamma(s+w-s_0)\Gamma(s+w+s_0-1)\Gamma(1-2w)}{\Gamma(1-p/4-w)\Gamma(s+w)} \\
&\qquad\qquad \times {}_3F_2\Big(s+w-s_0,\, s+w+s_0-1,\, w-\frac{p}{4};\, 2w,\, s+w;\, 1\Big) \\
&\quad + \frac{\Gamma(s-w+s_0)\Gamma(s-w-s_0+1)\Gamma(2w-1)}{\Gamma(w-p/4)\Gamma(s-w+1)} \\
&\quad \times {}_3F_2\Big(s-w+s_0,\, s-w-s_0+1,\, 1-\frac{p}{4}-w;\, 2-2w,\, s-w+1;\, 1\Big),
\end{aligned} \tag{A-1}$$

if $|\Re(s_0-\frac{1}{2})|+|\Re(w-\frac{1}{2})| < \Re s$. The generalized hypergeometric series that appear in (A-1) converge for $\Re s < 1+p/4$. We now set $s = \frac{1}{2}+it$ and $w = \frac{1}{2}-it$ and get

$$\begin{aligned}
&\int_0^\infty W_{0,s_0-\frac{1}{2}}(y) W_{p/4,-it}(y) y^{s-1} \frac{dy}{y} \\
&= \frac{\Gamma(1-s_0)\Gamma(s_0)\Gamma(2it)}{\Gamma(\frac{1}{2}-\frac{p}{4}+it)\Gamma(1)} {}_3F_2\Big(1-s_0,\, s_0,\, \frac{1}{2}-\frac{p}{4}-it;\, 1-2it,\, 1;\, 1\Big) \\
&\quad + \frac{\Gamma(s_0+2it)\Gamma(1-s_0+2it)\Gamma(-2it)}{\Gamma(\frac{1}{2}-\frac{p}{4}-it)\Gamma(1+2it)} \\
&\qquad\quad \times {}_3F_2\Big(s_0+2it,\, 1-s_0+2it,\, \frac{1}{2}-\frac{p}{4}+it;\, 1+2it,\, 1+2it;\, 1\Big).
\end{aligned}$$

Using [Bailey 1964, p. 18], we infer that (see also [Jakobson 1994, (2.9), p. 1491])

$$
\begin{aligned}
&{}_3F_2\Big(s_0+2it,\, 1-s_0+2it,\, \frac{1}{2}-\frac{p}{4}+it;\, 1+2it,\, 1+2it;\, 1\Big)\\
&\qquad = \frac{\Gamma(\frac{1}{2}+\frac{p}{4}-it)\Gamma(1+2it)}{\Gamma(\frac{1}{2}+\frac{p}{4}+it)\Gamma(1)}\, {}_3F_2\Big(\frac{1}{2}-\frac{p}{4}+it,\, 1-s_0,\, s_0;\, 1+2it,\, 1;\, 1\Big),
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\int_0^\infty W_{0,s_0-\frac{1}{2}}(y)W_{p/4,-it}(y)y^{s-1}\,\frac{dy}{y}\\
&= \frac{\Gamma(1-s_0)\Gamma(s_0)\Gamma(2it)}{\Gamma(\frac{1}{2}-\frac{p}{4}+it)}\, {}_3F_2\Big(1-s_0,\, s_0,\, \frac{1}{2}-\frac{p}{4}-it;\, 1-2it,\, 1;\, 1\Big)\\
&\qquad + \frac{\Gamma(s_0+2it)\Gamma(1-s_0+2it)\Gamma(-2it)\Gamma(\frac{1}{2}+\frac{p}{4}-it)}{\Gamma(\frac{1}{2}-\frac{p}{4}-it)\Gamma(\frac{1}{2}+\frac{p}{4}+it)}\\
&\qquad\qquad \times {}_3F_2\Big(1-s_0,\, s_0,\, \frac{1}{2}-\frac{p}{4}+it;\, 1+2it,\, 1;\, 1\Big). \qquad \text{(A-2)}
\end{aligned}
$$

We want to understand the asymptotic behavior of the hypergeometric series appearing in (A-2). Since $\Re(s)=\frac{1}{2}<1+\frac{p}{4}$, these converge absolutely. Moreover, the only difference between the two series is the sign of it, so that it suffices to treat the first series. The treatment of the second hypergeometric series appearing in (A-2) is similar. Using the series representation for ${}_3F_2$ we see that

$$
{}_3F_2\Big(s_0,\, 1-s_0,\, \frac{1}{2}-\frac{p}{4}-it;\, 1-2it,\, 1;\, 1\Big) = \sum_{n=0}^{\infty} \frac{(s_0)_n(1-s_0)_n(\frac{1}{2}-\frac{p}{4}-it)_n}{(1)_n(1-2it)_n}\frac{1}{n!}. \qquad \text{(A-3)}
$$

In order to determine its asymptotic behavior as $|t|\to\infty$ we want to interchange the summation with the limit, i.e., we want to take the limit $|t|\to\infty$ in each term of the series separately. For this, let $\epsilon\in(0;\frac{1}{4})$ be sufficiently small and rewrite the terms appearing in (A-3) as

$$
\left|\frac{(s_0)_n(1-s_0)_n(\frac{1}{2}-\frac{p}{4}-it)_n}{(1)_n(1-2it)_n}\right| = \left|\frac{(s_0)_n(1-s_0)_n}{(1+\epsilon)_n}\right|\left|\frac{(1+\epsilon)_n(\frac{1}{2}-\frac{p}{4}-it)_n}{(1)_n(1-2it)_n}\right|.
$$

For $0\le l\le n$ we have

$$
\left|\frac{(l+1+\epsilon)(l+\frac{1}{2}-\frac{p}{4}-it)}{(l+1)(l+1-2it)}\right|^2 = \frac{\big(l^2+(\frac{3}{2}-\frac{p}{4}+\epsilon)l+(1+\epsilon)(\frac{1}{2}-\frac{p}{4})\big)^2+t^2(l+1+\epsilon)^2}{(l+1)^4+4t^2(l+1)^2}.
$$

Since $2(l+1)>l+1+\epsilon$ and

$$
0\le l^2+\Big(\frac{3}{2}-\frac{p}{4}+\epsilon\Big)l+(1+\epsilon)\Big(\frac{1}{2}-\frac{p}{4}\Big)\le(l+1)^2, \qquad \text{(A-4)}
$$

this implies that

$$\left|\frac{(s_0)_n(1-s_0)_n(\frac{1}{2}-\frac{p}{4}-it)_n}{(1)_n(1-2it)_n}\right| \le \left|\frac{(s_0)_n(1-s_0)_n}{(1+\epsilon)_n}\right|$$

for all $n \ge 0$. Furthermore, the hypergeometric series

$${}_2F_1(s_0,\, 1-s_0;\, 1+\epsilon;\, 1) = \sum_{n=0}^{\infty} \frac{(s_0)_n(1-s_0)_n}{(1+\epsilon)_n}\frac{1}{n!}$$

converges absolutely and therefore, by the theorem of majorized convergence, we finally obtain

$$\lim_{|t|\to\infty} {}_3F_2\Big(s_0,\, 1-s_0,\, \frac{1}{2}-\frac{p}{4}-it;\, 1-2it,\, 1;\, 1\Big) = {}_2F_1\Big(s_0,\, 1-s_0;\, 1;\, \frac{1}{2}\Big).$$

Thus only the Gamma factors appearing in (A-2) determine the asymptotic behavior, and using Stirling's formula we see that

$$\int_0^\infty W_{0,s_0-\frac{1}{2}}(y)W_{p/4,-it}(y)y^{s-1}\,\frac{dy}{y} = O\big(|t|^{-(\frac{1}{2}-\frac{p}{4})}e^{-\frac{\pi}{2}|t|}\big)$$

as $|t| \to \infty$. This implies the desired bound. $\square$

Lemma A.3. *Let* $p \in \{\pm 1\}$. *We have*

$${}_3F_2\Big(\frac{1}{2}+\frac{p}{4}-it,\, \frac{1}{2}+iu,\, \frac{1}{2}-iu;\, 1,\, 1-2it;\, 1\Big) \ll e^{\pi|u|}|u|^{-2\epsilon}$$

as $|u| \to \infty$, *where the implied constant does not depend on* t. *Furthermore, there exists a constant* C *independent of* t *such that*

$${}_3F_2\Big(\frac{1}{2}+\frac{p}{4}-it,\, \frac{1}{2},\, \frac{1}{2};\, 1,\, 1-2it;\, 1\Big) \le C.$$

Proof. Since $\Re(2-2it-(1+\frac{1}{2}+\frac{p}{4}-it)) > 0$, the hypergeometric series ${}_3F_2(\frac{1}{2}+\frac{p}{4}-it,\, \frac{1}{2}+iu,\, \frac{1}{2}-iu;\, 1,\, 1-2it;\, 1)$ converges. By the definition of the hypergeometric series we have

$${}_3F_2\Big(\frac{1}{2}+\frac{p}{4}-it,\, s,\, 1-s;\, 1,\, 1-2it;\, 1\Big) = 1+\sum_{m=1}^{\infty}\frac{(s)_m(1-s)_m}{(1)_m m!}\frac{(\frac{1}{2}+\frac{p}{4}-it)_m}{(1-2it)_m}$$

with $s = \frac{1}{2}+iu$. We now determine the behavior of the series as $|u| \to \infty$. We use the same argumentation that was already useful in the proof of Lemma A.1. We write

$$\frac{(s)_m(1-s)_m}{(1)_m m!}\frac{(\frac{1}{2}+\frac{p}{4}-it)_m}{(1-2it)_m} = \frac{(s)_m(1-s)_m}{(1+\epsilon)_m m!}\frac{(\frac{1}{2}+\frac{p}{4}-it)_m(1+\epsilon)_m}{(1)_m(1-2it)_m} \tag{A-5}$$

with $\epsilon > 0$ sufficiently small. As before the second factor on the right-hand side can be bounded in norm by 1, and it is straightforward to see that the first factor is real and positive, so

$$\left|{}_3F_2\left(\frac{1}{2}+\frac{p}{4}-it, s, 1-s; 1, 1-2it; 1\right)\right| \le {}_2F_1(s, 1-s; 1+\epsilon; 1).$$

The last hypergeometric function equals (see [Bailey 1964, (1), p. 2])

$$\frac{\Gamma(1+\epsilon)\Gamma(\epsilon)}{\Gamma(\frac{1}{2}+\epsilon+iu)\Gamma(\frac{1}{2}+\epsilon-iu)},$$

and the first statement now follows from Stirling's formula. The second statement follows from plugging $u=0$ in the above argument. □

Remark A.4. A similar bound is given in [Jakobson 1994], Claim 3.4, p. 1499.

Lemma A.5. *Let $p \in \{\pm 1\}$. For $u, t \in \mathbb{R}$ we have*

$$\frac{1}{|\Gamma(\frac{1}{2}+\frac{p}{4}+it)|^2}\int_0^\infty y^{-\frac{1}{2}+iu}\left|W_{p/4,it}(y)\right|^2\frac{dy}{y} = O((1+|t|)^{-\frac{1}{2}})$$

as $|t| \to \infty$. The implied constant is uniform in u.

Proof. Set

$$I_{p,t}(u) := \int_0^\infty y^{-\frac{1}{2}+iu}\left|W_{p/4,it}(y)\right|^2\frac{dy}{y}.$$

Since $|I_{n,t}(u)| \le I_{n,t}(0)$, we assume that $u=0$. By [Gradshteyn and Ryzhik 2007, Formula 7.611 7., p. 821] we get

$$\begin{aligned}
&I_{n,t}(0)\\
&=\frac{\Gamma(\frac{1}{2}-2it)\Gamma(\frac{1}{2})\Gamma(2it)}{\Gamma(\frac{1}{2}-\frac{p}{4}+it)\Gamma(1-\frac{p}{4}-it)}\times {}_3F_2\left(\frac{1}{2}-2it, \frac{1}{2}, \frac{1}{2}-\frac{p}{4}-it; 1-2it, 1-\frac{p}{4}-it; 1\right)\\
&\quad+\frac{\Gamma(\frac{1}{2}+2it)\Gamma(\frac{1}{2})\Gamma(-2it)}{\Gamma(\frac{1}{2}-\frac{p}{4}-it)\Gamma(1-\frac{p}{4}+it)}\times {}_3F_2\left(\frac{1}{2}+2it, \frac{1}{2}, \frac{1}{2}-\frac{p}{4}+it; 1+2it, 1-\frac{p}{4}+it; 1\right).
\end{aligned}$$

It suffices to consider the first term since the second term differs from the first one only by the sign of t. Using the transformation formulae of [Bailey 1964, p. 18], as in the proof of Lemma A.1 we see that

$$\begin{aligned}
&{}_3F_2\left(\frac{1}{2}-2it, \frac{1}{2}, \frac{1}{2}-\frac{p}{4}-it; 1-2it, 1-\frac{p}{4}-it; 1\right)\\
&\qquad=\frac{\Gamma(1-\frac{p}{4}-it)\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}-\frac{p}{4}-it)}{}_3F_2\left(\frac{1}{2}+\frac{p}{4}-it, \frac{1}{2}, \frac{1}{2}; 1, 1-2it; 1\right).
\end{aligned}$$

By the second part of Lemma A.3 the hypergeometric series is bounded and we find — by bounding all the Gamma functions using Stirling — that

$$|I_{p,t}(0)| = O\left(\frac{\Gamma(1-\frac{p}{4}-it)}{\Gamma(\frac{1}{2}-\frac{p}{4}-it)}\frac{\Gamma(\frac{1}{2}-2it)\Gamma(2it)}{\Gamma(\frac{1}{2}-\frac{p}{4}+it)\Gamma(1-\frac{p}{4}-it)}\right) = O\left(e^{-\pi|t|}|t|^{-\frac{1}{2}+\frac{p}{2}}\right)$$

as $|t| \to \infty$, which gives the result. □

Acknowledgements

We thank Gautam Chinta, Adrian Diaconu and Valentin Blomer for useful discussions about multiple Dirichlet series. Also we thank the referees for several helpful comments.

References

[Bailey 1964] W. N. Bailey, *Generalized hypergeometric series*, Cambridge Tracts in Mathematics and Mathematical Physics **32**, Stechert-Hafner, New York, 1964. MR 32 #2625 Zbl 0011.02303

[Blomer 2011] V. Blomer, "Subconvexity for a double Dirichlet series", *Compos. Math.* **147**:2 (2011), 355–374. MR 2012f:11170 Zbl 1228.11140

[Blomer et al. 2014] V. Blomer, L. Goldmakher, and B. Louvel, "L-functions with n-th-order twists", *Int. Math. Res. Not.* **2014**:7 (2014), 1925–1955. MR 3190355

[Brubaker et al. 2004] B. Brubaker, A. Bucur, G. Chinta, S. Frechette, and J. Hoffstein, "Nonvanishing twists of GL(2) automorphic L-functions", *Int. Math. Res. Not.* **2004**:78 (2004), 4211–4239. MR 2005h:11099 Zbl 1083.11030

[Bump et al. 1996] D. Bump, S. Friedberg, and J. Hoffstein, "On some applications of automorphic forms to number theory", *Bull. Amer. Math. Soc.* (*N.S.*) **33**:2 (1996), 157–175. MR 97a:11072 Zbl 0865.11043

[Chinta and Diaconu 2005] G. Chinta and A. Diaconu, "Determination of a GL_3 cuspform by twists of central L-values", *Int. Math. Res. Not.* **2005**:48 (2005), 2941–2967. MR 2006k:11094 Zbl 1085.11026

[Davenport 2000] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics **74**, Springer, New York, 2000. MR 2001f:11001 Zbl 1002.11001

[Diaconu et al. 2003] A. Diaconu, D. Goldfeld, and J. Hoffstein, "Multiple Dirichlet series and moments of zeta and L-functions", *Compositio Math.* **139**:3 (2003), 297–360. MR 2005a:11124 Zbl 1053.11071

[Dutta Gupta 1997] S. Dutta Gupta, "On the Rankin–Selberg method for functions not of rapid decay on congruence subgroups", *J. Number Theory* **62**:1 (1997), 115–126. MR 98f:11041 Zbl 0872.11024

[Friedberg and Hoffstein 1995] S. Friedberg and J. Hoffstein, "Nonvanishing theorems for automorphic L-functions on GL(2)", *Ann. of Math.* (2) **142**:2 (1995), 385–423. MR 96e:11072 Zbl 0847.11026

[Gradshteyn and Ryzhik 2007] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., Elsevier/Academic Press, Amsterdam, 2007. MR 2008g:00005 Zbl 1208.65001

[Heath-Brown 1995] D. R. Heath-Brown, "A mean value estimate for real character sums", *Acta Arith.* **72**:3 (1995), 235–275. MR 96h:11081 Zbl 0828.11040

[Hoffstein and Kontorovich 2010] J. Hoffstein and A. Kontorovich, "The first non-vanishing quadratic twist of an automorphic L-series", preprint, 2010. arXiv 1008.0839

[Iwaniec 1997] H. Iwaniec, *Topics in classical automorphic forms*, Graduate Studies in Mathematics **17**, Amer. Math. Soc., Providence, RI, 1997. MR 98e:11051 Zbl 0905.11023

[Iwaniec 2002] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd ed., Graduate Studies in Mathematics **53**, Amer. Math. Soc., Providence, RI, 2002. MR 2003k:11085 Zbl 1006.11024

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications **53**, Amer. Math. Soc., Providence, RI, 2004. MR 2005h:11005 Zbl 1059.11001

[Jakobson 1994] D. Jakobson, "Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathrm{PSL}_2(\mathbf{R})$", *Ann. Inst. Fourier* (*Grenoble*) **44**:5 (1994), 1477–1504. MR 96b:11068 Zbl 0820.11040

[Koblitz 1984] N. Koblitz, *Introduction to elliptic curves and modular forms*, Graduate Texts in Mathematics **97**, Springer, New York, 1984. MR 86c:11040 Zbl 0553.10019

[Koyama 2000] S.-y. Koyama, "Quantum ergodicity of Eisenstein series for arithmetic 3-manifolds", *Comm. Math. Phys.* **215**:2 (2000), 477–486. MR 2001m:11085 Zbl 0982.11030

[Lindenstrauss 2006] E. Lindenstrauss, "Invariant measures and arithmetic quantum unique ergodicity", *Ann. of Math.* (2) **163**:1 (2006), 165–219. MR 2007b:11072 Zbl 1104.22015

[Luo and Sarnak 1995] W. Z. Luo and P. Sarnak, "Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$", *Inst. Hautes Études Sci. Publ. Math.* 81 (1995), 207–237. MR 97f:11037 Zbl 0852.11024

[Mizuno 2005] Y. Mizuno, "The Rankin–Selberg convolution for Cohen's Eisenstein series of half integral weight", *Abh. Math. Sem. Univ. Hamburg* **75** (2005), 1–20. MR 2006k:11081 Zbl 1082.11025

[Nelson et al. 2014] P. D. Nelson, A. Pitale, and A. Saha, "Bounds for Rankin–Selberg integrals and quantum unique ergodicity for powerful levels", *J. Amer. Math. Soc.* **27**:1 (2014), 147–191. MR 3110797 Zbl 06228355

[Petridis et al. 2013] Y. N. Petridis, N. Raulf, and M. S. Risager, "Quantum limits of Eisenstein series and scattering states", *Canad. Math. Bull.* **56**:4 (2013), 814–826. MR 3121690 Zbl 06231120

[Roelcke 1966] W. Roelcke, "Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene, I", *Math. Ann.* **167**:4 (1966), 292–337. MR 1513277 Zbl 0152.07705

[Sarnak 2011] P. Sarnak, "Recent progress on the quantum unique ergodicity conjecture", *Bull. Amer. Math. Soc.* (*N.S.*) **48**:2 (2011), 211–228. MR 2012e:58062 Zbl 1234.58007

[Selberg 1989] A. Selberg, *Collected papers, I*, Springer, Berlin, 1989. MR 92h:01083 Zbl 0675.10001

[Shahidi 1988] F. Shahidi, "On the Ramanujan conjecture and finiteness of poles for certain L-functions", *Ann. of Math.* (2) **127**:3 (1988), 547–584. MR 89h:11021 Zbl 0654.10029

[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

[Shimura 1975] G. Shimura, "On the holomorphy of certain Dirichlet series", *Proc. London Math. Soc.* (3) **31**:1 (1975), 79–98. MR 52 #3064 Zbl 0311.10029

[Shnirelman 1974] A. I. Shnirelman, "Ergodic properties of eigenfunctions", *Uspehi Mat. Nauk* **29**:6(180) (1974), 181–182. In Russian. MR 53 #6648 Zbl 0324.58020

[Soundararajan 2010a] K. Soundararajan, "Quantum unique ergodicity for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$", *Ann. of Math.* (2) **172**:2 (2010), 1529–1538. MR 2011j:11098 Zbl 1209.58019

[Soundararajan 2010b] K. Soundararajan, "Weak subconvexity for central values of L-functions", *Ann. of Math.* (2) **172**:2 (2010), 1469–1498. MR 2011i:11077 Zbl 1234.11066

[Soundararajan and Young 2010] K. Soundararajan and M. P. Young, "The second moment of quadratic twists of modular L-functions", *J. Eur. Math. Soc.* (*JEMS*) **12**:5 (2010), 1097–1116. MR 2011g:11097 Zbl 1213.11165

[Sturm 1980] J. Sturm, "Special values of zeta functions, and Eisenstein series of half integral weight", *Amer. J. Math.* **102**:2 (1980), 219–240. MR 82b:10033a Zbl 0433.10015

[Titchmarsh 1986] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd ed., Clarendon Press, New York, 1986. MR 88c:11049 Zbl 0601.10026

[Truelsen 2011] J. L. Truelsen, "Quantum unique ergodicity of Eisenstein series on the Hilbert modular group over a totally real field", *Forum Math.* **23**:5 (2011), 891–931. MR 2836373 Zbl 1282.11047

[Colin de Verdière 1985] Y. Colin de Verdière, "Ergodicité et fonctions propres du laplacien", *Comm. Math. Phys.* **102**:3 (1985), 497–502. MR 87d:58145 Zbl 0592.58050

[Zagier 1981] D. Zagier, "The Rankin–Selberg method for automorphic functions which are not of rapid decay", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **28**:3 (1981), 415–437. MR 83k:10056 Zbl 0505.10011

[Zelditch 1987] S. Zelditch, "Uniform distribution of eigenfunctions on compact hyperbolic surfaces", *Duke Math. J.* **55**:4 (1987), 919–941. MR 89d:58129 Zbl 0643.58029

[Zelditch 1991] S. Zelditch, "Mean Lindelöf hypothesis and equidistribution of cusp forms and Eisenstein series", *J. Funct. Anal.* **97**:1 (1991), 1–49. MR 92h:11046 Zbl 0743.58034

[Zelditch 1992] S. Zelditch, *Selberg trace formulae and equidistribution theorems for closed geodesics and Laplace eigenfunctions: Finite area surfaces*, Mem. Amer. Math. Soc. **465**, Amer. Math. Soc., Providence, RI, 1992. MR 93a:11047 Zbl 0753.11023

Communicated by Peter Sarnak
Received 2012-10-05 Revised 2014-08-03 Accepted 2014-09-08

i.petridis@ucl.ac.uk *Department of Mathematics, University College London, Gower street, London, WC1E 6BT, United Kingdom*

raulf@math.univ-lille1.fr *Laboratoire Paul Painlevé, U.F.R. de Mathématiques, Université Lille 1, Sciences et Technologies, 59 655 Villeneuve d'Asqc Cédex, France*

risager@math.ku.dk *Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark*

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1597

Monodromy and local-global compatibility for $l = p$

Ana Caraiani

We strengthen the compatibility between local and global Langlands correspondences for GL_n when n is even and $l = p$. Let L be a CM field and Π a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_L)$ which is conjugate self-dual and regular algebraic. In this case, there is an l-adic Galois representation associated to Π, which is known to be compatible with local Langlands in almost all cases when $l = p$ by recent work of Barnet-Lamb, Gee, Geraghty and Taylor. The compatibility was proved only up to semisimplification unless Π has Shin-regular weight. We extend the compatibility to Frobenius semisimplification in all cases by identifying the monodromy operator on the global side. To achieve this, we derive a generalization of Mokrane's weight spectral sequence for log crystalline cohomology.

1. Introduction

This paper is a continuation of [Caraiani 2012]. Here we extend our local-global compatibility result to the case $l = p$.

Theorem 1.1. *Let $n \in \mathbb{Z}_{\geq 2}$ be an integer and L be a CM field with complex conjugation c. Let l be a prime of $\mathbb{Q}$ and $\iota_l : \bar{\mathbb{Q}}_l \to \mathbb{C}$ be an isomorphism. Let Π be a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_L)$ satisfying*

- $\Pi^\vee \simeq \Pi \circ c$,

Partially supported by NSF Postdoctoral Fellowship DMS-1204465.
MSC2010: primary 11F80; secondary 11G18, 11R39.
Keywords: Galois representations, automorphic forms, local-global compatibility, monodromy operator, crystalline cohomology.

- Π *is cohomological for some irreducible algebraic representation* Ξ *of* $\mathrm{GL}_n(L \otimes_{\mathbb{Q}} \mathbb{C})$.

Let

$$R_l(\Pi) : \mathrm{Gal}(\bar{L}/L) \to \mathrm{GL}_n(\bar{\mathbb{Q}}_l)$$

be the Galois representation associated to Π *by* [Shin 2011; Chenevier and Harris 2013]. *Let* y *be a place of* L *above* l. *Then we have the isomorphism of Weil–Deligne representations*

$$WD(R_l(\Pi)|_{\mathrm{Gal}(\bar{L}_y/L_y)})^{\text{F-ss}} \simeq \iota_l^{-1} \mathscr{L}_{n,L_y}(\Pi_y).$$

Here $\mathscr{L}_{n,L_y}(\Pi_y)$ is the image of Π_y under the local Langlands correspondence, using the geometric normalization; i.e., $\mathscr{L}_{n,L_y}(\Pi_y) := \mathrm{rec}\big(\Pi_y^\vee \otimes |\det|^{\frac{1-n}{2}}\big)$, where rec is the local Langlands correspondence compatible with L- and ϵ-factors (see the introduction to [Harris and Taylor 2001] for more details). $WD(r)$ is the Weil–Deligne representation attached to a de Rham l-adic representation r of the absolute Galois group of an l-adic field; F-ss denotes Frobenius semisimplification. Note that we are assuming throughout that $n \geq 2$. The local-global compatibility of Langlands correspondences for GL_1 follows from the compatibility between local and global class field theory.

This theorem is proved in [Barnet-Lamb et al. 2012; Barnet-Lamb et al. 2011] in the case when Π has Shin-regular weight (either n is odd or if n is even then Π satisfies an additional regularity condition) and in general up to semisimplification. The strategy for obtaining the local-global compatibility of monodromy operators in these cases is to make use of the fact that the l-adic Galois representation associated to Π occurs in the cohomology of certain very special unitary Shimura varieties. These are associated to unitary similitude groups with signature $(1, n-1)$ (respectively, $(1, n)$ if n is even) at exactly one infinite place and signature $(0, n)$ (respectively, $(0, n+1)$) at all the other infinite places. The problem can be reduced to the case when Π_y has an Iwahori-fixed vector, in which case one has to compute the crystalline cohomology of a compact Shimura variety which is strictly semistable. This computation makes use of the weight spectral sequence for crystalline cohomology due to Mokrane [1993], which is shown to degenerate at the first page. We remark that the l-adic Galois representation associated to Π is only known to occur in the cohomology of a proper, smooth variety in the case when Π has Shin-regular weight.

Our goal in this paper is to match up the monodromy operators in the case when n is even and Π does not necessarily have Shin-regular weight. Following the conventions of [Taylor and Yoshida 2007], we call a Weil–Deligne representation *pure of weight* k if it admits a weight filtration, with all the weights in $k + \mathbb{Z}$, such that the (iterated) monodromy operator induces an isomorphism of the $(k+i)$-th and $(k-i)$-th graded pieces for all positive integers i. By Lemma 1.4 (4) of [Taylor and Yoshida 2007], given a semisimple representation of the Weil group of some l-adic field, there is at most one way to choose the monodromy operator such that the resulting Weil–Deligne representation is pure of some weight. By Theorem 1.2 of [Caraiani 2012], Π_y is tempered, so we know that $\iota_l^{-1}\mathscr{L}_{n,L_y}(\Pi_y)$ is pure of some weight.

By Theorem A of [Barnet-Lamb et al. 2011], we also know that we have an isomorphism up to semisimplification:

$$WD(R_l(\Pi)|_{\mathrm{Gal}(\bar{L}_y/L_y)})^{\mathrm{ss}} \simeq \iota_l^{-1}\mathscr{L}_{n,L_y}(\Pi_y)^{\mathrm{ss}}.$$

We note that Theorem A of [Barnet-Lamb et al. 2011] is stated for an imaginary CM field F. For our CM field L we proceed as on pages 230–231 of [Harris and Taylor 2001] to find a quadratic extension F/L which is an imaginary CM field, in which $y = y'y''$ splits, such that

$$[R_l(\Pi)|_{\mathrm{Gal}(\bar{L}/F)}] = [R_l(BC_{F/L}(\Pi))].$$

This together with Theorem A of [Barnet-Lamb et al. 2011] gives the compatibility up to semisimplification for the place y of L. Therefore, in order to complete the proof of Theorem 1.1, it suffices to show that $W := WD(R_l(\Pi)_{\mathrm{Gal}(\bar{L}_y/L_y)})^{\text{F-ss}}$ is pure of some weight when n is even. From now on we will let $n \in \mathbb{Z}_{\geq 2}$ be an even integer.

Our argument will follow the same general lines as that of [Taylor and Yoshida 2007], which is also the strategy followed by [Barnet-Lamb et al. 2012; Barnet-Lamb et al. 2011]. We reduce the problem to the case when Π_y has an Iwahori-fixed vector. In this case, we find not W itself, but rather the tensor square of W in the log crystalline cohomology of a compact Shimura variety with Iwahori-level structure, and finally compute a part of this cohomology explicitly. For the last step, however, we can not make use of the Mokrane spectral sequence, since our Iwahori-level Shimura variety is no longer semistable, but rather Zariski-locally étale over a product of strictly semistable schemes. Therefore, we need to derive a formula for the log crystalline cohomology of the special fiber of this Shimura variety in terms of the crystalline cohomology of closed Newton polygon strata in the special fiber. Deriving this formula constitutes the heart of this paper; we obtain it in the form of a generalization of the Mokrane spectral sequence or as a crystalline analogue of Corollary 4.28 of [Caraiani 2012].

We briefly outline the structure of our paper. In Section 2 we reduce to the case where Π has an Iwahori-fixed vector, we define an inverse system of compact Shimura varieties associated to a unitary group and we show that the crystalline cohomology of the Iwahori-level Shimura variety realizes the tensor square of W. The Shimura varieties we work with are the same as those studied in [Caraiani 2012], so in Section 2 we also recall the main results from [Caraiani 2012] concerning them. In Section 3 we recall and adapt to our situation some standard results from the theory of log crystalline cohomology and the de Rham–Witt complex; we define and study some slight generalizations of the logarithmic de Rham–Witt complex. In Section 4 we generalize the Mokrane spectral sequence to our geometric setting. The main technical result is Theorem 4.6. In Section 5 we prove Theorem 1.1.

2. Shimura varieties

Let L, Π, $R_l(\Pi)$ and y be as described in the introduction. Below, we show that we can understand the Weil–Deligne representation $W = WD(R_l(\Pi)_{\mathrm{Gal}(\bar{L}_y/L_y)})^{\text{F-ss}}$ by computing a part of the crystalline cohomology of an inverse system of Shimura varieties. In the first part we closely follow Sections 2 and 7 of [Caraiani 2012] and afterwards we use some results from Section 5 of the same work.

We claim first that we can reduce the problem to the case when Π has Iwahori-fixed vectors at y, and we can also put ourselves in a situation where the base change from unitary groups to GL_n is well understood. This means that we can reduce the problem to understanding the cohomology of certain Iwahori-level unitary Shimura varieties. More precisely, we can find a CM field extension F' of L such that:

- $F' = EF_1$, where E is an imaginary quadratic field in which l splits and $F_1 = (F')^{c=1}$ has $[F_1 : \mathbb{Q}] \geq 2$;
- F' is soluble and Galois over L;
- $\Pi^0_{F'} := BC_{F'/L}(\Pi)$ is a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_{F'})$; and
- there is a place $\mathfrak{p}$ above the place y of L such that $\Pi^0_{F',\mathfrak{p}}$ has a nonzero Iwahori-fixed vector;

and a CM field F which is a quadratic extension of F', such that:

- $\mathfrak{p} = \mathfrak{p}_1\mathfrak{p}_2$ splits in F;
- $\mathrm{Ram}_{F/\mathbb{Q}} \cup \mathrm{Ram}_{\mathbb{Q}}(\Pi) \subset \mathrm{Spl}_{F/F_2,\mathbb{Q}}$, where $F_2 := (F)^{c=1}$; and
- $\Pi^0_F = BC_{F/F'}(\Pi^0_{F'})$ is a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_F)$.

We can find F and F' as in the proof of Corollary 5.9 of [Caraiani 2012]. Since purity is preserved under finite extensions by Lemma 1.4 of [Taylor and Yoshida 2007], to show that W is pure it suffices to show that

$$W_{F'} := WD\big(R_l(\Pi^0_{F'})|_{\mathrm{Gal}(\overline{F'_{\mathfrak{p}}}/F'_{\mathfrak{p}})}\big)^{\text{F-ss}}$$

is pure. Note that in this new situation $\Pi^0_{F',\mathfrak{p}}$ has a nonzero Iwahori-fixed vector.

We can define an algebraic group G over $\mathbb{Q}$ and an inverse system of Shimura varieties over F' corresponding to a PEL Shimura datum $(F, *, V, \langle\cdot,\cdot\rangle, h)$. Here

F is the CM field defined above and $* = c$ is the involution corresponding to complex conjugation. We take V to be the F-vector space F^n. The pairing

$$\langle\cdot,\cdot\rangle : V \times V \to \mathbb{Q}$$

is a nondegenerate Hermitian pairing such that $\langle f v_1, v_2\rangle = \langle v_1, f^* v_2\rangle$ for all $f \in F$ and $v_1, v_2 \in V$. The last element we need is an $\mathbb{R}$-algebra homomorphism $h : \mathbb{C} \to \mathrm{End}_F(V) \otimes_{\mathbb{Q}} \mathbb{R}$ such that the bilinear pairing

$$(v_1, v_2) \to \langle v_1, h(i) v_2\rangle$$

is symmetric and positive definite. We define the algebraic group G over $\mathbb{Q}$ by

$$G(R) = \{(g,\lambda) \in \mathrm{End}_{F\otimes_{\mathbb{Q}}\mathbb{R}}(V \otimes_{\mathbb{Q}} R)^{\times} \times R^{\times} \mid \langle g v_1, g v_2\rangle = \lambda\langle v_1, v_2\rangle\}$$

for any $\mathbb{Q}$-algebra R.

We choose embeddings $\tau_i : F \hookrightarrow \mathbb{C}$ such that $\tau_2 = \tau_1 \circ \sigma$, where σ is an element of $\mathrm{Gal}(F/F')$ which takes $\mathfrak{p}_1$ to $\mathfrak{p}_2$. For $\sigma \in \mathrm{Hom}_{E,\tau_E}(F, \mathbb{C})$ we let (p_σ, q_σ) be the signature at σ of the pairing $\langle\cdot,\cdot\rangle$ on $V \otimes_{\mathbb{Q}} \mathbb{R}$. In particular, $\tau_E := \tau_1|_E = \tau_2|_E$ is well-defined. We claim that it is possible to choose a PEL datum as above such that $(p_\tau, q_\tau) = (1, n-1)$ for $\tau = \tau_1$ or τ_2 and $(p_\tau, q_\tau) = (0, n)$ otherwise and such that $G_{\mathbb{Q}_v}$ is quasisplit at every finite place v of $\mathbb{Q}$. This follows from Lemma 2.1 of [Caraiani 2012] and the discussion following it, and it depends crucially on the fact that n is even. We choose such a PEL datum and we let G be the corresponding algebraic group over $\mathbb{Q}$ with the prescribed signature at infinity and quasisplit at all the finite places.

Let $\Xi_F^0 := BC_{F/L}(\Xi)$ and $F_2 = F^{c=1}$. Lemma 7.2 of [Shin 2011] says that we can find a character $\psi : \mathbb{A}_E^{\times}/E^{\times} \to \mathbb{C}^{\times}$ and an algebraic representation $\xi_{\mathbb{C}}$ of G over $\mathbb{C}$ satisfying the following conditions:

- $\psi_{\Pi_F^0} = \psi^c/\psi$.
- Ξ_F^0 is isomorphic to the restriction of Ξ' to $\mathrm{Res}_{F/\mathbb{Q}}(\mathrm{GL}_n) \times_{\mathbb{Q}} \mathbb{C}$, where Ξ' is obtained from $\xi_{\mathbb{C}}$ by base change from G to $\mathbb{G}_n := \mathrm{Res}_{E/\mathbb{Q}}(G \times_{\mathbb{Q}} E)$.
- $\xi_{\mathbb{C}}|^{-1}_{E_\infty^{\times}} = \psi_\infty^c$.
- $\mathrm{Ram}_{\mathbb{Q}}(\psi) \subset \mathrm{Spl}_{F/F_2,\mathbb{Q}}$.
- $\psi|_{\mathcal{O}_{E_u}^{\times}} = 1$, where u is the place above l induced by $\iota_l^{-1}\tau_E$.

Define $\xi := \iota_l \xi_{\mathbb{C}}$, and define $\Pi^1 := \psi \otimes \Pi_F^0$, which is a cuspidal automorphic representation of $\mathrm{GL}_1(\mathbb{A}_E) \times \mathrm{GL}_n(\mathbb{A}_F)$.

Corresponding to the PEL datum $(F, *, V, \langle\cdot,\cdot\rangle, h)$, we have a PEL-type moduli problem of abelian varieties. This moduli problem is defined in Section 2.1 of [Caraiani 2012], and here we recall some facts about it. Since the reflex field of the PEL datum is F', the moduli problem for an open compact subgroup $U \subset G(\mathbb{A}^\infty)$ is representable by a Shimura variety X_U/F', which is a smooth and quasiprojective scheme of dimension $2n-2$. The inverse system of Shimura varieties X_U as U varies has an action of $G(\mathbb{A}^\infty)$. As in Section III.2 of [Harris and Taylor 2001], starting with ξ, which is an irreducible algebraic representation of G over $\overline{\mathbb{Q}}_l$, we can define a lisse $\overline{\mathbb{Q}}_l$-sheaf $\mathscr{L}_\xi$ over each X_U, and the action of $G(\mathbb{A}^\infty)$ extends to the inverse system of sheaves. The direct limit

$$H^i(X,\mathscr{L}_\xi) := \varinjlim H^i(X_U \times_{F'} \overline{F'}, \mathscr{L}_\xi)$$

is a semisimple admissible representation of $G(\mathbb{A}^\infty)$ with a continuous action of $\mathrm{Gal}(\overline{F'}/F')$. It can be decomposed as

$$H^i(X,\mathscr{L}_\xi) = \bigoplus_\pi \pi \otimes R^i_{\xi,l}(\pi),$$

where the sum runs over irreducible admissible representations π of $G(\mathbb{A}^\infty)$ over $\overline{\mathbb{Q}}_l$. The $R^i_{\xi,l}(\pi)$are finite-dimensional continuous representations of $\mathrm{Gal}(\overline{F'}/F')$ over $\overline{\mathbb{Q}}_l$. Let $\mathscr{A}_U$ be the universal abelian variety over X_U, to the inverse system of which the action of $G(\mathbb{A}^\infty)$ extends. To the irreducible representation ξ of G we can associate as in Section III.2 of [Harris and Taylor 2001] nonnegative integers m_ξ and t_ξ as well as an idempotent a_ξ of $H^*(\mathscr{A}_U^{m_\xi} \times_{F'} \overline{F'}, \overline{\mathbb{Q}}_l(t_\xi))$. (Here $\mathscr{A}_U^{m_\xi}$ denotes the m_ξ-fold product of $\mathscr{A}_U$ with itself over X_U and $\overline{\mathbb{Q}}_l(t_\xi)$ is a Tate twist.) We have an isomorphism

$$H^i(X_U \times_{F'} \overline{F'}, \mathscr{L}_\xi) \simeq a_\xi H^{i+m_\xi}(\mathscr{A}_U^{m_\xi} \times_{F'} \overline{F'}, \overline{\mathbb{Q}}_l(t_\xi)),$$

which commutes with the $G(\mathbb{A}^\infty)$-action.

For every finite place v of $\mathbb{Q}$ we can define a base-change morphism taking certain admissible $G(\mathbb{Q}_v)$-representations to admissible $\mathbb{G}(\mathbb{Q}_v)$-representations, as in Section 4.2 of [Shin 2011]. Recall that $\mathrm{Ram}_{F/\mathbb{Q}} \cup \mathrm{Ram}_{\mathbb{Q}}\,\Pi^1 \subset \mathrm{Spl}_{F/F_2,\mathbb{Q}}$. If $v \notin \mathrm{Spl}_{F/F_2,\mathbb{Q}}$ then we can define the morphism

$$BC : \mathrm{Irr}^{\mathrm{ur}}_{(l)}(G(\mathbb{Q}_v)) \to \mathrm{Irr}^{\mathrm{ur},\theta\text{-st}}_{(l)}(\mathbb{G}(\mathbb{Q}_v)),$$

taking unramified representations of $G(\mathbb{Q}_v)$ to unramified, θ-stable representations of $\mathbb{G}(\mathbb{Q}_v)$. If $v \in \mathrm{Spl}_{F/F_2,\mathbb{Q}}$ then the morphism

$$BC : \mathrm{Irr}_{(l)}(G(\mathbb{Q}_v)) \to \mathrm{Irr}^{\theta\text{-st}}_{(l)}(\mathbb{G}(\mathbb{Q}_v))$$

can be defined explicitly since $G(\mathbb{Q}_v)$ is split. Putting these maps together, we get, for any finite set of primes $\mathfrak{S}_{\mathrm{fin}}$ such that

$$\mathrm{Ram}_{F/\mathbb{Q}} \cup \mathrm{Ram}_{\mathbb{Q}}(\Pi) \subset \mathfrak{S}_{\mathrm{fin}} \subset \mathrm{Spl}_{F/F_2,\mathbb{Q}},$$

a base-change morphism

$$BC : \mathrm{Irr}^{\mathrm{ur}}_{(l)}\big(G(\mathbb{A}^{\mathfrak{S}_{\mathrm{fin}}\cup\{\infty\}})\big) \otimes \mathrm{Irr}_{(l)}\big(G(\mathbb{A}_{\mathfrak{S}_{\mathrm{fin}}})\big) \to \mathrm{Irr}^{\mathrm{ur},\theta\text{-st}}_{(l)}\big(\mathbb{G}(\mathbb{A}^{\mathfrak{S}_{\mathrm{fin}}\cup\{\infty\}})\big) \otimes \mathrm{Irr}^{\theta\text{-st}}_{(l)}\big(\mathbb{G}(\mathbb{A}_{\mathfrak{S}_{\mathrm{fin}}})\big).$$

Let p be a prime of $\mathbb{Q}$ which splits in E and such that there is a place of F' above p which splits in F. Let $\mathfrak{S}_{\mathrm{fin}}$ be a finite set of primes such that

$$\mathrm{Ram}_{F/\mathbb{Q}} \cup \mathrm{Ram}_{\mathbb{Q}}(\Pi) \cup \{p\} \subset \mathfrak{S}_{\mathrm{fin}} \subset \mathrm{Spl}_{F/F_2,\mathbb{Q}},$$

and set $\mathfrak{S} := \mathfrak{S}_{\mathrm{fin}} \cup \{\infty\}$. For any $R \in \mathrm{Groth}(G(\mathbb{A}^{\mathfrak{S}}) \times G(\mathbb{A}_{\mathfrak{S}_{\mathrm{fin}}}) \times \mathrm{Gal}(\overline{F'}/F'))$ (over $\overline{\mathbb{Q}}_l$) and $\pi^{\mathfrak{S}} \in \mathrm{Irr}^{\mathrm{ur}}(G(\mathbb{A}^{\mathfrak{S}}))$ define the $\pi^{\mathfrak{S}}$-isotypic part of R to be

$$R\{\pi^{\mathfrak{S}}\} := \sum_{\rho} n(\pi^{\mathfrak{S}} \otimes \rho)[\pi^{\mathfrak{S}}][\rho],$$

where ρ runs over $\mathrm{Irr}_l(G(\mathbb{A}_{\mathfrak{S}}) \times \mathrm{Gal}(\overline{F'}/F'))$. Also define

$$R[\Pi^{1,\mathfrak{S}}] := \sum_{\pi^{\mathfrak{S}}} R[\pi^{\mathfrak{S}}],$$

where each sum runs over $\pi^{\mathfrak{S}} \in \mathrm{Irr}^{\mathrm{ur}}_l(G(\mathbb{A}^{\mathfrak{S}}))$ such that $BC(\iota_l \pi^{\mathfrak{S}}) \simeq \Pi^{1,\mathfrak{S}}$.

Proposition 2.1. *Let $\mathfrak{S} = \mathfrak{S}_{\mathrm{fin}} \cup \{\infty\}$ be as above. We have the equality*

$$BC(H^{2n-2}(X, \mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}]) \simeq C_G[\iota_l^{-1}\Pi^{1,\infty}][R_l(\Pi^0_{F'})^{\otimes 2} \otimes \mathrm{rec}_{l,\iota_l}(\psi)]$$

of elements of $\mathrm{Groth}(G(\mathbb{A}^\infty) \times \mathrm{Gal}(\overline{F'}/F))$. *Here C_G is a positive integer and* $\mathrm{rec}_{l,\iota_l}(\psi)$ *is the continuous l-adic character* $\mathrm{Gal}(\bar{E}/E) \to \overline{\mathbb{Q}}_l^\times$ *associated to ψ by global class field theory, normalized so that it matches uniformizers with geometric Frobenius elements.*

Remark. Unlike in the classical situation of modular forms or in the case of Harris–Taylor-type Shimura varieties [Harris and Taylor 2001; Shin 2011], the cohomology of our inverse system of Shimura varieties realizes a twist of the tensor square of the l-adic Galois representation associated to Π, because we have chosen our unitary similitude groups to have signature $(1, n-1)$ two infinite places. One could use Matsushima's formula and $(\mathfrak{g}, K)$-cohomology to check that the dimension of the Galois representation seen by this cohomology is n^2, as predicted by the statement.

Proof. Let $p \in \mathfrak{S}_{\mathrm{fin}}$ be a prime which splits in E such that there is a place w of F' above the place induced by τ_E over p which splits in F, $w = w_1 w_2$. We start by recalling some constructions and results from Sections 2 and 5 of [Caraiani 2012]. It is possible to define an integral model of each X_U over the ring of integers $\mathcal{O}_K$ in $K := F_{w_1} \simeq F_{w_2}$, which itself represents a moduli problem of abelian varieties and to which the sheaf $\mathcal{L}_\xi$ extends. The special fiber Y_U of this integral model has a stratification by open Newton polygon strata $Y^\circ_{U,S,T}$, according to the formal (or étale) height of the p-divisible group of the abelian variety at w_1 and w_2. Each open Newton polygon stratum is covered by a tower of Igusa varieties $\mathrm{Ig}^{(h_1,h_2)}_{U^p,\vec{m}}$, where $0 \le h_1, h_2 \le n-1$ represent the étale heights of the p-divisible groups at w_1 and w_2, and $\vec{m}$ is a tuple of positive integers describing the level structure at p.

Define

$$J^{(h_1,h_2)}(\mathbb{Q}_p) := \mathbb{Q}_p^\times \times D^\times_{K,n-h_1} \times \mathrm{GL}_{h_1}(K) \times D^\times_{K,n-h_2} \times \mathrm{GL}_{h_2}(K) \times \prod_w \mathrm{GL}_n(F_w),$$

where $D_{K,n-h}$ is the division algebra over K of invariant $1/(n-h)$ and w runs over places of F above τ_E other than w_1 and w_2. The group $J^{(h_1,h_2)}(\mathbb{Q}_p)$ acts on the directed system of $H^j_c(\mathrm{Ig}^{(h_1,h_2)}_{U^p,\vec{m}}, \mathcal{L}_\xi)$, as U^p and $\vec{m}$ vary. Let

$$H_c(\mathrm{Ig}^{(h_1,h_2)}, \mathcal{L}_\xi) \in \mathrm{Groth}\big(G(\mathbb{A}^{\infty,p}) \times J^{(h_1,h_2)}\big)$$

be the alternating sum of the direct limit of $H^j_c(\mathrm{Ig}^{(h_1,h_2)}_{U^p,\vec{m}}, \mathcal{L}_\xi)$ as in Section 5.1 of [Caraiani 2012]. Let $\pi_p \in \mathrm{Irr}_l(G(\mathbb{Q}_p))$ be a representation such that $BC(\pi_p) \simeq \iota_l^{-1}\Pi^1_p$ (such a π_p is unique up to isomorphism since p splits in E). Theorem 5.6 of [Caraiani 2012] gives a formula for computing the cohomology of Igusa varieties, as elements of $\mathrm{Groth}\big(\mathbb{G}(\mathbb{A}^{\mathfrak{S}}) \times \mathbb{G}(\mathbb{A}_{\mathfrak{S}_{\mathrm{fin}}\setminus\{p\}}) \times J^{(h_1,h_2)}(\mathbb{Q}_p)\big)$:

$$\begin{aligned} &BC^p\big(H_c(\mathrm{Ig}^{(h_1,h_2)}, \mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}]\big) \\ &\qquad = e_0(-1)^{h_1+h_2} C_G [\iota_l^{-1}\Pi^{1,\mathfrak{S}}][\iota_l^{-1}\Pi^1_{\mathfrak{S}_{\mathrm{fin}}\setminus\{p\}}][\mathrm{Red}^{(h_1,h_2)}_n(\pi_p)] \end{aligned} \tag{2.1}$$

Here $e_0 = \pm 1$ independently of h_1, h_2 and $\mathrm{Red}^{(h_1,h_2)}_n$ is a group morphism from $\mathrm{Groth}(G(\mathbb{Q}_p))$ to $\mathrm{Groth}(J^{(h_1,h_2)}(\mathbb{Q}_p))$, defined explicitly above Theorem 5.6 of [Caraiani 2012].

We can combine the above formula with Mantovan's formula for the cohomology of Shimura varieties. This is the equality

$$H(X, \mathcal{L}_\xi) = \sum_{0 \le h_1, h_2 \le n-1} (-1)^{h_1+h_2} \mathrm{Mant}_{(h_1,h_2)}\big(H_c(\mathrm{Ig}^{(h_1,h_2)}, \mathcal{L}_\xi)\big) \tag{2.2}$$

of elements of $\mathrm{Groth}(G(\mathbb{A}^\infty) \times W_K)$. Here $H(X, \mathcal{L}_\xi)$ is the alternating sum of the direct limit of the cohomology of the Shimura fibers (generic fibers) and

$$\mathrm{Mant}_{(h_1,h_2)} : \mathrm{Groth}(J^{(h_1,h_2)}(\mathbb{Q}_p)) \to \mathrm{Groth}(G(\mathbb{Q}_p) \times W_K)$$

is the functor defined in [Mantovan 2005]. The formula (2.2) is what Theorem 22 of [Mantovan 2005] amounts to in our situation, where h_1 and h_2 are the parameters for the Newton stratification. The extra term $(-1)^{h_1+h_2}$ occurs on the right-hand side because we use the same convention for the alternating sum of cohomology as in [Caraiani 2012], which differs by a sign from the conventions used in [Mantovan 2005] and [Shin 2011].

By combining formulas (2.1) and (2.2) we get

$$BC^p\big(H(X,\mathscr{L}_\xi)[\Pi^{1,\mathfrak{S}}]\big) = e_0 C_G[\iota_l^{-1}\Pi^{1,\infty,p}]\Bigg(\sum_{0\le h_1,h_2\le n-1}[\mathrm{Mant}_{(h_1,h_2)}(\mathrm{Red}_n^{(h_1,h_2)}(\pi_p))]\Bigg)$$

in $\mathrm{Groth}(\mathbb{G}(\mathbb{A}^{\infty,p})\times G(\mathbb{Q}_p)\times W_K)$. We claim that

$$\sum_{0\le h_1,h_2\le n-1}[\mathrm{Mant}_{(h_1,h_2)}(\mathrm{Red}_n^{(h_1,h_2)}(\pi_p))] = [\pi_p]\big[(\pi_{p,0}\circ \mathrm{Art}_{\mathbb{Q}_p}^{-1})|_{W_K}\otimes \iota_l^{-1}\mathscr{L}_{K,n}(\Pi^0_{F',w})\big]. \quad (2.3)$$

By its definition above Theorem 5.6 of [Caraiani 2012], the morphism $\mathrm{Red}_n^{(h_1,h_2)}(\pi_p)$ breaks down as a product

$$(-1)^{h_1+h_2}\pi_{p,0}\otimes \mathrm{Red}^{n-h_1,h_1}(\pi_{w_1})\otimes \mathrm{Red}^{n-h_2,h_2}(\pi_{w_2})\otimes \bigotimes_{w\ne w_1,w_2}\pi_w,$$

where w runs over places above the place of p induced by τ_E other than w_1 and w_2. The morphism

$$\mathrm{Red}^{n-h,h}:\mathrm{Groth}(\mathrm{GL}_n(K))\to \mathrm{Groth}(D^\times_{K,n-h}\times \mathrm{GL}_h(K))$$

is also defined above Theorem 5.6 of [Caraiani 2012]. On the other hand, the functor $\mathrm{Mant}_{(h_1,h_2)}$ also decomposes as a product (see [Shin 2011, Formula 5.6]), into

$$\mathrm{Mant}_{(h_1,h_2)}(\rho) = \mathrm{Mant}_{1,0}(\rho_0)\otimes \mathrm{Mant}_{n-h_1,h_1}(\rho_{w_1})\otimes \mathrm{Mant}_{n-h_2,h_2}(\rho_{w_2})\otimes \bigotimes_{w\ne w_1,w_2}\mathrm{Mant}_{0,m}(\rho_w),$$

where w again runs over places above the place of p induced by τ_E other than w_1 and w_2. So

$$\sum_{0\le h_1,h_2\le n-1}[\mathrm{Mant}_{(h_1,h_2)}(\mathrm{Red}_n^{(h_1,h_2)}(\pi_p))] = [\mathrm{Mant}_{1,0}(\pi_{p,0})]\otimes \sum_{h_1=0}^{n-1}(-1)^{h_1}[\mathrm{Mant}_{n-h_1,h_1}(\mathrm{Red}^{n-h_1,h_1}(\pi_{w_1}))] \otimes \sum_{h_2=0}^{n-1}(-1)^{h_2}[\mathrm{Mant}_{n-h_2,h_2}(\mathrm{Red}^{n-h_2,h_2}(\pi_{w_2}))]\otimes \bigotimes_{w\ne w_1,w_2}[\pi_w].$$

Now by applying Propositions 2.2(i) and 2.3 of [Shin 2011] we get the desired result (note that the normalization used in their statements is slightly different than ours, but the relation between the two different normalizations is explained above the statement of Proposition 2.3).

Applying Equation (2.3), we first see that

$$BC\big(H(X,\mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}]\big) \\ = e_0 C_G[\iota_l^{-1}\Pi^{1,\infty}]\big[(\pi_{p,0}\circ \mathrm{Art}_{\mathbb{Q}_p}^{-1})|_{W_K} \otimes \iota_l^{-1}\mathcal{L}_{K,n}(\Pi^0_{F',w})\big] \quad (2.4)$$

in $\mathrm{Groth}(\mathbb{G}(\mathbb{A}^\infty)\times W_K)$, which means that

$$BC\big(H(X,\mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}]\big) = e_0[\iota_l^{-1}\Pi^{1,\infty}][R'(\Pi^1)],$$

for some $[R'(\Pi^1)] \in \mathrm{Groth}(\mathrm{Gal}(\bar{F}'/F))$. We show now that

$$[R'(\Pi^1)] = C_G[R(\Pi^0_{F'})^{\otimes 2}\otimes \mathrm{rec}_{l,\iota_l}(\psi)]$$

in $\mathrm{Groth}(\mathrm{Gal}(\bar{F}/F'))$, using the Cebotarev density theorem. Note first that $R'(\Pi^1)$ is simply the sum of (the alternating sum of) $R^k_{\xi,l}(\pi^\infty)$, where π^∞ runs over $\mathrm{Irr}_l(G(\mathbb{A}^\infty))$ such that

- $BC(\iota_l\pi^{\mathfrak{S}}) \simeq \Pi^{1,\mathfrak{S}}$,
- $BC(\iota_l\pi_{\mathfrak{S}_{\mathrm{fin}}}) \simeq \Pi_{\mathfrak{S}_{\mathrm{fin}}}$,
- $R^k_{\xi,l}(\pi^\infty) \neq 0$ for some k.

The set of such π doesn't depend on $\mathfrak{S}$ if $\mathfrak{S}$ is chosen as described above this proposition, so the Galois representation $R'(\Pi^1)$ is also independent of $\mathfrak{S}$. Therefore, for any prime w_1 of F where Π^1 is unramified and which is above a prime w of F' which splits in F and above a prime $p \neq l$ of $\mathbb{Q}$ which splits in E, we can choose a finite set of places $\mathfrak{S}$ containing p such that we get from Equation (2.4)

$$[R'(\Pi^1)|_{W_{F_{w_1}}}] = C_G\big[(R(\Pi^0_{F'})^{\otimes 2}\otimes \mathrm{rec}_{l,\iota_l}(\psi))_{W_{F_{w_1}}}\big].$$

By the Cebotarev density theorem (which tells us the Frobenius elements of primes w_1 are dense in $\mathrm{Gal}(\bar{F}'/F)$) we conclude that

$$[R'(\Pi^1)] = C_G[R(\Pi^0_{F'})^{\otimes 2}\otimes \mathrm{rec}_{l,\iota_l}(\psi)]$$

in $\mathrm{Groth}(\mathrm{Gal}(\bar{F}/F'))$.

It remains to see that $e_0 = 1$ and that $H^k(X,\mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}] = 0$ unless $k = 2n-2$. In fact, it suffices to show the latter, since then $H(X,\mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}]$ will have to be an actual representation, so that would force $e_0 = 1$. The fact that $H^k(X,\mathcal{L}_\xi)[\Pi^{1,\mathfrak{S}}] = 0$ for $k \neq 2n-2$ can be seen, as in the proof of Corollary 7.3 of [Caraiani 2012], by choosing a prime $p \neq l$ to work with and applying the spectral sequences in Proposition 7.2 of [ibid.], and noting that the terms of those spectral sequence are 0 outside the diagonal corresponding to $k = 2n-2$. □

Corollary 2.2. *By Lemmas* 1.4 *and* 1.7 *of* [Taylor and Yoshida 2007] *and by the same argument as in the proof of Theorem* 7.4 *of* [Caraiani 2012], *in order to show that*

$$WD\big(R_l(\Pi^0_{F'})|_{\mathrm{Gal}(\overline{F'_{\mathfrak{p}}}/F'_{\mathfrak{p}})}\big)^{\text{F-ss}}$$

is pure, *it suffices to show that*

$$WD\big(BC(H^{2n-2}(X, \mathscr{L}_\xi)[\Pi^{\mathfrak{S}}])|_{\mathrm{Gal}(\overline{F'_{\mathfrak{p}}}/F'_{\mathfrak{p}})}\big)^{\text{F-ss}}$$

is pure, *where* $\mathfrak{S}$ *is chosen such that it contains* l.

At this point, we've reduced the question of proving the local-global compatibility of monodromy operators when $l = p$ to proving that the $\Pi^{\mathfrak{S}}$-part of the cohomology of a system of proper, smooth Shimura varieties over F' gives rise to a pure Weil–Deligne representation. In the rest of this section, we shall describe integral models of these Shimura varieties which are no longer smooth but are log smooth and of Cartier type. We shall relate their log crystalline cohomology to the Weil–Deligne representation we are interested in. The upshot is that we reduce the question of local-global compatibility to proving the purity of (the $\Pi^{\mathfrak{S}}$-part of) certain log crystalline cohomology groups. This statement is made precise in Corollary 2.3 below.

Recall that $\mathfrak{p}$ is a place of F' above l such that $\mathfrak{p} = \mathfrak{p}_1\mathfrak{p}_2$. From now on, set $K := F_{\mathfrak{p}_1} \simeq F_{\mathfrak{p}_2}$, where the isomorphism is via σ. Let $\mathbb{O}_K$ be the ring of integers in K with uniformizer ϖ and residue field k. For $i = 1, 2$ let $\mathrm{Iw}_{n,\mathfrak{p}_i}$ be the subgroup of matrices in $\mathrm{GL}_n(\mathbb{O}_K)$ which reduce modulo $\mathfrak{p}_i$ to the Borel subgroup $B_n(k)$. Now we set

$$U_{\mathrm{Iw}} = U^l \times U_l^{\mathfrak{p}_1,\mathfrak{p}_2}(m) \times \mathrm{Iw}_{n,\mathfrak{p}_1} \times \mathrm{Iw}_{n,\mathfrak{p}_2} \subset G(\mathbb{A}^\infty),$$

for some $U^l \subset G(\mathbb{A}^\infty)$ compact open and $U_l^{\mathfrak{p}_1,\mathfrak{p}_2}$ a congruence subgroup at l away from $\mathfrak{p}_1$ and $\mathfrak{p}_2$. In Section 2.2 of [Caraiani 2012], an integral model for $X_{U_{\mathrm{Iw}}}/\mathbb{O}_K$ is defined. This is a proper scheme of dimension $2n-1$ with smooth generic fiber. The special fiber $Y_{U_{\mathrm{Iw}}}$ has a stratification by closed Newton polygon strata $Y_{U_{\mathrm{Iw}},S,T}$ with $S, T \subseteq \{1, \dots, n\}$ nonempty subsets. These strata are proper, smooth schemes over k of dimension $2n - \#S - \#T$. In fact,

$$Y_{U_{\mathrm{Iw}},S,T} = \Big(\bigcap_{i\in S} Y_{1,i}\Big) \cap \Big(\bigcap_{j\in T} Y_{2,j}\Big),$$

where each $Y_{i,j}$ for $i = 1, 2$ and $j = 1, \dots, n$ is cut out by one local equation. We can also define

$$Y_{U_{\mathrm{Iw}}}^{(l_1,l_2)} = \coprod_{\substack{S,T\subseteq\{1,\dots,n\}\\ \#S=l_1\\ \#T=l_2}} Y_{U_{\mathrm{Iw}},S,T}.$$

By Proposition 2.8 of [Caraiani 2012], the completed local rings of $X_{U_{\mathrm{Iw}}}$ at closed geometric points s of $X_{U_{\mathrm{Iw}}}$ are isomorphic to

$$\mathcal{O}^{\wedge}_{X_{U_{\mathrm{Iw}}},s} \simeq W_{(K)}[\![X_1,\dots,X_n,Y_1,\dots,Y_n]\!]/(X_{i_1}\cdots X_{i_r}-\varpi, Y_{j_1}\cdots Y_{j_s}-\varpi),$$

where $\{i_1,\dots,i_r\}\subseteq\{1,\dots,n\}$, $\{j_1,\dots,j_r\}\subseteq\{1,\dots,n\}$ and $W_{(K)}$ is the ring of integers in the completion of the maximal unramified extension of K. The closed subscheme Y_{1,i_l} is cut out in $\mathcal{O}^{\wedge}_{X_{U_{\mathrm{Iw}}},s}$ by $X_{i_l}=0$ and Y_{2,j_l} is cut out by $Y_{j_l}=0$.

The action of $G(\mathbb{A}^{\infty,p})$ extends to the inverse system $X_{U_{\mathrm{Iw}}}/\mathcal{O}_K$. There is a universal abelian variety $\mathcal{A}_{U_{\mathrm{Iw}}}/\mathcal{O}_K$ and the actions of $G(\mathbb{A}^{\infty})$ and a_ξ extend to it. We can define a stratification of the special fiber of $\mathcal{A}_{U_{\mathrm{Iw}}}$ by

$$\mathcal{A}_{U_{\mathrm{Iw}},S,T} = \mathcal{A}_{U_{\mathrm{Iw}}} \times_{X_{U_{\mathrm{Iw}}}} X_{U_{\mathrm{Iw}},S,T}.$$

Moreover, $\mathcal{A}_{U_{\mathrm{Iw}}}$ satisfies the same geometric properties as $X_{U_{\mathrm{Iw}}}$ with respect to the above stratification of its special fiber and the analogous statement holds for $\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}$. In particular, we shall see in the next section (or it follows from Section 3 of [Caraiani 2012]) that $\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}$ can be endowed with a vertical logarithmic structure M such that

$$(\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}, M) \to (\operatorname{Spec}\mathcal{O}_K, \mathbb{N})$$

is log smooth, where $(\operatorname{Spec}\mathcal{O}_K,\mathbb{N})$ is the canonical log structure associated to the closed point. Also, we will see that its special fiber is of Cartier type. This means that we can define the log crystalline cohomology of $(\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}, M)$. Indeed, if $W=W(k)$ is the ring of Witt vectors of k, then we let

$$H^*_{\mathrm{cris}}(\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}/W)$$

be the log crystalline cohomology of $(\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}\times_{\mathcal{O}_K} k, M)$ (here we suppress M from the notation). This also has an action of a_ξ as an idempotent and of $G(\mathbb{A}^{\mathfrak{S}})$.

From the isomorphism

$$H^{2n-2}(X,\mathcal{L}_\xi) \simeq a_\xi H^{2n-2+m_\xi}(\mathcal{A}^{m_\xi}, \overline{\mathbb{Q}}_l(t_\xi))$$

and Corollary 2.2, we see that it is enough to show that

$$a_\xi WD\big(H^{2n-2+m_\xi}(\mathcal{A}^{m_\xi}, \overline{\mathbb{Q}}_l(t_\xi)|_{\mathrm{Gal}(\overline{K}/K)})[\Pi^{1,\mathfrak{S}}]\big)$$

is pure. Let $\tau_0 : W \hookrightarrow \overline{\mathbb{Q}}_l$ be an embedding over $\mathbb{Z}_l$. By the semistable comparison theorem of [Nizioł 2008], we have

$$\begin{aligned}\varinjlim_{U_{\mathrm{Iw}}} a_\xi\big(H^{2n-2+m_\xi}_{\mathrm{cris}}(\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}\times_{\mathcal{O}_K} k/W)\otimes_{W,\tau_0}\overline{\mathbb{Q}}_l(t_\xi)\big)[\Pi^{1,\mathfrak{S}}]\\ \simeq \varinjlim_{U_{\mathrm{Iw}}} a_\xi WD\big(H^{2n-2+m_\xi}(\mathcal{A}^{m_\xi}\times_{\mathcal{O}_K}\overline{K}, \overline{\mathbb{Q}}_l(t_\xi)|_{\mathrm{Gal}(\overline{K}/K)})[\Pi^{1,\mathfrak{S}}]\big),\end{aligned}$$

where the crystalline cohomology on the left-hand side as constructed in [Hyodo and Kato 1994] has *a priori* the structure of a (φ, N)-module over W, but which gives rise to a Weil–Deligne representation (r, N) of W_K by setting $r(\sigma) := \varphi^{n[k:\mathbb{F}_p]}$ whenever $\sigma \in W_K$ is a lift of Frob_k^n. Therefore, it suffices to understand the (direct limit of the) log crystalline cohomology of the special fiber of $\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}}}$. Note that the semistable comparison theorem was first proved by Kato [1994a] and Tsuji [1999] for proper, vertical log schemes with semistable reduction; the reason for citing Niziol's work is that her main theorem applies to a general fine and saturated, log-smooth, proper, vertical $(\operatorname{Spec} \mathbb{O}_K, \mathbb{N})$-scheme with special fiber of Cartier type. The fact that $(\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}}}, M)$ satisfies all these properties follows immediately from the explicit description of the log structure M in Section 3.

We summarize the above discussion in the following corollary:

Corollary 2.3. *The Weil–Deligne representation*

$$WD(R_l(\Pi^0_{F'})|_{\mathrm{Gal}(\overline{F'_{\mathfrak{p}}}/F'_{\mathfrak{p}})})^{\text{F-ss}}$$

is pure if

$$\lim_{\overrightarrow{U_{\mathrm{Iw}}}} a_\xi\big(H^{2n-2+m_\xi}_{\mathrm{cris}}(\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}}} \times_{\mathbb{O}_K} k/W) \otimes_{W,\tau_0} \overline{\mathbb{Q}}_l(t_\xi)\big)[\Pi^{1,\mathfrak{S}}]$$

is pure, *where $\mathfrak{S}$ is chosen such that it contains l.*

3. Log crystalline cohomology

3A. ***Log structures.*** Let $\mathbb{O}_K$ be the ring of integers in a finite extension K of $\mathbb{Q}_p$ with uniformizer ϖ and residue field k. (Here, p is some prime number, which will be taken to equal l for our applications to local-global compatibility.) Let $W = W(k)$ be the ring of Witt vectors of k, with $W_n = W_n(k)$ referring to the Witt vectors of length n over k. Let $W_{(K)}$ be the ring of integers in the completion of the maximal unramified extension of K.

Let $X/\mathbb{O}_K$ be a scheme locally of finite type such that the completions of the strict henselizations $\mathbb{O}^\wedge_{X,s}$ at closed geometric points s of X are isomorphic to

$$W_{(K)}[\![X_1, \dots, X_n, Y_1, \dots, Y_n, Z_1, \dots, Z_m]\!]/(X_{i_1} \cdots X_{i_r} - \varpi, Y_{j_1} \cdots Y_{j_s} - \varpi)$$

for some indices $i_1, \dots, i_r, j_1, \dots, j_s \in \{1, \dots, n\}$ and some $1 \le r, s \le n$. Also assume that the special fiber Y is a union of closed subschemes $Y_{1,j}$ with $j \in \{1, \dots, n\}$ which are cut out by one local equation in $\mathbb{O}_X$, such that if s is a closed geometric point of $Y_{1,j}$, then $j \in \{i_1, \dots, i_r\}$ and $Y_{1,j}$ is cut out in $\mathbb{O}^\wedge_{X,s}$ by the equation $X_j = 0$. Similarly, assume that Y is a union of closed subschemes $Y_{2,j}$ with $j \in \{1, \dots, n\}$, which are cut out by one local equation in $\mathbb{O}_X$ such that if s is a closed geometric point of $Y_{2,j}$ then $j \in \{j_1, \dots, j_r\}$ and $Y_{2,j}$ is cut out in $\mathbb{O}^\wedge_{X',s}$ by the equation $Y_j = 0$. Then, by Lemma 2.9 of [Caraiani 2012], X is Zariski-locally étale over

$$X_{r,s,m} = \operatorname{Spec} \mathcal{O}_K[X_1, \dots, X_n, Y_1, \dots, Y_n, Z_1, \dots, Z_m]/(X_1 \cdots \varpi, Y_1 \cdots Y_s - \varpi).$$

The closed subschemes $Y_{i,j}$ for $i = 1, 2$ and $j = 1, \dots, n$ are Cartier divisors, which in the local model $X_{r,s,m}$ correspond to the divisors $X_j = 0$ or $Y_j = 0$.

Let Y/k be the special fiber of X. For $1 \le i, j \le n$ we define $Y^{(i,j)}$ to be the disjoint union of the closed subschemes of Y

$$(Y_{1,l_1} \cap \cdots \cap Y_{1,l_i}) \bigcap (Y_{2,m_1} \cap \cdots \cap Y_{2,m_j}),$$

as $\{l_1, \dots, l_i\}$ (resp. $\{m_1, \dots, m_j\}$) range over subsets of $\{1, \dots, n\}$ of cardinality i (resp. j). Each $Y^{(i,j)}$ is a proper smooth scheme over k of dimension $2n - i - j$.

Remark 3.1. Even though this section is general, we will only apply the results of this section in the case when X is $\mathcal{A}_{U_{\mathrm{Iw}}}$ for some compact open subgroup $U_{\mathrm{Iw}} \subset G(\mathbb{A}^\infty)$ with Iwahori-level structure at $\mathfrak{p}_1$ and $\mathfrak{p}_2$. $X_{U_{\mathrm{Iw}}}$ (and therefore $\mathcal{A}_{U_{\mathrm{Iw}}}$ as well) satisfies the above conditions by Proposition 2.8 of [Caraiani 2012]. The prime p is meant to be identified with l.

Let $(\operatorname{Spec} \mathcal{O}_K, \mathbb{N})$ be the log scheme corresponding to $\operatorname{Spec} \mathcal{O}_K$ endowed with the canonical log structure associated to the special fiber. This is given by the map $1 \in \mathbb{N} \mapsto \varpi \in \mathcal{O}_K$. We endow X with the log structure M associated to the special fiber Y. Let $j : X_K \to X$ be the open immersion and $i : Y \to X$ be the closed immersion. This log structure is defined by

$$M = j_*(\mathcal{O}_{X_K}^\times) \cap \mathcal{O}_X \to \mathcal{O}_X.$$

We have a map of log schemes $(X, M) \to (\operatorname{Spec} \mathcal{O}_K, \mathbb{N})$, given by sending $1 \in \mathbb{N}$ to $\varpi \in M$. Locally, we have a chart for this map, given by

$$\mathbb{N} \to \mathbb{N}^r \oplus \mathbb{N}^s/(1, \dots, 1, 0, \dots, 0) = (0, \dots, 0, 1, \dots, 1),$$
$$1 \mapsto (1, \dots, 1, 0, \dots, 0) = (0, \dots, 0, 1, \dots, 1).$$

It is easy to see from this that $(X, M)/(\operatorname{Spec} \mathcal{O}_K, \mathbb{N})$ is log smooth and that the log structure M on X is fine, saturated and vertical. We can pull back M to a log structure on Y, which we still denote M and then we get a log-smooth map of log schemes

$$(Y, M) \to (\operatorname{Spec} k, \mathbb{N}).$$

(Here we have the canonical log structure on k associated to $1 \in \mathbb{N} \mapsto 0 \in k$, which is the same as the pullback of the canonical log structure on $\operatorname{Spec} \mathcal{O}_K$.) Note that, since (X, M) is saturated over $(\operatorname{Spec} \mathcal{O}_K, \mathbb{N})$, its special fiber is of Cartier type (see [Tsuji 1997]).

We can also endow X with log structures $\widetilde{M}_1$, $\widetilde{M}_2$ and $\widetilde{M}$. Let $U_{i,j}$ be the complement of $Y_{i,,j}$ in X for $i = 1, 2$ and $j = 1, \dots, n$. Let

$$j_{i,j} : U_{i,j} \to X$$

denote the open immersion. We define $\widetilde{M}_1$, $\widetilde{M}_2$ and $\widetilde{M}$ as follows

$$\widetilde{M}_1 = \Big(\bigoplus_{j=1}^{n} \big(j_{1,j*}(\mathbb{O}_{U_{1,j}}^{\times}) \cap \mathbb{O}_X\big)\Big) / \sim,$$

$$\widetilde{M}_2 = \Big(\bigoplus_{j=1}^{n} \big(j_{1,j*}(\mathbb{O}_{U_{1,j}}^{\times}) \cap \mathbb{O}_X\big)\Big) / \sim,$$

$$\widetilde{M} = \Big(\bigoplus_{j=1}^{n} \big(j_{1,j*}(\mathbb{O}_{U_{1,j}}^{\times}) \cap \mathbb{O}_X\big) \oplus \bigoplus_{j=1}^{n} \big(j_{2,j*}(\mathbb{O}_{U_{2,j}}^{\times}) \cap \mathbb{O}_X\big)\Big) / \sim,$$

where $\sim$ signifies that we have identified the image of $\mathbb{O}_X^{\times}$ in all the terms of the direct sums (in other words, we are taking an amalgamated sum of the log structures associated to each of the $Y_{i,j}$). We have a map $\widetilde{M} \to M$ given by inclusion on each $\mathbb{O}_{U_{i,j}}^{\times}$.

Lemma 3.2. *Locally on X, we have a chart for $\widetilde{M}$ given by*

$$X \to \operatorname{Spec} \mathbb{O}_k[X_1, \dots, X_n, Y_1, \dots, Y_n, Z_1, \dots, Z_m]/(X_1 \cdots X_r - \varpi, Y_1 \cdots Y_r - \varpi) \to \operatorname{Spec} \mathbb{Z}[\mathbb{N}^r \oplus \mathbb{N}^s],$$

where $(0, \dots, 0, 1, 0, \dots, 0) \mapsto X_i$ *if the* 1 *is in the* i*-th position and* $1 \le i \le r$ *and* $(0, \dots, 0, 1, 0, \dots, 0) \mapsto Y_{i-r}$ *if the* 1 *is in the* i*-th position and* $r + 1 \le i \le r + s$.

Proof. We shall make use of Kato and Niziol's results on log smoothness and log regularity, namely:

- If $f : T \to S$ is a log smooth morphism of fs log schemes with S log regular then T is log regular (see 8.2 of [Kato 1994b]).
- If T is log regular, then $M_T = j_* \mathbb{O}_U^{\times} \cap \mathbb{O}_X$, where $j : U \hookrightarrow T$ is the inclusion of the open subset of triviality of T (see 8.6 of [Niziol 2006]).

Let us define the following log schemes over $(\operatorname{Spec} \mathbb{O}_K, \mathrm{triv})$:

$$\widetilde{U} := \operatorname{Spec} \mathbb{O}_K[X_1, \dots, X_n, \sigma]/(X_1 \cdots X_r - \sigma),$$

$$\widetilde{V} := \operatorname{Spec} \mathbb{O}_K[Y_1, \dots, Y_n, \tau]/(Y_1 \cdots Y_s - \tau),$$

$$W := \operatorname{Spec} \mathbb{O}_K[Z_1, \dots, Z_m],$$

$$Z := \widetilde{U} \times_{(\operatorname{Spec} \mathbb{O}_K, \mathrm{triv})} \widetilde{V} \times_{(\operatorname{Spec} \mathbb{O}_K, \mathrm{triv})} W.$$

Then Z, equipped with the product log structure L, is smooth over $\mathcal{O}_K$ and log smooth over $(\operatorname{Spec}\mathcal{O}_K[\sigma,\tau], \mathrm{triv})$. Therefore, Z is regular. The log structure L is given by the simple normal crossings divisor

$$D := \Big(\bigcup_{j=1}^{r}(X_j = 0)\Big) \cup \Big(\bigcup_{j=1}^{s}(Y_j = 0)\Big).$$

Since Z is regular, the log structure L is the same as the amalgamation of the log structures defined by the smooth divisors $(X_j = 0)$, $(Y_j = 0)$. Locally on X, we have a commutative diagram of schemes with a cartesian square

$$\begin{tikzcd} X \arrow[r] & X_{r,s,m} \arrow[r] \arrow[d] & Z \arrow[d] \\ & \operatorname{Spec}\mathcal{O}_K \arrow[r] & \operatorname{Spec}\mathcal{O}_K[\tau,\sigma] \end{tikzcd} \tag{3A.1}$$

where the inverse image of $(X_j = 0)$ in X is Y_j^1 and the inverse image of $(Y_j = 0)$ in X is Y_j^2. Therefore, the log structure on X induced by that of Z coincides with the log structure $\widetilde{M}$ defined as the amalgamated sum of the log structures induced by the Y_j^1 and Y_j^2. □

If we endow $\operatorname{Spec}\mathcal{O}_K$ with the log structure $\mathbb{N}^2$ associated to $(a,b) \in \mathbb{N}^2 \mapsto \pi^{a+b} \in \mathcal{O}_K$, then we claim that we have a log-smooth map of log schemes

$$(X, \widetilde{M}) \to (\operatorname{Spec}\mathcal{O}_K, \mathbb{N}^2) \tag{3A.2}$$

whose chart is given locally by

$$(a,b) \in \mathbb{N}^2 \mapsto (a,\dots,a,b,\dots b) \in \mathbb{N}^r \oplus \mathbb{N}^s.$$

By definition, $\widetilde{M}$ is the amalgamated sum of $\widetilde{M}_1$ and $\widetilde{M}_2$ as log structures on X (or, in other words, $\widetilde{M}$ is the log structure associated to the prelog structure $\widetilde{M}_1 \oplus \widetilde{M}_2 \to \mathcal{O}_X$). Therefore, it suffices to prove the following lemma:

Lemma 3.3. *We can define a global map of log schemes* $(X, \widetilde{M}_1) \to (\operatorname{Spec}\mathcal{O}_K, \mathbb{N})$ *which locally admits the chart given by the diagonal embedding* $\mathbb{N} \to \mathbb{N}^r$.

Proof. It suffices to show that ϖ is a global section of $\widetilde{M}_1$, since then we can simply map $1 \in \mathbb{N}$ to $\varpi \in \widetilde{M}_1$. For this, note that we have a natural map of log structures on X

$$\widetilde{M}_1 \to M,$$

since the open subset of triviality of $\widetilde{M}_1$ is the generic fiber of X and M is the log structure defined by the inclusion of the generic fiber. Moreover, we can check locally that this map is injective, since it can be described by the chart $\mathbb{N}^r \to \mathbb{N}^r \oplus \mathbb{N}^s \to (\mathbb{N}^r \oplus \mathbb{N}^s)/\mathbb{N}$ for $r, s \geq 1$, where the first map is the identity on the first factor. Now, locally on X we have the equation $X_1 \cdots X_r = \varpi$, where X_i are local equations defining the closed subschemes Y_i^1 of X. By definition, the X_i are local sections of $\widetilde{M}_1$, so ϖ is a local section of $\widetilde{M}_1$. But ϖ is also a global section of M and $\widetilde{M}_1 \hookrightarrow M$, so ϖ is a global section of $\widetilde{M}_1$. □

Lemma 3.4. *We have a cartesian diagram of maps of log schemes*

$$\begin{tikzcd}
(X, M) \arrow[r] \arrow[d] & (X, \widetilde{M}) \arrow[d] \\
(\operatorname{Spec} \mathcal{O}_K, \mathbb{N}) \arrow[r] & (\operatorname{Spec} \mathcal{O}_K, \mathbb{N}^2)
\end{tikzcd}$$

where the bottom horizontal arrow is the identity on the underlying schemes and maps $(a, b) \in \mathbb{N}^2$ *to* $a + b \in \mathbb{N}$.

Proof. We go back to the notation used in the proof of Lemma 3.2. Locally on X, we have the following commutative diagram of log schemes

$$\begin{tikzcd}
(X, M) \arrow[r] \arrow[d] & \widetilde{U} \times_{\operatorname{Spec} \mathcal{O}_K[u]} \widetilde{V} \times W \arrow[r] \arrow[d] & Z \arrow[d] \\
(\operatorname{Spec} \mathcal{O}_K, \mathbb{N}) \arrow[r] & (\operatorname{Spec} \mathcal{O}_K[u], \mathbb{N}) \arrow[r] & (\operatorname{Spec} \mathcal{O}_K[\tau, \sigma], \mathbb{N}^2)
\end{tikzcd}$$

where in the bottom row both τ and σ are mapped to u, which is in turn mapped to 0. The second square is cartesian and the horizontal maps in it are closed, but not exact, immersions. The first bottom map is an exact closed immersion, while the first top map is the composition of an étale morphism with an exact closed immersion. The lemma follows from the commutative diagram (3A.1) and the above diagram. □

3B. ***Variations on the logarithmic de Rham–Witt complex.*** Define the prelog structure $\mathbb{N}^2 \to W_n[\tau, \sigma]$ given by $(a, b) \mapsto \tau^a \sigma^b$. By abuse of notation, we write $(\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$ for the log scheme endowed with the associated log structure. We have the composite map of log schemes

$$(Y, \widetilde{M}) \to (\operatorname{Spec} k, \mathbb{N}^2) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2),$$

where $\mathbb{N}^2 \to \mathbb{N}^2$ is the obvious isomorphism. We shall call $(Z, \widetilde{N})$ a *lifting* for this morphism if $(Z, \widetilde{N})$ is a fine log scheme such that the composite map $(Y, \widetilde{M}) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$ factors through $f : (Y, \widetilde{M}) \to (Z, \widetilde{N})$ which is a closed immersion, and a map $(Z, \widetilde{N}) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$ which is log smooth. Such liftings always exists locally on Y and give rise to embedding systems as defined in Paragraph 2.18 of [Hyodo and Kato 1994]. If $(U, \widetilde{M}_U) \to (Y, \widetilde{M})$ is a covering and $(Z, \widetilde{N})$ is a lifting for $(U, \widetilde{M}_U) \to (\operatorname{Spec} W_n[\tau, \sigma]), \mathbb{N}^2)$, then we may define an embedding system $((U^i, \widetilde{M}^i_U), (Z^i, \widetilde{N}^i))$ for $(Y, \widetilde{M}) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$ by taking the fiber product of $i+1$ copies of U over Y and of $i+1$ copies of $(Z, \widetilde{N})$ over $(\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$. Since $(Y, \widetilde{M})$ is an fs log scheme, we may assume the same for the local lifting $(Z, \widetilde{N})$.

Let $C^\bullet_{(Y,\widetilde{M})/(W_n,\mathrm{triv})}$ be the crystalline complex associated to the embedding system obtained from local liftings $(Z^\bullet, \widetilde{N}^\bullet)$, and define

$$\tilde{\tilde{C}}^\bullet_Y := C_{(Y,\widetilde{M})^\bullet/(W_n,\mathrm{triv})} \otimes_{W_n\langle\tau,\sigma\rangle} W_n.$$

Let $\operatorname{Spec} W_n[u]$ be endowed with the log structure associated to $1 \in \mathbb{N} \mapsto u \in W_n[u]$. Consider the map of log schemes $G : (\operatorname{Spec} W_n[u], \mathbb{N}) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$ given by $\tau, \sigma \mapsto u$ and $(a, b) \in \mathbb{N}^2 \mapsto a + b \in \mathbb{N}$. The pullback of $(Y, \widetilde{M})$ along G is just (Y, M). Let (Z', N') be the pullback of $(Z, \widetilde{N})$ along G, equipped with a map $f' : (Y', M') \to (Z', N')$ which is the pullback of f. Then (Z', N') is a (local) lifting for $(Y, M) \to (\operatorname{Spec} W_n[u], \mathbb{N})$, and gives rise to an embedding system for this morphism. Indeed, what we need to check is that $(Z', N') \to (\operatorname{Spec} W_n[u], \mathbb{N})$ is log smooth and that f' is a closed immersion of log schemes. For the first we note that log-smoothness is preserved under base change in the category of log schemes, and that

$$(Z', N') = \big(((Z, \widetilde{N}) \times_G (\operatorname{Spec} W_n[u], \mathbb{N}))^{\mathrm{int}}\big)^{\mathrm{sat}} \to (Z, \widetilde{N}) \times_G (\operatorname{Spec} W_n[u], \mathbb{N})$$

is log smooth. We also note that $g : Y \to (Z \times_{\operatorname{Spec} W_n[\tau,\sigma]} \operatorname{Spec} W_n[u])$ is a closed immersion, since $Y \to Z$ is a closed immersion. The morphism of schemes $Z' \to (Z \times_{\operatorname{Spec} W_n[\tau,\sigma]} \operatorname{Spec} W_n[u])$ is a composition of a finite morphism with a closed immersion, so $Y \to Z'$ is a closed immersion as well. Also, $g^*(\widetilde{N} \oplus_{\mathbb{N}^2} \mathbb{N}) \to M$ is surjective and factors through $(f')^*(N') \to M$, so $(f')^*(N') \to M$ is surjective as well.

We now follow the constructions in Section 3.6 of [Hyodo and Kato 1994] using the embedding system obtained from the liftings (Z', N'). Let $C^\bullet_{(Y,M)/(W_n,\mathrm{triv})}$ be the crystalline complex associated to the composite $(Z', N') \to (W_n, \mathrm{triv})$. Define

$$\widetilde{C}^\bullet_Y := C^\bullet_{(Y,M)/(W_n,\mathrm{triv})} \otimes_{W_n\langle u\rangle} W_n.$$

On the other hand, let $Z'' = Z' \times_{\operatorname{Spec} W_n[u]} \operatorname{Spec} W_n\langle u\rangle$ be endowed with N'' the inverse image of the log structure N'. Let $\mathscr{L}$ be the log structure on $\operatorname{Spec} W_n\langle u\rangle$ obtained by taking the inverse image of (the log structure associated to) $\mathbb{N}$ on $\operatorname{Spec} W_n[u]$. Then (Z'', N'') gives rise to an embedding system for

$$(Y, M) \to (\operatorname{Spec} W_n\langle u\rangle, \mathscr{L}),$$

with crystalline complex $C^\bullet_{(Y,M)/(\operatorname{Spec} W_n\langle u\rangle,\mathscr{L})}$. Define

$$C^\bullet_Y := C^\bullet_{(Y,M)/(\operatorname{Spec} W_n\langle u\rangle,\mathscr{L})} \otimes_{W_n\langle u\rangle} W_n.$$

Note that $C_Y^\bullet$ is the crystalline complex $C^\bullet_{(Y,M)/(W_n,\mathbb{N})}$ with respect to the embedding system obtained from $(Z' \times_{\operatorname{Spec} W_n[u]} \operatorname{Spec} W_n, N''')$. As in Section 3.6 of [Hyodo and Kato 1994], we have an exact sequence of complexes

$$0 \to C_Y^\bullet[-1] \to \widetilde{C}_Y^\bullet \to C_Y^\bullet \to 0, \tag{3B.1}$$

where the second arrow is $\wedge(du/u)$ and the third arrow is the canonical projection. The monodromy operator on the crystalline cohomology of (Y, M) is induced by the connecting homomorphism of this exact sequence.

Lemma 3.5. *Let $C_Z^\bullet$ be one of the complexes $\widetilde{\widetilde{C}}{}^\bullet_Y$, $\widetilde{C}_Y^\bullet$ or $C_Y^\bullet$ obtained with respect to a lifting $(Z, \widetilde{N})$ of some cover $U \to Y$. In the derived category, $C_Z^\bullet$ is independent of the choice of lifting $(Z, \widetilde{N})$.*

Proof. We may work étale locally on Y, in which case we have to show that for any two liftings $(Z_1, \widetilde{N}_1)$ and $(Z_2, \widetilde{N}_2)$ we have a canonical quasi-isomorphism between the corresponding complexes and moreover, that these quasi-isomorphisms satisfy the obvious cocycle condition for three different liftings.

First, we show that the complexes corresponding to $(Z_1, \widetilde{N}_1)$ and $(Z_2, \widetilde{N}_2)$ are quasi-isomorphic. We may assume that $i_i : (Y, \widetilde{M}) \to (Z_i, \widetilde{N}_i)$ is an exact closed immersion for $i = 1, 2$. Let $i_{12} : (Y, \widetilde{M}) \to (Z_1 \times_{W_n} Z_2, \widetilde{N}_{1\times 2})$ be the diagonal immersion of $(Y, \widetilde{M})$ into the fiber product of $(Z_1, \widetilde{N}_1)$ and $(Z_2, \widetilde{N}_2)$ as fs log schemes over (W_n, triv). Let $(Z_{12}, \widetilde{N}_{12})$ be a log scheme such that étale locally on Y we have a factorization of i_{12}

$$(Y, \widetilde{M}) \xrightarrow{f} (Z_{12}, \widetilde{N}_{12}) \xrightarrow{g} (Z_1 \times Z_2, \widetilde{N}_{1\times 2}),$$

with g log étale and f an exact closed immersion. This factorization is possible by Lemma 4.10 of [Kato 1989]. Let D_i be the PD-envelope of Y in Z_i (again, for $i = 1, 2$ or 12). (Since we have exact closed immersions, the logarithmic PD-envelope coincides with the usual PD-envelope in these cases.) It suffices to show that the canonical map

$$\omega^\bullet_{(Z_1,\widetilde{N}_1)/W_n,\mathrm{triv}} \otimes_{\mathbb{O}_{Z_1}} \mathbb{O}_{D_1} \to \omega^\bullet_{(Z_{12},\widetilde{N}_{12})/W_n,\mathrm{triv}} \otimes_{\mathbb{O}_{Z_{12}}} \mathbb{O}_{D_{12}} \tag{3B.2}$$

is a quasi-isomorphism. This follows from Paragraph 2.21 of [Hyodo and Kato 1994]. For completeness, we sketch the proof here. Let $p_1 : (Z_{12}, N_{12}) \to (Z_1, N_1)$ be the log-smooth map induced by projection onto the first factor. For any geometric point $\bar{y}$ of Y, the stalks at $\bar{y}$ of N_{12} and $p_1^* N_1$ coincide, so by replacing (Z_{12}, N_{12}) with an étale neighborhood of $\bar{y} \to Z_{12}$, we may assume that $N_{12} = p_1^* N_1$. Then the map $p_1 : Z_{12} \to Z_1$ is smooth in the usual sense. Since the problem is étale local on Y, we may assume that $Z_{12} \simeq Z_1 \otimes_{W_n} W_n[t_1, \dots, t_r]$ for some positive integer r such that Y is contained in the closed subscheme of Z_{12} defined by $t_1 = \cdots = t_r = 0$. As in Proposition 6.5 of [Kato 1989], we also have $\mathbb{O}_{D_{12}} \simeq \mathbb{O}_{D_1}\langle t_1, \dots, t_r\rangle$, the PD-polynomial ring over $\mathbb{O}_{D_1}$ in r variables. The quasi-isomorphism (3B.2) is reduced then to the standard quasi-isomorphism

$$W_n \to \Omega_{W_n[t_1,\dots,t_r]} \otimes_{W_n[t_1,\dots,t_r]} W_n\langle t_1, \dots, t_r\rangle.$$

The quasi-isomorphism (3B.2) commutes with $\otimes_{W_n\langle\tau,\sigma\rangle} W_n$, so it induces a quasi-isomorphism

$$\tilde{\tilde{C}}^{\bullet}_{Z_1} \xrightarrow{\sim} \tilde{\tilde{C}}^{\bullet}_{Z_{12}}.$$

Now consider the morphism $Z'_{12} \to Z'_1$ obtained by pulling back $Z_{12} \to Z_1$ along G. We claim that the canonical morphisms $\tilde{C}^{\bullet}_{Z_{12}} \to \tilde{C}^{\bullet}_{Z_1}$ and $C^{\bullet}_{Z_{12}} \to C^{\bullet}_{Z_1}$ are quasi-isomorphisms as well. This is proved in the same way as in the case of $\tilde{\tilde{C}}^{\bullet}$ (for $C^{\bullet}_{Z_{12}} \to C^{\bullet}_{Z_1}$ this amounts to proving that the logarithmic de Rham–Witt complex is independent of the choice of embedding system). The quasi-isomorphisms are also compatible with the canonical maps $\tilde{\tilde{C}}^{\bullet}_Z \to \tilde{C}^{\bullet}_Z \to C^{\bullet}_Z$.

Note that the above result also implies that in the derived category, $C^{\bullet}$ commutes with étale base change. Indeed, if Y_2/Y_1 is étale and $(Z_1, \tilde{N}_1)$ is a lifting for $(Y_1, \tilde{M}) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$, then by [Grothendieck 1967, 18.1.1] we can find, Zariski locally on Y_2, an étale morphism $Z_2 \to Z_1$ such that the following diagram is cartesian

$$\begin{tikzcd} Y_2 \arrow[r] \arrow[d] & Z_2 \arrow[d] \\ Y_1 \arrow[r] & Z_1 \end{tikzcd}$$

We take $\tilde{N}_2$ on Z_2 to be the inverse image of $\tilde{N}_1$. Then $(Z_2, \tilde{N}_2)$ is a lifting for $(Y_2, \tilde{M}) \to (\operatorname{Spec} W_n[\tau, \sigma], \mathbb{N}^2)$ and, since log differentials commute with étale base change [Kato 1989, Proposition 3.12], $C^{\bullet}_{(Z_2)}$ on Y_2 is just the pullback of $C^{\bullet}_{(Z_2)}$ on Y_1.

We are left with verifying the cocycle condition. The canonical quasi-isomorphism $\gamma_{12} : C^{\bullet}_{Z_1} \xrightarrow{\sim} C_{\bullet}Z_2{}^{\bullet}$ factors through $C^{\bullet}_{Z_1 \times Z_2}$, since by construction Z_{12} is log étale over $Z_1 \times Z_2$ and so we have a quasi-isomorphism $C^{\bullet}_{Z_1 \times Z_2} \xrightarrow{\sim} C^{\bullet}_{Z_{12}}$. Let $(Z_3, \tilde{N}_3)$ be another lifting. Then we have the following commutative diagram of complexes:

$$\begin{tikzcd}
 & & C^{\bullet}_{Z_1\times Z_2\times Z_3} & & \\
C^{\bullet}_{Z_1\times Z_2} \arrow[urr] & & C^{\bullet}_{Z_1\times Z_3} \arrow[u] & & C^{\bullet}_{Z_2\times Z_3} \arrow[ull] \\
C^{\bullet}_{Z_1} \arrow[u] \arrow[urr] \arrow[rr, "\gamma_{12}"] & & C^{\bullet}_{Z_2} \arrow[ull] \arrow[urr] \arrow[rr, "\gamma_{23}"] & & C^{\bullet}_{Z_3} \arrow[ull] \arrow[u]
\end{tikzcd}$$

where all the maps are quasi-isomorphisms. This proves the cocycle condition.

□

Corollary 3.6. *The following sheaves on Y are well-defined and commute with étale base change*:

$$W_n\tilde{\tilde{\omega}}^q_Y := \mathcal{H}^q\big(\tilde{\tilde{C}}^{\bullet}_Y\big), \quad W_n\tilde{\omega}^q_Y := \mathcal{H}^q\big(\tilde{C}^{\bullet}_Y\big) \quad \textit{and} \quad W_n\omega^q_Y := \mathcal{H}^q(C^{\bullet}_Y).$$

The sheaves $W_n\omega^q_Y$ make up the q-th terms of the log de Rham–Witt complex associated to (Y, M). We have canonical morphisms of sheaves on Y:

$$W_n\tilde{\tilde{\omega}}^q_Y \to W_n\tilde{\omega}^q_Y \to W_n\omega^q_Y.$$

In order to understand the monodromy N, we will study the short exact sequence of complexes

$$0 \to W_n\omega^{\bullet}_Y[-1] \to W_n\tilde{\omega}^{\bullet}_Y \to W_n\omega^{\bullet}_Y \to 0,$$

which we obtain below from the short exact sequence (3B.1). In Section 4 we will construct a resolution of this short exact sequence in terms of some subquotients of $W_n\tilde{\tilde{\omega}}^{\bullet}_Y$. For now, since these complexes are independent of the choice of lifting, we will fix some specific kinds of liftings of $(Y, \widetilde{M})$ over $(W[\tau,\sigma], \mathbb{N}^2)$, which we call *admissible liftings,* following the terminology used in [Hyodo 1991] and [Mokrane 1993]. Since Y is locally étale over

$$Y_{r,s,m} = \operatorname{Spec} k[X_1, \dots, X_n, Y_1, \dots, Y_n, Z_1, \dots Z_m]/(X_1 \cdots X_r, Y_1 \cdots Y_s),$$

we consider the lifting

$$Z_{r,s,m} \\ = \operatorname{Spec} W[X_1, \dots, X_n, Y_1, \dots, Y_n, Z_1, \dots Z_m, \tau, \sigma]/(X_1 \cdots X_r - \tau, Y_1 \cdots Y_s - \sigma)$$

of $(Y_{r,s,m}, \mathbb{N}^r \oplus \mathbb{N}^s)/(W[\tau,\sigma], \mathbb{N}^2)$. The log structure on $Z_{r,s,m}$ is also induced from $\mathbb{N}^r \oplus \mathbb{N}^s$ (with the obvious structure map sending $\mathbb{N}^r$ to products of the X_i

and $\mathbb{N}^s$ to products of the Y_j). We let $Z/Z_{r,s,m}$ be étale and such that the diagram

$$\begin{tikzcd} (Y,\widetilde{M}) \arrow[r] \arrow[d] & (Z,\widetilde{N}) \arrow[d] \\ (Y_{r,s,m},\mathbb{N}^r\oplus\mathbb{N}^s) \arrow[r] & (Z_{r,s,m},\mathbb{N}^r\oplus\mathbb{N}^s) \end{tikzcd}$$

is cartesian, with the log structures on top obtained by pullback from the ones on the bottom. Then, locally on Y, the complexes $W_n\tilde{\tilde{\omega}}^\bullet_Y$, $W_n\tilde{\omega}^\bullet_Y$ and $W_n\omega^\bullet_Y$ are just pullbacks of the corresponding complexes on $Y_{r,s,m}$ with respect to the lifting $(Z_{r,s,m},\mathbb{N}^r\oplus\mathbb{N}^s)$. Note that admissible liftings exist locally on Y.

Now we will explain the relationships between $\tilde{\tilde{C}}^\bullet_Y$, $\tilde{C}^\bullet_Y$ and $C^\bullet_Y$. First, note that we have the functoriality map $G^*\omega_{(Z,\widetilde{N})/(W_n,\mathrm{triv})}\to\omega_{(Z',N')/(W_n,\mathrm{triv})}$, which induces a canonical map

$$C^\bullet_{(Y,\widetilde{M})/(W_n,\mathrm{triv})}\otimes_{W_n\langle\tau,\sigma\rangle}W_n\langle u\rangle\to C^\bullet_{(Y,M)/(W_n,\mathrm{triv})},$$

which in turn induces a canonical map $\tilde{\tilde{C}}^\bullet_Y\to\tilde{C}^\bullet_Y$. By composition, we also get a map $\tilde{\tilde{C}}^\bullet_Y\to C^\bullet_Y$. We claim that we can identify $\tilde{C}^\bullet_Y$ with $\tilde{\tilde{C}}^\bullet_Y/(d\tau/\tau-d\sigma/\sigma)\wedge\tilde{\tilde{C}}^\bullet_Y$ and $C^\bullet_Y$ with $\tilde{\tilde{C}}^\bullet_Y/((d\tau/\tau)\wedge\tilde{\tilde{C}}^\bullet_Y+(d\sigma/\sigma)\wedge\tilde{\tilde{C}}^\bullet_Y)$. We explain this in the case of $\tilde{C}^\bullet_Y$:

Lemma 3.7. *We have an isomorphism*

$$\tilde{\tilde{C}}^\bullet_Y/\Big(\frac{d\tau}{\tau}-\frac{d\sigma}{\sigma}\Big)\wedge\tilde{\tilde{C}}^{\bullet-1}_Y\xrightarrow{\sim}\tilde{C}^\bullet_Y.$$

Proof. Let $(Z,\widetilde{N})$ be an admissible lifting of $(Y,\widetilde{M})$ over $(\operatorname{Spec}W_n[\tau,\sigma],\mathbb{N}^2)$. Let $(D,\widetilde{M}_D)$ be the divided power envelope of $(Y,\widetilde{M})$ in $(Z,\widetilde{N})$. Note that the kernel of the map $\mathbb{O}_D\to\mathbb{O}_Y$ is generated by $\tau^{[n]}$ and $\sigma^{[n]}$. The divided power envelope (D',M'_D) of (Y,M) in (Z',N') satisfies the property

$$\mathbb{O}_{D'}\simeq\mathbb{O}_D\otimes_{W_n\langle\tau,\sigma\rangle}W_n\langle u\rangle,$$

where the map $W_n\langle\tau,\sigma\rangle\to W_n\langle u\rangle$ is given by $\tau^{[n]},\sigma^{[n]}\mapsto u^{[n]}$. The complexes $\tilde{\tilde{C}}^\bullet_Y$ and $\tilde{C}^\bullet_Y$ are defined by

$$\begin{aligned}\tilde{\tilde{C}}^\bullet_Y:=&\big(\omega^\bullet_{(Z,\widetilde{N})/(W_n,\mathrm{triv})}\otimes_{\mathbb{O}_Z}\mathbb{O}_D\big)\otimes_{W_n\langle\tau,\sigma\rangle}W_n\langle u\rangle\otimes_{W_n\langle u\rangle}W_n\\ =&\big(\omega^\bullet_{(Z,\widetilde{N})/(W_n,\mathrm{triv})}\otimes_{W_n[\tau,\sigma]}W_n[u]\big)\otimes_{\mathbb{O}_{Z'}}\mathbb{O}_{D'}\otimes_{W_n\langle u\rangle}W_n\end{aligned}$$

and

$$\tilde{C}^\bullet_Y=(\omega^\bullet_{Z',N'/W_n,\mathrm{triv}})\otimes_{\mathbb{O}_{Z'}}\mathbb{O}_{D'}\otimes_{W_n\langle u\rangle}W_n.$$

Note that since we have chosen an admissible lifting, (Z',N') has $Z\times_{W_n[\tau,\sigma]}W_n[u]$ as its underlying scheme because $\widetilde{N}\oplus_{\mathbb{N}^2}\mathbb{N}$ is already fine and saturated. It is enough to show that the sequence

$$\omega^{\bullet-1}_{(Z,\widetilde{N})/(W_n,\mathrm{triv})} \otimes_{W_n[\tau,\sigma]} W_n[u] \to \omega^{\bullet}_{(Z,\widetilde{N})/(W_n,\mathrm{triv})} \otimes_{W_n[\tau,\sigma]} W_n[u] \to \omega^{\bullet}_{(Z',N')/(W_n,\mathrm{triv})} \to 0 \quad \text{(3B.3)}$$

is exact, where the first map is $\wedge(d\tau/\tau - d\sigma/\sigma)$ and the second map is induced by functoriality. We denote by G^* the pullback along $\operatorname{Spec} W_n[u] \to \operatorname{Spec} W_n[\tau,\sigma]$ or along $Z' \to Z$. By Proposition 3.12 of [Kato 1989], we have the following diagram of (vertical) exact sequences of sheaves on Z':

$$\begin{tikzcd}
 & 0 \arrow[d] \\
G^*\omega^1_{(\operatorname{Spec} W_n[\tau,\sigma],\mathbb{N}^2)/(W_n,\mathrm{triv})} \otimes_{W_n[u]} \mathcal{O}_{Z'} \arrow[r] \arrow[d] & \omega^1_{(\operatorname{Spec} W_n[u],\mathbb{N})/(W_n,\mathrm{triv})} \otimes_{W_n[u]} \mathcal{O}_{Z'} \arrow[d] \\
G^*\omega^1_{(Z,\widetilde{N})/(W_n,\mathrm{triv})} \arrow[r] \arrow[d] & \omega^1_{(Z',N')/(W_n,\mathrm{triv})} \arrow[d] \\
G^*\omega^1_{(Z,\widetilde{N})/(\operatorname{Spec} W_n[\tau,\sigma],\mathbb{N}^2)} \arrow[r] \arrow[d] & \omega^1_{(Z',N')/(\operatorname{Spec} W_n[u],\mathbb{N})} \arrow[d] \\
0 & 0
\end{tikzcd}$$

The bottom horizontal arrow is an isomorphism, since (Z', N') was obtained by pullback from $(Z, \widetilde{N})$. In order to show that the middle horizontal arrow is a surjection, it is enough to check that du/u is in its image, but both $d\sigma/\sigma$ and $d\tau/\tau$ map to du/u. We also see similarly that the kernel of the middle horizontal arrow is generated by $d\tau/\tau - d\sigma/\sigma$. The exactness of (3B.3) follows. □

Corollary 3.8. *We have an isomorphism*

$$\tilde{\tilde{C}}^{\bullet}_Y \Big/ \Big(\frac{d\tau}{\tau} \wedge \tilde{\tilde{C}}^{\bullet-1}_Y + \frac{d\sigma}{\sigma} \wedge \tilde{\tilde{C}}^{\bullet-1}_Y \Big) \xrightarrow{\sim} C^{\bullet}_Y.$$

Proof. This follows from the exact sequence (3B.1) and Lemma 3.7. □

Lemma 3.9. *The sections $d\tau/\tau$, $d\sigma/\sigma \in W_n\tilde{\tilde{\omega}}^1_Y$ are global sections, independent of the choice of admissible lifting. The same holds for $du/u \in W_n\tilde{\omega}^1_Y$.*

Proof. We will explain the proof only for $d\tau/\tau$ since the same proof also works for $d\sigma/\sigma$ and du/u. We use basically the same argument as for part 3 of Lemma 3.4 of [Mokrane 1993]. We consider two admissible liftings of $(Y, \widetilde{M})$, $(Z_1, \widetilde{N}_1)$ and

$(Z_2, \tilde{N}_2)$, and we let $(Z_{12}, \tilde{N}_{12})$ be defined as in Lemma 3.5. It is enough to show that locally on Y

$$\frac{d\tau}{\tau} \in \omega^1_{(Z_1,\tilde{N}_1)/(W_n,\text{triv})} \otimes_{\mathbb{O}_{Z_1}} \mathbb{O}_{D_1}$$

and

$$\frac{d\tau'}{\tau'} \in \omega^1_{(Z_2,\tilde{N}_2)/(W_n,\text{triv})} \otimes_{\mathbb{O}_{Z_2}} \mathbb{O}_{D_2}$$

have the same image in $\mathcal{H}^1(\omega^\bullet_{(Z_{12},\tilde{N}_{12})/(W_n,\text{triv})} \otimes_{\mathbb{O}_{Z_{12}}} \mathbb{O}_{D_{12}})$.

Note that $d\tau/\tau \in \tilde{N}_1$ and $d\tau'/d\tau' \in \tilde{N}_2$ have the same image in $\tilde{M}$. This is because locally on Y we have commutative diagrams

$$\begin{tikzcd} (Y, \tilde{M}) \arrow[r] \arrow[d] & (Z_i, \tilde{N}_i) \arrow[d] \\ (k, \mathbb{N}^2) \arrow[r] & (W_n[\tau,\sigma], \mathbb{N}^2) \end{tikzcd}$$

for $i = 1, 2$, so both $d\tau/\tau$ and $d\tau'/\tau'$ map to the image of $(1, 0) \in \mathbb{N}^2$ in $\tilde{M}$. By the construction of $(Z_{12}, \tilde{N}_{12})$ (see the proof of Proposition 4.10 of [Kato 1989]), we know that $d\tau/\tau - d\tau'/\tau' = m \in \tilde{N}_{12}$. Moreover, if $\alpha_{12} : N_{12} \to \mathbb{O}_{Z_{12}}$ is the map defining the log structure of Z_{12} then m maps to $0 \in \tilde{M}$, so $v = \alpha_{12}(m) \in \mathbb{O}^\times_{Z_{12}}$ maps to $1 \in \mathbb{O}_Y$. Therefore,

$$\frac{d\tau}{\tau} - \frac{d\tau'}{\tau'} = \frac{dv}{v}$$

for some $v \in \mathbb{O}_{D_{12}}$ for which $W_n\langle v - 1\rangle \subseteq \mathbb{O}_{D_{12}}$. But then we see that $dv/v \in d(W_n\langle v - 1\rangle)$, using the fact that the power series expansion of $\log(v)$ around 1 belongs to $W_n\langle v-1\rangle$. Therefore, $d\tau/\tau - d\tau'/\tau'$ is exact and the lemma follows. □

As in the classical case [Illusie and Raynaud 1983; Hyodo and Kato 1994], we can define operators $F : W_{n+1}\tilde{\tilde{\omega}}^q \to W_n\tilde{\tilde{\omega}}^q$, $V : W_n\tilde{\tilde{\omega}}^q \to W_{n+1}\tilde{\tilde{\omega}}^q$ and the differential $d : W_n\tilde{\tilde{\omega}}^q \to W_n\tilde{\tilde{\omega}}^{q+1}$, which satisfy

$$d^2 = 0, \quad FV = VF = p, \quad dF = pFd, \quad Vd = pdV \quad \text{and} \quad FdV = d.$$

Indeed, fix local liftings $(Z_n, \tilde{N}_n)$ of $(Y, \tilde{M}) \to (\operatorname{Spec} W_n[\tau,\sigma], \mathbb{N}^2)$ and denote the crystalline complex $\tilde{\tilde{C}}^\bullet_{Z_n}$ by $\tilde{\tilde{C}}^\bullet_n$. We can see that $\tilde{\tilde{C}}^\bullet_n$ is flat over W_n in the same way as in Lemma 2.22 of [Hyodo and Kato 1994] (using an admissible lifting), and we have

$$\tilde{\tilde{C}}^\bullet_n \otimes_{\mathbb{Z}/p^n\mathbb{Z}} \mathbb{Z}/p^m\mathbb{Z} \xrightarrow{\sim} \tilde{\tilde{C}}^\bullet_m$$

for $m \le n$. We let $F : W_{n+1}\tilde{\tilde{\omega}}^\bullet \to W_n\tilde{\tilde{\omega}}^\bullet$ be the map induced by $\tilde{\tilde{C}}^\bullet_{n+1} \to \tilde{\tilde{C}}^\bullet_n$ and $V : W_n\tilde{\tilde{\omega}}^\bullet \to W_{n+1}\tilde{\tilde{\omega}}^\bullet$ be the map induced by $p : \tilde{\tilde{C}}^\bullet_n \to \tilde{\tilde{C}}^\bullet_{n+1}$. We define d to be the connecting homomorphism in the exact sequence of cohomology sheaves associated to the exact sequence of crystalline complexes

$$0 \to \tilde{\tilde{C}}^{\bullet}_{n} \xrightarrow{p^n} \tilde{\tilde{C}}^{\bullet}_{2n} \to \tilde{\tilde{C}}^{\bullet}_{n} \to 0.$$

The same operators can be defined for $W_{\bullet}\tilde{\omega}^{\bullet}_{Y}$ and $W_{\bullet}\omega^{\bullet}_{Y}$.

Lemma 3.10. *Let $n = 1$. Locally, fix an admissible lifting $(Z, \widetilde{N})$ as above. Let* Fr *be the relative Frobenius of Y/k. We have Cartier isomorphisms*

$$\begin{aligned}
C^{-1} &: \omega^q_Y \xrightarrow{\sim} \mathcal{H}^q(\mathrm{Fr}_* \, \omega^{\bullet}_Y),\\
\widetilde{C}^{-1} &: \omega^q_{(Z',N')/(k,\mathrm{triv})} \otimes_{k[u]} k \xrightarrow{\sim} \mathcal{H}^q(\mathrm{Fr}_*(\omega^{\bullet}_{(Z',N')/k,\mathrm{triv}} \otimes_{k[u]} k)),\\
\tilde{\tilde{C}}^{-1} &: \omega^q_{(Z,\widetilde{N})/(k,\mathrm{triv})} \otimes_{k[t,s]} k \xrightarrow{\sim} \mathcal{H}^q(\mathrm{Fr}_*(\omega^{\bullet}_{(Z,\widetilde{N})/k,\mathrm{triv}} \otimes_{k[t,s]} k)).
\end{aligned}$$

Proof. Note that $(Y, M)/(\mathrm{Spec}\, k, \mathbb{N})$ is log smooth of Cartier type. The Cartier isomorphism for $W_1\omega^q_Y$ is then defined in Section 2.12 of [Hyodo and Kato 1994]. Similarly, $(Z', N')/(\mathrm{Spec}\, k, \mathrm{triv})$ and $(Z, \widetilde{N})/(\mathrm{Spec}\, k, \mathrm{triv})$ are log smooth and of Cartier type. Thus, the morphisms $\widetilde{C}^{-1}$ and $\tilde{\tilde{C}}^{-1}$ for $\widetilde{C}^q_Y$ and $\tilde{\tilde{C}}^q_Y$ are induced from the Cartier isomorphisms for these schemes.

Since we are working locally on Y, we may assume that $Y = Y_1 \times_k Y_2$ and that the lifting $Z = Z_1 \times Z_2$, where Z_1, Z_2 are smooth over k and Y_i is a reduced normal crossings divisor in Z_i. Let $\mathcal{I}_i$ be the ideal defining $Y_i \times_k Z_{3-i}$ in Z for $i = 1, 2$. Define $\omega^{\bullet}_{1,2} := \omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1 \oplus \omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_2$ To check that $\tilde{\tilde{C}}^{-1}$ is an isomorphism, we use the following commutative diagram of exact sequences:

$$\begin{tikzcd}
\omega^q_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1\mathcal{I}_2 \arrow[r] \arrow[d] & \omega^q_{1,2} \arrow[r] \arrow[d] & \omega^q_{(Z,\widetilde{N})/k} \arrow[r] \arrow[d] & \tilde{\tilde{C}}^q_Y \arrow[r] \arrow[d] & 0 \\
\mathcal{H}^q(\mathrm{Fr}_* \, \omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1\mathcal{I}_2) \arrow[r] & \omega^q_{1,2} \arrow[r] & \mathcal{H}^q(\mathrm{Fr}_* \, \omega^{\bullet}_{(Z,\widetilde{N})/k}) \arrow[r] & \mathcal{H}^q(F_* \tilde{\tilde{C}}^{\bullet}_Y) \arrow[r] & 0
\end{tikzcd}$$

The complex $\omega^{\bullet}_{(Z,\widetilde{M})/k,\mathrm{triv}}$ is the same as $\Omega^{\bullet}_{Z_1/k}(\log Y_1) \otimes_k \Omega^{\bullet}_{Z_2/k}(\log Y_2)$, so it does satisfy a Cartier isomorphism, by [Deligne and Illusie 1987, 4.2.1.1]. Similarly, the complexes on its left are (sums of) products of complexes of the form $\Omega^{\bullet}_{Z_i/k}(\pm \log Y_i)$ for $i = 1, 2$, which also satisfy a Cartier isomorphism, by [Deligne and Illusie 1987, 4.2.1.3]. Therefore, the first three vertical arrows are isomorphisms. Once we know the exactness of the top and bottom sequence we can also deduce that the rightmost vertical arrow is an isomorphism. The exactness of the top row follows from the definition of $\tilde{\tilde{C}}^q_Y$.

The exactness of the bottom row follows from the cohomology long exact sequence associated to the short exact sequences obtained from the top row combined with the Cartier isomorphisms for the first three vertical arrows, which tell us that the coboundary morphisms of these long exact sequences are all 0. Indeed, if we let $\bar{\omega}^{\bullet}_{(Z,\widetilde{N})}$ be the complex obtained by completing the inclusion of complexes

$$\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1\mathcal{I}_2 \to \omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1 \oplus \omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_2$$

to a distinguished triangle, then we get a long exact sequence

$$\cdots \to \mathcal{H}^q(\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1\mathcal{I}_2) \to \mathcal{H}^q(\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1) \oplus \mathcal{H}^q(\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_2) \\ \to \mathcal{H}^q(\bar{\omega}^{\bullet}_{(Z,\widetilde{N})}) \to \cdots .$$

From the Cartier isomorphisms for $\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1\mathcal{I}_2$ and $\omega^{\bullet}_{1,2}$, we deduce that

$$\mathcal{H}^q(\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1\mathcal{I}_2) \hookrightarrow \mathcal{H}^q(\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_1) \oplus \mathcal{H}^q(\omega^{\bullet}_{(Z,\widetilde{N})/k} \otimes \mathcal{I}_2),$$

so the coboundaries of the long exact sequence are all 0. By continuing this argument, we deduce the exactness of the entire bottom row, and this proves that $\tilde{\tilde{C}}^{-1}$ is an isomorphism.

Now we prove that $\widetilde{C}^{-1}$ is an isomorphism. We will show that $\widetilde{C}^{-1}$ is an isomorphism in degree q as well. From the short exact sequence (3B.1), we get the following commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & C_Y^{q-1} \arrow[r] \arrow[d] & \widetilde{C}_Y^q \arrow[r] \arrow[d] & C_Y^q \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathcal{H}^{q-1}(\mathrm{Fr}_* C_Y^{\bullet}) \arrow[r] & \mathcal{H}^q(\mathrm{Fr}_* \widetilde{C}_Y^{\bullet}) \arrow[r] & \mathcal{H}^q(\mathrm{Fr}_* C_Y^{\bullet}) \arrow[r] & 0
\end{tikzcd}$$

To see that the bottom row is exact, we have to check that in the long exact cohomology sequence associated to the top row the coboundaries are all 0, which is equivalent to showing surjectivity of $\mathcal{H}^q(\mathrm{Fr}_* \widetilde{C}_Y^{\bullet}) \to \mathcal{H}^q(\mathrm{Fr}_* C_Y^{\bullet})$. However, by the top row and the Cartier isomorphism C^{-1}, the composite

$$\widetilde{C}_Y^q \to C_Y^q \to \mathcal{H}^q(\mathrm{Fr}_* C_Y^{\bullet})$$

is surjective, so the desired map is surjective as well. Now we have a map of short exact sequences, where the left and right vertical maps are isomorphisms, so the middle one must be as well. $\square$

We can define canonical projections $\pi : W_{n+1}\tilde{\tilde{\omega}}^{\bullet}_Y \to W_n\tilde{\tilde{\omega}}^{\bullet}_Y$ using the Cartier isomorphisms. The construction works in the same way for $W_n\tilde{\omega}^{\bullet}_Y$. The definition of π for $W_n\omega^{\bullet}_Y$ can be found in Section 1 of [Hyodo 1991] in the semistable case and in Section 4 of [Hyodo and Kato 1994] in general. The constructions in [Hyodo 1991] and in [Hyodo and Kato 1994] are the same, although they are formulated slightly differently. Our construction follows that in Section 1 of [Hyodo 1991], by first defining a map $\mathfrak{p} : W_n\tilde{\tilde{\omega}}_Y^\bullet \to W_{n+1}\tilde{\tilde{\omega}}_Y^\bullet$ and then showing that $\mathfrak{p}$ is injective and its image coincides with the image of multiplication by p on $W_{n+1}\tilde{\tilde{\omega}}_Y^\bullet$. The projection π will then be the unique map which makes the following diagram commute:

$$\begin{tikzcd}
W_n\tilde{\tilde{\omega}}_Y^\bullet \arrow[d, "\mathfrak{p}"] & W_{n+1}\tilde{\tilde{\omega}}_Y^\bullet \arrow[l, "\pi"'] \arrow[dl, "p"] \\
W_{n+1}\tilde{\tilde{\omega}}_Y^\bullet &
\end{tikzcd}$$

The map $\mathfrak{p} : W_n\tilde{\tilde{\omega}}_Y^i \to W_{n+1}\tilde{\tilde{\omega}}_Y^i$ is induced from $p^{-i+1}\,\mathrm{Fr}^* : \tilde{\tilde{C}}{}_Y^i \to \tilde{\tilde{C}}{}_Y^i$, where $\mathrm{Fr} : (Z, \widetilde{N}) \to (Z, \widetilde{N})$ is a lifting of the Frobenius endomorphism of $(Z, \widetilde{N}) \times_W k$ such that $\mathrm{Fr}^*(W[\tau, \sigma]) \subset W[\tau, \sigma]$. The injectivity of $\mathfrak{p}$ and the fact that its image coincides with that of multiplication by p are deduced as in Section 2 of [Hyodo 1991] (or as in Lemma 6.8 of [Nakkajima 2005]) from the Cartier isomorphism and from the fact that $\tilde{\tilde{C}}{}_Y^\bullet$ is W-torsion-free (when we take $\tilde{\tilde{C}}{}_Y^\bullet$ to be the crystalline complex associated to an embedding system for $(Y, \widetilde{M})$ over W).

Now we will consider a different interpretation of the monodromy operator N. Taking the cohomology sheaves of the short exact sequence

$$0 \to C_Y^\bullet[-1] \to \widetilde{C}_Y^\bullet \to C_Y^\bullet \to 0,$$

we get a long exact sequence of sheaves on Y

$$\cdots \to W_n\omega_Y^{q-1} \to W_n\tilde{\omega}_Y^q \to W_n\omega_Y^q \to \cdots$$

whose coboundaries are actually all 0. This can be checked as in Lemma 1.4.3 of [Hyodo 1991], since it suffices to see that the induced map on cocycles $Z^q(\widetilde{C}_Y) \to Z^q(C_Y)$ modulo p^n is surjective, and we can use the Cartier isomorphisms in Lemma 3.10 to give an explicit formula for cocycles modulo p^n. So we have a short exact sequence of sheaves on Y

$$0 \to W_n\omega_Y^{q-1} \to W_n\tilde{\omega}_Y^q \to W_n\omega_Y^q \to 0, \tag{3B.4}$$

which is compatible with operators π, F, V and d. We have a morphism of distinguished triangles in the derived category $D(Y_{\text{ét}}, W)$ of sheaves of W-modules on Y:

$$\begin{tikzcd}
C_Y^\bullet[-1] \arrow[r] \arrow[d] & \widetilde{C}_Y^\bullet \arrow[r] \arrow[d] & C_Y^\bullet \arrow[r] \arrow[d] & C_Y^\bullet \arrow[d] \\
W_n\omega_Y^\bullet[-1] \arrow[r] & W_n\tilde{\omega}_Y^\bullet \arrow[r] & W_n\omega_Y^\bullet \arrow[r] & W_n\omega_Y^\bullet
\end{tikzcd}$$

The left and right vertical maps are defined in the proof of Theorem 4.19 of [Hyodo and Kato 1994], and the middle one can be defined in exactly the same way. Note that the definition of the maps in Theorem 4.19 has a gap which is corrected in Lemma 7.18 of [Nakkajima 2005], namely, checking that they commute with the transition morphisms $\pi : W_{n+1}\omega_Y^\bullet \to W_n\omega_Y^\bullet$. The fact that the middle map commutes with the transition morphisms $\pi : W_{n+1}\tilde{\omega}_Y^\bullet \to W_n\tilde{\omega}_Y^\bullet$ can be checked in the same way as in Lemma 7.18 of [Nakkajima 2005], using the corresponding Cartier isomorphism to check that the complexes $W_n\tilde{\omega}_Y^\bullet$ give rise to formal de Rham–Witt complexes as in Definition 6.1 of [ibid.] and thus applying Corollary 6.28(8). We also need to check that $\varprojlim W_n\tilde{\omega}_Y^1$ is torsion-free, but we can use the fact that this is known for $\varprojlim W_n\omega_Y^1$ and the exact sequence (3B.4). The first and third vertical maps are quasi-isomorphisms by Theorem 4.19 of [Hyodo and Kato 1994], so we get an isomorphism of distinguished triangles. Thus, the exact sequence (3B.4) induces the monodromy operator N on cohomology.

Assume that Y has an admissible lifting $\underline{Z}$ over $(W[t,s],\mathbb{N}^2)$, and set $Z = \underline{Z}\otimes_W k$. We consider a few more variations on the de Rham–Witt complex, which we will only define locally on Z. Let $W_n\Omega_Z^\bullet$ be the de Rham–Witt complex of Z. Let

$$Y^1 = \operatorname{Spec} k[X_1,\dots,X_n,Y_1,\dots,Y_n,Z_1,\dots,Z_m]/X_1\cdots X_r$$

and

$$Y^2 = \operatorname{Spec} k[X_1,\dots,X_n,Y_1,\dots,Y_n,Z_1,\dots,Z_m]/Y_1\cdots Y_s.$$

Each Y^i is a normal crossings divisor in $Z_{r,s,m}\times_W k$. Let $\mathcal{D}_n^i$ be the structure sheaf of the divided power envelope of Y^i in $Z_{r,s,m}$ and $\mathcal{I}\mathcal{D}_n^i = \ker(\mathcal{D}_n^i \to \mathcal{O}_{Y^i})$. For $i=1,2$, let $W_n\Omega_Z^\bullet(-\log Y^i)$ be the (pullback to Z) of the "compact support" version of the de Rham–Witt complex of $Z_{r,s,m}$ with respect to Y^i. This complex was introduced by Hyodo [1991, Section 1] and is defined by

$$W_n\Omega^q_{Z_{r,s,m}}(-\log Y^i) = \mathcal{H}^q(\Omega^\bullet_{Z/W_n}(\log Y^i)\otimes_{\mathcal{O}_{Z_{r,s,m}}}\mathcal{I}\mathcal{D}_n^i).$$

Let $W_n\Omega_Z^\bullet(-\log Y^1-\log Y^2)$ be the pullback from $Z_{r,s}$ to Z of the complex defined by

$$W_n\Omega^q_{Z_{r,s,m}}(-\log Y^1-\log Y^2) := \mathcal{H}^q(\omega^\bullet_{Z_{r,s,m},\mathbb{N}^r\oplus\mathbb{N}^s/W_n}\otimes_{\mathcal{O}_Z}\mathcal{I}\mathcal{D}_1\mathcal{I}\mathcal{D}_2).$$

This third complex is meant to approximate a product of complexes of the form $W_n\Omega_Z(-\log Y)$. When $n=1$, consider $Z^1=\operatorname{Spec} k[X_1,\dots,X_n,t]/(X_1\cdots X_r-t)$, $Z^2=\operatorname{Spec} k[Y_1,\dots,Y_n,u]/(Y_1\cdots Y_s-u)$ and $Z^3=\operatorname{Spec} k[Z_1,\dots,Z_m]$. Then

$$\begin{aligned} &W_1\Omega^\bullet_{Z_{r,s,m}}(-\log Y^1-\log Y^2)\\ &\qquad\simeq \Omega^\bullet_{Z^1/k}(-\log Y^1)\otimes_k\Omega^\bullet_{Z^2/k}(-\log Y^2)\otimes_k\Omega^\bullet_{Z^3/k}. \end{aligned}\tag{3B.5}$$

All these also are endowed with operators F, V, differential d and projection π, and they also satisfy a Cartier isomorphism.

Let $\mathbb{R}$ be Raynaud's ring, introduced in [Illusie and Raynaud 1983], i.e., the graded W-algebra generated by F, V in degree 0 and d in degree 1, subject to the usual relations

$$d^2 = 0, \quad FV = VF = p, \quad dF = pFd, \quad Vd = pdV \quad \text{and} \quad FdV = d.$$

Let $\mathbb{R}_n$ be the right $\mathbb{R}$-module $\mathbb{R}/(V^n\mathbb{R} + dV^n\mathbb{R})$.

Lemma 3.11. *Let $W_n\Omega^\bullet$ be one of the complexes $W_n\Omega^\bullet_Z$, $W_n\Omega^\bullet_Z(-\log Y^i)$ for $i = 1, 2$ or $W_n\Omega^\bullet_Z(-\log Y^1 - \log Y^2)$. Let*

$$W\Omega^\bullet = \varprojlim W_n\Omega^\bullet.$$

Then $W\Omega^\bullet \otimes^L_{\mathbb{R}} \mathbb{R}_n = W_n\Omega^\bullet$.

Proof. For $n = 1$, and $W_n\Omega^\bullet_Z$ and $W_n\Omega^\bullet_Z(-\log Y^i)$, we have Cartier isomorphisms

$$W_1\Omega^i \xrightarrow{\sim} \mathcal{H}^i(F_* W_1\Omega^\bullet),$$

by [Deligne and Illusie 1987, Result 4.2.1.3]. For $W_n\Omega^\bullet_Z(-\log Y^1 - \log Y^2)$ the Cartier isomorphism follows from the product formula (3B.5) and from the Cartier isomorphisms above. Let $\mathfrak{X}_n = \underline{Z} \times_W W_n$. By abuse of notation, we write $\Omega^\bullet_{Z_n}$ for the complex of sheaves of W_n-modules such that

$$W_n\Omega^i = \mathcal{H}^i(\Omega^\bullet_{\mathfrak{X}_n}).$$

In fact, we have complexes $\Omega^\bullet_{\underline{Z}}$, $\Omega^\bullet_{\underline{Z}}(-\log Y^i)$ or $\Omega^\bullet_{\underline{Z}}(-\log Y^1 - \log Y^2)$ which give the corresponding complexes $\Omega^\bullet_{\mathfrak{X}_n}$, $\Omega^\bullet_{\mathfrak{X}_n}(-\log Y^i)$ or $\Omega^\bullet_{\mathfrak{X}_n}(-\log Y^1 - \log Y^2)$ when reduced modulo p^n. We also denote any of the initial complexes over W as $\Omega^\bullet_{\underline{Z}}$. Then there is an explicit description of cocycles modulo p^n given by

$$d^{-1}(p^n\Omega^{i+1}_{\underline{Z}}) = \sum_{k=0}^{n} p^k f^{n-k}\Omega^i_{\underline{Z}} + \sum_{k=0}^{n-1} f^k d\Omega^{i-1}_{\underline{Z}},$$

where $f : \Omega^i_{\underline{Z}} \to \Omega^i_{\underline{Z}}$ is defined by $f = \mathrm{Fr}/p^i$. This is the same as Formula A from Editorial Comment 11 in [Hyodo 1991] and is proven in the same way as in that paper and in the same way as in the classical crystalline cohomology case (see [Illusie 1979, 0.2.3.13]).

As in the case of $W_n\omega_Y$, $W_\bullet\Omega^\bullet$ (and $W\Omega^\bullet$) is endowed with a differential d, operators F, V satisfying the usual relations and a canonical projection $\pi_n : W_{n+1}\Omega^\bullet \to W_n\Omega^\bullet$ such that $\mathfrak{p} \circ \pi_n$ coincides with multiplication by p on $W_{n+1}\Omega^\bullet$.

We claim that the lemma follows from the Cartier isomorphism, from the description of cocycles modulo p^n in $\Omega^\bullet_{\underline{Z}}$ and from the formal properties of $W_n\Omega^\bullet$. The proof is the same as for Lemma 1.3.3 of [Mokrane 1993]. We outline the argument in order to show that it applies to our case as well. To prove the desired result, we use the flat resolution of $\mathbb{R}_n$ as an $\mathbb{R}$-module given by

$$0 \to \mathbb{R} \xrightarrow{(F^n,-F^n d)} \mathbb{R} \oplus \mathbb{R} \xrightarrow{dV^n+V^n} \mathbb{R} \to \mathbb{R}_n \to 0,$$

and it suffices by Corollary 1.3.3 of [Illusie and Raynaud 1983] to prove that the sequence

$$0 \to W\Omega^{i-1} \xrightarrow{(F^n,-F^n d)} W\Omega^{i-1} \oplus W\Omega^i \xrightarrow{dV^n+V^n} W\Omega^i \to W_n\Omega^i \to 0$$

is exact. The last map is the canonical projection $\pi : W\Omega^i \to W_n\Omega^i$.

Exactness at the first term follows from the fact that multiplication by p (and hence also F) is injective on $W\Omega^\bullet$. Indeed, multiplication by p on $W_n\Omega^\bullet$ factors as $\mathfrak{p} \circ \pi_n$ and $\mathfrak{p}$ is injective by definition, so if $p(x_n) = 0$ for all n then $\pi_n(x_n) = x_{n-1} = 0$ for all n, so $x = (x_n) = 0$.

Exactness at the last term is the statement that π is surjective, which follows by construction, since $p = \mathfrak{p} \circ \pi$, $\mathfrak{p}$ is injective and the image of $\mathfrak{p} : W_n\Omega^\bullet \to W_{n+1}\Omega^\bullet$ coincides with the image of multiplication by p.

Now we check that $\ker \pi = dV^n W\Omega^\bullet + V^n W\Omega^\bullet$. Recall that $\pi_n : W_{n+1} \to W_n$ is the canonical projection. It is enough to show that $\ker \pi_n = dV^n W_1\Omega^\bullet + V^n W_1\Omega^\bullet$. First, if $x = V^n a + dV^n b \in W_{n+1}\Omega$, it suffices to check that $px = 0$ and indeed $px = FV^{n+1}a + dFV^{n+1}b = 0$. Now, let $[x]_{n+1} \in \ker \pi_n$, where x is an element of $\Omega^\bullet_{\underline{Z}}$ modulo p^{n+1}. Then $[px]_{n+1} = p[x]_{n+1} = 0$, so it must be the case that $px = p^{n+1}a + db$. We get $db = 0 \mod p$, so by the description of cocycles mod p we have $b = pb' + Fb'' + db''$, so that $db = pdb' + pFdb''$. Thus,

$$\begin{aligned}
[x]_{n+1} &= [p^n a]_{n+1} + [db']_{n+1} + [Fdb'']_{n+1} \\
&= V^n[a]_{n+1} + d[p^n Fb'']_{n+1} = V^n[a] + dV^n[Fb''].
\end{aligned}$$

Now we check exactness at the second term. First, note that the sequence

$$W_{2n}\Omega^{q-1} \xrightarrow{F^n} W_n\Omega^{q-1} \xrightarrow{d} W_n\Omega^q$$

is exact, which is proved in the same way as Lemma 1.3.4 of [Mokrane 1993], by taking the long exact sequence of cohomology sheaves of the short exact sequence

$$0 \to \Omega^\bullet_{\underline{Z}}/p^n\Omega^\bullet_{\underline{Z}} \xrightarrow{p^n} \Omega^\bullet_{\underline{Z}}/p^{2n}\Omega^\bullet_{\underline{Z}} \to \Omega^\bullet_{\underline{Z}}/p^n\Omega^\bullet_{\underline{Z}} \to 0.$$

We note that the proof of the analogous statement in the classical case in [Illusie 1979, I(3.21)] is wrong and corrected in [Illusie and Raynaud 1983, II(1.3)]. Nakkajima [2005, 6.28(6)] proves this statement for formal de Rham–Witt complexes, using the same argument as Lemma 1.3.4 of [Mokrane 1993].

We now claim that the projection

$$W\Omega^\bullet/p^n W\Omega^\bullet \to W_n\Omega^\bullet$$

is a quasi-isomorphism. This implies that

$$d^{-1}(p^n W\Omega^q) = F^n W\Omega^{q-1},$$

so if $dV^n x + V^n y = 0$, then $dx + p^n y = 0$, which in turn implies $x = F^n z$ and $y = -F^n dz$ for some $z \in W\Omega^{q-1}$. This checks exactness at the second term. Moreover, the fact that

$$W\Omega^\bullet/p^n W\Omega^\bullet \to W_n\Omega^\bullet$$

is a quasi-isomorphism follows in the same way as Corollary 3.17 of [Illusie 1979], boiling down to the Cartier isomorphism and to the description of $\ker\pi$ as $dV^n + V^n$. □

Remark 3.12. One can use the Cartier isomorphisms to check Properties 6.0.1–6.0.5 of [Nakkajima 2005] for $\Omega^\bullet_{\underline{Z}}$, $\Omega^\bullet_{\underline{Z}}(-\log Y^i)$ and $\Omega^\bullet_{\underline{Z}}(-\log Y^1 - \log Y^2)$, thus proving the analogue of Proposition 6.27 there for all three complexes. Then Theorem 6.24 of [Nakkajima 2005] also implies Lemma 3.11.

3C. ***The weight filtration.*** The goal of this section is to define a double filtration $P_{k,l}$ on $W\tilde{\tilde{\omega}}^\bullet_Y$, which will be an analogue of the weight filtration defined by Mokrane on $W_n\tilde{\omega}^\bullet_Y$ in the semistable case (see [1993, Section 3]).

Let $(Z, \tilde{N})$ be an admissible lifting of $(Y, \tilde{M})$ over $(W[\tau,\sigma], \mathbb{N}^2)$. We know that such liftings exist étale locally. Let $\mathfrak{X}_n = Z \times_W W_n$. Let $\tilde{N}_1$ be the log structure on Z (or $\mathfrak{X}_n$) obtained by pulling back the log structure on $Z_{r,s,m}$ associated to

$$\begin{aligned} \mathbb{N}^r &\to W[X_1, \dots, X_n, Y_1, \dots, Y_n, Z_1, \dots, Z_m], \\ (0, \dots, 0, 1, 0, \dots, 0) &\mapsto X_i, \end{aligned}$$

where 1 is in the i-th position. Define $\tilde{N}_2$ analogously. The pullback of $\tilde{N}_i$ to Y is the same as $\tilde{M}_i$. For $i = 1, 2$, we have maps of sheaves of monoids $\tilde{N}_i \to \tilde{N}$.

We define the following filtration on $\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}$:

$$\begin{aligned} &P_{i,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \\ &\qquad := \mathrm{Im}\big(\omega^i_{(\mathfrak{X}_n,\tilde{N}_1)/(W_n,\mathrm{triv})} \otimes \omega^j_{(\mathfrak{X}_n,\tilde{N}_2)/(W_n,\mathrm{triv})} \otimes \Omega^{q-i-j}_{\mathfrak{X}_n/k} \to \omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}\big) \end{aligned}$$

for $i, j \geq 0$ and $i + j \leq q$. This filtration respects the differential and induces a filtration $P_{i,j}\tilde{\tilde{C}}^\bullet_Y$ on $\tilde{\tilde{C}}^\bullet_Y$ (which can be thought of as a quotient of $\omega^\bullet_{(\mathfrak{X}_n,\tilde{M})/(W_n,\mathrm{triv})}$, as in the proof of Lemma 3.10). Note that if we let

$$P_k\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} = \mathrm{Im}\big(\omega^k_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \otimes \Omega^{q-k}_{\mathfrak{X}_n/k} \to \omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}\big),$$

then P_k is the weight filtration defined in [Mokrane 1993, 1.1.1], and

$$P_{i,j}\omega^{\bullet}_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \subset P_{i+j}\omega^{\bullet}_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}.$$

For $i = 1, \dots, r$, let $D_{1,i}$ be the pullback to Z of the divisor of $Z_{r,s,m}$ obtained by setting $X_i = 0$. Similarly, for $i = 1, \dots, s$, let $D_{2,i}$ be the pullback to Z of the divisor of $Z_{r,s,m}$ obtained by setting $Y_i = 0$. For $i, j \geq 0$ let $D^{(i,j)}$ be the disjoint union of

$$D_{1,k_1} \times_Z \cdots \times_Z D_{1,k_i} \times_Z D_{2,l_1} \times_Z \cdots \times_Z D_{2,l_j},$$

over all $k_1, \dots, k_i \in \{1, \dots, r\}$ and $l_1, \dots, l_j \in \{1, \dots, s\}$. And let $\tau_{i,j} : D^{(i,j)} \to Z$ be the obvious morphism, with $\mathfrak{D}_n^{(i,j)}$, $\tau_{i,j}$ the pullbacks to $\mathfrak{X}_n$. Let

$$\begin{aligned}\mathrm{Gr}_{i,j}\,\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \\ := P_{i,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}\big/\big(P_{i-1,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} + P_{i,j-1}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}\big).\end{aligned}$$

For $i, j \geq 1$, we will define a morphism of sheaves

$$\mathrm{Res} : \mathrm{Gr}_{i,j}\,\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \to (\tau_{i,j})_*\Omega^{q-i-j}_{\mathfrak{D}_n^{(i,j)}/W_n},$$

which extends to a morphism of complexes. If

$$\omega = \alpha \wedge \frac{dX_{k_1}}{X_{k_1}} \wedge \cdots \wedge \frac{dX_{k_i}}{X_{k_i}} \wedge \frac{dY_{l_1}}{Y_{l_1}} \wedge \cdots \wedge \frac{dY_{k_j}}{Y_{k_j}}$$

is a local section of $P_{i,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}$ with $k_1 < \cdots < k_i$ and $l_1 < \cdots < l_j$, then

$$\mathrm{Res}(\omega) := \alpha|_{D_{1,k_1}\times_Z\cdots\times_Z D_{1,k_i}\times_Z D_{2,l_1}\times_Z\cdots\times_Z D_{2,l_j}}.$$

This factors through $P_{i-1,j} + P_{i,j-1}$ and extends to a global map of sheaves.

Alternatively, we can follow the construction in Section 3 of Chapter II of [Deligne 1970]. Let $\mathfrak{D}_n^k$ be the disjoint union of intersections of k divisors D_{j,k_i} with $j = 1, 2$ and $k_i \in \{1, \dots, n\}$. These intersections are in one-to-one correspondence with images of injections

$$f : \{1, \dots, k\} \to \{1, \dots, n\} \cup \{1, \dots, n\},$$

and so we denote one of these k intersections by $\mathfrak{D}_n^f$ (even though it only really depends on $\mathrm{Im}\, f$). We have

$$\mathfrak{D}_n^k = \bigsqcup_{\substack{i+j=k\\ i,j\geq 0}} \mathfrak{D}_n^{i,j} = \bigsqcup_{\mathrm{Im}\, f} \mathfrak{D}_n^f.$$

Let $\tau_f : \mathfrak{D}_n^f \to \mathfrak{X}_n$ be the closed immersion. In [Deligne 1970, 3.5.2], a morphism

$$\rho_1 : (\tau_f)_* \Omega^{q-k}_{\mathfrak{D}_n^f} \to P_k \omega^q_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} / P_{k-1}$$

(and then a morphism ρ_2, which depends on an ordering of $\{1, \ldots, n\} \cup \{1, \ldots, n\}$) is associated to each such injection, and the sum of ρ_2 over all injections f determines an isomorphism

$$\rho : (\tau_k)_* \Omega^\bullet_{\mathfrak{D}_n^k / W_n}[-k] \xrightarrow{\sim} P_k \omega^q_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} / P_{k-1}$$

by Proposition 3.6 of Chapter II of [Deligne 1970].

We are only interested in injections $q_{i,j} : \{1, \ldots, i+j\} \to \{1, \ldots, n\} \cup \{1, \ldots, n\}$ with image of cardinality i in the first $\{1, \ldots, n\}$ term and cardinality j in the second $\{1, \ldots, n\}$ term. We let Res^{-1} be the sum of the morphisms ρ_2 over all injections $q_{i,j}$. When we have an injection of type $q_{i,j}$, the image of the morphism ρ_2 defined by Deligne falls in

$$P_{i,j} \omega^q_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} / (P_{i-1,j} + P_{i,j-1}) \subset P_{i+j} \omega^q_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} / P_{i+j-1}.$$

For $k \geq 1$, we have the direct sum decompositions

$$P_k \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} / P_{k-1} = \bigoplus_{\substack{i+j=k \\ i,j \geq 0}} \mathrm{Gr}_{i,j}\, \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})},$$

$$(\tau_k)_* \Omega^{q-k}_{\mathfrak{D}_n^{(k)} / W_n} = \bigoplus_{\substack{i+j=k \\ i,j \geq 0}} (\tau_{i,j})_* \Omega^{q-i-j}_{\mathfrak{D}_n^{(i,j)} / W_n}.$$

It is easy to check that the isomorphism ρ matches up the (i, j) terms in each decomposition. Putting this discussion together, we get the following:

Lemma 3.13. *For $i, j \geq 1$, the map*

$$\mathrm{Res}^{-1} : (\tau_{i,j})_* \Omega^{q-i-j}_{\mathfrak{D}_n^{(i,j)} / W_n} \to \mathrm{Gr}_{i,j}\, \omega^q_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})}$$

is an isomorphism.

We also have the following analogue of Lemma 1.2 of [Mokrane 1993].

Lemma 3.14. *We have an exact sequence of complexes*

$$\begin{aligned} 0 \to P_{i-1,j-1} \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} &\to P_{i-1,j} \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} \oplus P_{i,j-1} \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} \\ &\to P_{i,j} \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} \to \mathrm{Gr}_{i,j}\, \omega^\bullet_{(\mathfrak{X}_n, \widetilde{N})/(W_n, \mathrm{triv})} \to 0. \end{aligned}$$

The long exact cohomology sequence(s) associated to this have all coboundaries 0, *so we get the exact sequence*

$$
\begin{aligned}
0 \to \mathscr{H}^q(P_{i-1,j-1}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}) \\
\to \mathscr{H}^q(P_{i-1,j}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}) \oplus \mathscr{H}^q(P_{i,j-1}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}) \\
\to \mathscr{H}^q(P_{i,j}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}) \to \mathscr{H}^q(\Omega^\bullet_{\mathfrak{D}_n^{(i,j)}/W_n}[-i-j]) \to 0.
\end{aligned}
$$

Proof. The first assertion is clear. In order to show that the second sequence is exact, it suffices to show the following two statements about cocycles:

(1) $ZP_{i,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \twoheadrightarrow Z\Omega^{q-i-j}_{\mathfrak{D}_n^{(i,j)}/W_n}$.

(2) $ZP_{i-1,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \oplus ZP_{j,i-1}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \twoheadrightarrow Z\big(P_{i-1,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} + P_{i,j-1}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}\big)$.

The first statement is proved in the same way as the main step in Lemma 1.1.2 of [Mokrane 1993]. If α is a local section of $Z\Omega^{q-i-j}_{\mathfrak{D}_n^{(i,j)}/W_n}$, assume that α is supported on some

$$D_{1,k_1} \times_Z \cdots \times_Z D_{1,k_i} \times_Z D_{2,l_1} \times_Z \cdots \times_Z D_{2,l_j},$$

for some $k_1,\ldots,k_i, l_1,\ldots,l_j \in \{1,\ldots,n\}$. Let

$$\rho : \mathfrak{X}_n \to D_{1,k_1} \times_Z \cdots \times_Z D_{1,k_i} \times_Z D_{2,l_1} \times_Z \cdots \times_Z D_{2,l_j}$$

be the retraction associated to the immersion

$$D_{1,k_1} \times_Z \cdots \times_Z D_{1,k_i} \times_Z D_{2,l_1} \times_Z \cdots \times_Z D_{2,l_j} \to \mathfrak{X}_n.$$

Then $\rho^*\alpha$ lifts α to a section of $Z\Omega^{q-i-j}_{\mathfrak{X}_n/W_n}$ and the section

$$\omega_\alpha = \rho^*\alpha \wedge \frac{dX_{k_1}}{X_{k_1}} \wedge \cdots \wedge \frac{dX_{k_i}}{X_{k_i}} \wedge \frac{dY_{l_1}}{Y_{l_1}} \wedge \cdots \wedge \frac{dY_{l_j}}{Y_{l_j}} \in P_{i,j}\omega^q_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}$$

satisfies $d\omega = 0$ and $\mathrm{Res}(\omega) = \alpha$. From this, we know that the coboundaries of the long exact sequence associated to

$$
\begin{aligned}
0 \to P_{i-1,j}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} + P_{i,j-1}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \\
\to P_{i,j}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \to \mathrm{Gr}_{i,j}\,\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} \to 0
\end{aligned}
$$

are 0, so we also know that

$$\mathscr{H}^q\big(P_{i-1,j}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})} + P_{i,j-1}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})}\big) \hookrightarrow \mathscr{H}^q(P_{i,j}\omega^\bullet_{(\mathfrak{X}_n,\tilde{N})/(W_n,\mathrm{triv})})$$

for every $i, j \geq 1$.

For the second statement, we have to prove that if $\alpha \in P_{i-1,j}\omega^q_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}$ and $\beta \in P_{i,j-1}\omega^q_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}$ satisfy $d(\alpha+\beta)=0$, then we can find $\alpha' \in ZP_{i-1,j}\omega^q_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}$ and $\beta' \in ZP_{i,j-1}\omega^q_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}$ such that $\alpha'+\beta' = \alpha+\beta$. If $\alpha \in P_{i-1,j-1}\omega^q_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}$, then we are done, since we can just take $\alpha'=0$, $\beta'=\alpha+\beta$. The same holds for β. Otherwise, we have $d\alpha \in P_{i-1,j-1}$, so by the injectivity proved in statement (1) for $(i-1, j)$, we know that $d\alpha = d\alpha_1 + d\alpha_2$ for some $\alpha_1 \in P_{i-1,j-1}$ and $\alpha_2 \in P_{i-2,j}$. Thus, we've reduced our problem from $(i-1, j)$ to $(i-2, j)$. Proceeding by induction, we may assume that $i = 0$. In that case $d\alpha_{2i} \in P_{0,j-1}$. By (the same argument as in the proof of) Lemma 1.1.2 of [Mokrane 1993], we have an injection

$$\mathcal{H}^q(P_{0,j-1}\omega^\bullet_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}) \hookrightarrow \mathcal{H}^q(P_{0,j}\omega^\bullet_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}),$$

which implies $d\alpha_{2i} = d\alpha_{2i+1}$ for some $\alpha_{2i+1} \in P_{0,j-1}$. Then

$$\alpha' := \alpha - \sum_{i'=0}^{i} \alpha_{2i'+1} \in ZP_{i-1,j}, \quad \beta' := \beta + \sum_{i'=0}^{i} \alpha_{2i'+1} \in ZP_{i,j-1}$$

satisfy the desired relations. □

The double filtration $P_{i,j}$ on $\omega^\bullet_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})}$ induces a double filtration $P_{i,j}$ on $\widetilde{\widetilde{C}}_{\mathfrak{X}_n}$, and, for $i, j \geq 1$, the residue morphism $\mathrm{Res} : P_{i,j}\omega^q_{(\mathfrak{X}_n,\widetilde{N})/(W_n,\mathrm{triv})} \to \Omega^{q-i-j}_{\mathfrak{D}_n^{(i,j)}/W_n}$ factors through $P_{i,j}\widetilde{\widetilde{C}}_{\mathfrak{X}_n}$.

Lemma 3.15. *For any two admissible liftings $(Z_1, \widetilde{N})$ and $(Z_2, \widetilde{N})$ of $(Y, \widetilde{M})$, we have a canonical isomorphism*

$$\alpha_{Z_1Z_2} : \mathcal{H}^q(P_{i,j}\widetilde{\widetilde{C}}^\bullet_{\mathfrak{X}_{1,n}}) \to \mathcal{H}^q(P_{i,j}\widetilde{\widetilde{C}}^\bullet_{\mathfrak{X}_{2,n}})$$

satisfying the cocycle condition for any three admissible liftings.

Moreover, the residue morphism

$$\mathrm{Res}_Z : \mathcal{H}^q(P_{i,j}\widetilde{\widetilde{C}}^\bullet_{\mathfrak{X}_n}) \to \mathcal{H}^{q-i-j}(\Omega^\bullet_{\mathfrak{D}_n^{(i,j)}/W_n}) \simeq W_n\Omega^q_{Y^{(i,j)}}$$

induced on cohomology satisfies the compatibility

$$\mathrm{Res}_{Z_1} = \mathrm{Res}_{Z_2} \circ \alpha_{Z_1Z_2}.$$

Proof. The proof of the first part is basically the same as the proof of Lemma 3.5. We take admissible lifts $(Z_1, \widetilde{N})$ and $(Z_2, \widetilde{N})$ (we denote the log structures on both simply by $\widetilde{N}$, as it will be understood from the context which is the underlying scheme). As in the proof of Lemma 3.5, we form $(Z_{12}, \widetilde{N})$, which is smooth over $(Z_i, \widetilde{N})$, even though it is not quite an admissible lift. However, Z_{12} is étale over

$$\operatorname{Spec} W[X_1,\dots,X_n,Y_1,\dots,Y_n,X'_1,\dots X'_n,Y'_1,\dots,Y'_n,v_1^{\pm 1},\dots,v_r^{\pm 1},u_1^{\pm 1},\dots u_s^{\pm 1}] \\ /(X_i v_i - X'_i . Y_j v_j - Y'_j).$$

So we can endow $\widetilde{\widetilde{C}}^{\bullet}_{\mathfrak{X}_{12,n}}$ with a filtration $P_{i,j}\widetilde{\widetilde{C}}^{\bullet}_{\mathfrak{X}_{12,n}}$ defined as above, in terms of log structures $\widetilde{N}_1$ and $\widetilde{N}_2$ (which come from formally "inverting" the X_i and X'_i or the Y_i and Y'_i). Then the same argument used in the proof of Lemma 3.5 gives us quasi-isomorphisms

$$P_{i,j}\widetilde{\widetilde{C}}_{\mathfrak{X}_{i,n}} \to P_{i,j}\widetilde{\widetilde{C}}_{\mathfrak{X}_{12,n}}$$

for $i = 1, 2$, which satisfy the right compatibility condition for three admissible lifts.

For the second part, we follow the argument in Lemma 3.4(2) of [Mokrane 1993]. We let

$$\omega = \alpha \wedge \frac{dX_{k_1}}{X_{k_1}} \wedge \dots \wedge \frac{dX_{k_i}}{X_{k_i}} \wedge \frac{dY_{l_1}}{Y_{l_1}} \wedge \dots \wedge \frac{dY_{l_i}}{Y_{l_i}}$$

be a section of $P_{i,j}\omega^q_{(\mathfrak{X}_{1,n},\widetilde{N})/(W_n,\mathrm{triv})}$ and

$$\omega' = \alpha' \wedge \frac{dX'_{k_1}}{X'_{k_1}} \wedge \dots \wedge \frac{dX'_{k_i}}{X'_{k_i}} \wedge \frac{dY'_{l_1}}{Y'_{l_1}} \wedge \dots \wedge \frac{dY'_{l_i}}{Y'_{l_i}}$$

be a section of $P_{i,j}\omega^q_{(\mathfrak{X}_{2,n},\widetilde{N})/(W_n,\mathrm{triv})}$ such that $\omega = \omega'$ in $P_{i,j}\omega^q_{(\mathfrak{X}_{12,n},\widetilde{N})/(W_n,\mathrm{triv})}$. We have to check that $\alpha|_{\mathfrak{D}^{(i,j)}_{12,n}} = \alpha'|_{\mathfrak{D}^{(i,j)}_{12,n}}$. But

$$\omega - \omega' = (\alpha - \alpha') \wedge \frac{dX_{k_1}}{X_{k_1}} \wedge \dots \wedge \frac{dX_{k_i}}{X_{k_i}} \wedge \frac{dY_{l_1}}{Y_{l_1}} \wedge \dots \wedge \frac{dY_{l_i}}{Y_{l_i}} + \Psi,$$

where $\Psi \in P_{i,j-1}\omega^q_{(\mathfrak{X}_{2,n},\widetilde{N})/(W_n,\mathrm{triv})} + P_{i-1,j}\omega^q_{(\mathfrak{X}_{2,n},\widetilde{N})/(W_n,\mathrm{triv})}$. This means that

$$(\alpha - \alpha') \wedge \frac{dX_{k_1}}{X_{k_1}} \wedge \dots \wedge \frac{dX_{k_i}}{X_{k_i}} \wedge \frac{dY_{l_1}}{Y_{l_1}} \wedge \dots \wedge \frac{dY_{l_i}}{Y_{l_i}}$$

is also a section of $P_{i,j-1}\omega^q_{(\mathfrak{X}_{2,n},\widetilde{N})/(W_n,\mathrm{triv})} + P_{i-1,j}\omega^q_{(\mathfrak{X}_{2,n},\widetilde{N})/(W_n,\mathrm{triv})}$, and so $(\alpha - \alpha')|_{\mathfrak{D}^{(i,j)}_{12,n}} = 0$. □

Corollary 3.16. *We can define the sheaves*

$$P_{i,j}W_n\tilde{\tilde{\omega}}^q_Y := \mathcal{H}^q(P_{i,j}\widetilde{\widetilde{C}}^{\bullet}_Y).$$

The complexes $P_{i,j} W_n \tilde{\tilde{\omega}}^\bullet_Y$ form an increasing double filtration of $W_n \tilde{\tilde{\omega}}^\bullet_Y$ such that the graded pieces for $i, j \geq 1$

$$\mathrm{Gr}_{i,j}\, W_n \tilde{\tilde{\omega}}^\bullet_Y := P_{i,j} W_n \tilde{\tilde{\omega}}^\bullet_Y / P_{i,j-1} + P_{i-1,j}$$

are canonically isomorphic to the de Rham–Witt complexes of the smooth subschemes $Y^{(i,j)}$:

$$\mathrm{Res} : \mathrm{Gr}_{i,j}\, W_n \tilde{\tilde{\omega}}^\bullet_Y \xrightarrow{\sim} W_n \Omega^\bullet_{Y^{(i,j)}}[-i-j](-i-j).$$

Lemma 3.17. *The constructions in Section 3C are compatible with the transition morphisms π, in the following way*:

(1) *The following diagrams are commutative*:

$$\begin{tikzcd} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\pi"] \arrow[d, "\wedge\frac{d\tau}{\tau}"'] & W_n\tilde{\tilde{\omega}}^q_Y \arrow[d, "\wedge\frac{d\tau}{\tau}"] \\ W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\pi"] & W_n\tilde{\tilde{\omega}}^q_Y \end{tikzcd} \quad \textit{and} \quad \begin{tikzcd} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\pi"] \arrow[d, "\wedge\frac{d\sigma}{\sigma}"'] & W_n\tilde{\tilde{\omega}}^q_Y \arrow[d, "\wedge\frac{d\sigma}{\sigma}"] \\ W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\pi"] & W_n\tilde{\tilde{\omega}}^q_Y \end{tikzcd}$$

(2) *The projection $\pi : W_{n+1}\tilde{\tilde{\omega}}^q_Y \to W_n\tilde{\tilde{\omega}}^q_Y$ preserves the weight filtration $P_{i,j}$ on $W_m\tilde{\tilde{\omega}}^q_Y$ for $m = n, n+1$.*

(3) *The morphism $\pi : P_{i,j} W_{n+1}\tilde{\tilde{\omega}}^q_Y \to P_{i,j} W_n\tilde{\tilde{\omega}}^q_Y$ is surjective.*

Proof. The first part follows in the same way as Proposition 8.1 of [Nakkajima 2005], by using a local admissible lifting $(Z, \widetilde{N})$ of $(Y, \widetilde{M})$ together with a lift of the Frobenius Φ. Then $\Phi^*(\tau) = \tau^p(1 + pu)$ for some

$$u \in \mathcal{O}_Z \otimes_{W[\tau,\sigma]} W_n\langle \tau, \sigma\rangle$$

and so $\Phi^*(d \log \tau)$ is equivalent to $pd \log \tau$ modulo an exact form. The same holds for σ.

The second part follows in the same way as Proposition 8.4 of [Nakkajima 2005]. The question is local, so we may assume that the admissible lift $(Z, \widetilde{N})$ is étale over $\mathrm{Spec}\, W[X_1, \dots, X_n, Y_1, \dots, Y_n]$, $\mathbb{N}^r \oplus \mathbb{N}^s$. First we see that, for a lift Φ of Frobenius we have that $\Phi^*(d \log X_i)$ is equivalent modulo an exact form to $pd \log X_i$ for $1 \leq i \leq r$ and that $\Phi^*(d \log Y_j)$ is equivalent modulo an exact form to $pd \log Y_j$ for $1 \leq j \leq s$. This implies that the map $\mathfrak{p} : W_n\tilde{\tilde{\omega}}^q_Y \to W_{n+1}\tilde{\tilde{\omega}}^q_Y$ preserves the weight filtration $P_{i,j}$.

In order to see that $\pi : W_{n+1}\tilde{\tilde{\omega}}^q_Y \to W_n\tilde{\tilde{\omega}}^q_Y$ also preserves $P_{i,j}$ we use a descending induction on (i, j) in lexicographic order. Note that $P_{r,s} W_n\tilde{\tilde{\omega}}^q_Y = W_n\tilde{\tilde{\omega}}^q_Y$, so there is nothing to prove in this case. We can prove the result for $(r, s-1)$ in the same way as Proposition 8.4(2) of [Nakkajima 2005], using the commutative diagrams

$$\begin{tikzcd} P_{i,j} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\mathrm{Res}"] \arrow[d, "\pi"'] & W_{n+1}\Omega^{q-i-j}_{Y^{(i,j)}} \arrow[d, "\pi"] \\ P_{i,j} W_n\tilde{\tilde{\omega}}^q_Y \arrow[r, "\mathrm{Res}"] & W_n\Omega^{q-i-j}_{Y^{(i,j)}} \end{tikzcd}$$

for (i, j) successively equal to $(r, s), (r-1, s), \ldots, (1, s)$. At the last step we get a commutative diagram of (vertical) exact sequences

$$\begin{tikzcd} 0 \arrow[d] & 0 \arrow[d] \\ P_{r,s-1} W_{n+1}\tilde{\tilde{\omega}}^q_Y + P_{0,s} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[d] & P_{r,s-1} W_n\tilde{\tilde{\omega}}^q_Y + P_{0,s} W_n\tilde{\tilde{\omega}}^q_Y \arrow[d] \\ P_{r,s-1} W_{n+1}\tilde{\tilde{\omega}}^q_Y + P_{1,s} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\pi"'] \arrow[d] & P_{r,s-1} W_n\tilde{\tilde{\omega}}^q_Y + P_{1,s} W_n\tilde{\tilde{\omega}}^q_Y \arrow[d] \\ W_{n+1}\Omega^{q-s-1}_{Y^{(1,s)}} \arrow[r, "\pi"'] \arrow[d] & W_n\Omega^{q-s-1}_{Y^{(1,s)}} \arrow[d] \\ 0 & 0 \end{tikzcd}$$

which means there is an induced morphism $\pi : P_{r,s-1} W_{n+1}\tilde{\tilde{\omega}}^q_Y + P_{0,s} W_{n+1}\tilde{\tilde{\omega}}^q_Y \to P_{r,s-1} W_n\tilde{\tilde{\omega}}^q_Y + P_{0,s} W_n\tilde{\tilde{\omega}}^q_Y$.

At this stage, we note that we can define

$$Y^{(0,s)} = \coprod_{\substack{T\subseteq\{1,\ldots,n\}\\ \#T=s}} \Bigl(\bigcap_{i\in T} Y_i^2\Bigr).$$

This will be a simple reduced normal crossings divisor over k, and we can endow it with the pullback of the log structure $\widetilde{M}_1$ so that $(Y, \widetilde{M})$ is a $(k, \mathbb{N})$-semistable log scheme, in the terminology of Section 2.4 of [Mokrane 1993]. There is a surjective residue morphism obtained via the restriction

$$P_{i,j} W_n\tilde{\tilde{\omega}}^q_Y \xrightarrow{\mathrm{Res}} P_i W_n\tilde{\omega}^{q-j}_{Y^{(0,j)}},$$

which respects the weight filtrations. Just as the commutative diagram 8.4.3 of [Nakkajima 2005] is obtained, we can use the injectivity of $\mathfrak{p} : W_n\tilde{\omega}^q_{Y^{(0,s)}} \to W_{n+1}\tilde{\omega}^q_{Y^{(0,s)}}$ for $Y^{(0,s)}/k$ [Nakkajima 2005, Corollary 6.28(2)] to see that there is a commutative diagram

$$\begin{tikzcd}
P_{0,s} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\mathrm{Res}"] \arrow[d, "\pi"'] & P_0 W_{n+1}\tilde{\omega}^{q-s}_{Y(0,s)} \arrow[d, "\pi"] \\
P_{0,s} W_n\tilde{\tilde{\omega}}^q_Y \arrow[r, "\mathrm{Res}"] & P_0 W_n\tilde{\omega}^{q-s}_{Y(0,s)}
\end{tikzcd}$$

We therefore get a commutative diagram of (vertical) exact sequences:

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] \\
P_{r,s-1} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[d] & P_{r,s-1} W_n\tilde{\tilde{\omega}}^q_Y \arrow[d] \\
P_{r,s-1} W_{n+1}\tilde{\tilde{\omega}}^q_Y + P_{0,s} W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\pi"'] \arrow[d] & P_{r,s-1} W_n\tilde{\tilde{\omega}}^q_Y + P_{0,s} W_n\tilde{\tilde{\omega}}^q_Y \arrow[d] \\
P_0 W_{n+1}\tilde{\omega}^{q-s}_{Y(0,s)} \arrow[r, "\pi"'] \arrow[d] & P_0 W_n\tilde{\omega}^{q-s}_{Y(0,s)} \arrow[d] \\
0 & 0
\end{tikzcd}$$

so there is an induced morphism $\pi : P_{r,s-1} W_{n+1}\tilde{\tilde{\omega}}^q_Y \to P_{r,s-1} W_n\tilde{\tilde{\omega}}^q_Y$.

Finally, the third part follows in the same way as Corollary 8.6.4 of [Nakkajima 2005]. For an admissible lift $(Z, \widetilde{N})$, let $Z_1 := Z \times_W k$. We have surjective morphisms $W_n\Omega^q_{Z_1} \to P_{0,0} W_n\tilde{\tilde{\omega}}^q_Y$, which commute with the transition morphisms π. So π is surjective for $P_{0,0}$. Using the exact sequences of the form

$$0 \to P_{0,j-1} W_n\tilde{\tilde{\omega}}^q_Y \to P_{0,j} W_n\tilde{\tilde{\omega}}^q_Y \to P_0 W_n\tilde{\omega}^{q-j}_{Y(0,j)} \to 0$$

and the surjectivity of π on the third term, we prove by induction on j that π is surjective for $P_{0,j}$. The same statement holds for $P_{i,0}$. Then, we prove that π is surjective for a general $P_{i,j}$ by induction on $i + j$, using the exact sequences of the form

$$0 \to P_{i-1,j} W_n\tilde{\tilde{\omega}}^q_Y + P_{i,j-1} W_n\tilde{\tilde{\omega}}^q_Y \to P_{i,j} W_n\tilde{\tilde{\omega}}^q_Y \to W_n\Omega^{q-i-j}_{Y(i,j)} \to 0. \qquad \square$$

4. Generalizing the Mokrane spectral sequence

In this section, we derive a generalization of the Mokrane spectral sequence which will allow us to compute the log crystalline cohomology of the Shimura varieties we are interested in terms of the crystalline cohomology of certain proper smooth

Newton polygon strata in the special fiber. Mokrane's spectral sequence applies to the case of semistable reduction. Here we treat the case of a scheme whose singularities are locally those of a product of semistable schemes which is no longer semistable.

We define a double complex $W_n A^{\bullet\bullet}$ as follows. Its terms are

$$W_n A^{ij} := \bigoplus_{k=0}^{j} W_n \tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k,i+j+2} + P_{i+j+2,j-k}) \quad \text{for } i, j \geq 0,$$

and $W_n A^{ij} := 0$ otherwise. The operators d, π, F, V of $W_\bullet \tilde{\tilde{\omega}}^\bullet$ induce operators d', π, F, V of the procomplexes $W_\bullet A^{\bullet j}$. For x in the direct summand $W_n \tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k,i+j+2} + P_{i+j+2,j-k})$ of $W_n A^{ij}$, $d'x$ is the class of $(-1)^j d\tilde{x}$, where $\tilde{x}$ is a lift of x in $W_n \tilde{\tilde{\omega}}_Y^{i+j+2}$. We also have a differential $d'' : W_n A^{ij} \to W_n A^{ij+1}$ given by

$$d''x = (-1)^i \Big(\frac{d\tau}{\tau} \wedge x + \frac{d\sigma}{\sigma} \wedge x\Big),$$

where $d\tau/\tau$ and $d\sigma/\sigma$ are the global sections of $W_n \tilde{\tilde{\omega}}_Y^1$ defined in Lemma 3.9. We have $d'd'' = d''d'$, so we indeed get a double procomplex $(W_\bullet A^{\bullet\bullet}, d', d'')$. As in Lemma 3.9 of [Mokrane 1993], we can use dévissage by weights to see that the components of this procomplex are p-torsion-free. Let $W_\bullet A^\bullet$ be the simple procomplex associated to the double procomplex $W_\bullet A^{\bullet\bullet}$.

We define now an endomorphism ν of bidegree $(-1, 1)$ of $W_n A^{\bullet\bullet}$ which will induce the monodromy operator on cohomology. For each $k \in \{0, \ldots, j\}$ we have natural maps

$$\begin{aligned} W_n \tilde{\tilde{\omega}}_Y^{i+j+2}&/(P_{k,i+j+2} + P_{i+j+2,j-k}) \\ &\to W_n \tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k,i+j+2} + P_{i+j+2,j+1-k}) \\ &\qquad \oplus W_n \tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k+1,i+j+2} + P_{i+j+2,j-k}), \end{aligned}$$

which are sums of $(-1)^{i+j+1}$ proj on each factor. Summing over k we get maps $\nu : W_n A^{ij} \to W_n A^{i-1j+1}$, which induce an endomorphism ν of bidegree $(-1, 1)$.

The morphism of complexes $W_n \tilde{\tilde{\omega}}_Y^\bullet \to W_n A^{\bullet 0}$ given by

$$x \mapsto \frac{d\tau}{\tau} \wedge \frac{d\sigma}{\sigma} \wedge x$$

factors through $W_n \omega_Y^\bullet$. We get a morphism of complexes

$$\Theta : W_n \omega_Y^\bullet \to W_n A^\bullet.$$

The following lemma is analogous to Theorem 9.9 of [Nakkajima 2005]. It ensures that the resulting spectral sequence will be compatible with the Frobenius endomorphism (defined as an endomorphism of W_n-modules). We let $\Phi_n : W_n\omega_Y \to W_n\omega_Y$ be the Frobenius endomorphism induced by the absolute Frobenius endomorphism of (Y, M).

Lemma 4.1. *Let n be a positive integer. Then the following hold*:

(1) *There exists a unique endomorphism $\tilde{\tilde{\Phi}}_n^{\bullet,\bullet}$ of $W_nA^{\bullet\bullet}$ of double complexes, making the following diagram commutative*:

$$\begin{array}{ccc} W_{n+1}A^{qm} & \xrightarrow{\pi} & W_nA^{qm} \\ \Big\downarrow{\scriptstyle p^qF} & & \Big\downarrow{\scriptstyle \tilde{\tilde{\Phi}}_n^{qm}} \\ W_nA^{qm} & \xrightarrow{\mathrm{id}} & W_nA^{qm} \end{array}$$

(2) *The endomorphism $\tilde{\tilde{\Phi}}_n^{\bullet,\bullet}$ induces an endomorphism $\tilde{\tilde{\Phi}}_n$ of the complex $W_nA^\bullet$, fitting in a commutative diagram*

$$\begin{array}{ccc} W_n\omega_Y^\bullet & \xrightarrow{\Phi_n} & W_n\omega_Y^\bullet \\ \Big\downarrow{\scriptstyle \Theta} & & \Big\downarrow{\scriptstyle \Theta} \\ W_nA^\bullet & \xrightarrow{\tilde{\tilde{\Phi}}_n} & W_nA^\bullet \end{array}$$

(3) *Finally, the Poincaré residue isomorphism* Res *fits in the following commutative diagrams for $i, j \geq 1$*:

$$\begin{array}{ccc} Gr_{i,j}W_n\tilde{\tilde{\omega}}_Y^q & \xrightarrow{\mathrm{Res}} & W_n\Omega^{q-i-j}_{Y(i,j)} \\ \Big\downarrow{\scriptstyle \Psi_n} & & \Big\downarrow{\scriptstyle p^{i+j}\Phi_n} \\ Gr_{i,j}W_n\tilde{\tilde{\omega}}_Y^q & \xrightarrow{\mathrm{Res}} & W_n\Omega^{q-i-j}_{Y(i,j)} \end{array}$$

where Ψ_n is an endomorphism of $W_n\tilde{\tilde{\omega}}_Y^\bullet$ which respects the weight filtration $P_{i,j}$ and which induces $\tilde{\tilde{\Phi}}_n^{\bullet,\bullet}$ on $W_nA^{\bullet,\bullet}$.

Proof. The proof is essentially the same as that of Theorem 9.9 of [Nakkajima 2005]. We emphasize only the key points. We can define a morphism

$$\Psi_n^{j,q} : W_n\tilde{\tilde{\omega}}_Y^q \to W_n\tilde{\tilde{\omega}}_Y^q$$

via the composition

$$W_n\tilde{\tilde{\omega}}^q_Y \xrightarrow{\mathfrak{p}} W_{n+1}\tilde{\tilde{\omega}}^q_Y \xrightarrow{p^{j-1}} W_{n+1}\tilde{\tilde{\omega}}^q_Y \xrightarrow{F} W_n\tilde{\tilde{\omega}}^q_Y.$$

The fact that these morphisms commute with the maps $(d\tau/\tau)\wedge$ and $(d\sigma/\sigma)\wedge$ follows from the proof of the first part of Lemma 3.17. This implies that the second diagram is commutative. The fact that the $\Psi_n^{\bullet\bullet}$ respect the weight filtration follows from the analogous statement for $\mathfrak{p}$, which is proved in Lemma 3.17 as well. This means that we can use Ψ_n^{jj+q+2} to define endomorphisms $\tilde{\tilde{\Phi}}{}_n^{jq}$ of W_nA^{jq}, at least for $j \geq 1$. For $j = 0$ we use the Frobenius endomorphism Φ_n of $W_n(\mathbb{O}_{Y^{(k+1,j-k+1)}})$ together with the residue isomorphisms to define $\tilde{\tilde{\Phi}}{}_n^{0q}$. The commutativity of the first diagram now follows from the definitions, from the commutative diagram

$$\begin{tikzcd}
W_{n+1}\tilde{\tilde{\omega}}^{qm}_Y \arrow[r, "\pi"] \arrow[d, "p^qF"'] & W_n\tilde{\tilde{\omega}}^{qm}_Y \arrow[d, "\Psi_n^{qm}"] \\
W_nA^{qm} \arrow[r, "\mathrm{id}"] & W_nA^{qm}
\end{tikzcd}$$

(which is deduced from $\mathfrak{p}d = d\mathfrak{p}$ and $dF = pFd$) and from Diagram 9.2.2 of [Nakkajima 2005] in the case of a smooth morphism. The fact that the first diagram is commutative ensures the uniqueness of $\Phi_n^{q,m}$. Finally, the third commutative diagram follows from the surjectivity of π proved in Lemma 3.17, from Diagram 9.2.2 of [Nakkajima 2005] in the case of a smooth morphism and from the commutative diagrams

$$\begin{tikzcd}
P_{i,j}W_{n+1}\tilde{\tilde{\omega}}^q_Y \arrow[r, "\mathrm{Res}"] \arrow[d, "\pi"'] & W_{n+1}\Omega^{q-i-j}_{Y^{(i,j)}} \arrow[d, "\pi"] \\
P_{i,j}W_n\tilde{\tilde{\omega}}^q_Y \arrow[r, "\mathrm{Res}"] & W_n\Omega^{q-i-j}_{Y^{(i,j)}}
\end{tikzcd}$$

for $i, j \geq 1$. □

Proposition 4.2. *The sequence*

$$0 \to W_n\omega^\bullet_Y \xrightarrow{\Theta} W_nA^{\bullet 0} \xrightarrow{d''} W_nA^{\bullet 1} \xrightarrow{d''} \cdots$$

is exact.

Proof. We follow the proof of Proposition 3.15 of [Mokrane 1993]. Let $\theta : W_n\tilde{\tilde{\omega}}^{i-1}_Y \oplus W_n\tilde{\tilde{\omega}}^{i-1}_Y \to W_n\tilde{\tilde{\omega}}^i_Y$ be defined by

$$(x, y) \mapsto \frac{d\tau}{\tau} \wedge x + \frac{d\sigma}{\sigma} \wedge y.$$

It suffices to check that the sequence

$$\begin{aligned}W_n\tilde{\tilde{\omega}}_Y^{i-2} &\xrightarrow{(\frac{d\sigma}{\sigma}\wedge,\frac{d\tau}{\tau}\wedge)} W_n\tilde{\tilde{\omega}}_Y^{i-1}\oplus W_n\tilde{\tilde{\omega}}_Y^{i-1} \xrightarrow{\theta} W_n\tilde{\tilde{\omega}}_Y^{i}\\ &\xrightarrow{\frac{d\tau}{\tau}\wedge\frac{d\sigma}{\sigma}\wedge} W_n\tilde{\tilde{\omega}}_Y^{i+2}/(P_{0,i+2}+P_{i+2,0})\\ &\xrightarrow{d''} W_n\tilde{\tilde{\omega}}_Y^{i+3}/(P_{1,i+3}+P_{i+3,0})\oplus W_n\tilde{\tilde{\omega}}_Y^{i+3}/(P_{0,i+3}+P_{i+3,1})\\ &\xrightarrow{d''}\cdots \end{aligned}\tag{4.1}$$

is exact. We do this by using first a dévissage by weights, reducing to the case $n = 1$ and then using the fact that the scheme Y is Zariski-locally étale over a product of (the special fibers of) strictly semistable schemes.

We let

$$\begin{aligned}K_{-4} &= W_n\tilde{\tilde{\omega}}_Y^{i-2},\\ K_{-3} &= W_n\tilde{\tilde{\omega}}_Y^{i-1}\oplus W_n\tilde{\tilde{\omega}}_Y^{i-1},\\ K_{-2} &= W_n\tilde{\tilde{\omega}}_Y^{i},\\ K_j &= \bigoplus_{k=0}^{j} W_n\tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k,i+j+2}+P_{i+j+2,j-k}),\quad j\ge 0.\end{aligned}$$

For $j \ge -4$, $j \ne -1$ we define a double filtration of K_j as follows:

$$\begin{aligned}P_{l,m}K_{-4} &= P_{l-2,m-2}W_n\tilde{\tilde{\omega}}_Y^{i-2},\\ P_{l,m}K_{-3} &= P_{l-2,m-1}W_n\tilde{\tilde{\omega}}_Y^{i-1}\oplus P_{l-1,m-2}W_n\tilde{\tilde{\omega}}_Y^{i-1},\\ P_{l,m}K_{-2} &:= P_{l-1,m-1}W_n\tilde{\tilde{\omega}}_Y^{i},\\ P_{l,m}K_j &:= \bigoplus_{k=0}^{j} P_{l+k,m+j-k}W_n\tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k,i+j+2}+P_{i+j+2,j-k}),\quad j\ge 0.\end{aligned}$$

Here we set the convention $P_{l,m}W_n\tilde{\tilde{\omega}}^i = 0$ if either $l < 0$ or $m < 0$. The sequence (4.1) is a filtered sequence and to prove exactness it suffices to prove exactness for each graded piece

$$\mathrm{Gr}_{l,m} K_j := P_{l,m}K_j/(P_{l,m-1}K_j + P_{l-1,m}K_j).$$

For $l, m \ge 0$ we can rewrite the sequences of graded pieces as:

$$\begin{aligned}\mathrm{Gr}_{l-2,m-2} W_n\tilde{\tilde{\omega}}_Y^{i-2} &\to \mathrm{Gr}_{l-2,m-1} W_n\tilde{\tilde{\omega}}_Y^{i-1}\oplus \mathrm{Gr}_{l-1,m-2} W_n\tilde{\tilde{\omega}}_Y^{i-1}\\ &\to \mathrm{Gr}_{l-1,m-1} W_n\tilde{\tilde{\omega}}_Y^{i}\to \mathrm{Gr}_{l,m} W_n\tilde{\tilde{\omega}}_Y^{i+2}\\ &\to \mathrm{Gr}_{l+1,m} W_n\tilde{\tilde{\omega}}_Y^{i+3}\oplus \mathrm{Gr}_{l,m+1} W_n\tilde{\tilde{\omega}}_Y^{i+3}\to\cdots.\end{aligned}$$

For $l<0$ or $m<0$ the sequence is trivial.

It suffices to show that the sequence of complexes

$$\begin{aligned}\mathrm{Gr}_{l-2,m-2}\, W_n\tilde{\tilde{\omega}}^\bullet_Y[-2] &\to \mathrm{Gr}_{l-2,m-1}\, W_n\tilde{\tilde{\omega}}^\bullet_Y[-1]\oplus \mathrm{Gr}_{l-1,m-2}\, W_n\tilde{\tilde{\omega}}^\bullet_Y[-1]\\ &\to \mathrm{Gr}_{l-1,m-1}\, W_n\tilde{\tilde{\omega}}^\bullet_Y \xrightarrow{\iota} \mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet_Y[2]\\ &\to \mathrm{Gr}_{l+1,m}\, W_n\tilde{\tilde{\omega}}^\bullet_Y[3]\oplus \mathrm{Gr}_{l,m+1}\, W_n\tilde{\tilde{\omega}}^\bullet_Y[3]\to\cdots \end{aligned}\tag{4.2}$$

is exact. Note that we can check this locally. When $l,m\geq 1$, we know by Corollary 3.16 that

$$\mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet_Y \simeq W_n\Omega^\bullet_{Y^{(l,m)}}[-l-m](-l-m).$$

For $l=0$ and $m\geq 1$ let $Y_{D^{0,m}}$ be the normal crossing divisor of $D^{0,m}$ corresponding to $s=0$. In this case we have

$$\mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet_Y \simeq [W_n\Omega_{D^{0,m}}(-\log Y_{D^{o,m}})\to W_n\Omega_{D^{0,m}}],$$

and for $l=0, m=0$ we have the quasi-isomorphism

$$\begin{aligned}\mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet_Y \simeq [&W_n\Omega^\bullet_Z(-\log Y^1-\log Y^2)\\ &\to W_n\Omega^\bullet_Z(-\log Y^1)\oplus W_n\Omega^\bullet_Z(-\log Y^2)\to W_n\Omega^\bullet_Z],\end{aligned}$$

where $Z=\underline{Z}\otimes_W k$. In any case, $\mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet$ satisfies the property

$$(\lim \mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet)\otimes^L_{\mathbb{R}}\mathbb{R}_n\simeq \mathrm{Gr}_{l,m}\, W_n\tilde{\tilde{\omega}}^\bullet$$

by Lemma 1.3.3 of [Mokrane 1993] and Lemma 3.11. By Proposition 2.3.7 of [Illusie 1983], it suffices to check exactness of the sequence (4.2) for $n=1$.

For $n=1$ and working locally with our admissible lifts, we know that the exact sequence (4.2) is the pullback to Y of the corresponding exact sequence on $Y_1\times_k Y_2$. We can assume that $Y=Y_1\times_k Y_2$ and $Z=Z_1\times_k Z_2$. Each Y_i for $i=1,2$ is a reduced normal crossings divisor in Z_i, for which we know that

$$\mathrm{Gr}_{l_i-2}\, W_1\tilde{\omega}^\bullet_{Y_i}[-1]\to \mathrm{Gr}_{l_i-1}\, W_1\tilde{\omega}_{Y_i}\to \mathrm{Gr}_{l}\, W_1\tilde{\omega}^\bullet_{Y_i}[1]\to \mathrm{Gr}_{l+1}\, W_1\tilde{\omega}^\bullet_{Y_i}[2]\to\cdots$$

is exact, by the proof of Proposition 3.15 of [Mokrane 1993]. In other words, for $i=1,2$ we have quasi-isomorphisms between the top row and the bottom row. Multiplying the quasi-isomorphisms for $i=1,2$ gives us exactly the quasi-isomorphism ι needed to prove the exactness of (4.2) in the case $n=1$. Here, we use the Cartier isomorphisms for $W_1\tilde{\omega}_{Y_i}$ and for $W_1\tilde{\tilde{\omega}}_Y$ and the fact that

$$(\omega^\bullet_{(Z_1,\tilde{N}_1)/k}\otimes_{\mathbb{O}_{Z_1}}\mathbb{O}_{Y_1})\otimes_k(\omega^\bullet_{(Z_2,\tilde{N}_2)/k}\otimes_{\mathbb{O}_{Z_2}}\mathbb{O}_{Y_2})\simeq \omega^\bullet_{(Z,\tilde{N})}\otimes_{\mathbb{O}_Z}\mathbb{O}_Y,$$

where the two complexes on the left determine $W_1\tilde{\omega}^\bullet_{Y_i}$ for $i=1,2$ and the one on the right determines $W_1\tilde{\tilde{\omega}}^\bullet_Y$. □

Corollary 4.3. *The morphism of complexes $\Theta : W_n\omega_Y^\bullet \to W_nA^\bullet$ is a quasi-isomorphism. It induces a quasi-isomorphism $\Theta : W\omega_Y^\bullet \xrightarrow{\sim} WA^\bullet$.*

Proposition 4.4. *The endomorphism ν of $W_\bullet A^{\bullet\bullet}$ induces the monodromy operator N over $H^*_{\mathrm{cris}}((Y, M)/(W, \mathbb{N}))$.*

Proof. We define the double complex $B_n^{\bullet\bullet}$ as follows:

$$\begin{aligned} B_n^{\bullet\bullet} &= W_nA^{i-1j} \oplus W_nA^{ij}, \quad i, j \geq 0, \\ d'(x_1, x_2) &= (d'x_1, d'x_2), \\ d''(x_1, x_2) &= (d''x_1 + \nu(x_2), d''x_2). \end{aligned}$$

We have a morphism of complexes $\Psi : W_n\tilde{\omega}_Y^\bullet \to B_n^\bullet$ defined as follows: for $x \in W_n\tilde{\omega}_Y^i$,

$$\Psi(x) = \left(\left(\frac{d\sigma}{\sigma} - \frac{d\tau}{\tau}\right) \wedge x \ (\mathrm{mod}\ P_{0,i+1} + P_{i+1,0}), \frac{d\tau}{\tau} \wedge \frac{d\sigma}{\sigma} \wedge x \ (\mathrm{mod}\ P_{0,i+2} + P_{i+2,0})\right).$$

Thus we have a commutative diagram of exact sequences of complexes:

$$\begin{tikzcd} 0 \arrow[r] & W_n\omega_Y^\bullet[-1] \arrow[r] \arrow[d, "{\Theta[-1]}"] & W_n\tilde{\omega}_Y^\bullet \arrow[r] \arrow[d, "\Psi"] & W_n\omega_Y^\bullet \arrow[r] \arrow[d, "\Theta"] & 0 \\ 0 \arrow[r] & W_nA^\bullet[-1] \arrow[r] & B_n^\bullet \arrow[r] & W_nA^\bullet \arrow[r] & 0 \end{tikzcd}$$

where the left and right downward arrows are quasi-isomorphisms. Thus, Ψ is also a quasi-isomorphism and the commutative diagram defines an isomorphism of distinguished triangles. Thus the monodromy operator N on cohomology is induced by the coboundary operator of the bottom exact sequence, which by construction is ν. □

We can compute the monodromy filtration of the nilpotent operator N on cohomology from the monodromy filtration of ν on $W_nA^\bullet$. We will exhibit a filtration $P_k(W_nA^\bullet) = \bigoplus_{i,j\geq 0} P_k(W_nA^{ij})$ which satisfies the following:

(1) $\nu(P_k(W_\bullet A^\bullet)) \subset P_{k-2}(W_\bullet A^\bullet)(-1)$.

(2) For $k \geq 0$, the induced map $\nu^k : \mathrm{Gr}_k(W_\bullet A^\bullet) \to \mathrm{Gr}_{-k}(W_\bullet A^\bullet)(-k)$ is an isomorphism.

A filtration satisfying these two properties must be the monodromy filtration of ν.

Note 4.5. From now on, we will not work in the category $\mathfrak{C}$ of complexes of sheaves of W-modules but rather in $\mathbb{Q} \otimes \mathfrak{C}$, which is the category with the same set of objects as $\mathfrak{C}$, but with morphisms $\mathbb{Q} \otimes \mathrm{Hom}_{\mathfrak{C}}(A, B)$. We will in fact identify the monodromy filtration of ν on $\mathbb{Q} \otimes W_nA^\bullet$, but for simplicity of notation we still denote an object A of $\mathfrak{C}$ as A when we regard it as an object of $\mathbb{Q} \otimes \mathfrak{C}$.

Define $P_l(W_n A^{\bullet\bullet}) := \bigoplus_{i,j\geq 0} P_l(W_n A^{ij})$ for $l \geq 0$, where $P_l(W_n A^{ij})$ is 0 if $l < 2n-2-j$ and

$$\bigoplus_{k=0}^{j}\Bigg(\sum_{m=0}^{l-2n+2+j} P_{k+m+1,2j-k+l-2n-m+3} W_n \tilde{\tilde{\omega}}_Y^{i+j+2}/(P_{k,i+j+2}+P_{j-k,i+j+2})\Bigg)$$

if $l \geq 2n-2-j$. It is easy to check that $\nu(P_l(W_n A^{ij})) \subset P_{l-2} W_n A^{i+1,j-1}$. Moreover, we can also compute the graded pieces $\mathrm{Gr}_l(W_n A^{\bullet\bullet}) = \bigoplus_{i,j\geq 0} \mathrm{Gr}_l(W_n A^{ij})$, where

$$\mathrm{Gr}_l(W_n A^{ij}) = \begin{cases} 0 & \text{if } l < 2n-2-j, \\ \bigoplus_{k=0}^{j} \bigoplus_{m=0}^{l-2n+2+j} \mathrm{Gr}_{k+m+1,2j-k+l-2n-m+3} W_n \tilde{\tilde{\omega}}_Y^{i+j+2} & \text{if } l \geq 2n-2-j. \end{cases}$$

For $l = 2n-2+h$, with $h > 0$, we claim that ν induces an injection $\mathrm{Gr}_l(W_n A^{ij}) \hookrightarrow \mathrm{Gr}_{l-1}(W_n A^{ij})$. This can be verified through a standard combinatorial argument. We have

$$\mathrm{Gr}_l(W_n A^{ij}) = \bigoplus_{k=0}^{j} \bigoplus_{m=0}^{h+j} \mathrm{Gr}_{(k+m)+1,2j+h+1-(k+m)} W_n \tilde{\tilde{\omega}}_Y^{i+j+2}$$

and

$$\mathrm{Gr}_{l-1}(W_n A^{ij}) = \bigoplus_{k=0}^{j+1} \bigoplus_{m=0}^{h+j-1} \mathrm{Gr}_{(k+m)+1,2j+h+1-(k+m)} W_n \tilde{\tilde{\omega}}_Y^{i+j+2}.$$

The map ν sends the term corresponding to a pair (k,m) to the direct sum of terms corresponding to (k,m) and to $(k+1,m-1)$. Therefore, it is easy to see that ν restricted to the direct sum of terms for which $k+m$ is constant is injective, so ν is injective. Moreover, we see that ν^h induces an isomorphism $\mathrm{Gr}_{2n-2+h}(W_n A^{ij}) \simeq \mathrm{Gr}_{2n-2-h}(W_n A^{i-h,j+h})$, since the terms on the right-hand side are of the form

$$\bigoplus_{k=0}^{j} \bigoplus_{m=0}^{h+j} \mathrm{Gr}_{(k+m)+1,2j+h+1-(k+m)} W_n \tilde{\tilde{\omega}}_Y^{i+j+2}$$

and the terms on the left-hand side are of the form

$$\bigoplus_{m=0}^{j} \bigoplus_{k=0}^{h+j} \mathrm{Gr}_{(k+m)+1,2j+h+1-(k+m)} W_n \tilde{\tilde{\omega}}_Y^{i+j+2},$$

so on either side we have the same number of terms corresponding to $k+m$. Since the filtration $P_l(W_n A^{\bullet\bullet})$ satisfies the two properties above, it must be the monodromy filtration of ν.

Note that the differentials d'' on $\mathrm{Gr}_l(W_\bullet A^{\bullet\bullet})$ are always 0. Using the isomorphisms in Corollary 3.16 we can rewrite

$$\begin{aligned}\mathrm{Gr}_{2n-2+h}&(W_\bullet A^\bullet)\\ &\simeq \bigoplus_{j\geq 0, j\geq -h} \bigoplus_{k=0}^{j} \bigoplus_{m=0}^{j+h} (W_\bullet \Omega^\bullet_{Y^{(k+m+1,2j+h+1-(k+m))}})[-2j-h](-j-h).\end{aligned}$$

This leads to the main geometric result of the paper. Recall from Section 3.1 that Y/k is the special fiber of $X/\mathbb{O}_K$, which is Zariski-locally étale over a product of strictly semistable schemes. Recall also that Y is globally the union of certain proper, smooth $(2n-2)$-dimensional subschemes $Y_{i,j}$ with $i = 1, 2$, $j = 1, \dots, n$. Taking disjoint unions of intersections of these subschemes gives rise to schemes $Y^{(l_1,l_2)}/k$ for $1 \leq l_1, l_2 \leq n$, which cover closed strata in Y. Each $Y^{(l_1,l_2)}$ is proper, smooth and has dimension $2n - l_1 - l_2$.

Theorem 4.6. *There is a spectral sequence*

$$\begin{aligned}E_1^{-h,i+h} = \bigoplus_{j\geq 0, j\geq -h} \bigoplus_{k=0}^{j} \bigoplus_{m=0}^{j+h} H^{i-2j-h}_{\mathrm{cris}}(Y^{(k+m+1,2j+h+1-(k+m))}/W)(-j-h)\\ \Rightarrow H^i_{\mathrm{cris}}(Y/W).\end{aligned}$$

Remark 4.7. Note that the schemes $Y^{(l_1,l_2)}$ are proper and smooth so the $E^{-h,i+h}$ terms of the spectral sequence are strictly pure of weight $i+h$. If the above spectral sequence degenerates at the first page, then $H^i_{\mathrm{cris}}(Y/W)$ is pure of weight i.

5. Proof of the main theorem

In this section we prove the main theorem. By Corollary 2.3, its proof reduces to the following proposition.

Proposition 5.1. *Let $\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}$ be the universal abelian variety over $X^{m_\xi}_{U_{\mathrm{Iw}}}$. The direct limit of log crystalline cohomologies*

$$\varinjlim_{U_{\mathrm{Iw}}} a_\xi\big(H^{2n-2+m_\xi}_{\mathrm{cris}}(\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}} \times_{\mathbb{O}_K} k/W) \otimes_W \overline{\mathbb{Q}}_l(t_\xi)\big)[\Pi^{1,\mathfrak{S}}]$$

is pure of a certain weight.

Proof. Recall that we have chosen

$$U_{\mathrm{Iw}} = U^l \times U_l^{\mathfrak{p}_1,\mathfrak{p}_2}(m) \times \mathrm{Iw}_{n,\mathfrak{p}_1} \times \mathrm{Iw}_{n,\mathfrak{p}_2} \subset G(\mathbb{A}^\infty).$$

Pick m large enough such that $(\pi_l)^{U_l^{\mathfrak{p}_1,\mathfrak{p}_2}(m)\times \mathrm{Iw}_{n,\mathfrak{p}_1}\times \mathrm{Iw}_{n,\mathfrak{p}_2}}$ is nonzero, where $\pi_l \in \mathrm{Irr}_l(G(\mathbb{Q}_l))$ is such that $BC(\pi_l) = \iota_l^{-1}\Pi_l$. The results of Sections 3 and 4 apply to $\mathcal{A}^{m_\xi}_{U_{\mathrm{Iw}}}$. We have a stratification of its special fiber by closed Newton polygon strata $\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}},S,T}$ with $S, T \subseteq \{1, \dots, n\}$ nonempty. For brevity, let $K_1 := k+m+1$ and $K_2 := 2j+h+1-(k+m)$. By Theorem 4.6, we have a spectral sequence

$$E_1^{-h,i+h} = \bigoplus_{\substack{j\geq 0\\ j\geq -h}} \bigoplus_{k=0}^{j} \bigoplus_{m=0}^{j+h} \bigoplus_{\substack{\#S=K_1\\ \#T=K_2}} H^{i-2j-h}_{\mathrm{cris}}\big((\mathscr{A}^{m_\xi,}_{U_{\mathrm{Iw}},S,T}/W)(-j-h)\big) \Rightarrow H^i_{\mathrm{cris}}(\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}}} \times_{\mathbb{O}_K} k/W).$$

We replace the cohomology degree i by $i+m_\xi$, tensor with $\overline{\mathbb{Q}}_l(t_\xi)$, apply a_ξ (which is obtained from a linear combination of étale morphisms); passing to a direct limit over U^l and taking the $\Pi^{1,\mathfrak{S}}$-isotypic components we get a spectral sequence

$$\begin{aligned} E_1^{-h,i+h} &= \bigoplus_{\substack{j\geq 0\\ j\geq -h}} \bigoplus_{k=0}^{j} \bigoplus_{m=0}^{j+h} \bigoplus_{\substack{\#S=K_1\\ \#T=K_1}} \varinjlim_{U^l}\Big(a_\xi H^{i+m_\xi-2j-h}_{\mathrm{cris}}\big((\mathscr{A}^{m_\xi,}_{U_{\mathrm{Iw}},S,T}/W)(-j-h) \\ &\qquad\qquad \otimes_{W,\tau_0}\overline{\mathbb{Q}}_l(t_\xi)\big)[\Pi^{1,\mathfrak{S}}]\Big) \\ &\Rightarrow \varinjlim_{U^l}\big(a_\xi H^{i+m_\xi}_{\mathrm{cris}}(\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}}} \times_{\mathbb{O}_K} k/W) \otimes_{W,\tau_0} \overline{\mathbb{Q}}_l(t_\xi)\big)[\Pi^{1,\mathfrak{S}}]. \end{aligned}$$

For any compact open subgroup $U^l \subset G(\mathbb{A}^{\infty,l})$ and any prime $p \neq l$ with isomorphism $\iota_p : \overline{\mathbb{Q}}_p \xrightarrow{\sim} \mathbb{C}$, set $\xi' := (\iota_p)^{-1}\iota_l\xi$ and $\Pi' := (\iota_p)^{-1}\Pi^1$.

We have

$$\begin{aligned} &\dim_{\overline{\mathbb{Q}}_l}\big(\varinjlim_{U^l} a_\xi H^{i+m_\xi-2j-h}_{\mathrm{cris}}\big((\mathscr{A}^{m_\xi,}_{U_{\mathrm{Iw}},S,T}/W)(-j-h)\otimes_{W,\tau_0}\overline{\mathbb{Q}}_l\big)\big)[\Pi^{1,\mathfrak{S}}]^{U^l} \\ &\qquad = \dim_{\overline{\mathbb{Q}}_p}\big(\varinjlim_{U^l} a_{\xi'} H^{i+m_{\xi'}-2j-h}(\mathscr{A}^{m_\xi}_{U_{\mathrm{Iw}},S,T}, \overline{\mathbb{Q}}_p)\big)[(\Pi')^{\mathfrak{S}}]^{U^l} \\ &\qquad = \dim_{\overline{\mathbb{Q}}_p}(\varinjlim_{U^l} H^{i-2j-h}\big(X_{U_{\mathrm{Iw}},S,T}, \mathscr{L}_{\xi'}\big))[(\Pi')^{\mathfrak{S}}]^{U^l}. \end{aligned}$$

The first equality is a consequence of the main theorem of [Gillet and Messing 1987] and of Theorem 2(2) of [Katz and Messing 1974]. The former proves that crystalline cohomology is a Weil cohomology theory in the strong sense. The latter is the statement that the characteristic polynomial on $H^i(X)$ of an integrally algebraic cycle on $X \times X$ of codimension n, for a projective smooth variety X/k of dimension n, is independent of the Weil cohomology theory H.

The dimension in the third row is equal to 0 unless $i = 2n-2$ by Proposition 5.10 of [Caraiani 2012]. Therefore, $E_1^{-h,i+h} = 0$ unless $i = 2n-2$, so the E_1 page of the spectral sequence is concentrated on a diagonal. The spectral sequence degenerates at the E_1 page and the term corresponding to $E_1^{h,2n-2+h}$ is strictly pure of weight $h+2n-2+m_\xi-2t_\xi$, which shows that the abutment is pure. $\square$

Acknowledgements

I am very grateful to my advisor, Richard Taylor, for suggesting this problem and for his constant encouragement and advice. I would like to thank Kazuya Kato, Mark Kisin, Jay Pottharst and Claus Sorensen for many useful conversations and comments. I would also like to thank Luc Illusie for his detailed comments on an earlier draft of this paper and in particular for pointing me to [Nakkajima 2005].

References

[Barnet-Lamb et al. 2011] T. Barnet-Lamb, T. Gee, D. Geraghty, and R. Taylor, "Local-global compatibility for $l = p$, II", preprint, 2011, Available at https://www2.bc.edu/david-geraghty/files/lgcII.pdf. To appear in *Ann. scient. de l'E.N.S.*

[Barnet-Lamb et al. 2012] T. Barnet-Lamb, T. Gee, D. Geraghty, and R. Taylor, "Local-global compatibility for $l = p$, I", *Ann. Fac. Sci. Toulouse Math.* (6) **21**:1 (2012), 57–92. MR 2954105 Zbl 1259.11057

[Caraiani 2012] A. Caraiani, "Local-global compatibility and the action of monodromy on nearby cycles", *Duke Math. J.* **161**:12 (2012), 2311–2413. MR 2972460 Zbl 06095601

[Chenevier and Harris 2013] G. Chenevier and M. Harris, "Construction of automorphic Galois representations, II", *Cambridge Math. J.* **1**:1 (2013), 53–73. Zbl 06266535

[Deligne 1970] P. Deligne, *Équations différentielles à points singuliers réguliers*, Lecture Notes in Mathematics **163**, Springer, Berlin, 1970. MR 54 #5232 Zbl 0244.14004

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

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

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

[Harris and Taylor 2001] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies **151**, Princeton University Press, 2001. MR 2002m:11050 Zbl 1036.11027

[Hyodo 1991] O. Hyodo, "On the de Rham–Witt complex attached to a semi-stable family", *Compositio Math.* **78**:3 (1991), 241–260. MR 93c:14022 Zbl 0742.14015

[Hyodo and Kato 1994] O. Hyodo and K. Kato, "Semi-stable reduction and crystalline cohomology with logarithmic poles", pp. 221–268 in *Périodes p-adiques* (Bures-sur-Yvette, 1988), Astérisque **223**, Société Mathématique de France, Paris, 1994. MR 95k:14034 Zbl 0852.14004

[Illusie 1979] L. Illusie, "Complexe de de Rham–Witt et cohomologie cristalline", *Ann. Sci. École Norm. Sup.* (4) **12**:4 (1979), 501–661. MR 82d:14013 Zbl 0436.14007

[Illusie 1983] L. Illusie, "Finiteness, duality, and Künneth theorems in the cohomology of the de Rham–Witt complex", pp. 20–72 in *Algebraic geometry* (Tokyo/Kyoto, 1982), edited by M. Raynaud and T. Shioda, Lecture Notes in Math. **1016**, Springer, Berlin, 1983. MR 85m:14033 Zbl 0538.14013

[Illusie and Raynaud 1983] L. Illusie and M. Raynaud, "Les suites spectrales associées au complexe de de Rham–Witt", *Inst. Hautes Études Sci. Publ. Math.* 57 (1983), 73–212. MR 85e:14026 Zbl 0538.14012

[Kato 1989] K. Kato, "Logarithmic structures of Fontaine–Illusie", pp. 191–224 in *Algebraic analysis, geometry, and number theory* (Baltimore, MD, 1988), edited by J.-I. Igusa, Johns Hopkins Univ. Press, Baltimore, MD, 1989. MR 99b:14020 Zbl 0776.14004

[Kato 1994a] K. Kato, "Semi-stable reduction and p-adic étale cohomology", pp. 269–293 in *Périodes p-adiques* (Bures-sur-Yvette, 1988), Astérisque **223**, Société Mathématique de France, Paris, 1994. MR 95i:14020 Zbl 0847.14009

[Kato 1994b] K. Kato, "Toric singularities", *Amer. J. Math.* **116**:5 (1994), 1073–1099. MR 95g:14056 Zbl 0832.14002

[Katz and Messing 1974] N. M. Katz and W. Messing, "Some consequences of the Riemann hypothesis for varieties over finite fields", *Invent. Math.* **23** (1974), 73–77. MR 48 #11117 Zbl 0275.14011

[Mantovan 2005] E. Mantovan, "On the cohomology of certain PEL-type Shimura varieties", *Duke Math. J.* **129**:3 (2005), 573–610. MR 2006g:11122 Zbl 1112.11033

[Mokrane 1993] A. Mokrane, "La suite spectrale des poids en cohomologie de Hyodo–Kato", *Duke Math. J.* **72**:2 (1993), 301–337. MR 95a:14022 Zbl 0834.14010

[Nakkajima 2005] Y. Nakkajima, "p-adic weight spectral sequences of log varieties", *J. Math. Sci. Univ. Tokyo* **12**:4 (2005), 513–661. MR 2007e:14030 Zbl 1108.14015

[Nizioł 2006] W. Nizioł, "Toric singularities: Log-blow-ups and global resolutions", *J. Algebraic Geom.* **15**:1 (2006), 1–29. MR 2006i:14015 Zbl 1100.14011

[Nizioł 2008] W. Nizioł, "Semistable conjecture via K-theory", *Duke Math. J.* **141**:1 (2008), 151–178. MR 2008m:14035 Zbl 1157.14009

[Shin 2011] S. W. Shin, "Galois representations arising from some compact Shimura varieties", *Ann. of Math.* (2) **173**:3 (2011), 1645–1741. MR 2800722 Zbl 1269.11053

[Taylor and Yoshida 2007] R. Taylor and T. Yoshida, "Compatibility of local and global Langlands correspondences", *J. Amer. Math. Soc.* **20**:2 (2007), 467–493. MR 2007k:11193 Zbl 1210.11118

[Tsuji 1997] T. Tsuji, "Saturated morphisms of logarithmic schemes", preprint, 1997.

[Tsuji 1999] T. Tsuji, "p-adic étale cohomology and crystalline cohomology in the semi-stable reduction case", *Invent. Math.* **137**:2 (1999), 233–411. MR 2000m:14024 Zbl 0945.14008

Communicated by Brian Conrad
Received 2013-04-27 Revised 2014-03-31 Accepted 2014-05-18

caraiani@princeton.edu *Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544, United States*

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1647

Finite generation of the cohomology of some skew group algebras

Van C. Nguyen and Sarah Witherspoon

We prove that some skew group algebras have Noetherian cohomology rings, a property inherited from their component parts. The proof is an adaptation of Evens' proof of finite generation of group cohomology. We apply the result to a series of examples of finite-dimensional Hopf algebras in positive characteristic.

1. Introduction

The cohomology ring of a Hopf algebra encodes potentially useful information about its structure and representations. It is always graded commutative (see, for example, [Suarez-Alvarez 2004]). For many classes of finite-dimensional Hopf algebras, it is also known to be finitely generated: for example, cocommutative Hopf algebras [Friedlander and Suslin 1997], small quantum groups [Ginzburg and Kumar 1993], and small quantum function algebras [Gordon 2000]. Etingof and Ostrik [2004] conjectured that it is always finitely generated, as a special case of a conjecture about finite tensor categories. Snashall and Solberg [2004] made an analogous conjecture for Hochschild cohomology of finite-dimensional algebras that was seen to be false when Xu [2008] constructed a counterexample. In contrast, there is neither a counterexample nor a proof of the Hopf algebra conjecture. Each finite generation result so far has used, in crucial ways, known structure of a particular class of Hopf algebras. Further progress will require new ideas.

In this article, we present one technique for handling some types of algebras inductively. Many (Hopf) algebras of interest are skew group algebras (that is, smash products with group algebras). Under some conditions on a skew group algebra, we show that its cohomology is Noetherian if the same is true of the underlying algebra on which the group acts.

This material is based upon work done while Nguyen was a Texas A&M graduate student. It was supported by the National Science Foundation under grant No. 0932078000 while both Nguyen and Witherspoon were in residence at the Mathematical Sciences Research Institute (MSRI) in Berkeley, California, during the semester of Spring 2013. Both authors were also supported by NSF grant DMS-1101399.

MSC2010: primary 16E40; secondary 16T05.

Keywords: cohomology, Hopf algebras, skew group algebras.

Specifically, if A is a finite-dimensional augmented algebra over a field k, with an action of a finite group G by automorphisms, there is a spectral sequence relating the cohomology of the smash product $A\#kG$ (definition in Section 2) as an augmented algebra to that of each of A and G. (It is essentially the Lyndon–Hochschild–Serre spectral sequence.) This allows us to use the framework of Evens' classic proof [1961] of finite generation of group cohomology to prove that the cohomology rings of some smash products are Noetherian (Theorem 3.1). In order to do this, we need a particularly nice set of permanent cycles in the cohomology of A. In the finite group case, these cycles exist due to an application of Evens' norm map. In our setting, there may be no such norm map, and we instead hypothesize existence of these permanent cycles.

We focus on a class of examples (in Section 5) found by Cibils, Lauve, and the second author [Cibils et al. 2009] that satisfy our hypotheses. We prove finite generation of the cohomology of these noncommutative, noncocommutative Hopf algebras in positive characteristic. While our main theorem is tailored to suit these examples, we state and prove it in the abstract setting, in order to add one more tool to the collection of techniques available for proving finite generation. Our restrictive hypotheses serve to highlight the difficulty in adapting methods designed for the finite group setting, where serendipity reigns.

2. Definitions and notation

Throughout this article, let k be a field. All algebras will be associative algebras over k, and all modules will be left modules, finite-dimensional over k. Let $\otimes = \otimes_k$.

Let G be a finite group acting on a finite-dimensional augmented k-algebra A by automorphisms. Let $A\#kG$ be the resulting *smash product* (or *skew group algebra*), that is, $A\otimes kG$ as a vector space, with multiplication $(a\otimes g)(b\otimes h) = a({}^g b)\otimes gh$, for all $a, b\in A$ and $g, h\in G$. (For simplicity, we will drop tensor symbols in this notation from now on.) We assume the action of G preserves the augmentation of A, so that $A\#kG$ is also augmented with augmentation map $\varepsilon_{A\#kG} : A\#kG \to k$ defined by $\varepsilon_{A\#kG}(ag) = \varepsilon_A(a)$, for all $a\in A$, $g\in G$.

We use the symbol k also to denote the one-dimensional A-module (respectively, $A\#kG$-module) on which A (respectively, $A\#kG$) acts via its augmentation. Let

$$\mathrm{H}^*(A, k) := \mathrm{Ext}^*_A(k, k) \quad \text{and} \quad \mathrm{H}^*(A\#kG, k) := \mathrm{Ext}^*_{A\#kG}(k, k).$$

Both are algebras under Yoneda composition. The embedding of A into $A\#kG$ as a subalgebra induces a *restriction map*

$$\mathrm{res}_{A\#kG,A} : \mathrm{H}^*(A\#kG, k) \to \mathrm{H}^*(A, k)$$

on cohomology. There is an action of G on $\mathrm{H}^*(A, k)$ that may be defined for example via the diagonal action of G on the components of the bar resolution for A.

There is a similar action of G on $\mathrm{H}^*(A\,\#\,kG, k)$ that is trivial since it comes from inner automorphisms on $A\,\#\,kG$.

3. Finite generation of cohomology

In this section, we prove our main theorem that under certain hypotheses, the cohomology ring $\mathrm{H}^*(A\,\#\,kG, k)$ of $A\,\#\,kG$ is Noetherian:

Theorem 3.1. *Let G be a finite group acting on a finite-dimensional augmented algebra A, preserving the augmentation map. Assume that* $\mathrm{Im}(\mathrm{res}_{A\#kG,A})$ *contains a polynomial subalgebra over which* $\mathrm{H}^*(A, k)$ *is Noetherian and free as a module, with a free basis whose k-linear span is a kG-submodule of* $\mathrm{H}^*(A, k)$. *Then,* $\mathrm{H}^*(A\,\#\,kG, k)$ *is Noetherian.*

Remarks 3.2. (a) The hypothesis that $\mathrm{Im}(\mathrm{res}_{A\#kG,A})$ contains a polynomial subalgebra over which $\mathrm{H}^*(A, k)$ is Noetherian, together with the left module version of [Goodearl and Warfield 2004, Corollary 1.5], implies that $\mathrm{H}^*(A, k)$ is (left) Noetherian.

(b) We did not specify the characteristic of the base field k in the theorem. If the characteristic of k does not divide the order of G, then kG is semisimple and its cohomology is trivial except in degree zero. In this case, $\mathrm{H}^*(A\#kG, k) \cong \mathrm{H}^*(A, k)^G$, the invariant ring under the action of G. Here, one can use invariant ring theory in the noncommutative setting to show that the conclusion of the theorem holds. (See, for example, [Montgomery 1993, Corollary 4.3.5].) For the proof of Theorem 3.1, we assume the characteristic of k divides the order of G.

Proof. We use the Lyndon–Hochschild–Serre spectral sequence (see, for example, [Barnes 1985, Chapter VI] in a very general setting):

$$E_2^{p,q} = E_2^{p,q}(k) = \mathrm{H}^p(G, \mathrm{H}^q(A, k)) \Longrightarrow \mathrm{H}^{p+q}(A\,\#\,kG, k).$$

Let $E_r(k)$ denote the resulting r-th page, and note that for each q, $\mathrm{H}^q(A, k)$ is a finite-dimensional k-vector space.

Note that $E_\infty^{0,*}$ is a submodule of $E_2^{0,*}$, since no $d_r : E_r^{p,q} \to E_r^{p+r,q-r+1}$ ends on the vertical edge. It follows that the restriction map $\mathrm{H}^*(A\,\#\,kG, k) \to E_2^{0,*}(k)$ is part of the following commuting diagram:

$$\begin{tikzcd}
\mathrm{H}^*(A\,\#\,kG, k) \arrow[r, "\mathrm{res}_{A\#kG,A}"] \arrow[d] & \mathrm{H}^0(G, \mathrm{H}^*(A, k)) = \mathrm{H}^*(A, k)^G \arrow[d, equal] \\
E_\infty^{0,*}(k) \arrow[r, hook] & E_2^{0,*}(k)
\end{tikzcd}$$

We can identify $E_\infty^{0,*}$ with the image of the restriction map in $E_2^{0,*}$.

Let $T = k[\chi_1, \ldots, \chi_m]$ denote the polynomial subalgebra of $\mathrm{Im}(\mathrm{res}_{A\#kG,A})$ hypothesized in the statement of the theorem. The action of G on $\mathrm{H}^*(A, k)$ restricts to the trivial action on T since it is a subalgebra of $\mathrm{Im}(\mathrm{res}_{A\#kG,A})$. Therefore, by the universal coefficients theorem, $\mathrm{H}^*(G, T) \cong \mathrm{H}^*(G, k) \otimes T$, an isomorphism of graded algebras.

Let $S := \mathrm{H}^*(G, k) = E_2^{*,0}(k)$. Let R be the subring of $E_2(k)$ generated by S and T. By the above observations, $R \cong S[\chi_1, \ldots, \chi_m]$, a polynomial ring over S in m indeterminates (that we also denote by $\chi_1, \ldots, \chi_m$ for convenience). Since d_2 vanishes on the horizontal edge, $R \subseteq \mathrm{Ker}(d_2)$. So R projects onto a subring of $E_3(k) = \mathrm{H}(E_2(k), d_2)$. Similarly, R projects onto a subring of $E_r(k)$ for every $r > 0$ including ∞. Therefore, we may consider $E_r(k)$ to be a module over R, for every $r > 0$ including ∞.

Claim 1. *$E_2(k)$ is a Noetherian module over R.*

Proof of Claim 1. By hypothesis, there are (homogeneous) elements $\rho_1, \ldots, \rho_t \in \mathrm{H}^*(A, k)$ that form a free basis of $\mathrm{H}^*(A, k)$ as a T-module, and for which

$$V := \mathrm{Span}_k\{\rho_1, \ldots, \rho_t\}$$

is a kG-submodule of $\mathrm{H}^*(A, k)$. Let

$$L := \mathrm{H}^*(G, V).$$

Note that L contains a copy of $S = \mathrm{H}^*(G, k)$ as V must include an element in degree 0, that is, in $\mathrm{H}^0(A, k) \cong k$, which has trivial G-action. By hypothesis, $\mathrm{H}^*(A, k) = k[\chi_1, \ldots, \chi_m] \cdot V$, and so

$$E_2(k) = \mathrm{H}^*(G,\ k[\chi_1, \ldots, \chi_m] \cdot V).$$

Further, $k[\chi_1, \ldots, \chi_m]$ has trivial G-action and the module $\mathrm{H}^*(A, k)$ for this polynomial ring is free with free basis $\rho_1, \ldots, \rho_t$. It follows that, as a kG-module,

$$k[\chi_1, \ldots, \chi_m] \cdot V \cong \bigoplus_{i_1,\ldots,i_m \geq 0} \chi_1^{i_1} \cdots \chi_m^{i_m} \cdot V \cong \bigoplus_{i_1,\ldots,i_m \geq 0} V,$$

a direct sum of copies of the same kG-module, V. Therefore by the universal coefficients theorem, $E_2(k)$ is the image of

$$\mathrm{H}^0(G, k[\chi_1, \ldots, \chi_m]) \otimes \mathrm{H}^*(G, V) \cong k[\chi_1, \ldots, \chi_m] \otimes L,$$

under cup product. We thus identify $E_2^{*,*}(k)$ with $S[\chi_1, \ldots, \chi_m] \otimes_S L$.

Since G is a finite group and V is a finite-dimensional vector space over k, $L = \mathrm{H}^*(G, V)$ is Noetherian over $S = \mathrm{H}^*(G, k)$ [Evens 1961]. By the Hilbert basis theorem for graded commutative rings (see, for example, [Goodearl and Warfield 2004, Theorem 2.6]), $S[\chi_1, \ldots, \chi_m] \otimes_S L$ is Noetherian over $R = S[\chi_1, \ldots, \chi_m]$.

Therefore, $E_2^{*,*}(k)$ is Noetherian over R. We have proven Claim 1.

Claim 2. *The spectral sequence stops*; *i.e.*, $E_r = E_\infty$ *for some* $r < \infty$.

Proof of Claim 2. Let Z_i be the space of i-cocycles and B_i be the space of i-coboundaries in $E_i = E_i(k)$. Recall that $E_1 = Z_1$ and $E_2 = Z_2/B_2$. Consider the "pull back" B_r in E_2 of $d_r(E_r)$ as follows.

Each element of E_2 on which d_2 vanishes determines an element of E_3. Suppose d_3 vanishes on that element, so that it in turn determines an element of E_4. Continue placing such restrictions until we determine an element of E_r, and suppose that element is in the image of d_r. We define:

$$B_r := \{\tau \in E_2 : \tau \in \mathrm{Ker}(d_i) \text{ for } 2 \le i \le r-1 \text{ and } \tau \in \mathrm{Im}(d_r)\}.$$

Note that B_r is an R-submodule of E_2 since d_j is a derivation for all j, $2 \le j \le r$, and the image in each E_j of R consists of universal cycles. Moreover, $B_r \subseteq B_{r+1}$ so we obtain an ascending chain of R-submodules of E_2:

$$0 = B_1 \subseteq B_2 \subseteq \cdots$$

Since E_2 is Noetherian over R by Claim 1, this chain must stabilize by the ascending chain condition. Thus there exists some r_0 finite such that $B_{r_0} = B_{r_0+1} = B_{r_0+2} = \cdots$, and so $d_r = 0$ for all $r > r_0$. This implies $E_r = E_\infty$ for $r > r_0$, proving Claim 2.

We can put this together to finish the proof of the theorem: Each Z_r, B_r is a submodule of E_2 over $R = S[\chi_1, \dots, \chi_m]$. Thus, each E_r, which is a submodule of a quotient module of E_{r-1}, is Noetherian over R by Claim 1 and induction on r. By Claim 2, E_∞ is Noetherian over R, and so by [Goodearl and Warfield 2004, Corollary 1.5] it is a Noetherian ring.

Now, $\mathrm{H}^*(A \# kG, k)$ has a filtration whose filtered quotients are

$$E_\infty^{p,q}(k) \cong \frac{F^p\,\mathrm{H}^{p+q}(A \# kG, k)}{F^{p+1}\,\mathrm{H}^{p+q}(A \# kG, k)}.$$

Suppose that $\mathrm{H}^*(A\#kG, k)$ is not Noetherian and let $T_1 \subseteq T_2 \subseteq \cdots \subseteq \mathrm{H}^*(A\#kG, k)$ be an infinite ascending chain of ideals of $\mathrm{H}^*(A \# kG, k)$. Let

$$F^pT_i := T_i \cap F^p\,\mathrm{H}^*(A \# kG, k)$$

and

$$U_i := \bigoplus_{p \ge 0} F^pT_i/F^{p+1}T_i \subseteq E_\infty(k).$$

If $x \in T_{i+1} \setminus T_i$, then for some p, $x \in F^pT_{i+1}$ but $x \notin F^pT_i$ and $x \notin F^{p+1}T_{i+1}$, so $x + F^{p+1}T_{i+1}$ is not in the image of the inclusion

$$F^pT_i/F^{p+1}T_i \hookrightarrow F^pT_{i+1}/F^{p+1}T_{i+1},$$

that is, $x \in U_{i+1} \setminus U_i$. So U_{i+1} properly contains U_i, for all i. Therefore, we have an infinite ascending chain of ideals of $E_\infty(k)$:

$$U_1 \subsetneqq U_2 \subsetneqq \cdots$$

This contradicts the result that $E_\infty(k)$ is Noetherian. Therefore, $\mathrm{H}^*(A \# kG, k)$ is Noetherian. $\square$

Remark 3.3. Theorem 3.1 parallels the main step in Evens' proof of finite generation of group cohomology: Let H be a finite p-group (where k has characteristic p), $A = kZ$ is the group algebra of a central subgroup Z of H of order p, and $G = H/Z$. (In case Z is complemented in H, we obtain $kH \cong A \# kG$, whereas more generally, kH is a crossed product of A with G.) In this case, Evens' norm map is applied to show that $\mathrm{Im}(\mathrm{res}_{kH,kZ})$ contains a polynomial subalgebra $k[\zeta]$ (in one indeterminate). One observes that $\mathrm{H}^*(kZ, k)$ is a free module over $k[\zeta]$, and that the k-linear span of any free basis is a kG-submodule. This special case is somewhat simpler than our more general context as it uses a polynomial ring in one indeterminate.

We are particularly interested in those actions of finite groups G on algebras A for which $A \# kG$ is a Hopf algebra. We turn to a class of such examples in the remainder of the paper.

4. Examples: Nichols algebras in positive characteristic

In this section, we first recall the Nichols algebras A from [Cibils et al. 2009, Corollary 3.14] and the corresponding Hopf algebras $A \# kG$ from the same paper. We will prove that these Hopf algebras have finitely generated cohomology. This will follow from Theorem 3.1 and explicit calculation using Anick's resolution [1986]. In this section we explain these calculations for A, and in the next we complete the proof of finite generation of cohomology of $A \# kG$. The results of this section were anticipated by Ø. Solberg (personal communication, 2012) as a consequence of computer calculations (for small p) that gave the graded vector space structure and generators of cohomology.

In the remainder of the paper, k will be a field of characteristic $p > 2$. (The case $p = 2$ is included in [Cibils et al. 2009], but is different, and we will not consider that case here.) Let A be the augmented k-algebra generated by a, b, with relations

$$a^p = 0, \quad b^p = 0, \quad ba = ab + \tfrac{1}{2}a^2,$$

and augmentation $\varepsilon : A \to k$ given by $\varepsilon(a) = \varepsilon(b) = 0$. Let G be a cyclic group of order p with generator g, acting on A by

$$g(a) = a, \quad g(b) = b - a.$$

Then $A \# kG$ is a Hopf algebra with comultiplication given by

$$\Delta(g) = g \otimes g, \quad \Delta(a) = a \otimes 1 + g \otimes a, \quad \Delta(b) = b \otimes 1 + g \otimes b.$$

It is useful to consider A as a quotient of a larger algebra. Let

$$B := k\langle a, b\rangle / (ba - ab - \tfrac{1}{2}a^2), \tag{4-1}$$

so that $A \cong B/(a^p, b^p)$. We will show that B is a PBW algebra in the sense of [Bueso et al. 2003] or [Shroff 2013, Section 2], although we will not need this fact for our cohomology calculations.

Choose the lexicographic order on $\mathbb{N}^2$ for which $(0, 1) < (1, 0)$, and assign $\deg(a) = (0, 1)$, $\deg(b) = (1, 0)$. Then $ba - ab - \frac{1}{2}a^2$ is a Gröbner basis for the ideal of the free algebra $k\langle a, b\rangle$ that it generates. It follows that $\{a^i b^j \mid i, j \geq 0\}$ is a vector space basis of B. The relation $ab = ba - \frac{1}{2}a^2$ satisfies the required condition in the definition of a PBW algebra since $\deg(a^2) < \deg(ab)$, so B is a PBW algebra. Moreover, B is a Koszul algebra by Theorem 5.3 in [Priddy 1970].

Applying [Cibils et al. 2009, (3.9)], one finds that the elements a^p, b^p are in the center of B. We may thus apply Theorem 4.3 of [Shroff 2013] to the Nichols algebra A to conclude that the cohomology ring $\mathrm{H}^*(A, k)$ of A is Noetherian.

We will need some details about this cohomology of A for the next section. For this, we will construct Anick's resolution [1986] for A, and show that it is minimal. We use the combinatorial description of the resolution given by Cojocaru and Ufnarovski [1997], however we index differently, and use left modules instead of right. This is a free resolution of the trivial A-module k, of the form

$$\cdots \longrightarrow A \otimes kC_2 \longrightarrow A \otimes kC_1 \longrightarrow A \otimes kC_0 \longrightarrow k \to 0,$$

for (finite) sets C_n, where kC_n denotes the vector space with basis C_n. Let $C_0 := \{1\}$ and $C_1 := \{a, b\}$. Then $C_2 := \{a^p,\ b^p,\ ba\}$ is the set of "tips" or "obstructions." To define C_n in general, consider the graph

$$\begin{tikzcd}
& 1 \arrow[dl] \arrow[dr] & \\
a \arrow[d, bend left] & & b \arrow[ll] \arrow[d, bend left] \\
a^{p-1} \arrow[u, bend left] & & b^{p-1} \arrow[u, bend left] \arrow[ull]
\end{tikzcd}$$

The elements of C_n correspond to paths of length n that start at 1. We label such paths with the product of all elements through which the path passes (including the endpoint). In this way we obtain

$$C_3 = \{a^{p+1},\ b^{p+1}, b^p a,\ ba^p\},$$
$$C_4 = \{a^{2p},\ b^{2p}, b^{p+1}a,\ b^p a^p,\ ba^{p+1}\},$$

and in general

$$C_{2m-1} = \{b^{kp}a^{(m-1-k)p+1},\ b^{kp+1}a^{(m-1-k)p} \mid k = 0, 1, \dots, m-1\},$$
$$C_{2m} = \{b^{mp},\ b^{kp}a^{(m-k)p},\ b^{kp+1}a^{(m-1-k)p+1} \mid k = 0, 1, \dots, m-1\}.$$

For qualitative understanding of the differentials, give each of the generators a, b of A the degree 1. We claim that the differentials preserve degree, where the graded module structure of a tensor product $A \otimes kC_i$ is given by $\deg(a \otimes x) = \deg(a) + \deg(x)$ if a, x are homogeneous. This claim results from the recursive definition of the differential d in each homological degree: By construction, d applied to elements of $A \otimes kC_1$ is multiplication, and to $A \otimes kC_2$ takes each tip to the Gröbner basis element to which it corresponds, suitably expressed as an element of $A \otimes kC_1$. The remaining differentials are defined iteratively, via splitting maps in each homological degree that are also defined iteratively. Since the relations are homogeneous and differentials in low homological degrees preserve degrees of elements, the splitting maps and differentials in higher degrees may be chosen to have the same property.

Now note that C_{2m-1} consists of elements of degree $(m-1)p+1$, and C_{2m} consists of elements of degrees mp and $(m-1)p+2$. Therefore elements of C_n and of C_{n-1} never have the same degree. As a consequence the differential takes elements of C_n to elements of $A_+ \otimes C_{n-1}$, where A_+ denotes all elements of A of positive degree (and these are in the kernel of the augmentation map ε). When applying the functor $\mathrm{Hom}_A(-, k)$, then, the induced differentials all become 0. Therefore in this case, Anick's resolution is minimal, and for each n, the dimension of $\mathrm{H}^n(A, k)$ is $n+1$.

5. Examples: pointed Hopf algebras in positive characteristic

We wish to apply Theorem 3.1 to the Hopf algebras $A \# kG$ introduced in the previous section. In order to do this, we next give some of the details from [Shroff 2013, Section 4] as they apply to these examples in particular. Recall the PBW algebra B defined in (4-1). Let $\xi_a, \xi_b : B \otimes B \to k$ be the k-linear functions given by

$$\xi_a(r \otimes s) = \gamma_a, \quad \xi_b(r \otimes s) = \gamma_b,$$

where γ_a and γ_b are the scalar coefficients of a^p and b^p, respectively, in the product rs in B. (Shroff writes these functions $\widetilde{\zeta}_1, \widetilde{\zeta}_2$.) Extending to left B-module homomorphisms in $\mathrm{Hom}_B(B^{\otimes 3}, k)$ under the isomorphism $\mathrm{Hom}_B(B^{\otimes 3}, k) \cong \mathrm{Hom}_k(B^{\otimes 2}, k)$, the functions ξ_a, ξ_b are coboundaries on the bar resolution of B, as shown in [loc. cit.], and they factor through $A \cong B/(a^p, b^p)$. The resulting functions (which we will also denote ξ_a, ξ_b by abuse of notation) are no longer coboundaries. They represent nonzero elements in the cohomology of A, corresponding to permanent cycles in the May spectral sequence for A as a filtered algebra (see [May 1966, Theorem 3] or [Weibel 1994, Theorem 5.4.1]). On page E_1 of this spectral sequence, their counterparts generate a polynomial ring over which E_1 is finitely generated (by the elements $1, \eta_a, \eta_b, \eta_a\eta_b$, where η_a, η_b have cohomological degree 1, functions dual to a and b in $\mathrm{Hom}_k(\mathrm{gr}\, A, k) \cong \mathrm{Hom}_{\mathrm{gr}\, A}(\mathrm{gr}\, A \otimes \mathrm{gr}\, A, k)$). The cohomology $\mathrm{H}^*(A, k)$ is finitely generated over its subalgebra generated by ξ_a, ξ_b, as a consequence of the proof of [Shroff 2013, Theorem 4.3]. We will see below that the subalgebra generated by ξ_a, ξ_b is in fact a polynomial ring in ξ_a, ξ_b, which is Noetherian, so applying the left module version of [Goodearl and Warfield 2004, Corollary 1.5], $\mathrm{H}^*(A, k)$ is itself (left) Noetherian.

To verify the hypothesis of Theorem 3.1, we use the above information to define 2-cocycles representing elements in $\mathrm{H}^*(A\#kG, k)$: Note that a^p, b^p are G-invariant by [Cibils et al. 2009, (3.10)]. Thus, by the construction of ξ_a, ξ_b, these functions are also G-invariant, and so they in fact extend to 2-coboundaries on $B\#kG$, factoring through $A\#kG \cong B\#kG/(a^p, b^p)$. This also shows that ξ_a, ξ_b commute with each other in $\mathrm{H}^*(A, k)$, since $\mathrm{H}^*(A\#kG, k)$ is graded commutative and ξ_a, ξ_b each have even degree, so they are commuting elements in $\mathrm{Im}(\mathrm{res}_{A\#kG,A})$.

We next claim that ξ_a, ξ_b generate a polynomial subalgebra $k[\xi_a, \xi_b]$ of $\mathrm{H}^*(A, k)$ over which $\mathrm{H}^*(A, k)$ is free with free basis $\{1,\ \eta_a,\ \eta_b,\ \eta_a\eta_b\}$.[1] This will follow once we see that the set

$$\{\xi_a^i \xi_b^j \eta_a^l \eta_b^m \mid i, j \geq 0,\ l, m = 0, 1\}$$

represents a basis of $\mathrm{H}^*(A, k)$, since ξ_a, ξ_b commute with each other. Note that the cohomology of $S = \mathrm{gr}\, A$ is well known, and has a basis precisely of this form. Recall that Anick's resolution for A is minimal, and a comparison shows that in each degree, the dimensions of $\mathrm{H}^*(A, k)$ and of $\mathrm{H}^*(S, k)$ are the same. This forces the May spectral sequence [1966] for A to collapse at $E_1 = \mathrm{H}^*(S, k)$, and so $\mathrm{gr}\, \mathrm{H}^*(A, k) \cong \mathrm{H}^*(S, k)$, and $\mathrm{H}^*(A, k)$ has basis as claimed. This implies that ξ_a, ξ_b generate a polynomial subring (we already know they commute). Therefore $\mathrm{H}^*(A, k)$ is free as a $k[\xi_a, \xi_b]$-module, as claimed. Further, the k-linear span of $\{1,\ \eta_a,\ \eta_b,\ \eta_a\eta_b\}$ is indeed a kG-submodule of $\mathrm{H}^*(A, k)$: we compute

$$^g\eta_a = \eta_a + \eta_b, \quad ^g\eta_b = \eta_b, \quad ^g(\eta_a\eta_b) = \eta_a\eta_b.$$

[1] Since B is a Koszul algebra, $\mathrm{H}^*(B, k) \cong B^!$, the Koszul dual of B, which is generated by η_a, η_b (by abuse of notation) with relations dual to those of B, that is, $\eta_a^2 = \frac{1}{2}\eta_a\eta_b,\ \eta_b^2 = 0,\ \eta_b\eta_a = -\eta_a\eta_b$. These relations also hold in $\mathrm{H}^*(A, k)$, however we do not need this fact.

We have shown that the hypotheses of Theorem 3.1 are satisfied. Therefore, $\mathrm{H}^*(A \# kG, k)$ is Noetherian.

Question 5.1. Are there more examples of Nichols algebras in positive characteristic to which Theorem 3.1 applies?

Nichols algebras and their bosonizations, which are Hopf algebras, have only just begun to be explored in positive characteristic. There is a vast (and recent) literature on Nichols algebras in characteristic zero. See, for example, [Andruskiewitsch et al. 2011a; 2011b; Andruskiewitsch and Schneider 2010; Heckenberger 2006].

Acknowledgments

We thank D. Benson and P. Symonds for very insightful conversations and suggestions. We thank Ø. Solberg for computing the cohomology of some of the Nichols algebras in Section 4; these computations led us to our general result on this series of Nichols algebras and corresponding Hopf algebras.

References

[Andruskiewitsch and Schneider 2010] N. Andruskiewitsch and H.-J. Schneider, "On the classification of finite-dimensional pointed Hopf algebras", *Ann. of Math.* (2) **171**:1 (2010), 375–417. MR 2011j:16058 Zbl 1208.16028

[Andruskiewitsch et al. 2011a] N. Andruskiewitsch, F. Fantino, M. Graña, and L. Vendramin, "The logbook of pointed Hopf algebras over the sporadic simple groups", *J. Algebra* **325** (2011), 282–304. MR 2012d:16092 Zbl 1217.16025

[Andruskiewitsch et al. 2011b] N. Andruskiewitsch, F. Fantino, M. Graña, and L. Vendramin, "Finite-dimensional pointed Hopf algebras with alternating groups are trivial", *Ann. Mat. Pura Appl.* (4) **190**:2 (2011), 225–245. MR 2012c:16095 Zbl 1234.16019

[Anick 1986] D. J. Anick, "On the homology of associative algebras", *Trans. Amer. Math. Soc.* **296**:2 (1986), 641–659. MR 87i:16046 Zbl 0598.16028

[Barnes 1985] D. W. Barnes, *Spectral sequence constructors in algebra and topology*, Mem. Amer. Math. Soc. **317**, 1985. MR 86e:55032 Zbl 0557.18006

[Bueso et al. 2003] J. Bueso, J. Gómez-Torrecillas, and A. Verschoren, *Algorithmic methods in non-commutative algebra: Applications to quantum groups*, Mathematical Modelling: Theory and Applications **17**, Kluwer, Dordrecht, 2003. MR 2005c:16069 Zbl 1063.16054

[Cibils et al. 2009] C. Cibils, A. Lauve, and S. Witherspoon, "Hopf quivers and Nichols algebras in positive characteristic", *Proc. Amer. Math. Soc.* **137**:12 (2009), 4029–4041. MR 2010k:16047 Zbl 1191.16032

[Cojocaru and Ufnarovski 1997] S. Cojocaru and V. Ufnarovski, "BERGMAN under MS-DOS and Anick's resolution", *Discrete Math. Theor. Comput. Sci.* **1**:1 (1997), 139–147. MR 99e:16017 Zbl 0935.16032

[Etingof and Ostrik 2004] P. Etingof and V. Ostrik, "Finite tensor categories", *Mosc. Math. J.* **4**:3 (2004), 627–654. MR 2005j:18006 Zbl 1077.18005

[Evens 1961] L. Evens, "The cohomology ring of a finite group", *Trans. Amer. Math. Soc.* **101** (1961), 224–239. MR 25 #1191 Zbl 0104.25101

[Friedlander and Suslin 1997] E. M. Friedlander and A. Suslin, "Cohomology of finite group schemes over a field", *Invent. Math.* **127**:2 (1997), 209–270. MR 98h:14055a Zbl 0945.14028

[Ginzburg and Kumar 1993] V. Ginzburg and S. Kumar, "Cohomology of quantum groups at roots of unity", *Duke Math. J.* **69**:1 (1993), 179–198. MR 94c:17026 Zbl 0774.17013

[Goodearl and Warfield 2004] K. R. Goodearl and R. B. Warfield, Jr., *An introduction to noncommutative Noetherian rings*, 2nd ed., London Mathematical Society Student Texts **61**, Cambridge University Press, 2004. MR 2005b:16001 Zbl 1101.16001

[Gordon 2000] I. G. Gordon, "Cohomology of quantized function algebras at roots of unity", *Proc. London Math. Soc.* (3) **80**:2 (2000), 337–359. MR 2000m:20078 Zbl 1032.17025

[Heckenberger 2006] I. Heckenberger, "The Weyl groupoid of a Nichols algebra of diagonal type", *Invent. Math.* **164**:1 (2006), 175–188. MR 2007e:16047 Zbl 1174.17011

[May 1966] J. P. May, "The cohomology of restricted Lie algebras and of Hopf algebras", *J. Algebra* **3** (1966), 123–146. MR 33 #1347 Zbl 0163.03102

[Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, Amer. Math. Soc., Providence, RI, 1993. MR 94i:16019 Zbl 0793.16029

[Priddy 1970] S. B. Priddy, "Koszul resolutions", *Trans. Amer. Math. Soc.* **152** (1970), 39–60. MR 42 #346 Zbl 0261.18016

[Shroff 2013] P. Shroff, "Finite generation of the cohomology of quotients of PBW algebras", *J. Algebra* **390** (2013), 44–55. MR 3072110 Zbl 06266620

[Snashall and Solberg 2004] N. Snashall and Ø. Solberg, "Support varieties and Hochschild cohomology rings", *Proc. London Math. Soc.* (3) **88**:3 (2004), 705–732. MR 2005a:16014 Zbl 1067.16010

[Suarez-Alvarez 2004] M. Suarez-Alvarez, "The Hilton–Heckmann argument for the anti-commutativity of cup products", *Proc. Amer. Math. Soc.* **132**:8 (2004), 2241–2246. MR 2005a:18016 Zbl 1045.18002

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

[Xu 2008] F. Xu, "Hochschild and ordinary cohomology rings of small categories", *Adv. Math.* **219**:6 (2008), 1872–1893. MR 2009h:18025 Zbl 1156.18007

Communicated by Susan Montgomery
Received 2013-10-04 Accepted 2014-08-17

v.nguyen@neu.edu Department of Mathematics, Northeastern University, 567 Lake Hall, Boston, MA 02115, United States

sjw@math.tamu.edu Department of Mathematics, Texas A&M University, Mailstop 3368, College Station, TX 77843, United States

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1659

On the supersingular locus of the GU(2,2) Shimura variety

Benjamin Howard and Georgios Pappas

We describe the supersingular locus of a GU(2, 2) Shimura variety at a prime inert in the corresponding quadratic imaginary field.

1. Introduction

This paper contributes to the theory of integral models of Shimura varieties, and, in particular, to the problem of explicitly describing the basic locus in the reduction modulo p of a canonical integral model. In many cases where this integral model is a moduli space of abelian varieties with additional structures, the basic locus coincides with the supersingular locus, i.e., with the subset of the moduli in positive characteristic where the corresponding abelian variety is isogenous to a product of supersingular elliptic curves. The first investigations of a higher-dimensional supersingular locus were for the Siegel moduli space, and are due to Koblitz, Katsura and Oort, and Li and Oort. See the introduction of [Vollaard 2010] for these and other references. More recently, such explicit descriptions for certain unitary and orthogonal Shimura varieties have found applications to Kudla's program relating arithmetic intersection numbers of special cycles on Shimura varieties to Eisenstein series; this motivated further study, as in [Kudla and Rapoport 2009; 1999; 2000; 2011; Vollaard and Wedhorn 2011].

In this paper, we study the supersingular locus of the special fiber of a GU(2, 2) Shimura variety at an odd prime inert in the corresponding imaginary quadratic field. Our methods borrow liberally from [Vollaard 2010] and [Vollaard and Wedhorn 2011], which dealt with the GU(n, 1) Shimura varieties at inert primes, and from [Rapoport et al. 2014], which considered them at ramified primes. If one attempts to directly imitate the arguments in those papers to study the general GU(r, s) Shimura variety, the method breaks down at a crucial point. The key new idea for overcoming this obstacle is to exploit the linear algebra underlying a twisted

Pappas is partially supported by NSF grant DMS-1102208; Howard is partially supported by NSF grant DMS-1201480.
MSC2010: primary 11G18; secondary 14G35.
Keywords: Shimura varieties, Deligne–Lusztig varieties.

version of the exceptional isomorphism $\mathrm{SU}(2,2) \cong \mathrm{Spin}(4,2)$ corresponding to the Dynkin diagram identity $A_3 = D_3$. As such, we do not expect our methods to extend to unitary groups of other signatures (although we do hope that our result will eventually help to predict the shape of the answer in the general case). The problem of understanding the supersingular locus of the $\mathrm{GU}(3,2)$ Shimura variety, for example, remains open.

However, our methods should extend to the family of $\mathrm{GSpin}(n,2)$ Shimura varieties. Work of Kisin [2010] and Madapusi Pera [2012] (see also [Vasiu 1999]) provides us with a good theory of integral models for these Shimura varieties, and recent work of W. Kim [2013] gives a good theory of Rapoport–Zink spaces as well. An extension of our results in this direction would have applications to Kudla's program, for example by allowing one to generalize the work of Kudla and Rapoport [1999; 2000] from $\mathrm{GSpin}(2,2)$ and $\mathrm{GSpin}(3,2)$ Shimura varieties to the general $\mathrm{GSpin}(n,2)$ case. Using the isomorphism between $\mathrm{GSpin}(6,2)$ and the similitude group of a 4-dimensional symplectic module over the Hamiltonian quaternions [Freitag and Hermann 2000], one could also expect to generalize Bültel's results [2012] on the supersingular locus of the moduli space of polarized abelian eightfolds with an action of a definite quaternion algebra. More ambitiously, one could hope to exploit the connection between polarized K3 surfaces and the $\mathrm{GSpin}(19,2)$ Shimura variety in order to study the moduli space of supersingular K3 surfaces. Some of these topics will be pursued in subsequent papers.

As this paper was being prepared, Görtz and He were conducting a general study of basic minuscule affine Deligne–Lusztig varieties for equicharacteristic discrete valued fields. The preprint [Göertz and He 2013] provides a list of cases where these affine Deligne–Lusztig varieties can be expressed as a union of usual Deligne–Lusztig varieties, and that list contains an equicharacteristic analogue of the $\mathrm{GU}(2,2)$ Rapoport–Zink space considered here. These results of Görtz and He in the equicharacteristic case are analogous to our mixed characteristic results.

1.1. ***The local result.*** Our main result concerns the structure of the Rapoport–Zink space parametrizing quasi-isogenies between certain p-divisible groups with extra structure. Fix an algebraically closed field k of characteristic $p > 2$, let W be the ring of Witt vectors over k, and let $E/\mathbb{Q}_p$ be an unramified degree-two extension. Consider the family of triples (G, ι, λ), defined over W-schemes S on which p is locally nilpotent, consisting of a supersingular p-divisible group G with an action $\iota : \mathcal{O}_E \to \mathrm{End}(G)$ and a principal polarization $\lambda : G \to G^\vee$. We require that the action ι and the polarization λ be compatible in the sense of (2-1), and that the action of $\mathcal{O}_E$ on $\mathrm{Lie}(G)$ satisfy the signature-$(2,2)$ determinant condition of (2-2). A choice of one such triple $(\boldsymbol{G}, \boldsymbol{\iota}, \boldsymbol{\lambda})$ over k as a basepoint determines the *Rapoport–Zink space*, $\mathscr{M}$, parametrizing quadruples $(G, \iota, \lambda, \varrho)$ in which $\varrho : G \times_S S_0 \to \boldsymbol{G} \times_k S_0$ is an $\mathcal{O}_E$-linear quasi-isogeny under which $\boldsymbol{\lambda}$ pulls back to a $\mathbb{Q}_p^\times$-multiple $c(\varrho)\lambda$. Here, $S_0 = S \times_W k$. The Rapoport–Zink space $\mathscr{M}$ is a formal scheme over W, and admits a decomposition into open and closed formal subschemes $\mathscr{M} = \biguplus_{\ell\in\mathbb{Z}} \mathscr{M}^{(\ell)}$, where $\mathscr{M}^{(\ell)}$ is the locus where $\mathrm{ord}_p(c(\varrho)) = \ell$. Here and elsewhere, we use the symbol $\biguplus$ to denote disjoint union. The group $p^{\mathbb{Z}}$ acts on $\mathscr{M}$, where the action of p sends $(G, \iota, \lambda, \varrho) \mapsto (G, \iota, \lambda, p\varrho)$. This action has $\mathscr{M}^{(0)} \uplus \mathscr{M}^{(1)}$ as a fundamental domain. In fact, the action of $p^{\mathbb{Z}}$ extends to a larger group J which acts transitively on the set $\{\mathscr{M}^{(\ell)} : \ell \in \mathbb{Z}\}$. Define

$$\mathscr{N} = p^{\mathbb{Z}} \backslash \mathscr{M},$$

and let $\mathscr{N}^+$ and $\mathscr{N}^-$ be the images of $\mathscr{M}^{(0)}$ and $\mathscr{M}^{(1)}$, respectively, under the quotient map $\mathscr{M} \to \mathscr{N}$.

In Section 2, we construct a 6-dimensional $\mathbb{Q}_p$-vector space

$$\boldsymbol{L}_{\mathbb{Q}}^{\Phi} \subset \mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}$$

of *special quasi-endomorphisms* of $\boldsymbol{G}$ as the Φ-fixed vectors in a slope-0 isocrystal $(\boldsymbol{L}_{\mathbb{Q}}, \Phi)$. The vector space $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ is endowed with a $\mathbb{Q}_p$-valued quadratic form $Q(x) = x \circ x$, and we define a *vertex lattice* in $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ to be a $\mathbb{Z}_p$-lattice $\Lambda \subset \boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ such that

$$p\Lambda \subset \Lambda^\vee \subset \Lambda.$$

The *type* $t_\Lambda \in \{2, 4, 6\}$ of Λ is the dimension of $\Lambda/\Lambda^\vee$. To each point $(G, \iota, \lambda, \varrho)$ of $\mathscr{N}$, the quasi-isogeny ϱ allows us to view Λ as a lattice of quasi-endomorphisms of G. Let $\widetilde{\mathscr{N}}_\Lambda \subset \mathscr{N}$ be the locus of points where $\Lambda \subset \mathrm{End}(G)$ (i.e., the locus where these quasi-endomorphisms are integral). It is a closed formal subscheme of $\mathscr{N}$, whose underlying reduced k-scheme we denote by $\mathscr{N}_\Lambda$. We show that the underlying reduced subscheme $\mathscr{N}_{\mathrm{red}}$ of $\mathscr{N}$ is covered by these closed subschemes:

$$\mathscr{N}_{\mathrm{red}} = \bigcup_{\Lambda} \mathscr{N}_\Lambda,$$

and that

$$\mathscr{N}_{\Lambda_1} \cap \mathscr{N}_{\Lambda_2} = \begin{cases} \mathscr{N}_{\Lambda_1\cap\Lambda_2} & \text{if } \Lambda_1 \cap \Lambda_2 \text{ is a vertex lattice,} \\ \varnothing & \text{otherwise,} \end{cases}$$

where the left-hand side is understood to mean the reduced subscheme underlying the scheme-theoretic intersection (we suspect that the scheme-theoretic intersection is already reduced, but are unable to provide a proof).

Section 3 is devoted to understanding the structure of $\mathscr{N}_\Lambda^\pm = \mathscr{N}_\Lambda \cap \mathscr{N}^\pm$. Setting $d_\Lambda = t_\Lambda/2$, we prove that $\mathscr{N}_\Lambda^\pm$ is a projective, smooth, and irreducible k-scheme of dimension $d_\Lambda - 1$. In fact:

(1) If $d_\Lambda = 1$, then $\mathscr{N}_\Lambda^\pm$ is a single point.

(2) If $d_\Lambda = 2$, then $\mathcal{N}_\Lambda^\pm$ is isomorphic to $\mathbb{P}^1$.

(3) If $d_\Lambda = 3$, then $\mathcal{N}_\Lambda^\pm$ is isomorphic to the Fermat hypersurface

$$x_0^{p+1} + x_1^{p+1} + x_2^{p+1} + x_3^{p+1} = 0.$$

The irreducible components of $\mathcal{N}^\pm$ are precisely the closed subschemes $\mathcal{N}_\Lambda^\pm$ indexed by the type-6 vertex lattices. From this we deduce the following theorem:

Theorem A. *The underlying reduced scheme $\mathcal{M}_{\mathrm{red}}^{(\ell)}$ of $\mathcal{M}^{(\ell)}$ is connected. Every irreducible component of $\mathcal{M}_{\mathrm{red}}^{(\ell)}$ is a smooth k-scheme of dimension 2, isomorphic to the Fermat hypersurface*

$$x_0^{p+1} + x_1^{p+1} + x_2^{p+1} + x_3^{p+1} = 0.$$

If two irreducible components intersect nontrivially, the reduced scheme underlying their scheme-theoretic intersection is either a point or a projective line.

See Sections 3.5 and 3.6 for a more detailed description of $\mathcal{M}_{\mathrm{red}}$.

1.2. ***The global result.*** In Section 4, we consider the global situation. Let E be a quadratic imaginary field, and let $p > 2$ be inert in E. Let $\mathcal{O} \subset E$ be the integral closure of $\mathbb{Z}_{(p)}$, and let V be a free $\mathcal{O}$-module of rank 4 endowed with a perfect $\mathcal{O}$-valued Hermitian form of signature $(2,2)$. Let $G = \mathrm{GU}(V)$ be the group of unitary similitudes of V, a reductive group over $\mathbb{Z}_{(p)}$. Fix a compact open subgroup $U^p \subset G(\mathbb{A}_f^p)$, which we assume is sufficiently small, and define $U_p = G(\mathbb{Z}_p)$ and $U = U_p U^p \subset G(\mathbb{A}_f)$.

Using this data we define a scheme M_U, smooth of relative dimension 4 over $\mathbb{Z}_{(p)}$, as a moduli space of abelian fourfolds, up to prime-to-p-isogeny, with additional structure, in such a way that the complex fiber of M_U is the Shimura variety

$$M_U(\mathbb{C}) = G(\mathbb{Q}) \backslash (\mathcal{D} \times G(\mathbb{A}_f)/U).$$

Here, $\mathcal{D}$ is the Grassmannian of negative-definite planes in $V \otimes_{\mathcal{O}} \mathbb{C}$.

Let k be an algebraically closed field of characteristic p, and denote by M_U^{ss} the reduced supersingular locus of the geometric special fiber $M_U \times_{\mathbb{Z}_{(p)}} k$. The uniformization theorem of Rapoport and Zink expresses M_U^{ss} as a disjoint union of quotients of the scheme $\mathcal{M}_{\mathrm{red}}$ described above. As a consequence we obtain the following result:

Theorem B. *The k-scheme M_U^{ss} has pure dimension 2. For U^p sufficiently small, all irreducible components of M_U^{ss} are isomorphic to the Fermat hypersurface*

$$x_0^{p+1} + x_1^{p+1} + x_2^{p+1} + x_3^{p+1} = 0.$$

If two irreducible components intersect nontrivially, the reduced scheme underlying their scheme-theoretic intersection is either a point or a projective line.

1.3. ***Notation.*** We use the following notation throughout Sections 2 and 3. Fix an odd prime p and an unramified quadratic extension E of the field of p-adic numbers $\mathbb{Q}_p$. The nontrivial Galois automorphism of E is denoted by $\alpha \mapsto \bar{\alpha}$. Let k be an algebraically closed field of characteristic p. Its ring of Witt vectors $W = W(k)$ is a complete discrete valuation ring with residue field $k = W/pW$ and fraction field $W_{\mathbb{Q}}$. Label the two embeddings of $\mathcal{O}_E$ into W as

$$\psi_0 : \mathcal{O}_E \to W, \quad \psi_1 : \mathcal{O}_E \to W,$$

and denote by σ both the absolute Frobenius $x \mapsto x^p$ on k and its unique lift to a ring automorphism of W. Denote by $\epsilon_0, \epsilon_1 \in \mathcal{O}_E \otimes W$ the orthogonal idempotents characterized by

$$\epsilon_i M = \{x \in M : (\alpha \otimes 1) \cdot x = (1 \otimes \psi_i(\alpha)) \cdot x \text{ for all } \alpha \in \mathcal{O}_E\}$$

for any $\mathcal{O}_E \otimes W$-module M. For any $\mathbb{Z}$-module M, we abbreviate $M_{\mathbb{Q}} = M \otimes_{\mathbb{Z}} \mathbb{Q}$. In particular, $M_{\mathbb{Q}} = M \otimes_W W[1/p]$ for any W-module M.

2. Moduli spaces and lattices

In this section we recall the Rapoport–Zink space of a GU$(2,2)$ Shimura variety, and define a stratification of the underlying reduced scheme.

2.1. ***The Rapoport–Zink space.*** Let Nilp_W be the category of W-schemes on which p is locally nilpotent. We wish to parametrize triples (G, ι, λ) over objects S of Nilp_W in which

- G is a supersingular p-divisible group of dimension 4,
- $\iota : \mathcal{O}_E \to \mathrm{End}(G)$ is an action of $\mathcal{O}_E$ on G,
- $\lambda : G \to G^\vee$ is a principal polarization.

We further require that every $\alpha \in \mathcal{O}_E$ satisfy both the $\mathcal{O}_E$*-linearity condition*

$$\lambda \circ \iota(\bar{\alpha}) = \iota(\alpha)^\vee \circ \lambda \tag{2-1}$$

and the *signature*-$(2,2)$ *condition*

$$\det(T - \iota(\alpha); \mathrm{Lie}(G)) = (T - \psi_0(\alpha))^2 (T - \psi_1(\alpha))^2 \tag{2-2}$$

as sections of $\mathcal{O}_S[T]$. The signature-$(2,2)$ condition is equivalent to each of the $\mathcal{O}_S$-module direct summands in $\mathrm{Lie}(G) = \epsilon_0 \mathrm{Lie}(G) \oplus \epsilon_1 \mathrm{Lie}(G)$ being locally free of rank 2.

Fix one such triple $(\boldsymbol{G}, \boldsymbol{\iota}, \boldsymbol{\lambda})$ over k as a base point, and let $\mathscr{M}$ be the functor on Nilp_W sending S to the set of isomorphism classes of quadruples (G, λ, i, ϱ) over S, where (G, ι, λ) is as above and $\varrho : G_{/S_0} \to \boldsymbol{G}_{/S_0}$ is an $\mathcal{O}_E$-linear quasi-isogeny such that $\varrho^* \boldsymbol{\lambda} = c(\varrho)\lambda$ for some $c(\varrho) \in \mathbb{Q}_p^\times$. Here, S_0 is the k-scheme $S \otimes_W k$.

The functor $\mathscr{M}$ is represented by a formal scheme locally of finite type over $\mathrm{Spf}(W)$ by [Rapoport and Zink 1996]. There is a decomposition $\mathscr{M} = \biguplus_{\ell \in \mathbb{Z}} \mathscr{M}^{(\ell)}$ into open and closed formal subschemes, where $\mathscr{M}^{(\ell)}$ is the locus of points where $\mathrm{ord}_p(c(\varrho)) = \ell$.

Let $J \subset \mathrm{End}(\boldsymbol{G})^\times_{\mathbb{Q}}$ denote the subgroup of E-linear elements such that $g^*\boldsymbol{\lambda} = \nu(g)\boldsymbol{\lambda}$ for some $\nu(g) \in \mathbb{Q}_p^\times$. The group J acts on $\mathscr{M}$ in an obvious way:

$$g \cdot (G, \iota, \lambda, \varrho) = (G, \iota, \lambda, g \circ \varrho).$$

As usual, the group J is the $\mathbb{Q}_p$-points of a reductive group over $\mathbb{Q}_p$. In fact, by [Vollaard 2010, Remark 1.16], this reductive group is the group of unitary similitudes of the split Hermitian space of dimension 4 over E. In particular, the derived subgroup J^{der} is isomorphic to the special unitary group, and the similitude character $\nu : J \to \mathbb{Q}_p^\times$ is surjective. Note that the action of any $g \in J$ with $\mathrm{ord}_p(\nu(g)) = 1$ defines an isomorphism $\mathscr{M}^{(\ell)} \cong \mathscr{M}^{(\ell+1)}$.

As a special case of this action, the group $p^{\mathbb{Z}}$ acts on $\mathscr{M}$ by

$$p \cdot (G, \iota, \lambda, \varrho) = (G, \iota, \lambda, p\varrho),$$

and the quotient $\mathscr{N} = p^{\mathbb{Z}} \backslash \mathscr{M}$ has $\mathscr{M}^{(0)} \uplus \mathscr{M}^{(1)}$ as a fundamental domain. Let $\mathscr{N}^+ \cong \mathscr{M}^{(0)}$ and $\mathscr{N}^- \cong \mathscr{M}^{(1)}$ be the open and closed formal subschemes of $\mathscr{N}$ on which $\mathrm{ord}_p(c(\varrho))$ is even and odd, respectively. By the previous paragraph there is an isomorphism $\mathscr{N}^+ \cong \mathscr{N}^-$, and we will see later in Theorem 3.12 that $\mathscr{N}^+$ and $\mathscr{N}^-$ are precisely the connected components of $\mathscr{N}$.

2.2. ***Special endomorphisms.*** In this subsection we will define a $\mathbb{Q}_p$-subspace

$$\boldsymbol{L}^{\Phi}_{\mathbb{Q}} \subset \mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}$$

of *special quasi-endomorphisms* of $\boldsymbol{G}$ in such a way that $x \mapsto x \circ x$ defines a $\mathbb{Q}_p$-valued quadratic form on $\boldsymbol{L}^{\Phi}_{\mathbb{Q}}$. The subspace $\boldsymbol{L}^{\Phi}_{\mathbb{Q}}$ is not quite canonical; it will depend on the auxiliary choice of a certain tensor $\boldsymbol{\omega}$ in the top exterior power of the Dieudonné module of $\boldsymbol{G}$.

Denote by $\boldsymbol{D}$ the covariant Dieudonné module of $\boldsymbol{G}$, with its induced action of $\mathcal{O}_E$ and induced alternating form $\lambda : \bigwedge^2_W \boldsymbol{D} \to W$ satisfying $\lambda(Fx, y) = \lambda(x, Vy)^\sigma$. Under the covariant conventions, $\mathrm{Lie}(\boldsymbol{G}) \cong \boldsymbol{D}/V\boldsymbol{D}$ as k-vector spaces with $\mathcal{O}_E$-actions. Abbreviate $\bigwedge^\ell_E \boldsymbol{D} = \bigwedge^\ell_{\mathcal{O}_E \otimes W} \boldsymbol{D}$. Once we fix a $\delta \in \mathcal{O}_E^\times$ satisfying $\delta^\sigma = -\delta$, there is a unique Hermitian form

$$\langle \cdot, \cdot \rangle : \boldsymbol{D} \times \boldsymbol{D} \to \mathcal{O}_E \otimes W$$

satisfying

$$\lambda(x, y) = \mathrm{Tr}_{E/\mathbb{Q}_p} \delta^{-1} \langle x, y \rangle, \tag{2-3}$$

which in turn induces a Hermitian form on every exterior power $\bigwedge_E^\ell \boldsymbol{D}$ by

$$\langle x_1 \wedge \cdots \wedge x_\ell, y_1 \wedge \cdots \wedge y_\ell \rangle = \sum_{\pi \in S_\ell} \operatorname{sgn}(\pi) \prod_{i=1}^{\ell} \langle x_i, y_{\pi(i)} \rangle.$$

This Hermitian form identifies each lattice $\bigwedge_E^\ell \boldsymbol{D}$ with its dual lattice in $\left(\bigwedge_E^\ell \boldsymbol{D}\right)_{\mathbb{Q}}$.

In order to make explicit calculations, we now put coordinates on $\boldsymbol{D}_{\mathbb{Q}}$.

Lemma 2.1. *There are $W_{\mathbb{Q}}$-bases*

$$e_1, e_2, e_3, e_4 \in \epsilon_0 \boldsymbol{D}_{\mathbb{Q}},$$
$$f_1, f_2, f_3, f_4 \in \epsilon_1 \boldsymbol{D}_{\mathbb{Q}},$$

such that

$$\langle e_i, f_j \rangle = \begin{cases} \epsilon_0 & \text{if } i = j, \\ 0 & \text{otherwise}, \end{cases} \tag{2-4}$$

and the σ-semilinear operator F satisfies

$$\begin{aligned} Fe_1 &= f_1, & Fe_2 &= f_2, & Fe_3 &= pf_3, & Fe_4 &= pf_4, \\ Ff_1 &= pe_1, & Ff_2 &= pe_2, & Ff_3 &= e_3, & Ff_4 &= e_4. \end{aligned}$$

Proof. Denote by $\boldsymbol{D}'_{\mathbb{Q}}$ the isocrystal with $W_{\mathbb{Q}}$-basis $\{e_1, \ldots, e_4, f_1, \ldots, f_4\}$ and by F the operator defined by the above relations. Endow $\boldsymbol{D}'_{\mathbb{Q}}$ with the E-action $\boldsymbol{\iota}'(\alpha)e_i = \psi_0(\alpha)e_i$ and $\boldsymbol{\iota}'(\alpha) f_i = \psi_1(\alpha) f_i$ and the unique Hermitian form satisfying (2-4). This Hermitian form determines a polarization $\boldsymbol{\lambda}'(x, y) = \operatorname{Tr}_{E/\mathbb{Q}_p} \delta^{-1}\langle x, y\rangle$. As $\boldsymbol{D}'_{\mathbb{Q}}$ is isoclinic of slope $1/2$, there is an isomorphism of isocrystals

$$\varrho : \boldsymbol{D}_{\mathbb{Q}} \cong \boldsymbol{D}'_{\mathbb{Q}}.$$

Any two embeddings of E into $\operatorname{End}(\boldsymbol{D}'_{\mathbb{Q}})$ are conjugate, by the Noether–Skolem theorem, and so ϱ may be modified to make it E-linear. Another application of Noether–Skolem shows that ϱ may be further modified to ensure that the polarizations on $\boldsymbol{D}_{\mathbb{Q}}$ and $\boldsymbol{D}'_{\mathbb{Q}}$ induce the same Rosati involution on

$$\operatorname{End}(\boldsymbol{D}_{\mathbb{Q}}) \cong \operatorname{End}(\boldsymbol{D}'_{\mathbb{Q}}).$$

This implies that ϱ identifies the polarizations, and hence the Hermitian forms, on $\boldsymbol{D}_{\mathbb{Q}}$ and $\boldsymbol{D}'_{\mathbb{Q}}$ up to scaling by an element $c(\varrho) \in \mathbb{Q}_p^\times$.

Finally, for every $c \in \mathbb{Q}_p^\times$ one can find an E-linear isocrystal automorphism g of $\boldsymbol{D}'_{\mathbb{Q}}$ such that g rescales the polarization of $\boldsymbol{D}'_{\mathbb{Q}}$ by the factor c. For example, if $\operatorname{ord}_p(c)$ is even then write $c = \alpha\bar{\alpha}$ with $\alpha \in E^\times$ and take $g = \boldsymbol{\iota}'(\alpha)$. If $c = p$ then take g to be

$$\begin{aligned} e_1 &\mapsto e_3, & e_2 &\mapsto e_4, & e_3 &\mapsto pe_1, & e_4 &\mapsto pe_2, \\ f_1 &\mapsto pf_3, & f_2 &\mapsto pf_4, & f_3 &\mapsto f_1, & f_4 &\mapsto f_2. \end{aligned}$$

Thus ϱ may be further modified to ensure that $c(\varrho) = 1$. □

Lemma 2.2. *There is an $\mathcal{O}_E \otimes W$-module generator $\boldsymbol{\omega} \in \bigwedge_E^4 \boldsymbol{D}$ such that $\langle \boldsymbol{\omega}, \boldsymbol{\omega} \rangle = 1$ and $F\boldsymbol{\omega} = p^2 \boldsymbol{\omega}$. If $\boldsymbol{\omega}' \in \bigwedge_E^4 \boldsymbol{D}$ is another such element, there is an $\alpha \in \mathcal{O}_E^\times$ such that $\alpha\overline{\alpha} = 1$ and $\boldsymbol{\omega}' = \alpha\boldsymbol{\omega}$.*

Proof. The W-module decomposition $\boldsymbol{D} = \epsilon_0 \boldsymbol{D} \oplus \epsilon_1 \boldsymbol{D}$ induces a corresponding decomposition $\bigwedge_E^4 \boldsymbol{D} = \bigwedge^4 \epsilon_0 \boldsymbol{D} \oplus \bigwedge^4 \epsilon_1 \boldsymbol{D}$. Fixing a basis as in Lemma 2.1, we must have

$$\bigwedge^4 \epsilon_0 \boldsymbol{D} = W \cdot p^{k_0} e_1 \wedge e_2 \wedge e_3 \wedge e_4,$$
$$\bigwedge^4 \epsilon_1 \boldsymbol{D} = W \cdot p^{k_1} f_1 \wedge f_2 \wedge f_3 \wedge f_4,$$

for some integers k_0 and k_1. The self-duality of $\bigwedge_E^4 \boldsymbol{D}$ under $\langle \cdot , \cdot \rangle$ implies $k_0 + k_1 = 0$. The signature-$(2,2)$ condition on

$$\mathrm{Lie}(\boldsymbol{G}) \cong \boldsymbol{D}/V\boldsymbol{D} = \epsilon_0 \boldsymbol{D}/V\epsilon_1 \boldsymbol{D} \oplus \epsilon_1 \boldsymbol{D}_1/V\epsilon_0 \boldsymbol{D}$$

implies that each of the summands on the right has dimension 2 over W/pW, and hence the cokernels of

$$V : \bigwedge^4 \epsilon_0 \boldsymbol{D} \to \bigwedge^4 \epsilon_1 \boldsymbol{D}, \quad V : \bigwedge^4 \epsilon_1 \boldsymbol{D} \to \bigwedge^4 \epsilon_0 \boldsymbol{D}$$

are each of length 2 as W-modules. Using

$$V(e_1 \wedge e_2 \wedge e_3 \wedge e_4) = p^2 f_1 \wedge f_2 \wedge f_3 \wedge f_4,$$
$$V(f_1 \wedge f_2 \wedge f_3 \wedge f_4) = p^2 e_1 \wedge e_2 \wedge e_3 \wedge e_4,$$

we deduce that k_1 and k_2 are equal, and hence both are equal to 0. It follows that

$$\boldsymbol{\omega} = e_1 \wedge e_2 \wedge e_3 \wedge e_4 + f_1 \wedge f_2 \wedge f_3 \wedge f_4 \tag{2-5}$$

generates $\bigwedge_E^4 \boldsymbol{D}$ as an $\mathcal{O}_E \otimes W$-module. A simple calculation shows that $\langle \boldsymbol{\omega}, \boldsymbol{\omega} \rangle = 1$ and $F\boldsymbol{\omega} = p^2\boldsymbol{\omega}$, proving the existence part of the lemma. The uniqueness part of the claim is obvious. □

Definition 2.3. For any $\boldsymbol{\omega}$ as in the lemma, define the *Hodge star operator* $x \mapsto x^\star$ on $\bigwedge_E^2 \boldsymbol{D}$ by the relation $y \wedge x^\star = \langle y, x \rangle \cdot \boldsymbol{\omega}$ for all $y \in \bigwedge_E^2 \boldsymbol{D}$.

The Hodge operator satisfies $(\alpha x)^\star = \overline{\alpha} x^\star$ for all $\alpha \in \mathcal{O}_E \otimes W$. Denote by

$$\boldsymbol{L} = \{x \in \bigwedge_E^2 \boldsymbol{D} : x^\star = x\}$$

the W-submodule of Hodge fixed vectors. The Hermitian form $\langle \cdot , \cdot \rangle$ on $\boldsymbol{D}$ determines an injection $\bigwedge_E^2 \boldsymbol{D} \to \mathrm{End}_W(\boldsymbol{D})$ by

$$(a \wedge b)(z) = \langle a, z \rangle b - \langle b, z \rangle a,$$

and we obtain inclusions $\boldsymbol{L} \subset \bigwedge_E^2 \boldsymbol{D} \subset \mathrm{End}_W(\boldsymbol{D})$. Note that both the Hodge star operator and the submodule $\boldsymbol{L}$ depend on the choice of $\boldsymbol{\omega}$.

Proposition 2.4. *For any choice of $\boldsymbol{\omega}$, the induced Hodge star operator has the following properties*:

(1) *Every $x \in \bigwedge_E^2 \boldsymbol{D}$ satisfies $(x^\star)^\star = x$.*

(2) *Every $x \in \boldsymbol{L}$, viewed as an endomorphism of $\boldsymbol{D}$, satisfies*

$$x \circ x = -\frac{\langle x, x\rangle}{2}. \tag{2-6}$$

In particular, $Q(x) = x \circ x$ defines a W-valued quadratic form on $\boldsymbol{L}$.

(3) *The W-quadratic space $\boldsymbol{L}$ is self-dual of rank* 6, *and*

$$\boldsymbol{L} = \{x \in \boldsymbol{L}_{\mathbb{Q}} : x\boldsymbol{D} \subset \boldsymbol{D}\}.$$

(4) *If $C(\boldsymbol{L})$ denotes the Clifford algebra of $\boldsymbol{L}$, the natural map*

$$C(\boldsymbol{L}) \to \mathrm{End}_W(\boldsymbol{D})$$

induced by the inclusion $\boldsymbol{L} \subset \mathrm{End}_W(\boldsymbol{D})$ is an isomorphism. Under this isomorphism, the even Clifford algebra is identified with the subalgebra of $\mathcal{O}_E$-linear endomorphisms in $\mathrm{End}_W(\boldsymbol{D})$.

Proof. Fix a basis of $\boldsymbol{D}_{\mathbb{Q}}$ as in Lemma 2.1, and suppose first that $\boldsymbol{\omega}$ is given by (2-5). An easy calculation shows that

$$\begin{aligned}
(e_1 \wedge e_2)^\star &= f_3 \wedge f_4, & (f_3 \wedge f_4)^\star &= e_1 \wedge e_2,\\
(e_1 \wedge e_3)^\star &= f_4 \wedge f_2, & (f_4 \wedge f_2)^\star &= e_1 \wedge e_3,\\
(e_1 \wedge e_4)^\star &= f_2 \wedge f_3, & (f_2 \wedge f_3)^\star &= e_1 \wedge e_4,\\
(e_2 \wedge e_3)^\star &= f_1 \wedge f_4, & (f_1 \wedge f_4)^\star &= e_2 \wedge e_3,\\
(e_2 \wedge e_4)^\star &= f_3 \wedge f_1, & (f_3 \wedge f_1)^\star &= e_2 \wedge e_4,\\
(e_3 \wedge e_4)^\star &= f_1 \wedge f_2, & (f_1 \wedge f_2)^\star &= e_3 \wedge e_4,
\end{aligned} \tag{2-7}$$

from which $(x^\star)^\star = x$ is obvious. Now set $\boldsymbol{\omega}' = \alpha\boldsymbol{\omega}$ with $\alpha \in \mathcal{O}_E^\times$ of norm 1, and denote by $x \mapsto x^{\star\prime}$ the Hodge star operator defined by $\boldsymbol{\omega}'$. It is related to the Hodge star operator for $\boldsymbol{\omega}$ by $x^{\star\prime} = \alpha x^\star$, and hence

$$(x^{\star\prime})^{\star\prime} = (\alpha(\alpha x^\star))^\star = \alpha\bar{\alpha}(x^\star)^\star = x.$$

This proves the first claim in full generality.

Keep $\boldsymbol{\omega}$ as in (2-5). For the second claim, one first checks that all $x, y \in \bigwedge_E^2 \boldsymbol{D}$ satisfy the relation

$$x \circ y + y^\star \circ x^\star = -\langle x, y\rangle \tag{2-8}$$

in $\mathrm{End}_W(\boldsymbol{D})$. Indeed, it suffices to prove this when x and y are pure tensors of the form $e_i \wedge e_j$ and $f_i \wedge f_j$, and this can be done by brute force. Of course (2-8) immediately implies (2-6) for all $x \in \boldsymbol{L}$, proving the second claim for $\boldsymbol{\omega}$. The validity of (2-8) for any other $\boldsymbol{\omega}'$ follows by the reasoning of the previous paragraph.

For the third claim, note that the quadratic form $Q(x) = -\langle x, x\rangle/2$ on $\boldsymbol{L}$ extends to a quadratic form on $\bigwedge_E^2 \boldsymbol{D}$ by the same formula (using the standing hypothesis that p is odd), with associated bilinear form

$$[x, y] = -\tfrac{1}{2} \cdot \mathrm{Tr}_{E/\mathbb{Q}_p} \langle x, y\rangle,$$

and that there is an orthogonal decomposition

$$\textstyle\bigwedge_E^2 \boldsymbol{D} = \boldsymbol{L} \oplus \{x \in \bigwedge_E^2 \boldsymbol{D} : x^\star = -x\}.$$

The self-duality of $\bigwedge_E^2 \boldsymbol{D}$ under $\langle\cdot,\cdot\rangle$ implies its self-duality under $[\cdot,\cdot]$, which then implies the self-duality of the orthogonal summand $\boldsymbol{L}$. The Hodge star operator acts on the W-module

$$\textstyle\bigwedge_E^2 \boldsymbol{D} = \bigwedge^2 \epsilon_0 \boldsymbol{D} \oplus \bigwedge^2 \epsilon_1 \boldsymbol{D}$$

of rank 12 by interchanging the two summands on the right, and hence its submodule of fixed points, $\boldsymbol{L}$, has rank 6. Finally, set $\boldsymbol{L}' = \{x \in \boldsymbol{L}_\mathbb{Q} : x\boldsymbol{D} \subset \boldsymbol{D}\}$. Certainly $\boldsymbol{L} \subset \boldsymbol{L}'$, and the quadratic form $Q(x) = x \circ x$ restricted to $\boldsymbol{L}'$ takes values in $W = W_\mathbb{Q} \cap \mathrm{End}_W(\boldsymbol{D})$. Therefore $(\boldsymbol{L}')^\vee \subset \boldsymbol{L}^\vee = \boldsymbol{L} \subset \boldsymbol{L}' \subset (\boldsymbol{L}')^\vee$, and so equality holds throughout.

For the fourth claim, the self-duality of $\boldsymbol{L}$ implies that $\boldsymbol{L}/p\boldsymbol{L}$ is the unique nondegenerate k-quadratic space of dimension 6, and so its Clifford algebra is isomorphic to $M_8(k)$. This means that the induced map

$$C(\boldsymbol{L}/p\boldsymbol{L}) \cong C(\boldsymbol{L}) \otimes_W k \to \mathrm{End}_W(\boldsymbol{D}) \otimes_W k$$

is a homomorphism between central simple k-algebras of the same dimension, and hence an isomorphism. It follows from Nakayama's lemma that $C(\boldsymbol{L}) \to \mathrm{End}_W(\boldsymbol{D})$ is an isomorphism. Every $x \in \boldsymbol{L}$ satisfies $x \circ \boldsymbol{\iota}(\alpha) = \boldsymbol{\iota}(\overline{\alpha}) \circ x$, and hence the composition of any two elements of $\boldsymbol{L}$ is $\mathcal{O}_E$-linear. This implies that the even Clifford algebra is contained in $\mathrm{End}_{\mathcal{O}_E \otimes W}(\boldsymbol{D})$, and equality holds because both are W-module direct summands of $C(\boldsymbol{L}) \cong \mathrm{End}_W(\boldsymbol{D})$ of the same rank. □

The operator

$$\Phi(a \wedge b) = p^{-1}(Fa) \wedge (Fb)$$

makes $\bigwedge_E^2 \boldsymbol{D}_\mathbb{Q}$ into a slope-0 isocrystal. In terms of the inclusion $\bigwedge_E^2 \boldsymbol{D}_\mathbb{Q} \subset \mathrm{End}_W(\boldsymbol{D})_\mathbb{Q}$, this operator is just

$$\Phi(a \wedge b) = F \circ (a \wedge b) \circ F^{-1}.$$

As Φ commutes with the Hodge star operator, it stabilizes the subspace $\boldsymbol{L}_{\mathbb{Q}}$ and makes $\boldsymbol{L}_{\mathbb{Q}}$ into a slope-0 isocrystal. In this way, we obtain inclusions of $\mathbb{Q}_p$-vector spaces

$$\boldsymbol{L}_{\mathbb{Q}}^{\Phi} \subset \left(\textstyle\bigwedge_E^2 \boldsymbol{D}_{\mathbb{Q}}\right)^{\Phi} \subset \mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}, \tag{2-9}$$

where the Φ superscripts denote the subspaces of Φ-fixed vectors. Endow $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ with the quadratic form $Q(x) = x \circ x$ and the associated bilinear form

$$[x, y] = x \circ y + y \circ x = -\tfrac{1}{2} \cdot \mathrm{Tr}_{E/\mathbb{Q}_p} \langle x, y \rangle.$$

Remark 2.5. The 6-dimensional E-vector space $(\bigwedge_E^2 \boldsymbol{D}_{\mathbb{Q}})^{\Phi}$ is characterized as the space of all Rosati-fixed $x \in \mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}$ satisfying $x \circ \boldsymbol{\iota}(\alpha) = \boldsymbol{\iota}(\overline{\alpha}) \circ x$ for all $\alpha \in E$. On the other hand, the 6-dimensional $\mathbb{Q}_p$-vector space $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ depends on the choice of $\boldsymbol{\omega}$, and so does not have a similar interpretation in terms of $\boldsymbol{\lambda}$ and $\boldsymbol{\iota}$ alone.

While the *subspace* $\boldsymbol{L}_{\mathbb{Q}}^{\Phi} \subset \mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}$ depends on the choice of $\boldsymbol{\omega}$, the following proposition shows that its isomorphism class as a quadratic space does not. Denote by $\mathbb{H}$ the hyperbolic $\mathbb{Q}_p$-quadratic space of dimension 2.

Proposition 2.6. *For any choice of $\boldsymbol{\omega}$, the quadratic space $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ has Hasse invariant -1 and determinant* $\det(\boldsymbol{L}_{\mathbb{Q}}^{\Phi}) = -\Delta$ *for any nonsquare $\Delta \in \mathbb{Z}_p^{\times}$. Furthermore, the special orthogonal group* $\mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$ *is quasi-split and splits over $\mathbb{Q}_{p^2}$, and the space $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ with the rescaled quadratic form $p^{-1}Q$ is isomorphic to $\mathbb{H}^2 \oplus \mathbb{Q}_{p^2}$, where $\mathbb{Q}_{p^2}$ is endowed with its norm form $x \mapsto$* $\mathrm{Norm}_{\mathbb{Q}_{p^2}/\mathbb{Q}_p}(x)$.

Proof. First suppose that $\boldsymbol{\omega}$ is defined by (2-5). In this case the relations (2-7) show that the vectors

$$\begin{aligned} x_1 &= e_1 \wedge e_2 + f_3 \wedge f_4, & x_2 &= e_3 \wedge e_4 + f_1 \wedge f_2, \\ x_3 &= e_1 \wedge e_3 + f_4 \wedge f_2, & x_4 &= e_4 \wedge e_2 + f_1 \wedge f_3, \\ x_5 &= e_1 \wedge e_4 + f_2 \wedge f_3, & x_6 &= e_2 \wedge e_3 + f_1 \wedge f_4 \end{aligned}$$

form a basis of $\boldsymbol{L}_{\mathbb{Q}}$. In this basis the operator Φ takes the block-diagonal form

$$\Phi = \begin{pmatrix} 0 & p & & & & \\ p^{-1} & 0 & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & 0 & 1 \\ & & & & 1 & 0 \end{pmatrix} \circ \sigma,$$

and the matrix of Q is

$$([x_i, x_j]) = \begin{pmatrix} 0 & -1 & & & & \\ -1 & 0 & & & & \\ & & 0 & -1 & & \\ & & -1 & 0 & & \\ & & & & 0 & -1 \\ & & & & -1 & 0 \end{pmatrix}.$$

Fix any nonsquare $\Delta \in \mathbb{Z}_p^\times$, and let $u \in W^\times$ be a square root of Δ. The vectors

$$\begin{aligned} y_1 &= px_1 + x_2, & y_2 &= u(px_1 - x_2), \\ y_3 &= x_3 + x_4, & y_4 &= u(x_3 - x_4), \\ y_5 &= x_5 + x_6, & y_6 &= u(x_5 - x_6) \end{aligned}$$

form an orthogonal basis of $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ with

$$\left(\frac{[y_i, y_j]}{2}\right) = \begin{pmatrix} -p & & & & & \\ & p\Delta & & & & \\ & & -1 & & & \\ & & & \Delta & & \\ & & & & -1 & \\ & & & & & \Delta \end{pmatrix}, \tag{2-10}$$

from which one computes the determinant $-\Delta$ and Hasse invariant $(-p, p\Delta) = -1$ of $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$. As a nondegenerate quadratic space over $\mathbb{Q}_p$ is determined by its rank, determinant, and Hasse invariant, the remaining claims are easily checked for this special choice of $\boldsymbol{\omega}$.

Now suppose $\boldsymbol{\omega}' = \alpha\boldsymbol{\omega}$ for some $\alpha \in \mathcal{O}_E^\times$ of norm 1. Hilbert's Theorem 90 implies that there is some $\eta \in \mathcal{O}_E^\times$ satisfying $\eta\bar{\eta}^{-1} = \alpha$. Denote by $x \mapsto x^{\star\prime}$ the Hodge star operator defined by $\boldsymbol{\omega}'$, by $\boldsymbol{L}' \subset \bigwedge_E^2 \boldsymbol{D}$ the submodule of Hodge fixed vectors, and by Q' the quadratic form $x \circ x$ on $\boldsymbol{L}'$. Using the relation $x^{\star\prime} = \alpha x^\star$, it is easy to see that the function $x \mapsto \eta x$ defines an isomorphism of quadratic spaces

$$(\boldsymbol{L}_{\mathbb{Q}}^{\Phi}, \eta\bar{\eta}Q) \cong (\boldsymbol{L}_{\mathbb{Q}}^{\prime\Phi}, Q').$$

In particular, there is a basis of $\boldsymbol{L}_{\mathbb{Q}}^{\prime\Phi}$ such that the quadratic form Q' is given by $\eta\bar{\eta}$ times the matrix of (2-10). The Hasse invariant and determinant (modulo squares) of the matrix in (2-10) are unchanged if the matrix is multiplied by any element of $\mathbb{Z}_p^\times$, and so $\boldsymbol{L}_{\mathbb{Q}}^{\prime\Phi}$ has the same determinant and Hasse invariant as $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$. $\square$

From now on we fix, once and for all, any $\boldsymbol{\omega}$ as in Lemma 2.2.

2.3. ***An exceptional isomorphism.*** Define the unitary similitude group

$$\mathrm{GU}(\boldsymbol{D}_{\mathbb{Q}}) = \{g \in \mathrm{Aut}_{E\otimes W}(\boldsymbol{D}_{\mathbb{Q}}) : g^*\boldsymbol{\lambda} = \nu(g)\boldsymbol{\lambda} \text{ for some } \nu(g) \in W_{\mathbb{Q}}^{\times}\},$$

and set

$$\mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}}) = \{g \in \mathrm{GU}(\boldsymbol{D}_{\mathbb{Q}}) : \nu(g)^2 = \det(g)\}.$$

The action $\bullet$ of $\mathrm{GU}(\boldsymbol{D}_{\mathbb{Q}})$ on $\mathrm{End}_W(\boldsymbol{D})_{\mathbb{Q}}$ defined by $g \bullet x = g \circ x \circ g^{-1}$ leaves invariant the subspace $\bigwedge_E^2 \boldsymbol{D}_{\mathbb{Q}}$, and satisfies

$$g \bullet (a \wedge b) = \nu(g)^{-1} \cdot (ga) \wedge (gb). \tag{2-11}$$

Using this formula one checks that the action of the subgroup $\mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}})$ commutes with the Hodge star operator on $\bigwedge_E^2 \boldsymbol{D}_{\mathbb{Q}}$, and so preserves the subspace $\boldsymbol{L}_{\mathbb{Q}}$.

The *canonical involution* $x \mapsto x'$ on the Clifford algebra $C(\boldsymbol{L})$ is the unique W-linear endomorphism satisfying $(x_1 \cdots x_k)' = x_k \cdots x_1$ for all $x_1, \dots, x_k \in \boldsymbol{L}$, and the spinor similitude group of $\boldsymbol{L}_{\mathbb{Q}}$ is

$$\mathrm{GSpin}(\boldsymbol{L}_{\mathbb{Q}}) = \{g \in C_0(\boldsymbol{L})_{\mathbb{Q}}^{\times} : g\boldsymbol{L}_{\mathbb{Q}}g^{-1} = \boldsymbol{L}_{\mathbb{Q}} \text{ and } g'g \in W_{\mathbb{Q}}^{\times}\}.$$

Here, $C_0(\boldsymbol{L})$ is the even Clifford algebra. From [Bass 1974] or [Shimura 2010] we have the exact sequence

$$1 \longrightarrow W_{\mathbb{Q}}^{\times} \longrightarrow \mathrm{GSpin}(\boldsymbol{L}_{\mathbb{Q}}) \longrightarrow \mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}) \longrightarrow 1.$$

Proposition 2.7. *There is an isomorphism*

$$\mathrm{GSpin}(\boldsymbol{L}_{\mathbb{Q}}) \cong \mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}}) \tag{2-12}$$

compatible with the action of both groups on $\boldsymbol{L}_{\mathbb{Q}}$. *In particular, the action of* $\mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}})$ *on* $\boldsymbol{L}_{\mathbb{Q}}$ *determines an exact sequence*

$$1 \longrightarrow W_{\mathbb{Q}}^{\times} \longrightarrow \mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}}) \xrightarrow{g\mapsto g\bullet} \mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}) \longrightarrow 1.$$

Proof. By Proposition 2.4 the inclusion of $\boldsymbol{L}$ into $\mathrm{End}_W(\boldsymbol{D})$ induces an isomorphism $C(\boldsymbol{L}) \cong \mathrm{End}_W(\boldsymbol{D})$, under which $C_0(\boldsymbol{L}) \cong \mathrm{End}_{\mathcal{O}_E\otimes W}(\boldsymbol{D})$. We will prove that the induced isomorphism

$$C_0(\boldsymbol{L})_{\mathbb{Q}}^{\times} \cong \mathrm{Aut}_{E\otimes W}(\boldsymbol{D}_{\mathbb{Q}})$$

restricts to an isomorphism (2-12). Note that every element $x \in \boldsymbol{L}$, viewed as an endomorphism of $\boldsymbol{D}$, satisfies $\langle xa, b\rangle = -\overline{\langle a, xb\rangle}$ (indeed, this already holds for every $x \in \bigwedge_E^2 \boldsymbol{D}$). Thus $\langle ga, b\rangle = \langle a, g'b\rangle$ for every $g \in C_0(\boldsymbol{L})$ and $a, b \in \boldsymbol{D}$.

One inclusion of (2-12) is obvious: if $g \in \mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}})$ then, as noted above, the conjugation action of g on $C(\boldsymbol{L})_{\mathbb{Q}} \cong \mathrm{End}_W(\boldsymbol{D})_{\mathbb{Q}}$ preserves the subspace $\boldsymbol{L}_{\mathbb{Q}}$. The relation $\langle ga, gb\rangle = \langle a, g'gb\rangle$ implies that $\nu(g) = g'g$, and so $g \in \mathrm{GSpin}(\boldsymbol{L}_{\mathbb{Q}})$.

For the other inclusion, start with a $g \in \mathrm{GSpin}(\boldsymbol{L}_{\mathbb{Q}})$. The relation $\langle ga, gb\rangle = (g'g)\langle a, b\rangle$ shows that $g \in \mathrm{GU}(\boldsymbol{D}_{\mathbb{Q}})$. To show that $\nu(g)^2 = \det(g)$, fix any $x \in \boldsymbol{L}$ and any $y \in \bigwedge_E^2 \boldsymbol{D}$ for which $\langle y, x\rangle \neq 0$. As $g \bullet x = gxg^{-1}$ lies in $\boldsymbol{L}_{\mathbb{Q}}$ by assumption, the Hodge star operator fixes $g \bullet x$. Thus

$$(g \bullet y) \wedge (g \bullet x) = \langle g \bullet y, g \bullet x\rangle \boldsymbol{\omega} = \langle y, x\rangle \boldsymbol{\omega},$$

where the second equality follows from (2-11). On the other hand, the Hodge star operator fixes x, and so

$$(g \bullet y) \wedge (g \bullet x) = \nu(g)^{-2} \det(g)(y \wedge x) = \nu(g)^{-2} \det(g)\langle y, x\rangle \boldsymbol{\omega}.$$

This proves that $g \in \mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}})$, and completes the proof of (2-12). □

The similitude character $\nu : \mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}}) \to W_{\mathbb{Q}}^\times$ restricts to $x \mapsto x^2$ on the subgroup $W_{\mathbb{Q}}^\times$, and so descends to the *spinor norm*

$$\tilde{\nu} : \mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}) \to W_{\mathbb{Q}}^\times/(W_{\mathbb{Q}}^\times)^2.$$

Remark 2.8. The group J defined in Section 2.1 is characterized by

$$J = \{g \in \mathrm{GU}(\boldsymbol{D}_{\mathbb{Q}}) : g \circ F = F \circ g\},$$

and we define a subgroup

$$J^0 = \{g \in \mathrm{GU}^0(\boldsymbol{D}_{\mathbb{Q}}) : g \circ F = F \circ g\}.$$

The isomorphism (2-12) restricts to an isomorphism $\mathrm{GSpin}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi}) \cong J^0$, and hence there is an exact sequence

$$1 \longrightarrow \mathbb{Q}_p^\times \longrightarrow J^0 \longrightarrow \mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi}) \longrightarrow 1,$$

which identifies J^{der} with $\mathrm{Spin}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$. See [Knus et al. 1998, Proposition IV.15.27] for similar exceptional isomorphisms.

2.4. ***Dieudonné lattices and special lattices.*** In this subsection we show that the k-points of $\mathcal{N}$ can be identified with the set of homothety classes of certain lattices in $\boldsymbol{D}_{\mathbb{Q}}$, which we call *Dieudonné lattices*. We then use the inclusion

$$\boldsymbol{L}_{\mathbb{Q}} \subset \mathrm{End}_W(\boldsymbol{D}_{\mathbb{Q}})$$

to construct a bijection between the set of homothety classes of Dieudonné lattices and a set of *special lattices* in the slope-0 isocrystal $\boldsymbol{L}_{\mathbb{Q}}$. Thus the points of $\mathcal{N}(k)$ are parametrized by these special lattices.

In fact, the proof of Theorem 3.9 below requires that we establish such a bijection not just over k, but over any extension field $k' \supset k$. Let W' be the Cohen ring of k'. Thus W' is the unique, up to isomorphism, complete discrete valuation ring of mixed characteristic with residue field $W'/pW' \cong k'$. The inclusion $k \to k'$

induces an injective ring homomorphism $W \to W'$, and we set $\boldsymbol{D}' = \boldsymbol{D} \otimes_W W'$ and $\boldsymbol{L}' = \boldsymbol{L} \otimes_W W'$. There is a continuous ring homomorphism $\sigma : W' \to W'$ reducing to the Frobenius on k', and the σ-semilinear operators F and Φ on $\boldsymbol{D}_{\mathbb{Q}}$ and $\boldsymbol{L}_{\mathbb{Q}}$ have σ-semilinear extensions to $\boldsymbol{D}'_{\mathbb{Q}}$ and $\boldsymbol{L}'_{\mathbb{Q}}$. Similarly the symplectic and Hermitian forms on $\boldsymbol{D}_{\mathbb{Q}}$ and the quadratic form on $\boldsymbol{L}_{\mathbb{Q}}$ have natural extensions to $\boldsymbol{D}'_{\mathbb{Q}}$ and $\boldsymbol{L}'_{\mathbb{Q}}$.

Note that the operators F and Φ are surjective on $\boldsymbol{D}_{\mathbb{Q}}$ and $\boldsymbol{L}_{\mathbb{Q}}$, respectively, but this need not be true of their extensions to $\boldsymbol{D}'_{\mathbb{Q}}$ and $\boldsymbol{L}'_{\mathbb{Q}}$. If $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ is a W'-submodule then so is its preimage $F^{-1}(D)$, but its image $F(D)$ need not be. Denote by $F_*(D)$ the W'-submodule generated by $F(D)$. Similarly, denote by $\Phi_*(L)$ the W'-submodule generated by $\Phi(L)$ for a W'-submodule $L \subset \boldsymbol{L}'_{\mathbb{Q}}$.

For any W'-lattice $D \subset \boldsymbol{D}'_{\mathbb{Q}}$, set $D_1 = F^{-1}(pD)$.

Definition 2.9. A *Dieudonné lattice* in $\boldsymbol{D}'_{\mathbb{Q}}$ is an $\mathcal{O}_E$-stable W'-lattice $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ such that

(1) $pD \subset D_1 \subset D$,

(2) $D^\vee = cD$ for some $c \in \mathbb{Q}_p^\times$,

(3) $D = F_*(F^{-1}(D))$.

Here, the superscript $\vee$ denotes the dual lattice with respect to the symplectic form $\boldsymbol{\lambda}$, or, equivalently, with respect to the Hermitian form $\langle \cdot , \cdot \rangle$.

The *volume* of a lattice $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ is the W'-submodule

$$\mathrm{Vol}(D) = \textstyle\bigwedge^8 D \subset \bigwedge^8 \boldsymbol{D}'_{\mathbb{Q}},$$

By considering the slopes of the isocrystal $\boldsymbol{D}_{\mathbb{Q}}$, one can show that $\mathrm{Vol}(F_*(D)) = p^4 \cdot \mathrm{Vol}(D)$. However, taking preimages of lattices may change volumes in unexpected ways: a lattice $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ satisfies

$$\mathrm{Vol}(F^{-1}(D)) \subset p^{-4} \cdot \mathrm{Vol}(D),$$

but equality holds if and only if $F_*(F^{-1}(D)) = D$. In particular, the condition $D = F_*(F^{-1}(D))$ in Definition 2.9 is equivalent to $\mathrm{Vol}(D_1) = p^4 \,\mathrm{Vol}(D)$, and so one could replace (3) in the definition of a Dieudonné lattice by

(3′) $\dim_{k'}(D_1/pD) = 4$.

The volume of a lattice in $\boldsymbol{L}'_{\mathbb{Q}}$ is defined in the analogous way, but in this case $\mathrm{Vol}(\Phi_*(L)) = \mathrm{Vol}(L)$ for any lattice $L \subset \boldsymbol{L}'_{\mathbb{Q}}$.

Proposition 2.10. *Suppose D is a Dieudonné lattice. The $\mathcal{O}_E$-stable k'-subspace $D_1/pD \subset D/pD$ is Lagrangian with respect to the nondegenerate symplectic form $c\boldsymbol{\lambda}$, and every $\alpha \in \mathcal{O}_E$ acts on D/D_1 with characteristic polynomial*

$$\det(T - \iota(\alpha); D/D_1) = (T - \psi_0(\alpha))^2 (T - \psi_1(\alpha))^2. \tag{2-13}$$

Proof. For any $a, b \in D_1$ we have

$$c\boldsymbol{\lambda}(a,b)^\sigma = p^{-1}c\boldsymbol{\lambda}(Fa, Fb) \in p\boldsymbol{\lambda}(cD, D) = pW.$$

This shows that D_1/pD is isotropic. It is maximal isotropic, as D_1/pD has dimension 4. Lemma 2.1 implies that

$$\textstyle\bigwedge^4 F_*(\epsilon_0 M) = p^2 \cdot \bigwedge^4 \epsilon_1 M,$$

as submodules of $\bigwedge^4 \epsilon_1 \boldsymbol{D}'_{\mathbb{Q}}$, for any lattice $M \subset \boldsymbol{D}'_{\mathbb{Q}}$. Applying this with $M = D_1$ shows that $\epsilon_1 D/\epsilon_1 D_1$ has dimension 2. The same argument shows that $\epsilon_0 D/\epsilon_0 D_1$ has dimension 2, and (2-13) follows. □

Corollary 2.11. *There is a bijection $\mathscr{M}(k') \cong \{$Dieudonné lattices in $\boldsymbol{D}'_{\mathbb{Q}}\}$.*

Proof. If $k = k'$ then this is immediate from the equivalence of categories between Dieudonné modules and p-divisible groups: for any point $(G, \iota, \lambda, \varrho) \in \mathscr{M}(k)$ we let D be the Dieudonné module of G, viewed as a lattice in $\boldsymbol{D}_{\mathbb{Q}}$ using the isomorphism of isocrystals $\varrho : D_{\mathbb{Q}} \cong \boldsymbol{D}_{\mathbb{Q}}$. For general k' the argument is the same, using Zink's theory of windows [2001] in place of Dieudonné modules. □

Theorem 2.12. *Given a Dieudonné lattice D, set*

$$L = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD_1 \subset D_1\} \quad \text{and} \quad L^\sharp = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD \subset D\}.$$

The rule $D \mapsto (L, L^\sharp)$ defines a bijection from $p^{\mathbb{Z}}\backslash\{$Dieudonné lattices in $\boldsymbol{D}'_{\mathbb{Q}}\}$ to the set of all pairs of self-dual lattices $(L, L^\sharp)$ in $\boldsymbol{L}'_{\mathbb{Q}}$ such that

(1) $\Phi_*(L) = L^\sharp$,

(2) $(L + L^\sharp)/L$ *has length* 1.

Moreover, $L + L^\sharp = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD_1 \subset D\}$.

The proof of Theorem 2.12 will be given in the next subsection.

Definition 2.13. A *special lattice* is a self-dual W'-lattice $L \subset \boldsymbol{L}'_{\mathbb{Q}}$ such that

$$\mathrm{length}\big((L + \Phi_*(L))/L\big) = 1.$$

Obviously any pair of self-dual lattices $(L, L^\sharp)$ appearing in Theorem 2.12 is determined by its first element, and in fact the function $L \mapsto (L, \Phi_*(L))$ establishes a bijection between the set of special lattices and the set of pairs of self-dual lattices $(L, L^\sharp)$ such that $\Phi_*(L) = L^\sharp$ and $(L+L^\sharp)/L$ has length 1. The only thing to check is the self-duality of $\Phi_*(L)$ for a special lattice L. The inclusion $\Phi_*(L) \subset \Phi_*(L)^\vee$ is clear from the self-duality of L and the relation $[\Phi x, \Phi y] = [x, y]^\sigma$. Equality holds because $\mathrm{Vol}(\Phi_*(L)) = \mathrm{Vol}(L)$ and L is self-dual. The following corollary is now simply a restatement of Theorem 2.12:

Corollary 2.14. *The rule* $D \mapsto \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD_1 \subset D_1\}$ *defines a bijection*

$$p^{\mathbb{Z}} \backslash \{\textit{Dieudonné lattices in } \boldsymbol{D}'_{\mathbb{Q}}\} \cong \{\textit{special lattices in } \boldsymbol{L}'_{\mathbb{Q}}\}.$$

2.5. ***Proof of Theorem 2.12.*** In this subsection we prove Theorem 2.12. Say that a W'-lattice $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ is *nearly self-dual* if $D^\vee = cD$ for some $c \in \mathbb{Q}_p^\times$.

Lemma 2.15. *The construction* $D \mapsto \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD \subset D\}$ *establishes a bijection*

$$p^{\mathbb{Z}} \backslash \{\textit{nearly self-dual lattices } D \subset \boldsymbol{D}'_{\mathbb{Q}}\} \cong \{\textit{self-dual lattices } L^\sharp \subset \boldsymbol{L}'_{\mathbb{Q}}\}.$$

Proof. Start with a nearly self-dual lattice D, and set $L^\sharp = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD \subset D\}$. The condition $D^\vee = cD$ implies that there is some $g \in \mathrm{GU}^0(\boldsymbol{D}'_{\mathbb{Q}})$ such that $D = g\boldsymbol{D}'$, and hence $L^\sharp = g \bullet \boldsymbol{L}'$. As $g\bullet$ respects the quadratic form Q, the self-duality of $\boldsymbol{L}'$ implies the self-duality of $L^\sharp$. Conversely, if we start with a self-dual $L^\sharp \subset \boldsymbol{L}'_{\mathbb{Q}}$, the Clifford algebra $C(L^\sharp)$ is a maximal order in $C(\boldsymbol{L}'_{\mathbb{Q}}) \cong \mathrm{End}_{W'}(\boldsymbol{D}'_{\mathbb{Q}})$, and so there is, up to scaling, a unique lattice $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ satisfying

$$C(L^\sharp) = \mathrm{End}_W(D). \tag{2-14}$$

Choose any $h \in \mathrm{SO}(\boldsymbol{L}'_{\mathbb{Q}})$ such that $L^\sharp = h\boldsymbol{L}'$, and lift h to an element $g \in \mathrm{GU}^0(\boldsymbol{D}'_{\mathbb{Q}})$. By rescaling g we may arrange to have $D = g\boldsymbol{D}'$, and the self-duality of $\boldsymbol{D}'$ implies $D^\vee = \nu(g)^{-1}D$. ☐

Lemma 2.16. *Suppose* $D \subset \boldsymbol{D}'_{\mathbb{Q}}$ *is nearly self-dual,* $L^\sharp \subset \boldsymbol{L}'_{\mathbb{Q}}$ *is self-dual, and* $L^\sharp$ *and* D *are related by* $L^\sharp = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD \subset D\}$. *If* $x \in L^\sharp/pL^\sharp$ *is any nonzero isotropic vector, viewed as an endomorphism of* D/pD *using* (2-14), *the kernel of* x *is an* $\mathcal{O}_E$*-stable Lagrangian subspace with respect to* $c\boldsymbol{\lambda}$. *Conversely, if* $\mathscr{D}_1 \subset D/pD$ *is an* $\mathcal{O}_E$*-stable Lagrangian subspace then* $\{x \in L^\sharp/pL^\sharp : x\mathscr{D}_1 = 0\}$ *is an isotropic line in* $L^\sharp/pL^\sharp$. *This construction establishes a bijection*

$$\{\textit{isotropic lines in } L^\sharp/pL^\sharp\} \cong \{\mathcal{O}_E\textit{-stable Lagrangian subspaces in } D/pD\}.$$

If $\mathscr{L}_1 \subset L^\sharp/pL^\sharp$ *corresponds to* $\mathscr{D}_1 \subset D/pD$ *under this bijection, then*

$$\mathscr{L}_1^\perp = \{x \in L^\sharp/pL^\sharp : x \cdot \mathscr{D}_1 \subset \mathscr{D}_1\}. \tag{2-15}$$

Proof. Abbreviate $\mathscr{L} = L^\sharp/pL^\sharp$ and $\mathscr{D} = D/pD$, so that $\mathscr{D}$ is the unique simple left module over the Clifford algebra $C(\mathscr{L}) \cong M_8(k')$. In particular $C(\mathscr{L}) \cong \mathscr{D}^8$ as left $C(\mathscr{L})$-modules. If $x \in \mathscr{L}$ is any nonzero isotropic vector, the kernel and image of left multiplication by x on $C(\mathscr{L})$ are equal, and hence the kernel and image of $x \in \mathrm{End}(\mathscr{D})$ are also equal. In particular $\ker(x)$ has dimension 4. The relation $\alpha x = x\bar{\alpha}$ for all $\alpha \in \mathcal{O}_E$ shows that $\ker(x)$ is $\mathcal{O}_E$-stable, and the relation $(c\boldsymbol{\lambda})(xs, t) = (c\boldsymbol{\lambda})(s, xt)$ implies that $\ker(x) = x\mathscr{D}$ is totally isotropic.

If $x, y \in \mathscr{L}$ are nonzero isotropic vectors with $\ker(x) = \ker(y)$ then, from the discussion above, $\ker(x) = y\mathscr{D}$ and $\ker(y) = x\mathscr{D}$. In particular $[x, y] =$

$x \circ y + y \circ x = 0$. If x and y are not colinear then (after possibly extending scalars) we can find a $z \in \mathscr{L}$ such that $k'x + k'y + k'z$ is a maximal isotropic subspace of $\mathscr{L}$. The left ideal $C(\mathscr{L})xyz$ has dimension 8 as a k-vector space, and so we must have $\mathscr{D} \cong C(\mathscr{L})xyz$ as left $C(\mathscr{L})$-modules. But it is easy to see by direct calculation that the kernels of left multiplication by x and y on $C(\mathscr{L})xyz$ are different. This contradiction shows that x and y are colinear, and so $x \mapsto \ker(x)$ establishes an injection $\mathscr{L}_1 \mapsto \mathscr{D}_1$ from the set of isotropic lines in $\mathscr{L}$ to the set of $\mathcal{O}_E$-stable Lagrangian subspaces in $\mathscr{D}$.

Endow $\mathscr{D}$ with the $\mathcal{O}_E \otimes_{\mathbb{Z}_p} k'$-valued Hermitian form induced by $c\langle \cdot, \cdot \rangle$. Exactly as in Proposition 2.7, there is an isomorphism of k'-groups $\mathrm{GSpin}(\mathscr{L}) \cong \mathrm{GU}^0(\mathscr{D})$. This isomorphism is compatible, in the obvious sense, with the map $\mathscr{L}_1 \mapsto \mathscr{D}_1$, and so the image of the map is stable under the action of $\mathrm{GU}^0(\mathscr{D})$. But $\mathrm{GU}^0(\mathscr{D})$ acts transitively on the set of $\mathcal{O}_E$-stable Lagrangian subspaces in $\mathscr{D}$, proving surjectivity.

Finally, we verify (2-15). If $\mathscr{L}_1$ corresponds to $\mathscr{D}_1$ under our bijection, then $\mathscr{D}_1 = \ker(y) = y\mathscr{D}$ for any nonzero $y \in \mathscr{L}_1$, and an elementary argument (using $k' \cap yC(\mathscr{L}) = 0$ for the middle $\Leftarrow$) shows that

$$x \perp y \iff xy + yx = 0 \iff xyC(\mathscr{L}) \subset yC(\mathscr{L}) \iff xy\mathscr{D} \subset y\mathscr{D}_1. \qquad \square$$

Proof of Theorem 2.12. Suppose first that D is a Dieudonné lattice. Using the relations $D = F_*(F^{-1}(D))$ and $\langle Fv, Fw \rangle = p\langle v, w \rangle^\sigma$, one can show that the near self-duality of D implies that $D_1 = F^{-1}(pD)$ is also nearly self-dual. Lemma 2.15 then implies that the lattices

$$L = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD_1 \subset D_1\} \quad \text{and} \quad L^\sharp = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD \subset D\} \tag{2-16}$$

are self-dual. The relation $\Phi(x) \circ F = F \circ x$ for all $x \in \boldsymbol{L}'_{\mathbb{Q}}$ implies that $\Phi_*(L) \subset L^\sharp$, and equality must hold as

$$\mathrm{Vol}(\Phi_*(L)) = \mathrm{Vol}(L) = \mathrm{Vol}(L^\sharp).$$

By Proposition 2.10 the k-subspace $D_1/pD \subset D/pD$ is $\mathcal{O}_E$-stable and Lagrangian, and so it follows from Lemma 2.16 that

$$\mathscr{L}_1 = \{x \in L^\sharp/pL^\sharp : x(D_1/pD) = 0\}$$

is an isotropic line in $L^\sharp/pL^\sharp$ with orthogonal complement

$$\mathscr{L}_1^\perp = \{x \in L^\sharp/pL^\sharp : x(D_1/pD) \subset D_1/pD\}.$$

On the other hand, $L \cap L^\sharp = \{x \in L^\sharp : xD_1 \subset D_1\}$, and so

$$(L + L^\sharp)/L \cong L^\sharp/(L \cap L^\sharp) \cong (L^\sharp/pL^\sharp)/\mathscr{L}_1^\perp$$

has length 1.

Now suppose we start with a pair of self-dual lattices $(L, L^\sharp)$ such that $\Phi_*(L) = L^\sharp$ and $(L+L^\sharp)/L$ has length 1. By Lemma 2.15 there are unique (up to scaling) nearly self-dual lattices D_1 and D in $\boldsymbol{D}'_{\mathbb{Q}}$ satisfying (2-16). Set $L_0 = L \cap L^\sharp$, so that $L^\sharp/L_0$ has length 1, and pick any nonzero $y \in L^\sharp/L_0$. The Clifford algebra $C(L^\sharp)$ satisfies

$$C(L^\sharp) = C(L_0) + yC(L_0),$$

where $C(L_0) \subset C(L^\sharp)$ is the W-subalgebra generated by L_0, and so

$$C(L^\sharp)D_1 = C(L_0)D_1 + yC(L_0)D_1 = D_1 + yD_1.$$

This implies $C(L^\sharp)pD_1 \subset D_1 \subset C(L^\sharp)D_1$. The self-duality of $L^\sharp$ implies that $C(L^\sharp)$ is a maximal order in $C(\boldsymbol{L}'_{\mathbb{Q}}) = \mathrm{End}_{W'}(\boldsymbol{D}'_{\mathbb{Q}})$, and so we must have $C(L^\sharp) = \mathrm{End}_{W'}(D)$. As the lattice $C(L^\sharp)D_1$ is obviously stabilized by $C(L^\sharp)$, it must have the form $C(L^\sharp)D_1 = p^k D$ for some integer k. Thus after rescaling D_1 we may assume that $C(L^\sharp)D_1 = D$ and

$$pD \subset D_1 \subset D.$$

The relation $\Phi(x) \circ F = F \circ x$ implies

$$C(L^\sharp)F_*(D_1) = C(\Phi_*(L))F_*(D_1) = F_*(C(L)D_1) = F_*(D_1),$$

and so $F_*(D_1) = p^k D$ for some k. Combining

$$p^{8k} \cdot \mathrm{Vol}(D) = \mathrm{Vol}(F_*(D_1)) = p^4 \cdot \mathrm{Vol}(D_1)$$

and

$$p^8 \cdot \mathrm{Vol}(D) \subset \mathrm{Vol}(D_1) \subset \mathrm{Vol}(D)$$

shows that in fact $F_*(D_1) = pD$. The relations $D_1 = F^{-1}(pD)$ and $F_*(F^{-1}(D)) = D$ follow easily from this, proving that D is a Dieudonné lattice.

It only remains to prove that $L + L^\sharp = \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD_1 \subset D\}$. The inclusion $L + L^\sharp \subset \{x \in \boldsymbol{L}'_{\mathbb{Q}} : xD_1 \subset D\}$ is obvious from (2-16). On the other hand, each side contains $L^\sharp$ with quotient of length 1 (for the right-hand side this follows from Proposition 2.10 and Lemma 2.16). Thus equality holds. □

2.6. ***Vertex lattices and the Bruhat–Tits stratification.*** If we start with a k-point $(G, \lambda, i, \varrho) \in \mathscr{M}(k)$ and let D be the covariant Dieudonné module of G, then $\varrho(D) \subset \boldsymbol{D}_{\mathbb{Q}}$ is a Dieudonné lattice. This construction is simply the $k' = k$ case of the bijection

$$\mathscr{M}(k) \cong \{\text{Dieudonné lattices in } \boldsymbol{D}_{\mathbb{Q}}\}$$

of Corollary 2.11. Combining this with Corollary 2.14 yields a bijection

$$\mathscr{N}(k) \cong \{\text{special lattices in } \boldsymbol{L}_{\mathbb{Q}}\} \tag{2-17}$$

defined by

$$(G, \lambda, i, \varrho) \mapsto \{x \in \boldsymbol{L}_{\mathbb{Q}} : x\varrho(D_1) \subset \varrho(D_1)\},$$

where $D_1 = VD$. Moreover, Theorem 2.12 implies that the special lattice

$$L = \{x \in \boldsymbol{L}_{\mathbb{Q}} : x\varrho(D_1) \subset \varrho(D_1)\}$$

satisfies

$$\Phi(L) = \{x \in \boldsymbol{L}_{\mathbb{Q}} : x\varrho(D) \subset \varrho(D)\}.$$

The next step is to show that the special lattices come in natural families, indexed by certain *vertex lattices* in the $\mathbb{Q}_p$-quadratic space $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$. Using this and the bijection (2-17), we will then express the reduced scheme underlying $\mathscr{N}$ as a union of closed subvarieties indexed by vertex lattices.

Definition 2.17. A *vertex lattice* is a $\mathbb{Z}_p$-lattice $\Lambda \subset \boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ such that

$$p\Lambda \subset \Lambda^{\vee} \subset \Lambda.$$

The *type* of Λ is $t_{\Lambda} = \dim_k(\Lambda/\Lambda^{\vee})$.

Lemma 2.18. *The type of a vertex lattice is either* 2, 4, *or* 6.

Proof. Let Λ be a vertex lattice. Proposition 2.6 implies that $\mathrm{ord}_p(\det(\Lambda))$ is even, from which it follows that the type of Λ is also even. If Λ has type 0 then Λ is self-dual, and hence admits a basis such that the matrix of Q is diagonal with diagonal entries in $\mathbb{Z}_p^{\times}$. But this implies that $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ has Hasse invariant 1, contradicting Proposition 2.6. □

The proof of the following proposition is identical to that of Proposition 4.1 of [Rapoport et al. 2014]. See also Lemma 2.1 of [Vollaard 2010].

Proposition 2.19. *Let $L \subset \boldsymbol{L}_{\mathbb{Q}}$ be a special lattice, and define*

$$L^{(r)} = L + \Phi(L) + \cdots + \Phi^r(L).$$

There is an integer $d \in \{1, 2, 3\}$ such that

$$L = L^{(0)} \subsetneq L^{(1)} \subsetneq \cdots \subsetneq L^{(d)} = L^{(d+1)}.$$

For each $L^{(r)} \subsetneq L^{(r+1)}$ with $0 \le r < d$, the quotient $L^{(r+1)}/L^{(r)}$ is annihilated by p and satisfies $\dim_k(L^{(r+1)}/L^{(r)}) = 1$. Moreover,

$$\Lambda_L = \{x \in L^{(d)} : \Phi(x) = x\}$$

is a vertex lattice of type $2d$ and satisfies $\Lambda_L^{\vee} = \{x \in L : \Phi(x) = x\}$.

By (2-9), each vertex lattice Λ determines a collection of quasi-endomorphisms $\Lambda^\vee \subset \mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}$. Define a closed formal subscheme $\widetilde{\mathscr{M}}_\Lambda \subset \mathscr{M}$ as the locus of points $(G, \iota, \lambda, \varrho)$ such that

$$\varrho^{-1}\Lambda^\vee\varrho = \{\varrho^{-1} \circ x \circ \varrho : x \in \Lambda^\vee\} \subset \mathrm{End}(G).$$

In other words, the locus where the quasi-endomorphisms $\varrho^{-1}\Lambda^\vee\varrho$ of G are actually integral. Set $\widetilde{\mathscr{N}}_\Lambda = p^{\mathbb{Z}}\backslash\widetilde{\mathscr{M}}_\Lambda$, and let $\mathscr{N}_\Lambda$ be the reduced k-scheme underlying $\widetilde{\mathscr{N}}_\Lambda$. The bijection (2-17) identifies

$$\begin{aligned}\mathscr{N}_\Lambda(k) &= \{\text{special lattices } L \text{ such that } \Lambda^\vee \subset \Phi(L)\}\\ &= \{\text{special lattices } L \text{ such that } \Lambda^\vee \subset L\}\\ &= \{\text{special lattices } L \text{ such that } \Lambda_L \subset \Lambda\}.\end{aligned} \tag{2-18}$$

The same proof used in [Rapoport et al. 2014, Proposition 4.3] shows that

$$\mathscr{N}_{\Lambda_1} \cap \mathscr{N}_{\Lambda_2} = \begin{cases}\mathscr{N}_{\Lambda_1\cap\Lambda_2} & \text{if } \Lambda_1 \cap \Lambda_2 \text{ is a vertex lattice,}\\ \varnothing & \text{otherwise,}\end{cases}$$

where the left-hand side is understood to mean the reduced subscheme underlying the scheme-theoretic intersection.

Proposition 2.20. *Each k-scheme $\mathscr{N}_\Lambda$ is projective.*

Proof. Let R_Λ be the W-subalgebra of $\mathrm{End}_W(\boldsymbol{D}_{\mathbb{Q}})$ generated by $\Lambda^\vee$, and let $\widetilde{R}_\Lambda$ be a maximal order in $\mathrm{End}_W(\boldsymbol{D}_{\mathbb{Q}})$ containing R_Λ. It follows from the isomorphism $C(\boldsymbol{L}_{\mathbb{Q}}) \cong \mathrm{End}_W(\boldsymbol{D}_{\mathbb{Q}})$ of Proposition 2.4 that R_Λ is a W-lattice in $\mathrm{End}_W(\boldsymbol{D}_{\mathbb{Q}})$, and hence $\widetilde{R}_\Lambda/R_\Lambda$ is killed by some power of p, say p^M. Up to scaling by powers of p, there is a unique W-lattice $\widetilde{D} \subset \boldsymbol{D}_{\mathbb{Q}}$ such that $\widetilde{R}_\Lambda\widetilde{D} = \widetilde{D}$.

Now suppose $(G, \iota, \lambda, \varrho)$ is a k-point of $\mathscr{N}_\Lambda$. The quasi-isogeny ϱ determines (up to scaling) a W-lattice $D \subset \boldsymbol{D}_{\mathbb{Q}}$ satisfying $R_\Lambda D = D$. It follows from $\widetilde{R}_\Lambda D = \widetilde{D}$ that

$$p^M\widetilde{D} \subset D \subset \widetilde{D},$$

after possibly rescaling D, and so there are integers $a < b$, independent of the point $(G, \iota, \lambda, \varrho)$, such that $p^a\boldsymbol{D} \subset D \subset p^b\boldsymbol{D}$. It follows from this bound and [Rapoport and Zink 1996, Corollary 2.29] that $\mathscr{N}_\Lambda$ is a closed subscheme of a projective scheme, hence is projective. □

An obvious corollary of Proposition 2.19 is that every special lattice L contains some $\Lambda^\vee$ (take $\Lambda = \Lambda_L$), and hence

$$\mathscr{N}_{\mathrm{red}} = \bigcup_\Lambda \mathscr{N}_\Lambda,$$

where the subscript red indicates the underlying reduced scheme. This union is not disjoint, as $\mathcal{N}^{\pm}_{\Lambda_1} \subset \mathcal{N}^{\pm}_{\Lambda_2}$ whenever $\Lambda_1 \subset \Lambda_2$. Define

$$\mathcal{N}^{\circ}_{\Lambda} = \mathcal{N}_{\Lambda} \setminus \bigcup_{\Lambda' \subsetneq \Lambda} \mathcal{N}_{\Lambda'},$$

so that (2-17) identifies

$$\mathcal{N}^{\circ}_{\Lambda}(k) = \{\text{special lattices } L \text{ such that } \Lambda_L = \Lambda\}.$$

It follows easily that $\mathcal{N}_{\Lambda} = \biguplus_{\Lambda' \subset \Lambda} \mathcal{N}^{\circ}_{\Lambda'}$, and that

$$\mathcal{N}_{\mathrm{red}} = \biguplus_{\Lambda} \mathcal{N}^{\circ}_{\Lambda}. \tag{2-19}$$

Abbreviate $\mathcal{N}^{\pm}_{\Lambda} = \mathcal{N}_{\Lambda} \cap \mathcal{N}^{\pm}$ and $\mathcal{N}^{\pm\circ}_{\Lambda} = \mathcal{N}^{\circ}_{\Lambda} \cap \mathcal{N}^{\pm}$. By analogy with [Rapoport et al. 2014; Vollaard 2010; Vollaard and Wedhorn 2011], we call the decomposition (2-19) the *Bruhat–Tits stratification* of $\mathcal{N}_{\mathrm{red}}$. This terminology should be taken with a grain of salt: unlike the situation in those references, the strata in (2-19) are not in bijection with the vertices in the Bruhat–Tits building of the group J^{der}. See Sections 2.7 and 3.6 below.

Remark 2.21. One could also define an *E-vertex lattice* to be an $\mathcal{O}_E$-lattice $\Lambda_E \subset (\bigwedge^2_E \boldsymbol{D}_{\mathbb{Q}})^{\Phi}$ such that $p\Lambda_E \subset \Lambda_E^{\vee} \subset \Lambda_E$, where the dual lattice is taken with respect to $\langle\cdot,\cdot\rangle$. The rule $\Lambda \mapsto \mathcal{O}_E \cdot \Lambda$ establishes a bijection between vertex lattices and E-vertex lattices, with inverse $\Lambda_E \mapsto \{x \in \Lambda_E : x^{\star} = x\}$. The action (2-11) of $\mathrm{GU}(\boldsymbol{D}_{\mathbb{Q}})$ on $\bigwedge^2_E \boldsymbol{D}_{\mathbb{Q}}$ restricts to an action of J on $(\bigwedge^2_E \boldsymbol{D}_{\mathbb{Q}})^{\Phi}$, and induces an action of J on the set of all E-vertex lattices. In particular, J acts on the set of all vertex lattices. This action is compatible with the action of J on $\mathcal{N}$ defined in Section 2.1, in the obvious sense: $g\mathcal{N}_{\Lambda} = \mathcal{N}_{g\bullet\Lambda}$. The restriction of this action to the subgroup J^0 of Remark 2.8 factors through the surjection $J^0 \to \mathrm{SO}(\boldsymbol{L}^{\Phi}_{\mathbb{Q}})$, and agrees with the obvious action of $\mathrm{SO}(\boldsymbol{L}^{\Phi}_{\mathbb{Q}})$ on the set of vertex lattices.

2.7. ***The Bruhat–Tits building.*** In [Garrett 1997, § 20.3] one finds a description of the Bruhat–Tits building of $\mathrm{SO}(\boldsymbol{L}^{\Phi}_{\mathbb{Q}})$ in terms of lattices. See also [Tits 1979, §1.16]. We will translate this description into the language of our vertex lattices. Consider the set $\mathcal{V}^{\mathrm{adm}}$ of all vertex lattices Λ of type 2 or 6. We call such vertex lattices *admissible*, and define an adjacency relation $\sim$ in $\mathcal{V}^{\mathrm{adm}}$ as follows: distinct admissible vertex lattices are adjacent ($\Lambda \sim \Lambda'$) if either:

(1) $\Lambda' \subset \Lambda$ or $\Lambda' \subset \Lambda$.

(2) Λ and Λ' are both type-6 and

$$\dim_{\mathbb{F}_p}(\Lambda/\Lambda \cap \Lambda') = \dim_{\mathbb{F}_p}(\Lambda'/\Lambda \cap \Lambda') = 1,$$
$$\dim_{\mathbb{F}_p}(\Lambda + \Lambda'/\Lambda) = \dim_{\mathbb{F}_p}(\Lambda + \Lambda'/\Lambda') = 1.$$

If Λ and Λ' are of type 6 and are adjacent, then $\Lambda \cap \Lambda'$ is a vertex lattice of type 4 (so is not admissible). We construct an abstract simplicial complex with set of vertices $\mathcal{V}^{\mathrm{adm}}$ as follows: An m-simplex $(0 \le m \le 2)$ of $\mathcal{V}^{\mathrm{adm}}$ is a subset of $m+1$ admissible vertex lattices $\Lambda_0, \Lambda_1, \dots, \Lambda_m$ which are mutually adjacent. The group $\mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$ acts simplicially on $\mathcal{V}^{\mathrm{adm}}$ by $g \in \mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$ taking Λ to $g \cdot \Lambda$.

Now consider the Bruhat–Tits building $\mathcal{BT}$ of $\mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$. We will use the same symbol $\mathcal{BT}$ to denote the underlying simplicial complex.

Proposition 2.22. *There is an* $\mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$*-equivariant simplicial bijection* $\mathcal{BT} \cong \mathcal{V}^{\mathrm{adm}}$.

Furthermore, every vertex lattice of type 4 *is contained in precisely two vertex lattices of type* 6, *and is equal to their intersection.*

Proof. Define a new quadratic space $(V_0, Q_0) = (\boldsymbol{L}_{\mathbb{Q}}^{\Phi}, p^{-1}Q)$, and note that, by Proposition 2.6, $V_0 \cong \mathbb{H}^2 \oplus \mathbb{Q}_{p^2}$. The rule $\Lambda \mapsto p\Lambda$ defines a bijection from the set of vertex lattices in $\boldsymbol{L}_{\mathbb{Q}}^{\Phi}$ to the set of lattices $L \subset V_0$ satisfying

$$L \subset L^* \subset p^{-1}L.$$

Here L^* is the dual lattice of L with respect to the quadratic form Q_0. The isomorphism $\mathcal{BT} \cong \mathcal{V}^{\mathrm{adm}}$ now follows from the description and properties of the affine building of $\mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi}) \cong \mathrm{SO}(V_0)$ found in [Garrett 1997, Section 20.3].

If Λ is a type-4 vertex lattice, the lattice $L = p\Lambda$ in V_0 satisfies $\dim(L^*/L) = 2$. Moreover, the k-quadratic space L^*/L is a hyperbolic plane (choose a basis of L for which the bilinear form has diagonal matrix, and use the fact that V_0 has Hasse invariant 1), and so contains exactly two isotropic lines. Those lines have the form L_1/L and L_2/L, and $p^{-1}L_1$ and $p^{-1}L_2$ are the unique type-6 vertex lattices containing Λ. □

We can also construct a simplicial complex $\mathcal{V}$ with vertices the set of *all* vertex lattices as follows (compare to [Rapoport et al. 2014, §3]). We call two distinct vertex lattices Λ and Λ' neighbors if $\Lambda \subset \Lambda'$ or $\Lambda' \subset \Lambda$. An m-simplex $(m \le 2)$ in $\mathcal{V}$ is formed by vertex lattices $\Lambda_0, \Lambda_1, \dots, \Lambda_m$ such that any two members of this set are neighbors. The vertex lattices of type 4 are in bijection with pairs of adjacent type-6 vertex lattices. Hence a vertex lattice of type 4 corresponds to an edge in the Bruhat–Tits building between type-6 vertex lattices. From basic properties of the Bruhat–Tits building, we deduce the following:

Corollary 2.23. *The group* $\mathrm{SO}(\boldsymbol{L}_{\mathbb{Q}}^{\Phi})$ *acts transitively on the set of vertex lattices of a given type, and any two vertex lattices are connected by a sequence of adjacent vertices in* $\mathcal{V}$. *In particular, the group* J, *and even the subgroup* J^0, *act transitively on the set of vertex lattices of a given type* (*under the action of* Remark 2.21).

3. Deligne–Lusztig varieties and the Bruhat–Tits strata

In this section we show that for any vertex lattice Λ, the varieties

$$\mathscr{N}_\Lambda = \mathscr{N}_\Lambda^+ \uplus \mathscr{N}_\Lambda^- \quad \text{and} \quad \mathscr{N}_\Lambda^\circ = \mathscr{N}_\Lambda^{\circ+} \uplus \mathscr{N}_\Lambda^{\circ-}$$

of Section 2.6 can be identified with varieties over k defined purely in terms of the linear algebra of the k-quadratic space $\Omega = (\Lambda/\Lambda^\vee) \otimes_{\mathbb{F}_p} k$.

3.1. ***Deligne–Lusztig varieties.*** Let us recall the general definition of Deligne–Lusztig varieties. Suppose that G_0 is a connected reductive group over the finite field $\mathbb{F}_p$, and set $G = G_0 \otimes_{\mathbb{F}_p} k$. We will also use the symbol G to denote the abstract group of k-valued points of G_0. Denote by $\Phi : G \to G$ the Frobenius morphism. By Lang's theorem, G_0 is quasi-split, and so we may choose a maximal torus $T \subset G$ and a Borel subgroup containing T, both defined over $\mathbb{F}_p$. The Weyl group W that corresponds to the pair (T, B) is acted upon by Φ, and the group W with its Φ-action does not depend on our choices. In fact, in [Deligne and Lusztig 1976] a Weyl group W with Φ-action is defined as a projective limit over all choices of pairs (T, B), without having to assume that these pairs are Φ-stable.

Let $\Delta^* = \{\alpha_1, \dots, \alpha_n\}$ be the set of simple roots corresponding to the pair (T, B), and consider the corresponding simple reflections $s_i = s_{\alpha_i}$ in the Weyl group W. For $I \subset \Delta^*$, let W_I be the subgroup of W generated by $\{s_i : i \in I\}$, and consider the corresponding parabolic subgroup $P_I = BW_IB$. The quotient G/P_I parametrizes parabolic subgroups of G of type I. Suppose $J \subset \Delta^*$ is another subset with corresponding standard parabolic P_J. Since

$$G = \biguplus_{w \in W_I \backslash W / W_J} P_I w P_J,$$

we have a bijection

$$P_I \backslash G / P_J \cong W_I \backslash W / W_J.$$

Composing this with $G/P_I \times G/P_J \to P_I \backslash G/P_J$ given by $(g_1, g_2) \mapsto g_1^{-1} g_2$ defines the *relative position invariant*

$$\mathrm{inv} : G/P_I \times G/P_J \to W_I \backslash W / W_J.$$

The Frobenius $\Phi : G \to G$ induces $\Phi : G/P_I \to G/P_{\Phi(I)}$.

Definition 3.1. For $w \in W_I \backslash W / W_{\Phi(I)}$, the *Deligne–Lusztig variety* $X_{P_I}(w)$ is the locally closed reduced subscheme of G/P_I with k-points

$$X_{P_I}(w) = \{gP_I \in G/P_I : \mathrm{inv}(g, \Phi(g)) = w\}.$$

The variety $X_{P_I}(w)$ is actually defined over the unique extension of degree r of $\mathbb{F}_p$ in k, where r is the smallest positive integer for which $\Phi^r(I) = I$.

Proposition 3.2. *The Deligne–Lusztig variety $X_{P_I}(w)$ is smooth of pure dimension*

$$\dim X_{P_I}(w) = \ell(w) + \dim(G/P_{I\cap\Phi(I)}) - \dim(G/P_I).$$

If $I = \Phi(I)$ then $\dim X_{P_I}(w) = \ell_I(w) - \ell(w_I)$, where w_I is the longest element in W_I and $\ell_I(w)$ is the maximal length of an element in $W_I w W_I$. Taking $I = \varnothing$, the variety $X_B(w)$ is irreducible of dimension $\dim X_B(w) = \ell(w)$.

Proof. This is standard. See, e.g., [Vollaard and Wedhorn 2011, Section 3.4]. □

Remark 3.3. If $w = 1$, then $X_{P_I}(1)$ can be identified with the intersection of the image of the closed immersion $G/P_{I\cap\Phi(I)} \hookrightarrow G/P_I \times G/P_{\Phi(I)}$ with the graph of the Frobenius $\Phi : G/P_I \to G/P_{\Phi(I)}$. In particular $X_{P_I}(1)$ is projective. If

$$\bigcup_{r\geq 0} \Phi^r(I) = \Delta^*$$

then $X_{P_I}(1)$ is also irreducible, by [Bonnafé and Rouquier 2006].

3.2. ***An even orthogonal group.*** We now consider the case that G_0 is a nonsplit special orthogonal group in an even number of variables. Let Ω_0 be an $\mathbb{F}_p$-vector space of dimension $2d$ equipped with a nondegenerate nonsplit quadratic form. There is a basis $\{e_1, \dots, e_d, f_1, \dots, f_d\}$ of $\Omega = \Omega_0 \otimes_{\mathbb{F}_p} k$ such that $\langle e_1, \dots, e_d\rangle$ and $\langle f_1, \dots, f_d\rangle$ are isotropic, $[e_i, f_j] = \delta_{ij}$, and the Frobenius

$$\Phi = \mathrm{id} \otimes \sigma$$

acting on Ω fixes e_i and f_i for $1 \leq i \leq d-1$, and interchanges e_d with f_d. Note that Ω contains no Φ-invariant Lagrangian subspaces. Abbreviate $G_0 = \mathrm{SO}(\Omega_0)$ and $G = \mathrm{SO}(\Omega)$.

Denote by $\mathrm{OGr}(r)$ the scheme whose functor of points assigns to a k-scheme S the set of all totally isotropic local $\mathcal{O}_S$-module direct summands $\mathcal{L} \subset \Omega \otimes_k \mathcal{O}_S$ of rank r. In particular, $\mathrm{OGr}(d)$ is the moduli space of Lagrangian subspaces of Ω. Denote by $\mathrm{OGr}(d-1, d)$ the scheme whose functor of points assigns to a k-scheme S the set of all flags of totally isotropic local $\mathcal{O}_S$-module direct summands $\mathcal{L}_{d-1} \subset \mathcal{L}_d \subset \Omega \otimes_k \mathcal{O}_S$ of rank $d-1$ and d, respectively. The following lemma is elementary:

Lemma 3.4. *For each totally isotropic local $\mathcal{O}_S$-module direct summand*

$$\mathcal{L}_{d-1} \subset \Omega \otimes_k \mathcal{O}_S$$

of rank $d-1$, there are exactly two totally isotropic local $\mathcal{O}_S$-module direct summands of rank d containing $\mathcal{L}_{d-1}$.

In other words, the forgetful map $\mathrm{OGr}(d-1, d) \to \mathrm{OGr}(d-1)$ is a two-to-one cover. In fact, the Grassmannian $\mathrm{OGr}(d-1, d)$ has two connected components, which are interchanged by the action of any orthogonal transformation of determinant -1. Each of the two components maps isomorphically to $\mathrm{OGr}(d-1)$ under the forgetful map. Label the two components as

$$\mathrm{OGr}(d-1,d) = \mathrm{OGr}^{+}(d-1,d) \uplus \mathrm{OGr}^{-}(d-1,d)$$

in such a way that the flags

$$\langle e_1,\dots,e_{d-1}\rangle \subset \langle e_1,\dots,e_{d-1},e_d\rangle \tag{3-1}$$

and

$$\langle e_1,\dots,e_{d-1}\rangle \subset \langle e_1,\dots,e_{d-1},f_d\rangle \tag{3-2}$$

define k-points of $\mathrm{OGr}^{+}(d-1,d)$ and $\mathrm{OGr}^{-}(d-1,d)$, respectively.

Denote by $\mathscr{X} \subset \mathrm{OGr}(d)$ the reduced closed subscheme with k-points

$$\mathscr{X} = \{\mathcal{L} \in \mathrm{OGr}(d) : \dim(\mathcal{L} + \Phi(\mathcal{L})) = d+1\}.$$

There is a closed immersion $\mathscr{X} \to \mathrm{OGr}(d-1,d)$ sending

$$\mathcal{L} \mapsto \mathcal{L} \cap \Phi(\mathcal{L}) \subset \mathcal{L},$$

and the open and closed subvariety $\mathscr{X}^{\pm} = \mathscr{X} \cap \mathrm{OGr}^{\pm}(d-1,d)$ of $\mathscr{X}$ is identified with

$$\mathscr{X}^{\pm} = \{\mathcal{L}_{d-1} \subset \mathcal{L}_d \in \mathrm{OGr}^{\pm}(d-1,d) : \mathcal{L}_{d-1} \subset \Phi(\mathcal{L}_d)\}. \tag{3-3}$$

Remark 3.5. Although we have defined $\mathrm{OGr}(d-1,d)$ and $\mathscr{X}$ as k-schemes, they both have natural $\mathbb{F}_p$-structures. The Frobenius morphism from $\mathrm{OGr}(d-1,d)$ to itself interchanges the flags (3-1) and (3-2), and hence interchanges the two connected components. It follows that $\mathscr{X}^{\pm} \cong \sigma^*\mathscr{X}^{\mp}$, and that the individual components $\mathscr{X}^{+}$ and $\mathscr{X}^{-}$ have natural $\mathbb{F}_{p^2}$-structures.

Our choice of basis of Ω determines a maximal Φ-stable torus $T \subset G$. Set

$$\begin{aligned} \mathcal{F}_i^{\pm} &= \langle e_1,\dots,e_i\rangle \ \text{ for } 1 \le i \le d-1, \\ \mathcal{F}_d^{+} &= \langle e_1,\dots,e_{d-1},e_d\rangle, \\ \mathcal{F}_d^{-} &= \langle e_1,\dots,e_{d-1},f_d\rangle. \end{aligned} \tag{3-4}$$

This gives two "standard" isotropic flags $\mathcal{F}_\bullet^{+}$ and $\mathcal{F}_\bullet^{-}$ in Ω satisfying $\mathcal{F}_\bullet^{\pm} = \Phi(\mathcal{F}_\bullet^{\mp})$. These flags have the same stabilizer $B \subset G$, which is a Φ-stable Borel subgroup containing T. The corresponding set of simple reflections in the Weyl group is

$$\{s_1,\dots,s_{d-2},t^{+},t^{-}\}$$

where:

- s_i interchanges e_i with e_{i+1}, f_i with f_{i+1}, and fixes the other basis elements.

- t^+ interchanges e_{d-1} with e_d, f_{d-1} with f_d, and fixes the other basis elements.
- t^- interchanges e_{d-1} with f_d, f_{d-1} with e_d, and fixes the other basis elements.

Notice that $\Phi(s_i) = s_i$ and $\Phi(t^\pm) = t^\mp$, and so the products

$$w^\pm = t^\mp s_{d-2} \cdots s_2 s_1$$

are *Coxeter elements*: products of exactly one representative from each Φ-orbit in the set of simple reflections above. More generally, define $w_0 = 1$, $w_1^\pm = t^\mp$, and

$$w_r^\pm = t^\mp \cdot s_{d-2} \cdots s_{d-r}$$

for $2 \le r \le d-1$. In particular, $w_{d-1}^\pm = w^\pm$. Define parabolic subgroups

$$B = P_{d-1} \subset P_{d-2} \subset \cdots \subset P_0 \subset P^\pm$$

of G as follows: set $P_{d-1} = P_{d-2} = B$, and for $0 \le r \le d-2$ let P_r be the parabolic corresponding to the set $\{s_1, \ldots, s_{d-(r+2)}\}$. Define $P^\pm$ to be the maximal parabolic corresponding to $\{s_1, \ldots, s_{d-2}, t^\pm\}$. One can easily check that P_0 is the stabilizer in G of $\mathcal{F}_{d-1}^\pm$, and so

$$G/P_0 \cong \mathrm{OGr}(d-1). \tag{3-5}$$

More generally, P_r is the stabilizer of the standard isotropic flag $\mathcal{F}_{d-r-1}^\pm \subset \cdots \subset \mathcal{F}_d^\pm$. Similarly, $P^\pm$ is the stabilizer of the Lagrangian subspace $\mathcal{F}_d^\pm$, and so

$$G/P^\pm \cong \mathrm{OGr}(d). \tag{3-6}$$

Proposition 3.6. *The isomorphism (3-6) identifies $\mathscr{X}^\pm$ with the Deligne–Lusztig variety $X_{P^\pm}(1)$. In particular, $\mathscr{X}^\pm$ is projective, irreducible, and smooth of dimension $d-1$.*

Proof. Note that $P_0 = P^+ \cap P^-$, and that the Frobenius Φ interchanges P^+ and P^-. The two projections $G/P_0 \to G/P^\pm$ combine to give closed immersions

$$i^+ : G/P_0 \to G/P^+ \times G/P^- \quad \text{and} \quad i^- : G/P_0 \to G/P^- \times G/P^+,$$

while the Frobenius induces morphisms $\Phi : G/P^+ \to G/P^-$ and $\Phi : G/P^- \to G/P^+$ with graphs $\Gamma_\Phi^+ \subset G/P^+ \times G/P^-$ and $\Gamma_\Phi^- \subset G/P^- \times G/P^+$, respectively. The isomorphisms (3-5) and (3-6) identify the intersection of $\Gamma_\Phi^\pm$ and the image of $i^\pm$ with the set of flags $\mathcal{L}_{d-1} \subset \mathcal{L}_d \in \mathrm{OGr}^\pm(d-1, d)$ such that $\mathcal{L}_{d-1} \subset \Phi(\mathcal{L}_d)$. By (3-3), this intersection is isomorphic to $\mathscr{X}^+$. All of the claims now follow from Remark 3.3, together with the dimension formula

$$\dim X_{P^\pm}(1) = \dim(G/P_0) - \dim(G/P^\pm) = d-1$$

of Proposition 3.2. □

It is also useful to view $\mathscr{X}^{\pm}$ as the closure of a Deligne–Lusztig variety in the flag variety G/P_0.

Lemma 3.7. *The isomorphism* $\mathrm{OGr}^{\pm}(d-1,d) \cong \mathrm{OGr}(d-1) \cong G/P_0$ *identifies* $\mathscr{X}^{\pm}$ *with the closure in* G/P_0 *of the Deligne–Lusztig variety*

$$X_{P_0}(t^{\mp}) = \{g \in G/P_0 : \mathrm{inv}(g, \Phi(g)) = t^{\mp}\}.$$

Proof. Using (3-3), we may characterize the k-points of $\mathscr{X}^{\pm}$ by

$$\mathscr{X}^{\pm} = \bigl\{\mathcal{L}_{d-1} \subset \mathcal{L}_d \in \mathrm{Gr}^{\pm}(d-1,d) : \\ \Phi(\mathcal{L}_{d-1}) = \mathcal{L}_{d-1} \text{ or } \mathcal{L}_{d-1} + \Phi(\mathcal{L}_{d-1}) = \Phi(\mathcal{L}_d)\bigr\}.$$

Recalling the standard isotropic flags of (3-4), the rule $g \mapsto g\mathcal{F}^{\pm}_{d-1} \subset g\mathcal{F}^{\pm}_d$ defines an isomorphism

$$X_{P_0}(1) \cong \{\mathcal{L}_{d-1} \subset \mathcal{L}_d \in \mathrm{Gr}^{\pm}(d-1,d) : \mathcal{L}_{d-1} = \Phi(\mathcal{L}_{d-1})\},$$

while the same rule defines an isomorphism

$$X_{P_0}(t^{\mp}) \cong \{\mathcal{L}_{d-1} \subset \mathcal{L}_d \in \mathrm{Gr}^{\pm}(d-1,d) : \mathcal{L}_{d-1} + \Phi(\mathcal{L}_{d-1}) = \Phi(\mathcal{L}_d)\}.$$

Thus $\mathscr{X}^{\pm}$ is the disjoint union of $X_{P_0}(1)$ and $X_{P_0}(t^{\mp})$. Elementary properties of the Bruhat order (see Section 8.5 of [Springer 1998], for example) imply that

$$\overline{X_{P_0}(t^{\mp})} = X_{P_0}(1) \uplus X_{P_0}(t^{\mp}),$$

completing the proof. □

The following proof is essentially the same as [Rapoport et al. 2014, Proposition 5.5].

Proposition 3.8. *There is a stratification*

$$\mathscr{X}^{\pm} = \biguplus_{r=0}^{d-1} X_{P_r}(w_r^{\pm})$$

of $\mathscr{X}^{\pm}$ *into a disjoint union of locally closed subvarieties. The stratum* $X_{P_r}(w_r^{\pm})$ *is smooth of pure dimension* r, *and has closure*

$$\overline{X_{P_r}(w_r^{\pm})} = X_{P_0}(w_0^{\pm}) \uplus \cdots \uplus X_{P_r}(w_r^{\pm}).$$

The highest-dimensional stratum $X_{P_{d-1}}(w_{d-1}^{\pm}) = X_B(w^{\pm})$ *is irreducible, open, and dense.*

Proof. Suppose $\mathcal{L}$ is a k-point of $\mathscr{X}^{\pm} \subset \mathrm{Gr}(d)$, and define

$$\mathcal{L}^{(i)} = \mathcal{L} \cap \Phi(\mathcal{L}) \cap \cdots \cap \Phi^i(\mathcal{L}).$$

An inductive argument, using

$$\mathcal{L}^{(i)} \cap \Phi(\mathcal{L}^{(i)}) = \mathcal{L}^{(i-1)} \cap \Phi(\mathcal{L}^{(i-1)}) \cap \Phi^2(\mathcal{L}^{(i-1)}),$$

shows that $\mathcal{L}^{(i+1)}$ has codimension at most 1 in $\mathcal{L}^i$. Denote by $\mathscr{X}_r^\pm \subset \mathscr{X}^\pm$ the reduced closed subscheme whose k-valued points are those $\mathcal{L}$ satisfying $\mathcal{L}^{(r+2)} = \mathcal{L}^{(r+1)}$. The complement $\mathscr{X}_r^\pm \setminus \mathscr{X}_{r-1}^\pm$ is the locally closed subvariety of $\mathscr{X}^\pm$ consisting of those $\mathcal{L}$ for which

$$\mathcal{L}^{(r+2)} = \mathcal{L}^{(r+1)} \subsetneq \cdots \subsetneq \mathcal{L}^{(1)} \subsetneq \mathcal{L}^{(0)} = \mathcal{L}. \tag{3-7}$$

Recalling that the parabolic subgroup P_r is the stabilizer of the standard isotropic flag $\mathcal{F}_{d-r-1}^\pm \subset \cdots \subset \mathcal{F}_d^\pm$ of length $r+2$, we obtain a morphism $\mathscr{X}_r^\pm \setminus \mathscr{X}_{r-1}^\pm \to G/P_r$ by sending $\mathcal{L}$ to the flag (3-7). By similar reasoning as [Rapoport et al. 2014, Proposition 5.5], this defines an isomorphism $\mathscr{X}_r^\pm \setminus \mathscr{X}_{r-1}^\pm \cong X_{P_r}(w_r^\pm)$ with inverse $g \mapsto g\mathcal{F}_d^\pm$. All claims now follow easily. □

3.3. ***A few special cases.*** We continue to let $G_0 = \mathrm{SO}(\Omega_0)$, where Ω_0 is the nonsplit quadratic space over $\mathbb{F}_p$ of dimension $2d$, $\Omega = \Omega_0 \otimes k$, and $G = \mathrm{SO}(\Omega)$. In the applications we will only need to consider $d \le 3$, and in these cases the structure of the k-variety $\mathscr{X}$ (with its $\mathbb{F}_{p^2}$-structure of Remark 3.5) can be made more explicit.

(a) First suppose $d = 1$. In this case $\mathscr{X}^\pm$ is a single point, defined over $\mathbb{F}_{p^2}$.

(b) Now suppose $d = 2$. In this case $P_1 = P_0 = B$, and the stratification of Proposition 3.8 is

$$\mathscr{X}^\pm = X_B(1) \uplus X_B(t^\mp),$$

where $X_B(1)$ is a 0-dimensional closed subvariety and the open stratum $X_B(t^\mp)$ has dimension 1. The Dynkin diagram identity $D_2 = A_1 \times A_1$ corresponds to an exceptional isomorphism $\mathrm{Spin}(\Omega) \cong \mathrm{SL}_2 \times \mathrm{SL}_2$. Indeed, the even Clifford algebra $C_0(\Omega_0)$ is isomorphic to $M_2(\mathbb{F}_{p^2})$, and hence $C_0(\Omega) \cong M_2(k) \times M_2(k)$. This isomorphism restricts to an isomorphism of algebraic groups

$$\mathrm{GSpin}(\Omega) \cong \{(x, y) \in \mathrm{GL}_2 \times \mathrm{GL}_2 : \det(x) = \det(y)\}$$

over k, which in turn determines an isomorphism of k-varieties $G/P_0 \cong \mathbb{P}^1 \times \mathbb{P}^1$ in such a way that the Frobenius morphism on the left corresponds to $(x, y) \mapsto (\Phi(y), \Phi(x))$ on the right. The subvarieties $\mathscr{X}^\pm \subset G/P_0$ are identified with

$$\mathscr{X}^+ = \{(\Phi(x), x) : x \in \mathbb{P}^1\},$$
$$\mathscr{X}^- = \{(x, \Phi(x)) : x \in \mathbb{P}^1\}.$$

Therefore, both $\mathscr{X}^+$ and $\mathscr{X}^-$ are isomorphic (over $\mathbb{F}_{p^2}$) to $\mathbb{P}^1$. The closed stratum $X_B(1)$ corresponds to the $\mathbb{F}_{p^2}$-rational points of $\mathbb{P}^1$.

(c) Finally, suppose $d = 3$. In this case

$$\mathscr{X}^{\pm} = X_{P_0}(1) \uplus X_B(t^{\mp}) \uplus X_B(t^{\mp}s_1). \tag{3-8}$$

The open stratum $X_B(t^{\mp}s_1)$ has dimension 2, the stratum $X_B(t^{\mp})$ is locally closed of dimension 1, and the closed stratum $X_{P_0}(1)$ has dimension 0. To continue, we will use the Dynkin diagram isomorphism $D_3 = A_3$, corresponding to an isomorphism between the adjoint forms of G and a unitary group in 4 variables, as in Proposition 2.7 and Remark 2.8.

Let $\mathcal{V}_0$ be the 4-dimensional $\mathbb{F}_{p^2}$-vector space with basis e_1, e_2, e_3, e_4, endowed with the split $\mathbb{F}_{p^2}/\mathbb{F}_p$-Hermitian form defined by $\langle e_i, e_{5-j}\rangle = \delta_{ij}$. Denote by U_0 the unitary group of $\mathcal{V}_0$, an algebraic group over $\mathbb{F}_p$, so that $U = U_0 \times_{\mathbb{F}_p} k$ acts on $\mathcal{V} = \mathcal{V}_0 \otimes_{\mathbb{F}_p} k$. Recall the isomorphism $\mathbb{F}_{p^2} \otimes_{\mathbb{F}_p} k \simeq k \oplus k$ defined by $x \otimes a \mapsto (xa, x^p a)$, and denote by ϵ_0 and ϵ_1 the idempotents that correspond to $(1, 0)$ and $(0, 1)$, so that

$$\epsilon_0\mathcal{V} \cong \mathcal{V}_0 \otimes_{\mathbb{F}_{p^2}} k. \tag{3-9}$$

The action of U on $\epsilon_0\mathcal{V}$ defines an isomorphism $U \cong \mathrm{GL}_4$. The diagonal torus and the standard Borel subgroup of upper-triangular matrices in GL_4 give a maximal torus and a Borel subgroup of U, both defined over $\mathbb{F}_p$. Given $r, s \geq 0$ with $r+s = 4$, the pair (r, s), viewed as an ordered partition of 4, defines a parabolic subgroup $P_{(r,s)}$ containing B with Levi component $\mathrm{GL}_r \times \mathrm{GL}_s$. The parabolic subgroup $P_{(r,s)}$ is defined over $\mathbb{F}_{p^2}$ and satisfies $\Phi(P_{(r,s)}) = P_{(s,r)}$.

Let $\mathrm{Gr}(r)$ be the Grassmanian of r-planes in $\epsilon_0\mathcal{V}$. The above isomorphism $U \cong \mathrm{GL}_4$ induces an isomorphism $U/P_{(r,s)} \cong \mathrm{Gr}(r)$ defined over $\mathbb{F}_{p^2}$, and the Frobenius morphism $\Phi : U/P_{(r,s)} \to U/P_{(s,r)}$ corresponds to a morphism

$$\Phi : \mathrm{Gr}(r) \to \mathrm{Gr}(s) \tag{3-10}$$

which can be described, as in [Vollaard 2010], as follows. Consider the k-valued form $\langle\langle \cdot, \cdot \rangle\rangle$ on (3-9) defined by

$$\langle\langle x \otimes a, y \otimes b\rangle\rangle = \langle x, y\rangle \otimes ab^p.$$

It is k-linear in the first variable but Frobenius-semilinear in the second. For a subspace $\mathcal{U} \subset \epsilon_0\mathcal{V}$, denote by $\mathcal{U}^{\llcorner}$ the perpendicular of $\mathcal{U}$ for the form $\langle\langle \cdot, \cdot \rangle\rangle$. If $\mathcal{U}$ is $\mathbb{F}_{p^2}$-rational then $\mathcal{U}^{\llcorner} = \mathcal{U}^{\perp}$. If $\dim_k(\mathcal{U}) = r$ then $\dim_k(\mathcal{U}^{\llcorner}) = s$, and, according to [Vollaard 2010, Lemma 2.12], the morphism (3-10) is given by $\Phi(\mathcal{U}) = \mathcal{U}^{\llcorner}$ on k-valued points. Using this description of Φ, the unitary Deligne–Lusztig variety $X_{P_{(r,s)}}(1) \subset \mathrm{Gr}(r)$ is seen to be

$$X_{P_{(r,s)}}(1) = \begin{cases} \{\mathcal{U} \subset \epsilon_0\mathcal{V} : \dim_k(\mathcal{U}) = r,\ \mathcal{U} \subset \mathcal{U}^{\llcorner}\} & \text{if } r \leq s, \\ \{\mathcal{U} \subset \epsilon_0\mathcal{V} : \dim_k(\mathcal{U}) = r,\ \mathcal{U}^{\llcorner} \subset \mathcal{U}\} & \text{if } s \leq r. \end{cases}$$

By Remark 3.3 these Deligne–Lusztig varieties are projective and smooth.

By inspecting the Dynkin diagram identity $D_3 = A_3$ we can see that the exceptional isomorphism between the adjoint forms of G and U can be chosen so that the parabolic subgroups P^+, P^- and P_0 of G correspond to $P_{(1,3)}$, $P_{(3,1)}$ and $P_{(1,3)} \cap P_{(3,1)}$ of $U \cong \mathrm{GL}_4$ respectively. Therefore, the Deligne–Lusztig variety $\mathscr{X}^+$ is isomorphic to the unitary Deligne–Lusztig variety

$$X_{P_{(1,3)}}(1) = \{\mathcal{U} \subset k^4 : \dim_k(\mathcal{U}) = 1,\ \mathcal{U} \subset \mathcal{U}^{\llcorner}\}.$$

Similarly $\mathscr{X}^-$ is isomorphic to the unitary Deligne–Lusztig variety

$$X_{P_{(3,1)}}(1) = \{\mathcal{U} \subset k^4 : \dim_k(\mathcal{U}) = 3,\ \mathcal{U}^{\llcorner} \subset \mathcal{U}\}.$$

Therefore, both $\mathscr{X}^+$ and $\mathscr{X}^-$ are isomorphic over $\mathbb{F}_{p^2}$ to the smooth hypersurface in $\mathbb{P}^3$ given by the homogeneous equation

$$x_1x_4^p + x_2x_3^p + x_3x_2^p + x_4x_1^p = 0.$$

In fact, since all nondegenerate Hermitian forms on $\mathcal{V}_0 = \mathbb{F}_{p^2}^4$ are isomorphic we can also determine equations for the unitary Deligne–Lusztig varieties using the Hermitian form given by the identity matrix I_4. This gives the Fermat hypersurface

$$x_0^{p+1} + x_1^{p+1} + x_2^{p+1} + x_3^{p+1} = 0,$$

which is isomorphic to the surface above.

The stratification (3-8) of $\mathscr{X}^+$ now corresponds to the stratification of the unitary Deligne–Lusztig variety $X_{P_{(1,3)}}(1)$ studied in [Vollaard 2010, Theorem 2.15]. The Frobenius $\sigma : k \to k$ defines an operator on $\mathcal{V}$ which interchanges the two summands $\mathcal{V} = \epsilon_0\mathcal{V} \oplus \epsilon_1\mathcal{V}$. Thus we obtain an operator $\tau = \sigma^2$ on $\epsilon_0\mathcal{V}$. Any k-subspace $\mathcal{U} \subset \epsilon_0\mathcal{V}$ satisfies $\tau(\mathcal{U}) = (\mathcal{U}^{\llcorner})^{\llcorner}$. The open 2-dimensional stratum of $X_{P_{(1,3)}}(1)$ has k-valued points corresponding to lines $\mathcal{U}$ such that

$$\dim_k(\mathcal{U} + \tau(\mathcal{U})) = 2,$$
$$\dim_k(\mathcal{U} + \tau(\mathcal{U}) + \tau^2(\mathcal{U})) = 3.$$

The 1-dimensional stratum has k-valued points corresponding to lines $\mathcal{U}$ such that $\dim_k(\mathcal{U} + \tau(\mathcal{U})) = 2$ and $\mathcal{U} + \tau(\mathcal{U})$ is τ-invariant (i.e., $\mathbb{F}_{p^2}$-rational). Finally, the 0-dimensional stratum consists of k-valued points corresponding to lines $\mathcal{U}$ which are τ-invariant. In other words, the 0-dimensional stratum of $\mathscr{X}^+$ is just the set of $\mathbb{F}_{p^2}$-rational points. For a k-valued point $\mathcal{U}$ on the 1-dimensional stratum, set $\mathcal{U}' = \mathcal{U} + \tau(\mathcal{U})$. This is an $\mathbb{F}_{p^2}$-rational plane with $\mathcal{U}'^{\llcorner} = \mathcal{U}'^{\perp} = \mathcal{U}'$. The irreducible components of the 1-dimensional stratum are parametrized by such planes. Indeed, conversely, given an $\mathbb{F}_{p^2}$-rational plane $\mathcal{U}'$ which is isotropic ($\mathcal{U}' = \mathcal{U}'^{\perp}$), we obtain a closed subscheme of $X_{P_{(1,3)}}(1)$ with points corresponding to all lines $\mathcal{U}$ with

$\mathcal{U} \subset \mathcal{U}'$. This subscheme is isomorphic to $\mathbb{P}^1$ and gives the Zariski closure of the corresponding irreducible component of the 1-dimensional stratum.

We can now also determine the number of components of the strata:

- The 0-dimensional stratum consists of $(p^3+1)(p^2+1)$ points. Indeed, observe that, as in [Vollaard and Wedhorn 2011, Example 5.6], we can calculate that the number of $\mathbb{F}_{p^2}$-valued points of the Fermat surface above is equal to $(p^3+1)(p^2+1)$. (Note that there is a typographical error in [loc. cit.]: the summation for Σ_l should start at $j = 0$.) This is equal to the number of $\mathbb{F}_{p^2}$-rational lines $\mathcal{U} \subset \mathbb{F}_{p^2}^4$ such that $\mathcal{U} \subset \mathcal{U}^\perp$, where the orthogonal is with respect to the (standard) Hermitian form $\langle\cdot,\cdot\rangle$ on $\mathbb{F}_{p^2}^4$. Therefore, we have $(p^3+1)(p^2+1)$ components of the 0-dimensional stratum.
- The 1-dimensional stratum has $(p^3+1)(p+1)$ components. Note that by the above, the Zariski closure of each component is isomorphic to a projective line $\mathbb{P}^1$ over $\mathbb{F}_{p^2}$ and the corresponding component is the complement of all $\mathbb{F}_{p^2}$-rational points in this line. To determine the number of irreducible components of the 1-dimensional stratum, we start by counting the number of such components whose closure passes a given $\mathbb{F}_{p^2}$-rational point, i.e., the number of copies of $\mathbb{P}^1$ in our configuration that cross at that point: By the above, this count is given by the number of $\mathbb{F}_{p^2}$-rational planes $\mathcal{U}'$ which are isotropic and satisfy $\mathcal{U} \subset \mathcal{U}' \subset \mathcal{U}^\perp$. These are given by $\mathbb{F}_{p^2}$-rational lines in $\mathcal{U}^\perp/\mathcal{U}$ which are isotropic for the induced hermitian form. We can easily see that there are exactly $p+1$ such lines. Since there are a total of $(p^3+1)(p^2+1)$ $\mathbb{F}_{p^2}$-rational points, each belonging on $p+1$ projective lines which each have p^2+1 points, we conclude that there are exactly $(p^3+1)(p+1)$ projective lines in our configuration.

The same discussion applies to $\mathscr{X}^-$.

3.4. ***Deligne–Lusztig varieties and the Bruhat–Tits stratification.*** Now we relate the varieties $\mathscr{X}$ studied above to the varieties $\mathscr{N}_\Lambda^\circ \subset \mathscr{N}_\Lambda$ of Section 2.6. Fix a vertex lattice Λ of type $2d \in \{2,4,6\}$, and endow the $2d$-dimensional $\mathbb{F}_p$-vector space

$$\Omega_0 = \Lambda/\Lambda^\vee$$

with the nondegenerate $\mathbb{F}_p$-valued quadratic form $q(x) = pQ(x)$ induced by the quadratic form Q on $\boldsymbol{L}_{\mathbb{Q}}^\Phi$. Set $\Omega = \Omega_0 \otimes_{\mathbb{F}_p} k$, and let G be the special orthogonal group of Ω. Note that Ω_0 is nonsplit: the existence of a d-dimensional totally isotropic subspace in Ω_0 would imply the existence of a self-dual lattice in $\boldsymbol{L}_{\mathbb{Q}}^\Phi$, contradicting the Hasse invariant calculation of Proposition 2.6.

Recall from Section 3.2 the reduced closed subscheme $\mathscr{X} = \mathscr{X}_\Lambda \subset \mathrm{OGr}(d)$ whose k-points are the Lagrangian subspaces $\mathcal{L} \subset \Omega$ with

$$\dim_k(\mathcal{L} + \Phi(\mathcal{L})) = d+1. \tag{3-11}$$

The Lagrangian subspaces of Ω are in bijection with the W-lattices $L \subset \boldsymbol{L}_{\mathbb{Q}}$ satisfying $L = L^\vee$ and $\Lambda^\vee \subset L$, and the condition (3-11) is equivalent to L being a special lattice. Combining this with (2-18), we obtain bijections

$$\mathscr{X}_\Lambda(k) \cong \{\text{special lattices } L \subset \boldsymbol{L}_{\mathbb{Q}} \text{ such that } \Lambda^\vee \subset L\} \cong \mathscr{N}_\Lambda(k).$$

Theorem 3.9. *There is an isomorphism of k-schemes $\mathscr{N}_\Lambda \cong \mathscr{X}_\Lambda$ inducing the above bijection on k-points. After possibly relabeling the two connected components of $\mathscr{X}_\Lambda = \mathscr{X}_\Lambda^+ \uplus \mathscr{X}_\Lambda^-$, we may assume that this isomorphism identifies $\mathscr{N}_\Lambda^\pm \cong \mathscr{X}_\Lambda^\pm$.*

Proof. Let R be a reduced k-algebra of finite type. Given an R-valued point $(G, \iota, \lambda, \varrho) \in \mathscr{N}_\Lambda(R)$, there is an induced map of $\mathbb{Z}_p$-modules

$$\Lambda^\vee \to \operatorname{End}(G)$$

defined by $x \mapsto \varrho^{-1} \circ x \circ \varrho$. Let $\mathcal{D}$ be the covariant Grothendieck–Messing crystal of G, evaluated at the trivial divided power thickening $\operatorname{Spec}(R) \to \operatorname{Spec}(R)$, so that $\mathcal{D}$ is a locally free R-module sitting in an exact sequence

$$0 \longrightarrow \mathcal{D}_1 \longrightarrow \mathcal{D} \longrightarrow \operatorname{Lie}(G) \longrightarrow 0.$$

The formation of the pair $\mathcal{D}_1 \subset \mathcal{D}$ is functorial in G, so there are induced morphisms of R-modules

$$\psi : (\Lambda^\vee / p\Lambda^\vee) \otimes_{\mathbb{F}_p} R \to \operatorname{End}_R(\mathcal{D})$$

and

$$\psi_1 : (\Lambda^\vee / p\Lambda^\vee) \otimes_{\mathbb{F}_p} R \to \operatorname{End}_R(\mathcal{D}_1)$$

with $\ker(\psi) \subset \ker(\psi_1)$. Given $x \in \ker(\psi_1)$ and $y \in (\Lambda^\vee / p\Lambda^\vee) \otimes_{\mathbb{F}_p} R$, the endomorphism

$$[x, y] = x \circ y + y \circ x \in \operatorname{End}_R(\mathcal{D}_1)$$

is trivial. Thus the kernel of ψ_1 is contained in the radical of the quadratic space $(\Lambda^\vee / p\Lambda^\vee) \otimes_{\mathbb{F}_p} R$, which is $(p\Lambda / p\Lambda^\vee) \otimes_{\mathbb{F}_p} R$. Let $\mathcal{L}^\sharp \subset \mathcal{K}$ be the images of $\ker(\psi) \subset \ker(\psi_1)$ under the obvious isomorphism

$$(p\Lambda / p\Lambda^\vee) \otimes_{\mathbb{F}_p} R \cong (\Lambda / \Lambda^\vee) \otimes_{\mathbb{F}_p} R \cong \Omega \otimes_k R.$$

When $R = k$, the point $(G, \iota, \lambda, \varrho)$ corresponds to some Dieudonné lattice D with $D_1 = VD$, and $\mathcal{D}_1 \subset \mathcal{D}$ is canonically identified with $D_1 / pD \subset D / pD$. Under these identifications,

$$\ker(\psi) = \{x \in (p\Lambda / p\Lambda^\vee) \otimes_{\mathbb{F}_p} k : xD \subset pD\},$$
$$\ker(\psi_1) = \{x \in (p\Lambda / p\Lambda^\vee) \otimes_{\mathbb{F}_p} k : xD_1 \subset pD\},$$

and so

$$\mathcal{L}^\sharp = \{x \in (\Lambda / \Lambda^\vee) \otimes k : xD \subset D\},$$
$$\mathcal{K} = \{x \in (\Lambda / \Lambda^\vee) \otimes k : xD_1 \subset D\}.$$

If we identify a subspace of $(\Lambda/\Lambda^\vee)\otimes k$ with the lattice in $\boldsymbol{L}_{\mathbb{Q}}$ that it generates, then $\mathcal{L}^\sharp$ corresponds to the lattice $L^\sharp=\{x\in \boldsymbol{L}_{\mathbb{Q}} : xD\subset D\}$ of Theorem 2.12, and $\mathcal{K}$ corresponds to $L+L^\sharp=\{x\in \boldsymbol{L}_{\mathbb{Q}} : xD_1\subset D\}$. In particular, $\mathcal{L}^\sharp$ is totally isotropic of dimension d and $\mathcal{K}$ has dimension $d+1$. Moreover, the quadratic space $\mathcal{K}/\mathcal{K}^\perp$ is a hyperbolic plane, and so has precisely two isotropic lines. One of them is $\mathcal{L}^\sharp$, and the other is the subspace $\mathcal{L}$ corresponding to $L=\{x\in \boldsymbol{L}_{\mathbb{Q}} : xD_1\subset D_1\}$.

For a general reduced R of finite type, it follows from the previous paragraph (use Exercise X.16 of [Lang 2002] and the fact that R is a Jacobson ring) that $\mathcal{L}^\sharp$ is a totally isotropic rank-d local direct summand of $\Omega\otimes_k R$ and $\mathcal{K}$ is a rank-$(d+1)$ local direct summand. By Lemma 3.4 there is a unique totally isotropic rank-d local direct summand $\mathcal{L}\neq\mathcal{L}^\sharp$ of $\Omega\otimes_k R$ contained in $\mathcal{K}$. As $\mathscr{N}_\Lambda$ is itself reduced and locally of finite type, the construction $(G,\iota,\lambda,\varrho)\mapsto\mathcal{L}$ defines a morphism of k-schemes

$$\mathscr{N}_\Lambda\to \mathrm{OGr}(d)$$

inducing the desired bijection $\mathscr{N}_\Lambda(k)\cong\mathscr{X}_\Lambda(k)$ on k-valued points. As the arguments of Section 2.4 were all done over an arbitrary extension of k, the above morphism induces a bijection $\mathscr{N}_\Lambda(k')\cong\mathscr{X}_\Lambda(k')$ for every field extension k'/k. The morphism $\mathscr{N}_\Lambda\to\mathscr{X}_\Lambda$ is therefore birational, quasi-finite, and proper (by Proposition 2.20). As $\mathscr{X}_\Lambda$ is smooth (and therefore normal), Zariski's main theorem implies $\mathscr{N}_\Lambda\cong\mathscr{X}_\Lambda$. The claim about connected components is obvious. □

3.5. *The main results.* Now we state our main results about the structure of the underlying reduced subscheme $\mathscr{N}_{\mathrm{red}}=\mathscr{N}^+_{\mathrm{red}}\sqcup\mathscr{N}^-_{\mathrm{red}}$ of $\mathscr{N}$. Recall from Section 2.6 that $\mathscr{N}^\pm_{\mathrm{red}}$ is covered by the closed subschemes $\mathscr{N}^\pm_\Lambda$ as Λ runs over the vertex lattices of type $t_\Lambda=2d_\Lambda\in\{2,4,6\}$ in the 6-dimensional $\mathbb{Q}_p$-quadratic space $\boldsymbol{L}^\Phi_{\mathbb{Q}}$, and that their intersections are given by the simple rule

$$\mathscr{N}^\pm_{\Lambda_1}\cap\mathscr{N}^\pm_{\Lambda_2}=\begin{cases}\mathscr{N}^\pm_{\Lambda_1\cap\Lambda_2} & \text{if } \Lambda_1\cap\Lambda_2 \text{ is a vertex lattice,}\\ \varnothing & \text{otherwise,}\end{cases}$$

where, as before, the left-hand side is understood to mean the reduced scheme underlying the scheme-theoretic intersection. In other words, the combinatorics of the intersections are controlled by the combinatorics of the simplicial complex $\mathcal{V}$ of Section 2.7.

Theorem 3.10. *The k-variety $\mathscr{N}^\pm_\Lambda$ is projective, smooth, and irreducible of dimension $d_\Lambda-1$. Moreover*:

(1) *If $d_\Lambda=1$, then $\mathscr{N}^\pm_\Lambda$ is a single point.*

(2) *If $d_\Lambda=2$, then $\mathscr{N}^\pm_\Lambda$ is isomorphic to $\mathbb{P}^1$.*

(3) *If $d_\Lambda=3$, then $\mathscr{N}^\pm_\Lambda$ is isomorphic to the Fermat hypersurface*

$$x_0^{p+1}+x_1^{p+1}+x_2^{p+1}+x_3^{p+1}=0.$$

Proof. Combine Theorem 3.9 with the discussion of Section 3.3. □

Theorem 3.11. *Under the isomorphism $\mathscr{X}_\Lambda^\pm \cong \mathscr{N}_\Lambda^\pm$, the stratification of Proposition 3.8 and the stratification*

$$\mathscr{N}_\Lambda^\pm = \biguplus_{\Lambda' \subset \Lambda} \mathscr{N}_{\Lambda'}^{\pm\circ}$$

of Section 2.6 are related by

$$X_{P_r}(w_r^\pm) \cong \biguplus_{\substack{\Lambda' \subset \Lambda \\ d_{\Lambda'} = r+1}} \mathscr{N}_{\Lambda'}^{\pm\circ} \tag{3-12}$$

for all $0 \le r \le d_\Lambda - 1$. In particular (by taking $r = d_\Lambda - 1$), the dense open subvariety $\mathscr{N}_\Lambda^{\pm\circ}$ is isomorphic to the Deligne–Lusztig variety $X_B(w^\pm)$ associated with a Coxeter element.

Proof. For each special lattice L, we defined in Proposition 2.19 a sequence of lattices

$$L = L^{(0)} \subsetneq L^{(1)} \subsetneq \cdots \subsetneq L^{(d)} = L^{(d+1)}$$

by $L^{(r)} = L + \Phi(L) + \cdots + \Phi^r(L)$, and a type-$2d$ vertex lattice

$$\Lambda_L = \{x \in L^{(d)} : \Phi(x) = x\}.$$

The bijection (2-18) identifies $\mathscr{N}_\Lambda^\pm(k)$ with the set of special lattices L with $\Lambda_L \subset \Lambda$, and the k-points of the right-hand side of (3-12) correspond to those L for which Λ_L has type $2r+2$; in other words, those L for which

$$L = L^{(0)} \subsetneq L^{(1)} \subsetneq \cdots \subsetneq L^{(r+1)} = L^{(r+2)}.$$

If we instead define $\underline{L}^{(r)} = L \cap \Phi(L) \cap \cdots \cap \Phi^r(L)$, this condition is equivalent to

$$\underline{L}^{(r+2)} = \underline{L}^{(r+1)} \subsetneq \cdots \subsetneq \underline{L}^{(1)} \subsetneq \underline{L}^{(0)} = L.$$

In the proof of Proposition 3.8, this is the same as the condition defining the strata $X_{P_r}(w_r^\pm)$. □

Theorem 3.12. *The reduced k-scheme $\mathscr{N}_{\mathrm{red}}$ is equidimensional of dimension two. It has two connected components, $\mathscr{N}_{\mathrm{red}}^+$ and $\mathscr{N}_{\mathrm{red}}^-$, and these connected components are isomorphic. The irreducible components of $\mathscr{N}_{\mathrm{red}}$ are precisely the closed subschemes $\mathscr{N}_\Lambda^\pm$ as Λ varies over the type-6 vertex lattices. Furthermore:*

(1) *For each irreducible component $\mathscr{N}_\Lambda$, there are exactly $(p^3+1)(p+1)$ irreducible components $\mathscr{N}_{\Lambda'}$ such that $\mathscr{N}_\Lambda \cap \mathscr{N}_{\Lambda'} \cong \mathbb{P}^1$, and $(p^3+1)(p^2+1)$ irreducible components $\mathscr{N}_{\Lambda'}$ such that $\mathscr{N}_\Lambda \cap \mathscr{N}_{\Lambda'}$ consists of a single point.*

(2) *For each type-4 vertex lattice Λ, the closed subscheme $\mathscr{N}_\Lambda \cong \mathbb{P}^1$ is contained in exactly two irreducible components, and is equal to their intersection.*

Proof. The isomorphism $\mathscr{N}^+_{\mathrm{red}} \cong \mathscr{N}^-_{\mathrm{red}}$ follows from the isomorphism $\mathscr{N}^+ \cong \mathscr{N}^-$ of Section 2.1. The connectedness of $\mathscr{N}^\pm_{\mathrm{red}}$ follows from Corollary 2.23. The remaining claims are clear from the theorems above and the discussion of Section 3.3. □

3.6. *Hermitian vertex lattices.* As in [Rapoport et al. 2014; Vollaard 2010; Vollaard and Wedhorn 2011], it is possible to describe the stratification of $\mathscr{N}$ in terms of the Bruhat–Tits building of the special unitary group J^{der}, although in our setting the description in these terms is slightly convoluted. Recall from Remark 2.8 the central isogeny $J^{\mathrm{der}} \to \mathrm{SO}(\boldsymbol{L}^\Phi_{\mathbb{Q}})$. Using [Bruhat and Tits 1984, § 4.2.15], we see that this gives an identification of the building $\mathcal{BT}$ of $\mathrm{SO}(\boldsymbol{L}^\Phi_{\mathbb{Q}})$, which was described in Section 2.7, with the building of J^{der}. Therefore, using [Vollaard 2010] and $J^{\mathrm{der}} \cong \mathrm{SU}(T)$, we can see that the underlying simplicial complex of the building $\mathcal{BT}$ can also be described using $\mathcal{O}_E$-lattices in the split Hermitian space T of dimension 4 over E.

We say that an $\mathcal{O}_E$-lattice $\Xi \subset T$ is a *Hermitian vertex lattice* if

$$\Xi \subset \Xi^\vee \subset p^{-1}\Xi.$$

The *type* of Ξ is $\dim_{\mathbb{F}_{p^2}}(\Xi^\vee/\Xi)$; the type can be 0, 2 or 4. As in [Vollaard 2010], these Hermitian vertex lattices correspond bijectively to the vertices of the Bruhat–Tits building of $\mathrm{SU}(T)$. The action of the group $\mathrm{SU}(T)$ preserves the vertex type and is transitive on the set of vertices of a given type. The simplicial structure of the building of $\mathrm{SU}(T)$ is generated, as above, using a notion of adjacency, in which Ξ and Ξ' are adjacent if either $\Xi \subset \Xi'$ or $\Xi' \subset \Xi$. Consider now the identification of the buildings given by the central isogeny $\mathrm{SU}(T) \to \mathrm{SO}(\boldsymbol{L}^\Phi_{\mathbb{Q}})$. We can see by looking at the local Dynkin diagrams that Hermitian vertex lattices Ξ of type 0 and 4 are sent to vertex lattices Λ of type 6, and Hermitian vertex lattices Ξ of type 2 are sent to vertex lattices Λ of type 2. Note that $\mathrm{SO}(\boldsymbol{L}^\Phi_{\mathbb{Q}})$ acts transitively on the set of vertex lattices of type 6, but the map $\mathrm{SU}(T) \to \mathrm{SO}(\boldsymbol{L}^\Phi_{\mathbb{Q}})$ is not surjective on $\mathbb{Q}_p$-points: its image is the kernel of the spinor norm.

Consider the set $\mathcal{S}$ which is defined as the disjoint union of the set of Hermitian vertex lattices Ξ with the set of all pairs $\{\Xi, \Xi'\}$ consisting of adjacent Hermitian vertex lattices of types 0 and 4. Note that there is a natural bijection between the set $\mathcal{S}$ and the set of all vertex lattices Λ. Hermitian vertex lattices of type 0 and 4 in $\mathcal{S}$ correspond to vertex lattices of type 6, Hermitian vertex lattices of type 2 in $\mathcal{S}$ correspond to vertex lattices of type 2, and finally the pairs $\{\Xi, \Xi'\}$ correspond to vertex lattices of type 4.

We define a partial order on $\mathcal{S}$ as follows: For two Hermitian vertex lattices we define $\Xi < \Xi'$ if either

(1) Ξ is of type 2, Ξ' is of type 0, and $\Xi \subset \Xi'$,

(2) Ξ is type 2, Ξ' is of type 4, and $\Xi' \subset \Xi$

(so Hermitian vertex lattices of type 0 and 4 are not comparable). Two pairs $\{\Xi_1, \Xi_1'\}$ and $\{\Xi_2, \Xi_2'\}$ in $\mathcal{S}$ are not compared. If Ξ is a Hermitian vertex lattice then $\Xi < \{\Xi_1, \Xi_2\}$ if Ξ, Ξ_1, and Ξ_2 form a simplex in the building of J^{der} (which requires that Ξ have type 2). Finally, $\{\Xi_1, \Xi_2\} < \Xi$ if $\Xi \in \{\Xi_1, \Xi_2\}$. Under the bijection between $\mathcal{S}$ and the set of vertex lattices, this partial order corresponds to inclusion of vertex lattices. Define an adjacency relation in $\mathcal{S}$ by $x \sim_{\mathcal{S}} y$ if either $x < y$ or $y < x$. We also define a dimension function $d : \mathcal{S} \to \{0, 1, 2\}$ by $d(x) = 0$ if x is a Hermitian vertex lattice of type 2, $d(x) = 2$ if x is a Hermitian vertex lattice of type 0 or 4, and $d(x) = 1$ if x is a pair $\{\Xi, \Xi'\}$.

The following theorems are simply restatements in this new language of some results of the previous subsection:

Theorem 3.13. *Writing the reduced k-scheme as a union*

$$\mathscr{M}_{\mathrm{red}} = \biguplus_{\ell \in \mathbb{Z}} \mathscr{M}_{\mathrm{red}}^{(\ell)}$$

gives the decomposition of $\mathscr{M}_{\mathrm{red}}$ into its connected components $\mathscr{M}_{\mathrm{red}}^{(\ell)}$. These connected components are all isomorphic and are of pure dimension 2*.*

(1) *There is a stratification of $\mathscr{M}_{\mathrm{red}}^{(0)}$ by locally closed smooth subschemes given by*

$$\mathscr{M}_{\mathrm{red}}^{(0)} = \biguplus_{x \in \mathcal{S}} \mathscr{M}_x^{\circ}.$$

Each stratum $\mathscr{M}_x^{\circ}$ is isomorphic to $\mathscr{N}_{\Lambda}^{+\circ}$, where Λ is the vertex lattice that corresponds to x, and is therefore isomorphic to a Deligne–Lusztig variety of dimension $d(x)$. The closure $\mathscr{M}_x$ of any $\mathscr{M}_x^{\circ}$ in $\mathscr{M}_{\mathrm{red}}^{(0)}$ is

$$\mathscr{M}_x = \biguplus_{y \le x} \mathscr{M}_y^{\circ}.$$

(2) *We have $\mathscr{M}_y \subset \mathscr{M}_x$ if and only if $y \le x$. In particular, the irreducible components of $\mathscr{M}_{\mathrm{red}}^{(0)}$ are precisely the closed subschemes $\mathscr{M}_{\Xi}$ for $\Xi \in \mathcal{S}$ a Hermitian vertex lattice of type* 0 *or* 4*.*

(3) *The schemes $\mathscr{M}_x$ are as follows*:

   (a) *If $d(x) = 0$, then $\mathscr{M}_x$ is a single point.*
   (b) *If $d(x) = 1$, then $\mathscr{M}_x$ is isomorphic to $\mathbb{P}^1$.*
   (c) *If $d(x) = 2$, then $\mathscr{M}_x$ is isomorphic to the Fermat hypersurface*
$$x_0^{p+1} + x_1^{p+1} + x_2^{p+1} + x_3^{p+1} = 0.$$

Theorem 3.14. *The irreducible components of $\mathscr{M}_{\mathrm{red}}^{(0)}$ are parametrized by vertices of type* 0 *and* 4 *in the Bruhat–Tits building of J^{der}. Two irreducible components $\mathscr{M}_{\Xi}$ and $\mathscr{M}_{\Xi'}$ intersect if and only if Ξ and Ξ' are either adjacent, or are adjacent* *to a common element of $\mathcal{S}$. If they are adjacent then one is type* 0, *the other of type* 4, *and they intersect along a* $\mathbb{P}^1$. *If they are not adjacent but have a common adjacent point $y \in \mathcal{S}$, then y is a Hermitian vertex lattice of type* 2, *and* $\mathscr{M}_\Xi \cap \mathscr{M}_{\Xi'} = \mathscr{M}_y$ *is a single point.*

4. Applications to Shimura varieties

We now use our explicit description of the Rapoport–Zink space $\mathscr{N} = p^{\mathbb{Z}} \backslash \mathscr{M}$ to describe the supersingular locus of a $\mathrm{GU}(2,2)$-Shimura variety. With the results of Section 3.5 in hand, this is exactly as in the $\mathrm{GU}(n-1,1)$ cases studied in [Rapoport et al. 2014; Vollaard and Wedhorn 2011]. Accordingly, our discussion will be brief.

4.1. ***The Shimura variety.*** Let $E \subset \mathbb{C}$ be a quadratic imaginary field, fix a prime $p > 2$ inert in E, and let $\mathcal{O} \subset E$ be the integral closure of $\mathbb{Z}_{(p)}$ in E. Let V be a free $\mathcal{O}$-module of rank 4 endowed with a perfect $\mathcal{O}$-valued Hermitian form $\langle \cdot, \cdot \rangle$ of signature $(2,2)$, and denote by $G = \mathrm{GU}(V)$ the group of unitary similitudes of V. It is a reductive group over $\mathbb{Z}_{(p)}$. Fix a compact open subgroup $U^p \subset G(\mathbb{A}_f^p)$, and define $U_p = G(\mathbb{Z}_p)$ and $U = U_p U^p \subset G(\mathbb{A}_f)$.

The Grassmannian $\mathcal{D}$ of negative-definite planes in $V \otimes_{\mathcal{O}} \mathbb{C}$ is a smooth complex manifold of dimension 4, with an action of $G(\mathbb{R})$. Define

$$M_U(\mathbb{C}) = G(\mathbb{Q}) \backslash (\mathcal{D} \times G(\mathbb{A}_f)/U).$$

For sufficiently small U^p, this is a smooth complex manifold parametrizing prime-to-p isogeny classes of quadruples $(A, \iota, \lambda, [\eta^p])$, in which A is an abelian variety of dimension 4, $\iota : \mathcal{O} \to \mathrm{End}(A)_{(p)}$ is a ring homomorphism such that

$$\det(T - \iota(\alpha); \mathrm{Lie}(A)) = (T - \alpha)^2 (T - \overline{\alpha})^2$$

for all $\alpha \in \mathcal{O}$, $\lambda \in \mathrm{Hom}(A, A^\vee)_{(p)}$ is a prime-to-p-quasi-polarization satisfying

$$\lambda \circ \iota(\overline{\alpha}) = \iota(\alpha)^\vee \circ \lambda$$

for all $\alpha \in \mathcal{O}$, and $[\eta^p]$ is the U^p-orbit of an $\mathcal{O} \otimes \mathbb{A}_f^p$-linear isomorphism

$$\eta^p : \widehat{\mathrm{Ta}}^p(A) \otimes \mathbb{A}_f^p \to V \otimes \mathbb{A}_f^p$$

respecting the Hermitian forms up to scaling by $(\mathbb{A}_f^p)^\times$ (the Hermitian form on the source is determined by λ, as in (2-3)). A *prime-to-p-isogeny* between two such pairs $(A, \iota, \lambda, [\eta^p])$ and $(A', \iota', \lambda', [\eta^{p\prime}])$ is an $\mathcal{O}$-linear quasi-isogeny in $\mathrm{Hom}(A, A')_{(p)}$ of degree prime to p that respects the level structures, and such that λ' pulls back to a $\mathbb{Z}_{(p)}^\times$-multiple of λ.

The parametrization is similar to the constructions found in [Kudla and Rapoport 2009], and can be described as follows. For each triple $(A, i, \lambda, [\eta])$ above, the existence of η^p implies that $H_1(A,\mathbb{Q})$ and $V\otimes\mathbb{Q}$ are isomorphic, as Hermitian spaces, locally at all places $v\nmid p$. But this implies that they are also isomorphic at p, and hence there is a global isomorphism

$$\beta : H_1(A,\mathbb{Q}) \to V\otimes\mathbb{Q}.$$

As $\mathrm{Ta}_p(A)\otimes\mathbb{Q}_p \cong V\otimes\mathbb{Q}_p$, a result of Jacobowitz, stated in [Kudla and Rapoport 2009, Proposition 2.14], shows that there is a unique U_p-orbit of isomorphisms $\mathrm{Ta}_p(A)\cong V\otimes\mathbb{Z}_p$ compatible with the $\mathcal{O}$-actions and Hermitian forms. Thus there is a unique way to extend η^p to a U-orbit of isomorphisms

$$\eta : \widehat{\mathrm{Ta}}(A)\otimes\mathbb{A}_f \cong V\otimes\mathbb{A}_f$$

compatible with the $\mathcal{O}$-actions and the symplectic forms, and identifying $\mathrm{Ta}_p(A)$ with $V\otimes\mathbb{Z}_p$. The composition

$$V\otimes\mathbb{A}_f \xrightarrow{\eta^{-1}} H_1(A,\mathbb{A}_f) \xrightarrow{\beta} V\otimes\mathbb{A}_f$$

defines an element $g\in G(\mathbb{A}_f)/U$, and the Hodge structure on $V\otimes\mathbb{R}$ induced by the isomorphism β corresponds to a point of $\mathcal{D}$, as in [Kudla and Rapoport 2009, Section 3].

4.2. ***The uniformization theorem.*** Let k be an algebraic closure of the field of p elements.

Extending the moduli problem of the previous subsection to $\mathbb{Z}_{(p)}$-schemes in the obvious way yields a scheme M_U, smooth of relative dimension 4 over $\mathbb{Z}_{(p)}$. Denote by M_U^{ss} the reduced supersingular locus of the geometric special fiber $M_U\times_{\mathbb{Z}_{(p)}}k$. A choice of geometric point $(\boldsymbol{A},\boldsymbol{\iota},\boldsymbol{\lambda},[\boldsymbol{\eta}])\in M_U^{\mathrm{ss}}(k)$ determines a base point $(\boldsymbol{G},\boldsymbol{\iota},\boldsymbol{\lambda})$ with $\boldsymbol{G}=\boldsymbol{A}[p^\infty]$, and so defines a Rapoport–Zink space $\mathscr{M}$ as in Section 2.1, endowed with an action of the subgroup $J\subset\mathrm{End}(\boldsymbol{G})_{\mathbb{Q}}^\times$. Denote by $I(\mathbb{Q})\subset\mathrm{End}(A)_{\mathbb{Q}}^\times$ the subgroup of $\mathcal{O}$-linear quasi-automorphisms that preserve the $\mathbb{Q}^\times$-span of $\boldsymbol{\lambda}$. It is the group of $\mathbb{Q}$-points of an algebraic group I over $\mathbb{Q}$ satisfying $I(\mathbb{Q}_p)\cong J$, and the orbit $[\boldsymbol{\eta}]$ determines a right U^p-orbit of isomorphisms $I(\mathbb{A}_f^p)\cong G(\mathbb{A}_f^p)$. In particular, $I(\mathbb{Q})$ acts on both $\mathscr{M}$ and on $G(\mathbb{A}_f^p)/U^p$.

Theorem 4.1 (Rapoport–Zink). *There is an isomorphism of k-schemes*

$$M_U^{\mathrm{ss}}\cong I(\mathbb{Q})\backslash(\mathscr{M}_{\mathrm{red}}\times G(\mathbb{A}_f^p)/U^p).$$

As in [Vollaard 2010, Corollary 6.2], combining the above uniformization theorem with the results of Section 3.5 yields the following corollary:

Corollary 4.2. *The k-scheme M_U^{ss} has pure dimension 2. For U^p sufficiently small, all irreducible components of M_U^{ss} are isomorphic to the Fermat hypersurface*

$$x_0^{p+1}+x_1^{p+1}+x_2^{p+1}+x_3^{p+1}=0,$$

and any two irreducible components either intersect trivially, intersect at a single point, or their intersection is isomorphic to $\mathbb{P}^1$. Here "intersection" is understood to mean the reduced scheme underlying the scheme-theoretic intersection.

Acknowledgements

The authors would like to thank M. Rapoport for many useful suggestions, and X. He for communicating his joint results with U. Görtz.

References

[Bass 1974] H. Bass, "Clifford algebras and spinor norms over a commutative ring", *Amer. J. Math.* **96** (1974), 156–206. MR 50 #13092 Zbl 0344.15017

[Bonnafé and Rouquier 2006] C. Bonnafé and R. Rouquier, "On the irreducibility of Deligne–Lusztig varieties", *C. R. Math. Acad. Sci. Paris* **343**:1 (2006), 37–39. MR 2007a:14055 Zbl 1099.14040

[Bruhat and Tits 1984] F. Bruhat and J. Tits, "Groupes réductifs sur un corps local, II: Schémas en groupes. Existence d'une donnée radicielle valuée", *Inst. Hautes Études Sci. Publ. Math.* 60 (1984), 197–376. MR 86c:20042 Zbl 0597.14041

[Bültel 2012] O. Bültel, "On the supersingular loci of quaternionic Siegel space", preprint, 2012. arXiv 1209.3548

[Deligne and Lusztig 1976] P. Deligne and G. Lusztig, "Representations of reductive groups over finite fields", *Ann. of Math.* (2) **103**:1 (1976), 103–161. MR 52 #14076 Zbl 0336.20029

[Freitag and Hermann 2000] E. Freitag and C. F. Hermann, "Some modular varieties of low dimension", *Adv. Math.* **152**:2 (2000), 203–287. MR 2001j:11023 Zbl 0974.11028

[Garrett 1997] P. Garrett, *Buildings and classical groups*, Chapman & Hall, London, 1997. MR 98k:20081 Zbl 0933.20019

[Göertz and He 2013] U. Göertz and X. He, "Basic loci in Shimura varieties of Coxeter type", preprint, 2013. arXiv 1311.6263

[Kim 2013] W. Kim, "Rapoport–Zink spaces of Hodge type", preprint, 2013. arXiv 1308.5537

[Kisin 2010] M. Kisin, "Integral models for Shimura varieties of abelian type", *J. Amer. Math. Soc.* **23**:4 (2010), 967–1012. MR 2011j:11109 Zbl 1280.11033

[Knus et al. 1998] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, American Mathematical Society Colloquium Publications **44**, American Mathematical Society, Providence, RI, 1998. MR 2000a:16031 Zbl 0955.16001

[Kudla and Rapoport 1999] S. S. Kudla and M. Rapoport, "Arithmetic Hirzebruch–Zagier cycles", *J. Reine Angew. Math.* **515** (1999), 155–244. MR 2002e:11076a Zbl 1048.11048

[Kudla and Rapoport 2000] S. S. Kudla and M. Rapoport, "Cycles on Siegel threefolds and derivatives of Eisenstein series", *Ann. Sci. École Norm. Sup.* (4) **33**:5 (2000), 695–756. MR 2002e:11076b Zbl 1045.11044

[Kudla and Rapoport 2009] S. Kudla and M. Rapoport, "Special cycles on unitary Shimura varieties II: global theory", preprint, 2009. arXiv 0912.3758

[Kudla and Rapoport 2011] S. Kudla and M. Rapoport, "Special cycles on unitary Shimura varieties, I: Unramified local theory", *Invent. Math.* **184**:3 (2011), 629–682. MR 2012k:14033 Zbl 1229.14020

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics **211**, Springer, New York, 2002. MR 2003e:00003 Zbl 0984.00001

[Madapusi Pera 2012] K. Madapusi Pera, "Integral canonical models for Spin Shimura varieties", preprint, 2012. arXiv 1212.1243

[Rapoport and Zink 1996] M. Rapoport and T. Zink, *Period spaces for p-divisible groups*, Annals of Mathematics Studies **141**, Princeton University Press, 1996. MR 97f:14023 Zbl 0873.14039

[Rapoport et al. 2014] M. Rapoport, U. Terstiege, and S. Wilson, "The supersingular locus of the Shimura variety for $\mathrm{GU}(1, n-1)$ over a ramified prime", *Math. Z.* **276**:3-4 (2014), 1165–1188. MR 3175176 Zbl 06304108

[Shimura 2010] G. Shimura, *Arithmetic of quadratic forms*, Springer Monographs in Mathematics, Springer, New York, 2010. MR 2011m:11003 Zbl 1202.11026

[Springer 1998] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progress in Mathematics **9**, Birkhäuser, Boston, 1998. MR 99h:20075 Zbl 0927.20024

[Tits 1979] J. Tits, "Reductive groups over local fields", pp. 29–69 in *Automorphic forms, representations and L-functions, Part 1* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math., XXXIII **33**, Amer. Math. Soc., Providence, R.I., 1979. MR 80h:20064 Zbl 0415.20035

[Vasiu 1999] A. Vasiu, "Integral canonical models of Shimura varieties of preabelian type", *Asian J. Math.* **3**:2 (1999), 401–518. MR 2002b:11087 Zbl 1002.11052

[Vollaard 2010] I. Vollaard, "The supersingular locus of the Shimura variety for $\mathrm{GU}(1, s)$", *Canad. J. Math.* **62**:3 (2010), 668–720. MR 2011j:14059 Zbl 1205.14028

[Vollaard and Wedhorn 2011] I. Vollaard and T. Wedhorn, "The supersingular locus of the Shimura variety of $\mathrm{GU}(1, n-1)$ II", *Invent. Math.* **184**:3 (2011), 591–627. MR 2012j:14035 Zbl 1227.14027

[Zink 2001] T. Zink, "Windows for displays of p-divisible groups", pp. 491–518 in *Moduli of abelian varieties* (Texel, 1999), edited by C. Faber et al., Progr. Math. **195**, Birkhäuser, Basel, 2001. MR 2002c:14073 Zbl 1099.14036

Communicated by Barry Mazur
Received 2013-10-16 Revised 2014-02-17 Accepted 2014-05-23

howardbe@bc.edu *Department of Mathematics, Boston College, Chestnut Hill, MA 02467-3806, United States*

pappas@math.msu.edu *Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, United States*

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1701

Poincaré–Birkhoff–Witt deformations of smash product algebras from Hopf actions on Koszul algebras

Chelsea Walton and Sarah Witherspoon

Let H be a Hopf algebra and let B be a Koszul H-module algebra. We provide necessary and sufficient conditions for a filtered algebra to be a Poincaré–Birkhoff–Witt (PBW) deformation of the smash product algebra $B \# H$. Many examples of these deformations are given.

A correction was posted on 3 May 2017 in an online supplement.

Introduction

Given a Hopf algebra (H-)action on a Koszul algebra B, the aim of this work is to provide necessary and sufficient conditions for a certain filtered algebra, namely $\mathcal{D}_{B,\kappa}$ in Notation 0.3 below, to be a *Poincaré–Birkhoff–Witt (PBW) deformation* of the smash product algebra $B \# H$, i.e., $\operatorname{gr} \mathcal{D}_{B,\kappa} \cong B \# H$ (Definition 0.1). One well-studied case is that of group actions on polynomial rings, where many algebras of interest arise as such deformations; see for example [Crawley-Boevey and Holland 1998; Drinfeld 1986; Etingof and Ginzburg 2002; Lusztig 1989; Ram and Shepler 2003; Shepler and Witherspoon 2012a]. For group actions on other Koszul algebras, see [Levandovskyy and Shepler 2014; Naidu and Witherspoon 2014; Shepler and Witherspoon 2012b; Shroff 2014]. There are some results involving Hopf algebra actions, such as those of Khare [2007], when H is cocommutative and B is a polynomial algebra. More specifically, the case when $H = U(\mathfrak{g})$, with $\mathfrak{g}$ the Lie algebra of a (not necessarily connected) reductive algebraic group, was studied by Etingof, Gan, and Ginzburg [Etingof et al. 2005], and by Khare and Tikaradze [2010] where $\mathfrak{g} = \mathfrak{sl}_2$. Results for an action of the quantized enveloping algebra $H = U_q(\mathfrak{sl}_2)$ on the quantum plane are provided by Gan and Khare [2007].

The goal of this paper is to provide a general theorem encompassing all of the above known classes of examples from the literature. Specifically, Theorem 3.1 gives PBW deformation conditions for $B \# H$, and it only requires the following

MSC2010: primary 16S80; secondary 16E40, 16S37, 16S40.
Keywords: Hopf algebra, Koszul algebra, PBW deformation, smash product algebra.

of H and B: (1) the antipode of the Hopf algebra H is bijective, (2) the Koszul H-module algebra B is connected ($B_0 = k$), and (3) the H-action on B preserves the grading of B. We then apply our theorem to several different choices of Hopf algebras acting on Koszul algebras to obtain nontrivial PBW deformations, both known and new. Our work indicates that such examples abound.

Many ring-theoretic properties are preserved under PBW deformation. To discuss this, let us consider the following definition:

Definition 0.1. Let $D = \bigcup_{i \geq 0} F_i$ be a filtered algebra with $\{0\} \subseteq F_0 \subseteq F_1 \subseteq \cdots \subseteq D$. We say that D is a *Poincaré–Birkhoff–Witt (PBW) deformation* of an $\mathbb{N}$-graded algebra A if A is isomorphic to the associated graded algebra $\operatorname{gr}_F D = \bigoplus_{i \geq 0} F_i/F_{i-1}$, as $\mathbb{N}$-graded algebras.

Now if $\operatorname{gr}_F D$ is an integral domain, prime, or (right) noetherian, then so is D. Moreover, if D is affine with the standard filtration F', then the Gelfand–Kirillov (GK) dimensions of D and of $\operatorname{gr}_{F'} D$ are equal; $\operatorname{GKdim}(\operatorname{gr}_F D) \leq \operatorname{GKdim}(D)$ for a general filtration F of D. The Krull dimension and global dimension of $\operatorname{gr}_F D$ serve as upper bounds for the corresponding dimensions of D. These ring-theoretic results can all be found in [McConnell and Robson 2001]. Homological properties preserved under PBW deformation have also been investigated; see [Berger and Taillefer 2007] and [Wu and Zhu 2013] regarding the Calabi–Yau property, for instance. The representation theory of some classes of PBW deformations of smash product algebras has been thoroughly studied in the literature and still remains an active area of research. Some examples of PBW deformations whose representation theory is of interest include rational Cherednik algebras, symplectic reflection algebras, and various types of Hecke algebras (see, for example, [Drinfeld 1986; Etingof et al. 2005; Etingof and Ginzburg 2002; Lusztig 1989; Ram and Shepler 2003], and for more recent work, see [Ding and Tsymbaliuk 2013; Losev and Tsymbaliuk 2014; Tikaradze 2010; Tsymbaliuk 2014]).

In order to state the main result, we need the following notation and terminology. Let k be a field of arbitrary characteristic and let an unadorned $\otimes$ mean $\otimes_k$. Let $\mathbb{N}$ denote the natural numbers, including 0. Recall that an $\mathbb{N}$-graded algebra is *Koszul* if its trivial module k admits a linear minimal graded free resolution; see [Polishchuk and Positselski 2005, Chapter 2] for more details.

Notation 0.2 $(H, B, I, \kappa, \kappa^C, \kappa^L)$. Let V be a finite-dimensional vector space over k.

(i) Let H be a Hopf algebra with the standard structure notation $(H, m, \Delta, u, \epsilon, S)$. Here, we assume that the antipode S of H is bijective.

(ii) Let $B = T_k(V)/(I)$ be an $\mathbb{N}$-graded, Koszul, left H-module algebra $B = \bigoplus_{j \geq 0} B_j$ with $B_0 = k$ and $I \subseteq V \otimes V$. We assume that the action of H preserves the grading and the subspace I of $V \otimes V$. So in this case, V is an H-module.

(iii) Take $\kappa : I \to H \oplus (V \otimes H)$ to be a k-bilinear map, where κ is the sum of its *constant* and *linear* parts $\kappa^C : I \to H$ and $\kappa^L : I \to V \otimes H$, respectively.

Notation 0.3 ($\mathcal{D}_{B,\kappa}$). Let $\mathcal{D}_{B,\kappa}$ be the filtered k-algebra given by

$$\mathcal{D}_{B,\kappa} = \frac{T_k(V) \# H}{(r - \kappa(r))_{r \in I}}.$$

Here, we assign the elements of H degree 0.

Our main result is given as follows:

Theorem 3.1. *The algebra $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B \# H$ if and only if the following conditions hold*:

(a) *κ is H-invariant* (Definition 1.4); *and*

If $\dim_k V \geq 3$, *then the following equations hold for the maps* $\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C$ *and* $\kappa^L \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^L$, *which are defined on the intersection* $(I \otimes V) \cap (V \otimes I)$:

(b) $\mathrm{Im}(\kappa^L \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^L) \subseteq I$;

(c) $\kappa^L \circ (\kappa^L \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^L) = -(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)$;

(d) $\kappa^C \circ (\mathrm{id} \otimes \kappa^L - \kappa^L \otimes \mathrm{id}) \equiv 0$.

In the case that H is cocommutative and B is the symmetric algebra $S(V)$, this result was proven by Khare [2007, Theorem 2.1], via the diamond lemma. Our proof is a generalization of that of [Braverman and Gaitsgory 1996, Lemma 0.4, Theorem 0.5] (where $H = k$) and of [Shepler and Witherspoon 2012b, Theorem 5.4] (where H is a group algebra).

Background information on Hopf algebra (co)actions, Hochschild cohomology, and deformations of algebras are provided in Section 1. In Section 2, we produce a free resolution of the smash product algebra $B \# H$; see Construction 2.5 and Theorem 2.10. This resolution is adapted from Guccione and Guccione [2002]; Negron [2014] independently constructed a similar resolution. Our resolution is used in the proof of Theorem 3.1, which is given in Section 3. Many examples of PBW deformations of $B \# H$ are provided in Section 4, including/involving:

- (Example 4.1) the *Crawley-Boevey–Holland algebras*;
- (Examples 4.2 and 4.4) some actions of semisimple, noncommutative, noncocommutative Hopf algebras on skew polynomial rings;
- (Examples 4.13 and 4.16) actions of the Sweedler and the Taft algebras on the polynomial ring $k[u, v]$;
- (Example 4.18) the *quantized symplectic oscillator algebras of rank 1*.

All of the examples of $B \# H$ above have nontrivial PBW deformations.

1. Background material

We begin by discussing Hopf (co)actions on algebras and (co)modules and end with a discussion on deformations of algebras. For further background on these topics, we refer the reader to [Montgomery 1993] and [Braverman and Gaitsgory 1996; Gerstenhaber 1964], respectively.

1A. *Hopf algebra (co)actions.*

Definition 1.1. (i) For a left H-module M, we denote the H-action by $\cdot : H \otimes M \to M$, that is, by $h \cdot m \in M$ for all $h \in H$, $m \in M$. Similarly for all $h \in H$ and $m \in M$, we denote the right h-action on m by $m \cdot h$.

(ii) Given a Hopf algebra H and an algebra A, we say that H *acts on* A (from the left, as a Hopf algebra) if A is a left H-module and

$$h \cdot (ab) = \sum (h_1 \cdot a)(h_2 \cdot b) \quad \text{and} \quad h \cdot 1_A = \epsilon(h) 1_A$$

for all $h \in H$, $a, b \in A$, where the comultiplication is given by $\Delta(h) = \sum h_1 \otimes h_2$ (Sweedler's notation). In this case, we say that A is a *left H-module algebra*.

(iii) For any left H-comodule M, we denote the left H-coaction by $\rho(M) \subseteq H \otimes M$, where $\rho(m) = \sum m_{-1} \otimes m_0$ for $m_{-1} \in H$ and $m, m_0 \in M$. Likewise, the right H-coaction on a right H-comodule M is given by $\rho(m) = \sum m_0 \otimes m_1$ for $m, m_0 \in M$ and $m_1 \in H$.

Note that H is naturally an H-bimodule via left and right multiplication. This yields a *left H-adjoint action on H* given by

$$h \cdot \ell := \sum h_1 \ell S(h_2) \tag{1-2}$$

for $h, \ell \in H$. Moreover, if V is a left H-module, we give $V \otimes H$ an H-bimodule structure as follows: $h(v \otimes \ell) = \sum (h_1 \cdot v) \otimes h_2 \ell$ and $(v \otimes \ell)h = v \otimes \ell h$ for all $h, \ell \in H$ and $v \in V$. A *left H-adjoint action on $V \otimes H$* arises by combining these:

$$h \cdot (v \otimes \ell) := \sum (h_1 \cdot v) \otimes h_2 \ell S(h_3). \tag{1-3}$$

The left H-adjoint actions in (1-2) and (1-3) extend to the standard left H-adjoint action on $A = B \# H$ (where $B = T_k(V)/(I)$ as in Notation 0.2(ii)), via Definition 1.1, since the action of H preserves I.

Now we discuss the H-invariance of the map κ (Notation 0.2(iii)), which is one of the necessary conditions for the filtered algebra $\mathcal{D}_{B,\kappa}$ (Notation 0.3) to be a PBW deformation of $B \# H$.

Definition 1.4. Recall Notation 0.2. We say that the map κ is *H-invariant* if $h \cdot (\kappa(r)) = \kappa(h \cdot r)$ in $H \oplus (V \otimes H)$ for any relation $r \in I$ and $h \in H$.

1B. *Deformations of algebras and Hochschild cohomology.* In this part, we remind the reader of the notion of a deformation of a k-algebra A and how Hochschild cohomology plays a role in its construction. This is seminal work of Gerstenhaber [1964], adapted to our graded setting as in [Braverman and Gaitsgory 1996].

Definition 1.5 (A_t, $A_{(j)}$). Let A be an associative algebra and let t be an indeterminate. A *deformation* of A over $k[t]$ is an associative $k[t]$-algebra A_t over $k[t]$ which is isomorphic to $A[t]$ as k-vector spaces, with multiplication given by

$$a_1 * a_2 = \mu_0(a_1 \otimes a_2) + \mu_1(a_1 \otimes a_2)t + \mu_2(a_1 \otimes a_2)t^2 + \cdots$$

for all $a_1, a_2 \in A$. Here, $\mu_i : A \otimes A \to A$ is a k-linear map, referred to as the *i-th multiplication map*. Moreover, $\mu_0(a_1 \otimes a_2) = a_1 a_2$ is the usual product in A.

Now assume that A is graded by $\mathbb{N}$. A *graded deformation* of A over $k[t]$ is an algebra A_t as above, which is itself graded by $\mathbb{N}$, setting $\deg(t) = 1$. The map μ_i is homogeneous of degree $-i$ in this case. A *j-th-level graded deformation* of A is a graded associative algebra $A_{(j)}$ over $k[t]/(t^{j+1})$ that is isomorphic to $A[t]/(t^{j+1})$ as k-vector spaces, with multiplication given by

$$a_1 * a_2 = \mu_0(a_1 \otimes a_2) + \mu_1(a_1 \otimes a_2)t + \cdots + \mu_j(a_1 \otimes a_2)t^j.$$

The maps $\mu_i : A \otimes A \to A$ are extended to be linear over $k[t]/(t^{j+1})$.

The associativity of $*$ for the deformation A_t imposes conditions on the maps μ_i. Specifically, for each degree i, the following equation must hold for all $a_1, a_2, a_3 \in A$:

$$\sum_{j=0}^{i} \mu_j(\mu_{i-j}(a_1 \otimes a_2) \otimes a_3) = \sum_{j=0}^{i} \mu_j(a_1 \otimes \mu_{i-j}(a_2 \otimes a_3)). \tag{1-6}$$

We use Hochschild cohomology to study these equations.

Definition 1.7 ($\boldsymbol{B}_\bullet(A)$). Let A be a k-algebra and let M be an A-bimodule, or equivalently, an A^e-module. Here, $A^e := A \otimes A^{\mathrm{op}}$. The *Hochschild cohomology* of M is $\mathrm{HH}^n(A, M) := \mathrm{Ext}^n_{A^e}(A, M)$. Moreover, this cohomology may be derived from the *bar resolution* $\boldsymbol{B}_\bullet(A)$ of the A^e-module A:

$$\boldsymbol{B}_\bullet(A) : \quad \cdots \xrightarrow{\delta_3} A^{\otimes 4} \xrightarrow{\delta_2} A^{\otimes 3} \xrightarrow{\delta_1} A \otimes A \xrightarrow{\delta_0} A \longrightarrow 0,$$

where

$$\delta_n(a_0 \otimes \cdots \otimes a_{n+1}) := \sum_{i=0}^{n} (-1)^i a_0 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_{n+1}$$

for all $n \geq 0$ and $a_0, \ldots, a_{n+1} \in A$. When $M = A$, write $\mathrm{HH}^n(A)$ for $\mathrm{HH}^n(A, A)$. Moreover, if A is ($\mathbb{N}$-)graded, then $\mathrm{HH}^n(A)$ inherits the grading of A: If $A = \bigoplus_i A_i$, then $\mathrm{HH}^n(A) = \bigoplus_i \mathrm{HH}^{n,i}(A)$.

Note that $\mathrm{Hom}_k(A^{\otimes n}, A) \cong \mathrm{Hom}_{A^e}(A^{\otimes(n+2)}, A)$, since the A^e-module $A^{\otimes(n+2)}$ is induced from the k-module $A^{\otimes n}$. We will identify these two Hom spaces often without further comment. Now we make some remarks about the multiplication maps μ_i.

Remark 1.8. Using (1-6) for $i = 1$, we see that

$$\mu_1(a_1 \otimes a_2)a_3 + \mu_1(a_1 a_2 \otimes a_3) = a_1\mu_1(a_2 \otimes a_3) + \mu_1(a_1 \otimes a_2 a_3) \tag{1-9}$$

for all $a_1, a_2, a_3 \in A$. In other words, μ_1 is a Hochschild 2-cocycle on the bar resolution of A; that is, $\delta_3^*(\mu_1) := \mu_1 \circ \delta_3$ vanishes. (Here we have identified the input $a_1 \otimes a_2 \otimes a_3$ with $1 \otimes a_1 \otimes a_2 \otimes a_3 \otimes 1$ to apply δ_3, under the identification of Hom spaces described above.)

Next, using (1-6) for $i = 2$, we see that

$$\begin{aligned}\mu_2(a_1 \otimes a_2)a_3 + \mu_1(\mu_1(a_1 \otimes a_2) \otimes a_3) + \mu_2(a_1 a_2 \otimes a_3)\\ = a_1\mu_2(a_2 \otimes a_3) + \mu_1(a_1 \otimes \mu_1(a_2 \otimes a_3)) + \mu_2(a_1 \otimes a_2 a_3).\end{aligned}$$

Therefore

$$\delta_3^*(\mu_2)(a_1 \otimes a_2 \otimes a_3) = \mu_1(\mu_1(a_1 \otimes a_2) \otimes a_3) - \mu_1(a_1 \otimes \mu_1(a_2 \otimes a_3)) \tag{1-10}$$

for all $a_1, a_2, a_3 \in A$. In other words, μ_2 is a cochain on the bar resolution of A whose coboundary is given by the right-hand side of (1-10).

For all $i \geq 1$, (1-6) is equivalent to

$$\delta_3^*(\mu_i)(a_1 \otimes a_2 \otimes a_3) = \sum_{j=1}^{i-1} \mu_j(\mu_{i-j}(a_1 \otimes a_2) \otimes a_3) - \mu_j(a_1 \otimes \mu_{i-j}(a_2 \otimes a_3)). \tag{1-11}$$

That is, μ_i is a cochain on the bar resolution of A whose coboundary is given by the right-hand side of (1-11).

Definition 1.12. The right-hand side of (1-11) is the $(i-1)$*-th obstruction* of the deformation A_t of A from Definition 1.5. An $(i-1)$-th-level graded deformation (defined by maps $\mu_1, \dots, \mu_{i-1}$) *lifts* to an i-th-level graded deformation if there exists a map μ_i for which $\mu_1, \dots, \mu_{i-1}, \mu_i$ define an i-th-level graded deformation.

The next proposition makes clear the choice of terminology in the above definition. Ultimately, one is interested in a deformation of A over $k[t]$ and its specializations at particular values of t. The i-th-level graded deformations are steps in this direction.

Proposition 1.13 [Braverman and Gaitsgory 1996, Proposition 1.5]. *All obstructions to lifting an* $(i-1)$*-th-level graded deformation to the next level lie in* $\mathrm{HH}^{3,-i}(A)$. *An* $(i-1)$*-th-level deformation lifts to the* i*-th-level if and only if its* $(i-1)$*-th obstruction cocycle is zero in cohomology*, *i.e.*, *there is a map* μ_i *such that* (1-11) *holds for all* a_1, a_2, a_3 *in* A.

The connection between graded deformations and PBW deformations is well known; the following statement is a consequence of the canonical embedding of A as a k-linear direct summand of $A[t]$, with splitting map given by specialization at $t=0$.

Proposition 1.14 [Braverman and Gaitsgory 1996, Remark 1.4]. *Given a graded algebra A and a graded deformation A_t of A, then A_t specialized at $t=1$ is a PBW deformation of A.* □

Now we explain that the two notions of deformation of $B\#H$ coincide; recall Notations 0.2 and 0.3. The following result is well known in related contexts, but we include some details for the reader's convenience.

Proposition 1.15. *The following statements are equivalent.*

- *The algebra $\mathfrak{D}_{B,\kappa} := (T_k(V)\#H)/(r-\kappa(r))_{r\in I}$ is a PBW deformation of $B\#H$.*
- *The algebra $\mathfrak{D}_{B,\kappa,t} := (T_k(V)\#H)[t]/(r-\kappa^L(r)t-\kappa^C(r)t^2)_{r\in I}$ is a graded deformation of $B\#H$ over $k[t]$.*

Proof. Assume that $\mathfrak{D}_{B,\kappa}$ is a PBW deformation of $B\#H$. By its definition, $\mathfrak{D}_{B,\kappa,t}$ is an associative algebra, and so we need only see that it is isomorphic to $B\#H[t]$ as a vector space. To this end, use the PBW property to define a k-linear map $\pi : B\#H \to T_k(V)\#H$ whose composition with the quotient map onto $\mathfrak{D}_{B,\kappa}$ is an isomorphism of filtered vector spaces. Extend π to a $k[t]$-linear map from $B\#H[t]$ to $T_k(V)\#H[t]$. Its composition with the quotient map to $\mathfrak{D}_{B,\kappa,t}$ is an isomorphism of k-vector spaces; one sees this by a degree argument.

Conversely, assume that $\mathfrak{D}_{B,\kappa,t}$ is a graded deformation of $B\#H$ over $k[t]$. We may specialize to $t=1$ to obtain $\mathfrak{D}_{B,\kappa}$. Now apply Proposition 1.14 to conclude that $\mathfrak{D}_{B,\kappa}$ is a PBW deformation of $B\#H$. □

2. Resolutions for smash product algebras

In this section, let A denote the smash product $B\#H$, which is an $\mathbb{N}$-graded algebra: $A=\bigoplus_{j\geq 0}(B_j\otimes H)$. Thus $A_0\cong H$. The aim is to construct a free A^e-resolution $X_\bullet$ of the A^e-module A from resolutions of H and of B (denoted by $C_\bullet$ and $D_\bullet$, respectively). This construction simultaneously generalizes results of Guccione and Guccione [2002] and of Shepler and Witherspoon [2012b, Section 4]. A similar resolution was constructed independently by Negron [2014].

Definition 2.1 ($C_\bullet$, C_i, C_i'). For $i\geq 0$, let C_i denote the H^e-module $H^{\otimes(i+2)}$. The left H-comodule structure $\rho : C_i \to H\otimes C_i$ is given by

$$\rho(h^0\otimes h^1\otimes\cdots\otimes h^{i+1}) := \sum h_1^0\cdots h_1^{i+1}\otimes h_2^0\otimes\cdots\otimes h_2^{i+1}\in H\otimes C_i$$

for all $h^0,\dots,h^{i+1}\in H$. For $h\in H$, the left and right h-actions on an element $x\in C_i$ are given respectively by left and right multiplication by h in the leftmost and rightmost factors of x. Now, let

$$C_\bullet : \quad \cdots \longrightarrow C_1 \longrightarrow C_0 \longrightarrow H \longrightarrow 0$$

be the bar resolution $\boldsymbol{B}_\bullet(H)$ of H (Definition 1.7), which is an H^e-free resolution of H.

There is an isomorphism of free H^e-modules $C_i \cong H \otimes C_i' \otimes H$, where $C_i' = H^{\otimes i}$ if $i \geq 1$ and $C_0' = k$. We give each C_i' the H-comodule structure inducing that on C_i under the usual tensor product of comodules.

Remark 2.2. The resolution $C_\bullet$ satisfies the following conditions:

(i) The right H-action and left H-coaction on C_i commute in the sense that for all $x \in C_i$ and $h \in H$

$$\sum (x \cdot h)_{-1} \otimes (x \cdot h)_0 = \sum x_{-1} h_1 \otimes (x_0 \cdot h_2).$$

That is, each C_i is a Hopf module (for which the action is a left action and the coaction is a right coaction).

(ii) The differentials are left H-comodule homomorphisms.

Definition 2.3 ($D_\bullet$, D_i, D_i'). Recall that B is a Koszul algebra. Let

$$\cdots \longrightarrow D_1 \longrightarrow D_0 \longrightarrow B \longrightarrow 0$$

be the *Koszul resolution* of B as a B^e-module: $D_0 = B \otimes B$, $D_1 = B \otimes V \otimes B$, $D_2 = B \otimes I \otimes B$, and for each $n \geq 3$, $D_i = B \otimes D_i' \otimes B$, where

$$D_i' = \bigcap_{j=0}^{i-2} (V^{\otimes j} \otimes I \otimes V^{\otimes (i-2-j)}).$$

Each D_i is a subspace of $B^{\otimes (i+2)}$, and the differential on the Koszul resolution is the one induced by the canonical embedding of the Koszul resolution into the bar resolution of B.

Remark 2.4. The resolution $D_\bullet$ satisfies the following conditions:

(i) Each B^e-module D_i is a left H-module and the differentials are H-module homomorphisms.

(ii) The left actions of B and H on D_i are compatible in the sense that they induce a left action of $A = B \# H$ on D_i.

(iii) In addition, the right B-action on D_i is compatible with the left H-action on D_i in the sense that for all $h \in H$, $y = b^0 \otimes y' \otimes b^1 \in D_i$ for $y' \in D_i'$ and $b^0, b^1, b \in B$,

$$\begin{aligned} h \cdot (y \cdot b) &= \sum (h_1 \cdot y) \cdot (h_2 \cdot b) = \Big[\sum (h_1 \cdot b^0) \otimes (h_2 \cdot y') \otimes (h_3 \cdot b^1)\Big] \cdot (h_4 \cdot b) \\ &= \sum (h_1 \cdot b^0) \otimes (h_2 \cdot y') \otimes (h_3 \cdot b^1) b = (h \cdot y) \cdot b. \end{aligned}$$

(iv) Each D_i is considered to be a left H-comodule in a trivial way by requiring that it be H-coinvariant; that is, the comodule structure is given by maps $\rho_i : D_i \to H \otimes D_i$, where $\rho_i(y) = 1 \otimes y$ for all $y \in D_i$. The maps ρ_i are maps of left H-modules, if we give $H \otimes D_i$ the tensor product H-module structure, where the factor H has the adjoint H-module structure. See Section 1A.

Construction 2.5 ($X_\bullet$). We wish to combine the two resolutions, $C_\bullet$ and $D_\bullet$ from Definitions 2.1 and 2.3, to form a resolution $X_\bullet$ of $A = B \# H$ by A-bimodules, via a tensor product. To that end, we first apply $(A \otimes_H -)$ to $C_\bullet$. Note that A is free as a right H-module (under multiplication) and that $A \otimes_H H \cong A$. The following sequence of $A \otimes H^{\mathrm{op}}$-modules is therefore exact:

$$\cdots \longrightarrow A \otimes_H C_1 \longrightarrow A \otimes_H C_0 \longrightarrow A \longrightarrow 0.$$

Similarly, we apply $(- \otimes_B A)$ to $D_\bullet$. Note that A is free as a left B-module, and that $B \otimes_B A \cong A$. The following sequence of $B \otimes A^{\mathrm{op}}$-modules is therefore exact:

$$\cdots \longrightarrow D_1 \otimes_B A \longrightarrow D_0 \otimes_B A \longrightarrow A \longrightarrow 0.$$

We will next extend the actions on the modules in each of these two sequences so that they become A^e-modules. Then, we will take their tensor product over A.

We extend the right H-module structure on $A \otimes_H C_\bullet$ to a right A-module structure by defining a right action of B on $A \otimes_H C_\bullet$: for all $a \in A$, $x \in C_i$, $b \in B$, we set

$$(a \otimes_H x) \cdot b := \sum a(x_{-1} \cdot b) \otimes_H x_0. \tag{2-6}$$

This does indeed make $A \otimes_H C_i$ into a right B-module, and, by combining with the right action of H, gives a right action of A on $A \otimes_H C_i$. Note that for $x = x^0 \otimes \cdots \otimes x^{i+1} \in C_i$ (with $x^0, \ldots, x^{i+1} \in H$),

$$\begin{aligned}\rho(hx) &= \sum (hx)_{-1} \otimes (hx)_0 = \sum h_1 x_1^0 \cdots x_1^{i+1} \otimes h_2 x_2^0 \otimes \cdots \otimes x_2^{i+1} \\ &= \sum h_1 x_{-1} \otimes h_2 x_0.\end{aligned}$$

The action is well-defined: If $h \in H$, then

$$(ah \otimes_H x) \cdot b \overset{(2\text{-}6)}{=} \sum ah(x_{-1} \cdot b) \otimes_H x_0 = \sum a(h_1 x_{-1} \cdot b) \otimes_H h_2 x_0 \overset{(2\text{-}6)}{=} (a \otimes_H hx) \cdot b.$$

Since the differentials on $C_\bullet$ are H-comodule homomorphisms (Remark 2.2(ii)), this action commutes with the differentials.

We extend the left B-module structure on $D_i \otimes_B A$ to a left A-module structure by defining a left action of H by

$$h \cdot (y \otimes_B a) := \sum (h_1 \cdot y) \otimes_B h_2 a \tag{2-7}$$

for all $h \in H$, $y \in D_i$, $a \in A$. It is well-defined, since for all $h \in H$, $b \in B$, we have by the definitions in Section 1A that

$$\begin{aligned} h \cdot (yb \otimes_B a) &\overset{(2\text{-}7)}{=} \sum (h_1 \cdot (yb)) \otimes_B (h_2 a S(h_3)) = \sum (h_1 \cdot y)(h_2 \cdot b) \otimes_B h_3 a S(h_4) \\ &= \sum (h_1 \cdot y) \otimes_B (h_2 \cdot b) h_3 a S(h_4) = \sum (h_1 \cdot y) \otimes_B (h_2 \cdot (ba)) \\ &\overset{(2\text{-}7)}{=} h \cdot (y \otimes_B ba). \end{aligned}$$

The left H-action on D_i is compatible with the right B-action on D_i by Remark 2.4(iii). Again, this action commutes with the differentials, since the differentials on $D_\bullet$ are H-module homomorphisms (Remark 2.4(i)).

We may now consider $A \otimes_H C_\bullet$ and $D_\bullet \otimes_B A$ to be complexes of A^e-modules via the A-bimodule structure defined above. We take their tensor product over A, letting $X_{\bullet,\bullet} := (A \otimes_H C_\bullet) \otimes_A (D_\bullet \otimes_B A)$; that is, for all $i, j \geq 0$,

$$X_{i,j} := (A \otimes_H C_i) \otimes_A (D_j \otimes_B A), \tag{2-8}$$

with horizontal and vertical differentials

$$d^h_{i,j} : X_{i,j} \to X_{i-1,j} \quad \text{and} \quad d^v_{i,j} : X_{i,j} \to X_{i,j-1}$$

given by $d^h_{i,j} := d^{C_\bullet}_i \otimes \mathrm{id}$ and $d^v_{i,j} := (-1)^i \, \mathrm{id} \otimes d^{D_\bullet}_j$.

Finally, let $X_\bullet$ be the total complex of $X_{\bullet,\bullet}$:

$$\cdots \longrightarrow X_2 \longrightarrow X_1 \longrightarrow X_0 \longrightarrow A \longrightarrow 0, \tag{2-9}$$

with $X_n = \bigoplus_{i+j=n} X_{i,j}$.

Theorem 2.10. *We have the following statements*:

(a) *For each i, j, the A^e-module $X_{i,j}$ is isomorphic to $A \otimes C'_i \otimes D'_j \otimes A$.*

(b) *The complex $X_\bullet$ given in (2-9) is a free resolution of the A^e-module A.*

Proof. (a) Write $C_i \cong H \otimes C'_i \otimes H$ and $D_j \cong B \otimes D'_j \otimes B$ for vector spaces C'_i and D'_j, as in Definitions 2.1 and 2.3. Then

$$\begin{aligned} X_{i,j} &\cong (A \otimes_H H \otimes C'_i \otimes H) \otimes_A (B \otimes D'_j \otimes B \otimes_B A) \\ &\cong (A \otimes C'_i \otimes H) \otimes_A (B \otimes D'_j \otimes A). \end{aligned}$$

We will show that this is isomorphic to $A \otimes C'_i \otimes D'_j \otimes A$ as an A^e-module. First, define a map as follows:

$$\begin{aligned} &(A \otimes C'_i \otimes H) \times (B \otimes D'_j \otimes A) \to A \otimes C'_i \otimes D'_j \otimes A, \\ &\qquad (a \otimes x \otimes h, b \otimes y \otimes a') \mapsto \sum a(x_{-1} h_1 \cdot b) \otimes x_0 \otimes (h_2 \cdot y) \otimes h_3 a' \end{aligned} \tag{2-11}$$

for all $a, a' \in A$, $x \in C_i'$, $y \in D_j'$, $h \in H$, $b \in B$. This map is k-bilinear by definition, and we will check that it is A-balanced. First, let $b' \in B$. We rewrite $(a \otimes x \otimes h) \cdot b'$ as follows. First, using $A \otimes C_i' \otimes H \cong A \otimes_H C_i$, identify this element with $a \otimes_H (1 \otimes x \otimes h) \in A \otimes_H C_i$. By (2-6),

$$(a \otimes_H (1 \otimes x \otimes h)) \cdot b' = \sum a((1 \otimes x \otimes h)_{-1} \cdot b') \otimes_H (1 \otimes x \otimes h)_0.$$

By Definition 2.1, and by identifying $x \in C_i'$ with $x^1 \otimes x^2 \otimes \cdots \otimes x^i$, we have that

$$\begin{aligned}\rho(1 \otimes x \otimes h) &= \sum (1 \otimes x \otimes h)_{-1} \otimes (1 \otimes x \otimes h)_0 \\ &= \sum (x_1^1 x_1^2 \cdots x_1^i h_1) \otimes (1 \otimes x_2^1 \otimes x_2^2 \otimes \cdots \otimes x_2^i \otimes h_2).\end{aligned}$$

So, $(1 \otimes x \otimes h)_{-1} = x_{-1} h_1$ and $(1 \otimes x \otimes h)_0 = x_0 \otimes h_2$. Now $C_i \cong H \otimes C_i' \otimes H$ as an H-comodule, so

$$\begin{aligned}((a \otimes x \otimes h) \cdot b', b \otimes y \otimes a') &= \sum (a(x_{-1} h_1 \cdot b') \otimes x_0 \otimes h_2, b \otimes y \otimes a') \\ &\mapsto \sum a(x_{-2} h_1 \cdot b')(x_{-1} h_2 \cdot b) \otimes x_0 \otimes (h_3 \cdot y) \otimes h_4 a'.\end{aligned}$$

On the other hand,

$$\begin{aligned}((a \otimes x \otimes h, b' \cdot (b \otimes y \otimes a')) &= (a \otimes x \otimes h, b'b \otimes y \otimes a') \\ &\mapsto \sum a(x_{-1} h_1 \cdot (b'b)) \otimes x_0 \otimes (h_2 \cdot y) \otimes h_3 a'),\end{aligned}$$

which is the same as the previous image since B is an H-module algebra. Now let $\ell \in H$. Then

$$\begin{aligned}((a \otimes x \otimes h) \cdot \ell, b \otimes y \otimes a') &= (a \otimes x \otimes h\ell, b \otimes y \otimes a') \\ &\mapsto \sum a(x_{-1} h_1 \ell_1 \cdot b) \otimes x_0 \otimes (h_2 \ell_2 \cdot y) \otimes h_3 \ell_3 a'.\end{aligned}$$

On the other hand,

$$\begin{aligned}(a \otimes x \otimes h, \ell \cdot (b \otimes y \otimes a')) &= \sum (a \otimes x \otimes h, (\ell_1 \cdot b) \otimes (\ell_2 \cdot y) \otimes \ell_3 a') \\ &\mapsto \sum a(x_{-1} h_1 \ell_1 \cdot b) \otimes x_0 \otimes (h_2 \ell_2 \cdot y) \otimes h_3 \ell_3 a',\end{aligned}$$

which is the same as the previous image. Therefore, there is an induced map

$$(A \otimes C_i' \otimes H) \otimes_A (B \otimes D_j' \otimes A) \to A \otimes C_i' \otimes D_j' \otimes A.$$

Now, we verify that the map below is an inverse map of (2-11):

$$a \otimes x \otimes y \otimes a' \mapsto (a \otimes x \otimes 1) \otimes_A (1 \otimes y \otimes a'). \tag{2-12}$$

It is clear that first applying (2-12) and then (2-11) yields the identity map on $A \otimes C_i' \otimes D_j' \otimes A$. On the other hand, the image of first applying (2-11) then (2-12) to $(a \otimes x \otimes h, b \otimes y \otimes a')$ is

$$
\begin{aligned}
\sum (a(x_{-1}h_1 \cdot b) \otimes x_0 \otimes 1) \otimes_A (1 \otimes (h_2 \cdot y) \otimes h_3 a') \\
&= \sum (a(x_{-1}h_1 \cdot b) \otimes x_0 \otimes 1) \otimes_A (\epsilon(h_2) \otimes (h_3 \cdot y) \otimes h_4 a') \\
&= \sum (a(x_{-1}h_1 \cdot b) \otimes x_0 \otimes 1) \otimes_A (h_2 \cdot (1 \otimes y \otimes a')) \\
&= \sum (a(x_{-1}h_1 \cdot b) \otimes x_0 \otimes h_2) \otimes_A (1 \otimes y \otimes a') \\
&\overset{(2\text{-}6)}{=} ((a \otimes x \otimes h) \cdot b) \otimes_A (1 \otimes y \otimes a') \\
&= (a \otimes x \otimes h) \otimes_A (b \otimes y \otimes a').
\end{aligned}
$$

Therefore the two A^e-modules X_{ij} and $A \otimes C_i' \otimes D_i' \otimes A$ are isomorphic, as claimed.

(b) We wish to apply the Künneth theorem to show that the complex $X_\bullet$ is a free resolution of the A^e-module A. To that end, we check that each term in the complex $D_\bullet \otimes_B A$ is a free left A-module and that the image of each differential in the complex is also projective as a left A-module. First, write each $D_i \otimes_B A \cong (B \otimes D_i' \otimes B) \otimes_B A \cong B \otimes D_i' \otimes A$. Define a k-linear map $f : A \otimes D_i' \otimes B \to B \otimes D_i' \otimes A$ by

$$f(rh \otimes y \otimes b) = \sum r \otimes (h_1 \cdot y) \otimes h_2 b$$

for $h \in H$, $y \in D_i'$, and $r, b \in B$. Give $A \otimes D_i' \otimes B$ the structure of a left A-module by requiring A to act by left multiplication on the leftmost factor. Clearly this is a free left A-module. The map f is an A-module homomorphism by the definition of the left A-action on $B \otimes D_i' \otimes A$; see (2-7). We claim that the following map is an inverse map, so that f is an isomorphism of A-modules: let S^{-1} denote the (composition) inverse of the antipode S of H. Let $g : B \otimes D_i' \otimes A \to A \otimes D_i' \otimes B$ be the k-linear map defined by

$$g(r \otimes y \otimes hb) = \sum rh_2 \otimes (S^{-1}(h_1) \cdot y) \otimes b.$$

Since for each $h \in H$ we have $\sum h_2 S^{-1}(h_1) = \epsilon(h) = \sum S^{-1}(h_2) h_1$ (see, e.g., [Radford 2012, Proposition 7.1.10]), the function g is indeed the inverse of f. Thus, each term in the complex $D_\bullet \otimes_B A$ is a free left A-module.

That the image of each differential is projective as a left A-module may be proved inductively, starting on one end of the complex

$$\cdots \longrightarrow D_1 \otimes_B A \xrightarrow{d_1 \otimes \mathrm{id}} D_0 \otimes_B A \xrightarrow{d_0 \otimes \mathrm{id}} A \longrightarrow 0,$$

as follows. Since A is a projective left A-module and $d_0 \otimes \mathrm{id}$ is surjective, the map splits, implying that $\mathrm{Ker}(d_0 \otimes \mathrm{id}) = \mathrm{Im}(d_1 \otimes \mathrm{id})$ is a direct summand of the free left A-module $D_0 \otimes_B A$. Therefore it is projective. Again, since $\mathrm{Im}(d_1 \otimes \mathrm{id})$ is projective, the map $d_1 \otimes \mathrm{id}$ from $D_1 \otimes_B A$ to its image splits so that $\mathrm{Ker}(d_1 \otimes \mathrm{id}) = \mathrm{Im}(d_2 \otimes \mathrm{id})$

is a direct summand of the free left A-module $D_1 \otimes_B A$. Continuing in this way, we see that $\mathrm{Im}(d_i \otimes \mathrm{id})$ is a free left A-module for each i.

The Künneth theorem [Weibel 1994, Theorem 3.6.3] then gives for each n an exact sequence

$$0 \longrightarrow \bigoplus_{i+j=n} \mathrm{H}_i(A \otimes_H C_\bullet) \otimes_A \mathrm{H}_j(D_\bullet \otimes_B A) \longrightarrow \mathrm{H}_n((A \otimes_H C_\bullet) \otimes_A (D_\bullet \otimes_B A)) \\ \longrightarrow \bigoplus_{i+j=n-1} \mathrm{Tor}_1^A(\mathrm{H}_i(A \otimes_H C_\bullet), \mathrm{H}_j(D_\bullet \otimes_B A)) \to 0.$$

Now $A \otimes_H C_\bullet$ and $D_\bullet \otimes_B A$ are exact except in degree 0, where their homologies are each A; that is, $H_0(A \otimes_H C_\bullet) = A$ and $H_0(D_\bullet \otimes_B A) = A$. Therefore the only potentially nonzero Tor term is when $i = 0 = j$, or $\mathrm{Tor}_1^A(A, A)$, yet this equals 0 since A is flat over A. So for each n, we have

$$\mathrm{H}_n((A \otimes_H C_\bullet) \otimes_A (D_\bullet \otimes_B A)) \cong \bigoplus_{i+j=n} \mathrm{H}_j(A \otimes_H C_\bullet) \otimes_A \mathrm{H}_i(D_\bullet \otimes_B A).$$

Again the right side is only nonzero when $i = 0 = j$, so we have

$$\mathrm{H}_0((A \otimes_H C_\bullet) \otimes_A (D_\bullet \otimes_B A)) \cong \mathrm{H}_0(A \otimes_H C_\bullet) \otimes_A \mathrm{H}_0(D_\bullet \otimes_B A) \cong A \otimes_A A \cong A$$

and $\mathrm{H}_n((A \otimes_H C_\bullet) \otimes_A (D_\bullet \otimes_B A)) = 0$ for all $n > 0$. Thus we have proven that $X_\bullet$ is an A^e-free resolution of A. □

We next relate the resolution $X_\bullet$ of A (from Construction 2.5) to the bar resolution $\boldsymbol{B}_\bullet(A)$ of A:

Lemma 2.13. *There exist degree-preserving chain maps between $X_\bullet$ and the bar resolution $\boldsymbol{B}_\bullet(A)$ of A,*

$$\phi_\bullet : X_\bullet \longrightarrow \boldsymbol{B}_\bullet(A) \quad \textit{and} \quad \psi_\bullet : \boldsymbol{B}_\bullet(A) \longrightarrow X_\bullet,$$

such that $\psi_n \phi_n$ is the identity map on the A^e-submodule $X_{0,n}$ of X_n for each $n \geq 0$.

Proof. Recall by Notation 0.2 that B is generated by the vector space V, with quadratic relations $I \subseteq V \otimes V$. First we prove by induction on n that there are degree-preserving maps $\phi_n : X_n \to A^{\otimes(n+2)}$ and $\psi_n : A^{\otimes(n+2)} \to X_n$ commuting with the differentials. For clarity, we denote the differential on the bar resolution of A by δ. We have the diagram

$$\begin{tikzcd} X_\bullet : & \cdots \arrow[r] & X_3 \arrow[r, "d_3"] \arrow[d, shift left, "\phi_3"] & X_2 \arrow[r, "d_2"] \arrow[d, shift left, "\phi_2"] & X_1 \arrow[r, "d_1"] \arrow[d, shift left, "\phi_1"] & X_0 \arrow[r, "d_0"] \arrow[d, shift left, "\phi_0"] & A \arrow[r] & 0 \\ \boldsymbol{B}_\bullet(A) : & \cdots \arrow[r] & A^{\otimes 5} \arrow[r, "\delta_3"] \arrow[u, shift left, "\psi_3"] & A^{\otimes 4} \arrow[r, "\delta_2"] \arrow[u, shift left, "\psi_2"] & A^{\otimes 3} \arrow[r, "\delta_1"] \arrow[u, shift left, "\psi_1"] & A \otimes A \arrow[r, "\delta_0"] \arrow[u, shift left, "\psi_0"] & A \arrow[r] & 0 \end{tikzcd}$$

where $\boldsymbol{B}_n(A) = A^{\otimes(n+2)}$ and $X_n = \bigoplus_{i+j=n} X_{i,j}$, with $X_{i,j}$ defined in (2-8); see also Theorem 2.10(a).

Define $\phi_0 = \mathrm{id} \otimes \mathrm{id} = \psi_0$, the identity map from $A \otimes A$ to itself. We wish to define $\phi_\bullet$ so that when restricted to $X_{0,\bullet}$ it corresponds to the standard embedding of the Koszul complex into the bar complex: for $n = 1$, this is the embedding of $A \otimes V \otimes A$ into $A \otimes A \otimes A$ via the containment of V in A. We may define ϕ_1 on $X_1 = X_{0,1} \oplus X_{1,0} \cong (A \otimes V \otimes A) \oplus (A \otimes H \otimes A)$ by $\phi_1(1 \otimes v \otimes 1) = 1 \otimes v \otimes 1$ and $\phi_1(1 \otimes h \otimes 1) = 1 \otimes h \otimes 1$ for all $v \in V$, $h \in H$. Note that for $n \geq 2$,

$$X_{0,n} \cong A \otimes \left(\bigcap_{i=0}^{n-2} V^{\otimes i} \otimes I \otimes V^{\otimes (n-i-2)} \right) \otimes A, \tag{2-14}$$

which is a free A^e-submodule of $A^{\otimes(n+2)}$. For each i, j with $i + j = n$, choose a basis of the vector space $C'_i \otimes D'_j$ (whose elements are homogeneous of degree j, as H is declared to have degree 0). By hypothesis, ϕ_{n-1} is degree-preserving, and d_n is degree-preserving by construction. So, applying $\phi_{n-1} d_n$ to these basis elements of $C'_i \otimes D'_j$ produces elements of degree j in the kernel of δ_{n-1}, that is, the image of δ_n. We define ϕ_n by choosing (arbitrary) corresponding elements in the inverse image of $\mathrm{Im}(\delta_n)$. If we start with an element in $X_{0,n}$, we may choose its image in $A^{\otimes(n+2)}$ under the canonical embedding of $X_{0,n}$ into $A^{\otimes(n+2)}$ (see (2-14)). Given $X_{i,j}$ and $X_{i',j'}$ with $i + j = i' + j' = n$ and $i \neq 0$, $i' \neq 0$, elements of $X_{i,j}$ have degree j and elements of $X_{i',j'}$ have degree j'. So their images under ϕ_n may be chosen independently, and in particular, independently of those of $X_{0,n}$. Thus, we have the maps ϕ_n, as desired.

Now we show that ψ_n may be chosen so that $\psi_n \phi_n$ is the identity map on $X_{0,n}$. In degree 1, we have summands $X_{0,1} \cong A \otimes V \otimes A$ and $X_{1,0} \cong A \otimes H \otimes A$. Note that $V \oplus H$ is a direct summand of A as a vector space. We may therefore define $\psi_1(1 \otimes v \otimes 1) = 1 \otimes v \otimes 1$ in $X_{0,1}$ for all $v \in V$, and $\psi_1(1 \otimes h \otimes 1) = 1 \otimes h \otimes 1$ in $X_{1,0}$ for all $h \in H$. We also have that ψ_1 is the identity map on elements of the form $1 \otimes z \otimes 1$, for z ranging over a basis of a chosen complement of $V \oplus H$ as a vector subspace of A. This complement may be chosen arbitrarily, subject to the condition that $d_1 \psi_1(1 \otimes z \otimes 1) = \psi_0 \delta_1(1 \otimes z \otimes 1)$. Since ψ_0, d_1, δ_1 all have degree 0 as maps, one may also choose ψ_1 to have degree 0. In particular, note that $\psi_1 \phi_1$ is the identity map on $X_{0,1}$. Now let $n \geq 2$ and assume that ψ_{n-2} and ψ_{n-1} have been defined to be degree-0 maps for which $d_{n-1} \psi_{n-1} = \psi_{n-2} \delta_{n-1}$ and $\psi_{n-1} \phi_{n-1}$ is the identity map on $X_{0,n-1}$. To define ψ_n, first note that $A^{\otimes(n+2)}$ contains the space $X_{0,n}$ as an A^e-submodule (see (2-14)) and the image of each $X_{i,j}$ under ϕ_n ($n = i + j$, $i \geq 1$). By construction, their images intersect in 0, the image of $X_{0,n}$ under ϕ_n is free and, moreover, ϕ_n is injective on restriction to $X_{0,n}$. Choose a set of free generators of $\phi_n(X_{0,n})$, and choose a set of free generators of its complement in $A^{\otimes(n+2)}$. For each chosen generator x of $X_{0,n}$, we define $\psi_n(\phi_n(x))$ to be x. On the complement of $\phi_n(X_{0,n})$, define ψ_n arbitrarily, subject to being a chain map of degree 0. Thus,

$\psi_n\phi_n$ is the identity map on $X_{0,n}$. Now for all $x \in X_{0,n}$, since $d_n(x) \in X_{0,n-1}$, we have that $\psi_{n-1}\phi_{n-1}d_n(x) = d_n(x)$, by induction. As $\delta_n\phi_n(x) = \phi_{n-1}d_n(x)$, it follows that $d_n\psi_n\phi_n(x) = \psi_{n-1}\delta_n\phi_n(x)$. So ψ_n also extends the chain map from degree $n-1$ to degree n, as desired. □

3. Poincaré–Birkhoff–Witt theorem for Hopf algebra actions

Consider the algebra $\mathcal{D}_{B,\kappa}$ from Notation 0.3. The goal of this section is to prove our main result, Theorem 3.1. We provide necessary and sufficient conditions for $\mathcal{D}_{B,\kappa}$ to be a PBW deformation of $B \# H$ (Definition 0.1) as follows:

Theorem 3.1. *The algebra $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B \# H$ if and only if the following conditions hold*:

(a) *κ is H-invariant* (Definition 1.4);

If $\dim_k V \geq 3$, *then the following equations hold for the maps* $\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C$ *and* $\kappa^L \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^L$, *which are defined on the intersection* $(I \otimes V) \cap (V \otimes I)$:

(b) $\mathrm{Im}(\kappa^L \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^L) \subseteq I$;

(c) $\kappa^L \circ (\kappa^L \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^L) = -(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)$;

(d) $\kappa^C \circ (\mathrm{id} \otimes \kappa^L - \kappa^L \otimes \mathrm{id}) \equiv 0$.

Recall Notation 0.2: B is generated by the k-vector space V with quadratic relations $I \subset V \otimes V$, so $B = T_k(V)/(I)$. Moreover, consider:

Notation 3.2 (U, $T_H(U)$, R, P). Let $U := V \otimes H$, which is an H-bimodule under the actions defined in Section 1A. Set $R = I \otimes H$, similarly an H-bimodule, and an H-subbimodule of $U \otimes_H U$. Let $P = \{r \otimes 1 - \kappa(r) \mid r \in I\}$ be the relation space of $\mathcal{D}_{B,\kappa}$, generating an H-submodule of $H \oplus U \oplus (U \otimes_H U)$ in the tensor algebra

$$T_H(U) = H \oplus U \oplus (U \otimes_H U) \oplus (U \otimes_H U \otimes_H U) \oplus \cdots .$$

Note that $U^{\otimes_H^n} \cong V^{\otimes n} \otimes H$ as k-vector spaces. We see that $\pi(P) = R$, where the map π is the projection onto the homogeneous quadratic part of P.

Consider the following preliminary results:

Lemma 3.3. *Since $T_H(U)$ is canonically isomorphic to $T_k(V) \# H$, we have that*

$$T_H(U)/(P) \cong \mathcal{D}_{B,\kappa} \quad \textit{and} \quad T_H(U)/(R) \cong (T_k(V) \# H)/(I) \cong B \# H,$$

where (I) is identified with the ideal of $T_k(V) \# H$ generated by I.

Hence, $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B \# H$ if and only if $T_H(U)/(P)$ is a PBW deformation of $T_H(U)/(R)$. □

Lemma 3.4 [Shepler and Witherspoon 2012b, Lemma 5.2]. *If $T_H(U)/(P)$ is a PBW deformation of $T_H(U)/(R)$, then the following conditions hold for maps $\alpha : R \to U$ and $\beta : R \to H$ for which $P = \{x - \alpha(x) - \beta(x) \mid x \in R\}$:*

(i) $\mathrm{Im}(\alpha \otimes_H \mathrm{id} - \mathrm{id} \otimes_H \alpha) \subseteq R$.

(ii) $\alpha \circ (\alpha \otimes_H \mathrm{id} - \mathrm{id} \otimes_H \alpha) = -(\beta \otimes_H \mathrm{id} - \mathrm{id} \otimes_H \beta)$.

(iii) $\beta \circ (\mathrm{id} \otimes_H \alpha - \alpha \otimes_H \mathrm{id}) \equiv 0$.

Here, the maps $\alpha \otimes_H \mathrm{id} - \mathrm{id} \otimes_H \alpha$ and $\beta \otimes_H \mathrm{id} - \mathrm{id} \otimes_H \beta$ are defined on the subspace $(R \otimes_H U) \cap (U \otimes_H R)$ of $T_H(U)$. □

Remark 3.5. Given the maps $\kappa^L : I \to V \otimes H$ and $\kappa^C : I \to H$ as in Notation 0.2, we see that $\alpha = \kappa^L \otimes \mathrm{id}_H$ and $\beta = \kappa^C \otimes \mathrm{id}_H$.

Lemma 3.6. *Consider the algebra*

$$(T_H(U)/(P))_t := \frac{T_H(U)[t]}{(x - \alpha(x)t - \beta(x)t^2)_{x \in R}}.$$

We have that $(T_H(U)/(P))_t$ is a PBW deformation of $T_H(U)/(R)$ over $k[t]$ if and only if $\mathcal{D}_{B,\kappa,t}$ (of Proposition 1.15) is a PBW deformation of $B \# H$ over $k[t]$.

Proof. This follows from Lemma 3.3 and Remark 3.5. □

Now we provide the proof of Theorem 3.1. A somewhat shorter proof would suffice in case H is semisimple: The first proof of [Shepler and Witherspoon 2012a, Theorem 3.1] may be generalized from semisimple group algebras to semisimple Hopf algebras. In that context, one has on hand a much smaller resolution than that which we will use below.

Proof of Theorem 3.1. Note that we will employ the identifications given in the lemmas and remark above, sometimes without comment. Namely, results from Section 2 will be used here where, for instance, I is identified so that $R = I \otimes H$ and $B \# H$ is identified with $T_H(U)/(R)$.

Necessity of conditions (a)–(d). Let us first show that conditions (a)–(d) are necessary. Assume that $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B\#H$, and take Q to be the relation space of $\mathcal{D}_{B,\kappa}$. Then, for all $h \in H$ and $r \in I$, we have that $h \cdot r - h \cdot (\kappa(r)) \in Q$. (Refer to Section 1A for the definition of these actions.) We also have that $h \cdot r - \kappa(h \cdot r) \in Q$, so $h \cdot (\kappa(r)) - \kappa(h \cdot r) \in Q$. This implies that $h \cdot (\kappa(r)) = \kappa(h \cdot r)$ in $\mathcal{D}_{B,\kappa}$, since Q cannot contain nonzero elements in degree less than two. Thus, condition (a) holds. Moreover, by Lemma 3.3, $T_H(U)/(P)$ satisfies the PBW property.

Now by applying Lemma 3.4, we see that conditions (i), (ii), (iii) hold for $T_H(U)/(P)$. These conditions are equivalent to conditions (b), (c), (d) in Theorem 3.1 for the algebra $\mathcal{D}_{B,\kappa}$ by Notation 3.2 and Remark 3.5. Thus, if $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B \# H$, then conditions (a)–(d) of Theorem 3.1 must hold.

Sufficiency of conditions (a)–(d). Conversely, let us assume that conditions (a)–(d) of Theorem 3.1 hold for the algebra $\mathcal{D}_{B,\kappa}$. Equivalently, by Notation 3.2, Lemma 3.3, and Remark 3.5, we assume the following for the algebra $T_H(U)/(P)$:

- The maps α and β are H-invariant.
- Conditions (i), (ii), (iii) of Lemma 3.4 hold.

The goal is to show that $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B\#H$, which, by Proposition 1.15, is equivalent to showing that $\mathcal{D}_{B,\kappa,t}$ (of Proposition 1.15) is a graded deformation of $B\#H$ over $k[t]$. Hence, by Lemma 3.6, the goal is then equivalent to verifying that the algebra $(T_H(U)/(P))_t$ is a graded deformation of $T_H(U)/(R)$ over $k[t]$. We thus have the following strategy:

- Let A denote $T_H(U)/(R)$.
- Construct multiplication maps, $\mu_i : A \otimes A \to A$, as in Definition 1.5, subject to the restraints listed in Remark 1.8.
- Form the graded deformation A_t of A as in Definition 1.5.
- Conclude that $A' := T_H(U)/(P) \cong (A_t)|_{t=1}$ is a PBW deformation of A by Proposition 1.15.

We generalize the proof in [Shepler and Witherspoon 2012b] from group actions to Hopf algebra actions. Namely, we use the free resolution $X_\bullet$ of the A^e-module A in Construction 2.5 to define the maps μ_i. Recall that $X_\bullet$ is constructed from $C_\bullet = \boldsymbol{B}_\bullet(A)$, the bar resolution of H, and from $D_\bullet$, the Koszul resolution of B.

Extending α and β to be maps on $X_\bullet$. We first extend α and β to be maps on $X_\bullet$ as follows. In degree 2, X_2 contains as a direct summand $X_{0,2} \cong A \otimes I \otimes A$; see (2-14). As α, β are H-bilinear by H-invariance, we may extend them to A^e-module maps from $A \otimes R \otimes_H A \cong A \otimes I \otimes A$ to A by composing with the multiplication map. By abuse of notation, denote these extended maps by α, β as well. Extend α and β yet further by setting them equal to 0 on the summands $X_{2,0}$ and $X_{1,1}$ of X_2 so that they become maps $\alpha, \beta : X_2 \to A$. More precisely, $\alpha, \beta \in \operatorname{Hom}_{A^e}(X_2, A) \cong \operatorname{Hom}_k(X_2', A)$ for $X_2 \cong A \otimes X_2' \otimes A$.

Construction of the multiplication map μ_1. To build $\mu_1 \in \operatorname{Hom}_k(A \otimes A, A) \cong \operatorname{Hom}_{A^e}(A^{\otimes 4}, A)$, recall that it must be a Hochschild 2-cocycle as in (1-9). We will show that $\alpha : X_2 \to A$ is a Hochschild 2-cocycle on $X_\bullet$, that is, $d_3^*(\alpha) = 0$. Recall the chain maps of Lemma 2.13. We set $\mu_1 = \psi_2^*(\alpha)$, which will be a Hochschild 2-cocycle on $B_\bullet(A)$, that is, $\delta_3^*(\mu_1) = 0$.

To show that $d_3^*(\alpha) : X_3 \to A$ is the zero map, first note that $X_3 = X_{0,3} \oplus X_{1,2} \oplus X_{2,1} \oplus X_{3,0}$ from (2-9) and that the images of $X_{2,1}$ and $X_{3,0}$ under d_3 lie in $X_{1,1} \oplus X_{2,0}$. Since $\alpha|_{X_{1,1} \oplus X_{2,0}} \equiv 0$ by the extension above, it suffices to show that $d_3^*(\alpha)|_{X_{0,3}}$ and $d_3^*(\alpha)|_{X_{1,2}}$ are zero maps.

Rewriting condition (i) of Lemma 3.4, we see that it implies that α is 0 on the image of the differential on $X_{0,3}$ as follows: let $\sum_i 1\otimes x_i\otimes y_i\otimes z_i\otimes 1\in A\otimes((I\otimes V)\cap(V\otimes I))\otimes A=X_{0,3}$; see (2-14). Then

$$\begin{aligned}\alpha\Big(d_3\Big(\sum_i 1\otimes x_i\otimes y_i\otimes z_i\otimes 1\Big)\Big)\\ &=\alpha\Big(\sum_i x_i\otimes y_i\otimes z_i\otimes 1-\sum_i 1\otimes x_iy_i\otimes z_i\otimes 1\\ &\qquad+\sum_i 1\otimes x_i\otimes y_iz_i\otimes 1-\sum_i 1\otimes x_i\otimes y_i\otimes z_i\Big)\\ &=\sum_i(x_i\alpha(y_i\otimes z_i)-\alpha(x_i\otimes y_i)z_i).\end{aligned}$$

(To see this, note that applying the multiplication map of A to elements in I yields 0.) Thus $d_3^*(\alpha)=\mathrm{id}\otimes\alpha-\alpha\otimes\mathrm{id}$ on $X_{0,3}$; here, we identify $\mathrm{id}\otimes\alpha-\alpha\otimes\mathrm{id}$ with $m\circ(\mathrm{id}\otimes\alpha-\alpha\otimes\mathrm{id})$, where m is the multiplication map on A. We see that condition (i) indeed implies (in fact, is equivalent to) $d_3^*(\alpha)|_{X_{0,3}}\equiv 0$.

Next, we claim that α being H-invariant implies that α is also 0 on the image of the differential on $X_{1,2}$. Let $a,b\in A$, $h\in H$, and $r\in I$, and consider $a\otimes h\otimes r\otimes b$ as an element of $X_{1,2}\cong A\otimes H\otimes I\otimes A$ by Theorem 2.10(a). By the definition of the differential on $X_{1,2}$,

$$\begin{aligned}d(a\otimes h\otimes r\otimes b)&=d((a\otimes h\otimes 1)\otimes(1\otimes r\otimes b))\\ &=d(a\otimes h\otimes 1)\otimes(1\otimes r\otimes b)-(a\otimes h\otimes 1)\otimes d(1\otimes r\otimes b).\end{aligned}$$

The second term lies in $X_{1,1}$, but α is 0 on $X_{1,1}$ by definition. Therefore,

$$\begin{aligned}\alpha(d(a\otimes h\otimes r\otimes b))&=\alpha((ah\otimes 1-a\otimes h)\otimes(1\otimes r\otimes b))\\ &=\alpha\Big(ah\otimes r\otimes b-\sum a\otimes(h_1\cdot r)\otimes h_2b\Big)\\ &=ah\alpha(r)b-\sum a\alpha(h_1\cdot r)h_2b.\end{aligned}\tag{3-7}$$

Since α is H-invariant, we have that

$$h\alpha(r)=\sum h_1\epsilon(h_2)\alpha(r)=\sum h_1\alpha(r)\epsilon(h_2)=\sum h_1\alpha(r)S(h_2)h_3=\sum\alpha(h_1\cdot r)h_2,$$

where the last equality used the fact that α is H-invariant. Thus, α is zero on the image of $d=d_3$ on $X_{1,2}$ by (3-7). It follows that α is a Hochschild 2-cocycle on $X_\bullet$.

Now, let $\mu_1=\psi_2^*(\alpha)$, where $\psi_\bullet$ is a chain map satisfying the conditions of Lemma 2.13. We conclude that

$$\delta_3^*(\mu_1)=\delta_3^*(\psi_2^*(\alpha))=\psi_3^*(d_3^*(\alpha))\equiv 0,$$

as desired. So, we have a first-level graded deformation $A_{(1)}$ of A with first multiplication map $\mu_1 : A \otimes A \to A$.

As an aside, we also get that $\phi_2^*(\mu_1) = \alpha$ as cochains. To see this, first note that since α is homogeneous of degree -1 by its definition, so is μ_1. Let $x \in X_{0,2}$. By Lemma 2.13, $\psi_2\phi_2(x) = x$, and thus

$$\mu_1\phi_2(x) = \alpha\psi_2\phi_2(x) = \alpha(x).$$

Now let y be a free generator of $X_{1,1}$ or of $X_{2,0}$, which may always be chosen to have degree 1 or 0, respectively. Then $\psi_2\phi_2(y)$ has degree 1 or 0 respectively, implying that its component in $X_{0,2}$ is 0. It follows that $\mu_1\phi_2(y) = \alpha\psi_2\phi_2(y) = 0 = \alpha(y)$; the last equation follows from the extension of α to $X_\bullet$. Therefore $\phi_2^*(\mu_1) = \alpha$.

Construction of the multiplication map μ_2. Given μ_1 as above, note that the map μ_2 must satisfy (1-10); that is, $\delta_3^*(\mu_2) = \mu_1 \circ (\mu_1 \otimes \mathrm{id} - \mathrm{id} \otimes \mu_1)$ as cochains on the bar resolution $\boldsymbol{B}_\bullet(A)$ of A. We will show that a modification of $\psi_2^*(\beta)$ is such a map, as follows:

First, note that $\beta = \phi_2^*(\psi_2^*(\beta))$ as cochains, by an argument similar to that above for α. Moreover, condition (ii) implies that $d_3^*(\beta) = \alpha \circ (\alpha \otimes_H \mathrm{id} - \mathrm{id} \otimes_H \alpha)$ as cochains on $X_{0,3}$. Let

$$\gamma = \delta_3^*\psi_2^*(\beta) - \mu_1 \circ (\mu_1 \otimes \mathrm{id} - \mathrm{id} \otimes \mu_1). \tag{3-8}$$

Then $\phi_3^*(\gamma)$ is zero on $X_{0,3}$: $\phi_3^*\delta_3^*\psi_2^*(\beta) = d_3^*(\beta)$ and $\phi_3^*(\mu_1 \circ (\mu_1 \otimes \mathrm{id} - \mathrm{id} \otimes \mu_1)) = \alpha \circ (\alpha \otimes \mathrm{id} - \mathrm{id} \otimes \alpha)$ by Lemma 3.4(ii). To see the last statement, note that the image of ϕ_3 on $X_{0,3}$ is contained in $A \otimes ((I \otimes V) \cap (V \otimes I)) \otimes A$ with $\phi^*(\mu_1) = \alpha$. We also see that $\phi_3^*(\gamma)$ is 0 on $X_{2,1}$ and $X_{3,0}$ since it is a map of degree -2. We claim it is also 0 on $X_{1,2}$ as follows. As an A^e-module, the image of $X_{1,2}$ under ϕ_3 is generated by elements of degree 2. Since $\mu_1 = \psi_2^*(\alpha)$, it is zero on elements of degree less than two, and so the map $\mu_1 \circ (\mu_1 \otimes \mathrm{id} - \mathrm{id} \otimes \mu_1)$ must be 0 on the image of $X_{1,2}$ under ϕ_3. Since β is H-invariant, and thus β is a cocycle (see the argument following (3-7)) and $\phi_2^*\psi_2^*(\beta) = \beta$, we have that $\phi_3^*\delta_3^*\psi_2^*(\beta) = d_3^*\phi_2^*\psi_2^*(\beta) = d_3^*(\beta)$ is 0 on $X_{1,2}$. Therefore $\phi_3^*(\gamma)$ is 0 on $X_{1,2}$.

We have shown that $\phi_3^*(\gamma)$ is 0 on all of X_3, and so γ must be a coboundary on the bar resolution $\boldsymbol{B}_\bullet(A)$ of A. Thus there is a 2-cochain μ of degree -2 on the bar resolution with

$$\delta_3^*(\mu) = \gamma. \tag{3-9}$$

Consider $\psi_2^*(\beta) - \mu$, yet note that $\phi_2^*(\psi_2^*(\beta) - \mu)$ may not agree with β on X_2. We need such a statement for the next step of constructing μ_3. Now,

$$d_3^*\phi_2^*(\mu) = \phi_3^*\delta_3^*(\mu) = \phi_3^*(\gamma) = 0,$$

so the 2-cochain $\phi_2^*(\mu)$ is a *cocycle* on the complex $X_\bullet$. Thus, $\phi_2^*(\mu)$ lifts to a *cocycle* μ' of degree -2 on the bar complex $\boldsymbol{B}_\bullet(A)$. In other words, $\phi_2^*(\psi_2^*(\beta)-\mu+\mu')$ agrees with β on X_2.

Moreover, $\delta_3^*(\mu')=0$, and by (3-8) and (3-9), we have that $\delta_3^*(\psi_2^*(\beta)-\mu)=\mu_1\circ(\mu_1\otimes\mathrm{id}-\mathrm{id}\otimes\mu_1)$. So,

$$\delta_3^*(\psi_2^*(\beta)-\mu+\mu')-\mu_1\circ(\mu_1\otimes\mathrm{id}-\mathrm{id}\otimes\mu_1)=0 \tag{3-10}$$

on the bar resolution $\boldsymbol{B}_\bullet(A)$ of A.

Thus, setting μ_2 equal to $\psi_2^*(\beta)-\mu+\mu'$, we have maps μ_1,μ_2 to obtain a second-level graded deformation $A_{(2)}$ of A extending $A_{(1)}$.

Construction of the multiplication map μ_3. Recall the restraint on μ_3 given in (1-11): μ_3 is a cochain on $\boldsymbol{B}_\bullet(A)$ whose coboundary is given by $\mu_1\circ(\mu_2\otimes\mathrm{id}-\mathrm{id}\otimes\mu_2)+\mu_2\circ(\mu_1\otimes\mathrm{id}-\mathrm{id}\otimes\mu_1)$. We construct μ_3 as follows.

By (3-10) and condition (iii) of Lemma 3.4, we have that $\mu_2\circ(\mu_1\otimes\mathrm{id}-\mathrm{id}\otimes\mu_1)$ is 0 on the image of ϕ. By degree considerations, $\mu_1\circ(\mu_2\otimes\mathrm{id}-\mathrm{id}\otimes\mu_2)$ is always 0 on the image of ϕ. Therefore, the obstruction

$$\mu_2\circ(\mu_1\otimes\mathrm{id}-\mathrm{id}\otimes\mu_1)+\mu_1\circ(\mu_2\otimes\mathrm{id}-\mathrm{id}\otimes\mu_2)$$

is a coboundary. Thus there exists a 2-cochain μ_3, necessarily having degree -3, satisfying the restraint given above, and the deformation lifts to the third level.

Construction of the multiplication maps μ_i for $i\geq 4$. The obstruction for a third-level graded deformation $A_{(3)}$ of A to lift to the fourth level lies in $\mathrm{HH}^{3,-4}(A)$ by Proposition 1.13. We apply ϕ_3^* to this obstruction to obtain a cochain on X_3. Since there are no cochains of degree -4 on X_3 by definition (as it is generated by elements of degree 3 or less), ϕ_3^* applied to the obstruction is automatically zero. Therefore, the deformation may be continued to the fourth level. Similar arguments show that it can be continued to the fifth level, and so on.

Construction of A_t. Let A_t be the graded deformation of A that we obtain in this manner (Definition 1.5). Then, A_t is the k-vector space $A[t]$ with multiplication defined for all $a_1,a_2\in A$ by

$$a_1*a_2=a_1a_2+\mu_1(a_1\otimes a_2)t+\mu_2(a_1\otimes a_2)t^2+\mu_3(a_1\otimes a_2)t^3+\cdots,$$

where a_1a_2 is the product in A and each $\mu_i : A\otimes A\to A$ is a k-linear map of homogeneous degree $-i$. Now for any r in R, $\mu_1(r)=(\psi_2^*\alpha)(r)$ and $\mu_2(r)=(\psi_2^*(\beta)-\mu+\mu')(r)$ by construction, and $\mu_i(r)=0$ for $i\geq 3$ since $\deg(r)=2$.

Conclusion that $A':=T_H(U)/(P)$ is a PBW deformation of $A=T_H(U)/(R)$. Now we show that $A':=T_H(U)/(P)$ is isomorphic, as a filtered algebra, to the fiber of the deformation A_t at $t=1$ as follows. Let $A''=(A_t)|_{t=1}$. Then A'' is generated by V and H and one thus obtains a surjective algebra homomorphism

$T_H(U) \cong T_k(V) \# H \to A''$. The elements of P lie in the kernel (by definition of A''), and thus this map induces a surjective algebra homomorphism $A' = T_H(U)/(P) \twoheadrightarrow A''$. This map is in fact an isomorphism of filtered algebras by a dimension argument in each degree. Therefore A' is a PBW deformation of A, since A'' is one (Proposition 1.14). □

4. Examples

For our examples, we restrict k to be an algebraically closed field of characteristic zero. There are many interesting examples, both known and new, in this setting. Less is known about Hopf actions on Koszul algebras and corresponding deformations in positive characteristic.

As an application of Theorem 3.1, we provide various examples of PBW deformations $\mathcal{D}_{B,\kappa}$ of smash products $B \# H$; recall Notations 0.2 and 0.3. We do this by describing deformation parameter(s) $\kappa = \kappa^C + \kappa^L$ below. In particular, Examples 4.1, 4.2, and 4.4 involve semisimple Hopf actions, and Examples 4.13, 4.16, and 4.18 involve nonsemisimple Hopf actions on (skew) polynomial rings. Recall that skew polynomial rings are Koszul by [Polishchuk and Positselski 2005, Example 4.2.1 and Theorem 4.3.1].

4A. ***Semisimple Hopf actions.*** We begin by revisiting the well-known PBW deformations of Crawley-Boevey and Holland [1998].

Example 4.1. Take $H = k\Gamma$, for Γ a finite subgroup of $SL_2(k)$, and $B = k[u, v]$. For $g = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right) \in \Gamma$, let the action of g on B be given by $g \cdot u = au + cv$ and $g \cdot v = bu + dv$.

By [Crawley-Boevey and Holland 1998], the deformation parameter κ of the PBW deformation $\mathcal{D}_{B,\kappa}$ of $B \# H$ must be in the center of Γ, which we verify again with Theorem 3.1. We assume here that $\kappa^L \equiv 0$, as in that work.

Since $\dim_k V = 2$, only condition (a) of Theorem 3.1 applies. So we have for all $g \in \Gamma$ that $g \cdot (\kappa(uv - vu)) = \kappa(g \cdot (uv - vu))$. Now since the determinant of g is 1, $g \cdot (\kappa(uv - vu)) = \kappa(uv - vu)$, and the image of κ lies in the center of $k\Gamma$. That is,

$$\mathcal{D}_{B,\kappa} = \frac{k\langle u, v\rangle \# k\Gamma}{(uv - vu - \lambda)}$$

is a PBW deformation of $k[u, v] \# k\Gamma$ if and only if $\lambda \in Z(k\Gamma)$.

It is worth pointing out that there are analogues of Crawley-Boevey–Holland algebras when working in positive characteristic; see the work of Emily Norton [2013] for some examples that are quite different from those in characteristic zero.

The following two Hopf actions were produced by Walton in joint work with Kenneth Chan, Ellen Kirkman, and James Zhang [Chan et al. 2012]. We thank them for permitting us to use these results.

Example 4.2. Let $H := H_8$ be the unique noncommutative noncocommutative semisimple 8-dimensional Hopf algebra [Kac and Paljutkin 1966; Masuoka 1995], and let $B = k\langle u, v\rangle/(u^2+v^2)$ (which is isomorphic to the skew polynomial ring $k\langle u, v\rangle/(uv+vu)$). The Hopf algebra H_8 is generated by x, y, z, and the relations are

$$x^2 = y^2 = 1, \quad xy = yx, \quad zx = yz, \quad zy = xz, \quad z^2 = \tfrac{1}{2}(1+x+y-xy).$$

The rest of the structure of H_8 and the left H_8-action on B are given by

$$\Delta(x) = x\otimes x, \quad \Delta(y) = y\otimes y, \quad \Delta(z) = \tfrac{1}{2}(1\otimes 1+1\otimes x+y\otimes 1-y\otimes x)(z\otimes z),$$
$$\epsilon(x) = \epsilon(y) = \epsilon(z) = 1, \quad S(x) = x, \quad S(y) = y, \quad S(z) = z,$$
$$x\cdot u = -u, \quad x\cdot v = v, \quad y\cdot u = u, \quad y\cdot v = -v, \quad z\cdot u = v, \quad z\cdot v = u.$$

Let $r := u^2+v^2$ and note that

$$x\cdot r = r, \quad y\cdot r = r, \quad z\cdot r = r,$$

so the ideal of relations of B, $I = \langle r\rangle$, is H-stable.

Since $\dim_k V = 2$, only condition (a) of Theorem 3.1 applies. We begin by computing κ^C. Let

$$\kappa^C(r) = \gamma_0+\gamma_1 x+\gamma_2 y+\gamma_3 xy+\gamma_4 z+\gamma_5 xz+\gamma_6 yz+\gamma_7 xyz$$

for some scalars $\gamma_i\in k$. Since $h\cdot(\kappa^C(r)) = \sum h_1(\kappa^C(r))S(h_2)$ (see Section 1A), both $x\cdot(\kappa^C(r)) = \kappa^C(r)$ and $y\cdot(\kappa^C(r)) = \kappa^C(r)$ imply that $\gamma_7 = \gamma_4$ and $\gamma_6 = \gamma_5$. Moreover,

$$z\cdot(\kappa^C(r)) = \gamma_0+\gamma_2 x+\gamma_1 y+\gamma_3 xy+\gamma_4 z+\gamma_5 xz+\gamma_5 yz+\gamma_4 xyz = \kappa^C(r),$$

which implies that $\gamma_2 = \gamma_1$. Thus,

$$\begin{aligned}\kappa^C(r) &=: g(\gamma_0, \gamma_1, \gamma_3, \gamma_4, \gamma_5)\\ &= \gamma_0+\gamma_1(x+y)+\gamma_3 xy+\gamma_4(z+xyz)+\gamma_5(xz+yz).\end{aligned} \tag{4-3}$$

On the other hand, let $\kappa^L(r) = u\otimes f+v\otimes f'\in V\otimes H$ with

$$f = \delta_0+\delta_1 x+\delta_2 y+\delta_3 xy+\delta_4 z+\delta_5 xz+\delta_6 yz+\delta_7 xyz,$$
$$f' = \delta'_0+\delta'_1 x+\delta'_2 y+\delta'_3 xy+\delta'_4 z+\delta'_5 xz+\delta'_6 yz+\delta'_7 xyz,$$

for some scalars $\delta_i, \delta'_i\in k$. Note that $h\cdot(\kappa^L(r)) = \sum h_1\cdot u\otimes h_2 f S(h_3)+\sum h_1\cdot v\otimes h_2 f' S(h_3)$ (see Section 1A). Since $x\cdot(\kappa^L(r)) = \kappa^L(r)$, it follows that:

$$\delta_0 = \delta_1 = \delta_2 = \delta_3 = 0, \quad \delta_4 = -\delta_7, \quad \delta_5 = -\delta_6 \quad \delta'_4 = \delta'_7, \quad \text{and} \quad \delta'_5 = \delta'_6.$$

By considering the coefficient of u in the equation $y\cdot(\kappa^L(r)) = \kappa^L(r)$, we now find that $f = 0$. Similarly, by considering the coefficient of v in the equation $y\cdot(\kappa^L(r)) = \kappa^L(r)$, we find that $f' = 0$. Hence, $\kappa^L(r) = 0$.

Thus the deformation parameter κ of $\mathcal{D}_{B,\kappa}$ equals its constant part κ^C, which depends on five scalar parameters as described above. In short,

$$\mathcal{D}_{B,\kappa} = \frac{k\langle u, v\rangle \# H_8}{(u^2+v^2-\kappa(u^2+v^2))}$$

is a PBW deformation of $(k\langle u, v\rangle/(u^2+v^2)) \# H_8$ if and only if $\kappa(u^2+v^2) = g(\gamma_0, \gamma_1, \gamma_3, \gamma_4, \gamma_5)$ as given in (4-3). This yields a five-parameter family of PBW deformations of $B \# H_8$.

Example 4.4. Let H be $H_{a:1}$, one of the 16-dimensional semisimple Hopf algebras classified in [Kashina 2000], and let B be the skew polynomial ring

$$B = \frac{k\langle t, u, v, w\rangle}{\begin{pmatrix} r_{tu} := tu-ut, & r_{tv} := tv-vt, & r_{tw} := tw+wt \\ r_{uv} := uv-vu, & r_{uw} := uw+wu, & r_{vw} := vw-wv \end{pmatrix}}.$$

The Hopf algebra $H_{a:1}$ is generated by x, y, z, subject to the relations

$$x^4 = y^2 = z^2 = 1, \quad yx = xy, \quad zx = xyz, \quad zy = yz.$$

The rest of the structure of $H_{a:1}$ and the left $H_{a:1}$-action on B are given by

$$\Delta(x) = x\otimes x, \quad \Delta(y) = y\otimes y, \quad \Delta(z) = \tfrac{1}{2}(1\otimes 1+1\otimes x^2+y\otimes 1-y\otimes x^2)(z\otimes z),$$
$$\epsilon(x) = \epsilon(y) = \epsilon(z) = 1, \quad S(x) = x^3, \quad S(y) = y, \quad S(z) = \tfrac{1}{2}(1+x^2+y-x^2y)z,$$
$$x\cdot t = it, \quad y\cdot t = -t, \quad z\cdot t = u, \quad x\cdot u = -iu, \quad y\cdot u = -u, \quad z\cdot u = t,$$
$$x\cdot v = v, \quad y\cdot v = -v, \quad z\cdot v = w, \quad x\cdot w = -w, \quad y\cdot w = -w, \quad z\cdot w = v,$$

where i is a primitive fourth root of unity in k. Note that

$$\begin{aligned}
&x\cdot r_{tu} = r_{tu}, && x\cdot r_{tv} = ir_{tv}, && x\cdot r_{tw} = -ir_{tw},\\
&x\cdot r_{uv} = -ir_{uv}, && x\cdot r_{uw} = ir_{uw}, && x\cdot r_{vw} = -r_{vw},\\
&y\cdot r_{tu} = r_{tu}, && y\cdot r_{tv} = r_{tv}, && y\cdot r_{tw} = r_{tw},\\
&y\cdot r_{uv} = r_{uv}, && y\cdot r_{uw} = r_{uw}, && y\cdot r_{vw} = r_{vw},\\
&z\cdot r_{tu} = r_{tu}, && z\cdot r_{tv} = r_{uw}, && z\cdot r_{tw} = r_{uv},\\
&z\cdot r_{uv} = r_{tw}, && z\cdot r_{uw} = r_{tv}, && z\cdot r_{vw} = -r_{vw}.
\end{aligned}$$

So, the ideal of relations $I = \langle r_{tu}, r_{tv}, r_{tw}, r_{uv}, r_{uw}, r_{vw}\rangle$ of B is H-stable.

Now we compute the possible values $\kappa^C(r) \in H$ for all $r \in I$, under condition (a) of Theorem 3.1. Take $\kappa^C(r) = g(\underline{\gamma}) \in H$ given by

$$\begin{aligned} g(\underline{\gamma}) = {}&\gamma_0+\gamma_1 x+\gamma_2 x^2+\gamma_3 x^3+\gamma_4 y+\gamma_5 xy+\gamma_6 x^2 y+\gamma_7 x^3 y\\ &+\gamma_8 z+\gamma_9 xz+\gamma_{10} x^2 z+\gamma_{11} x^3 z+\gamma_{12} yz+\gamma_{13} xyz+\gamma_{14} x^2 yz+\gamma_{15} x^3 yz, \end{aligned}$$

where $\gamma_i \in k$. Note that $h \cdot g(\underline{\gamma}) = \sum h_1 g(\underline{\gamma}) S(h_2)$. With the assistance of *Affine*, a subpackage of *Maxima*, we have the following computations:

$$x \cdot g(\underline{\gamma}) = xg(\underline{\gamma})x^3 = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \gamma_3 x^3 + \gamma_4 y + \gamma_5 xy + \gamma_6 x^2 y + \gamma_7 x^3 y + \gamma_8 yz + \gamma_9 xyz + \gamma_{10} x^2 yz + \gamma_{11} x^3 yz + \gamma_{12} z + \gamma_{13} xz + \gamma_{14} x^2 z + \gamma_{15} x^3 z; \tag{4-5}$$

$$y \cdot g(\underline{\gamma}) = yg(\underline{\gamma})y = g(\underline{\gamma}); \tag{4-6}$$

$$\begin{aligned} z \cdot g(\underline{\gamma}) &= \tfrac{1}{2}(zg(\underline{\gamma})S(z) + zg(\underline{\gamma})S(x^2 z) + yzg(\underline{\gamma})S(z) - yzg(\underline{\gamma})S(x^2 z)) \\ &= \gamma_0 + \gamma_1 xy + \gamma_2 x^2 + \gamma_3 x^3 y + \gamma_4 y + \gamma_5 x \\ &\quad + \gamma_6 x^2 y + \gamma_7 x^3 + \gamma_8 z + \gamma_9 xyz + \gamma_{10} x^2 z \\ &\quad + \gamma_{11} x^3 yz + \gamma_{12} yz + \gamma_{13} xz + \gamma_{14} x^2 yz + \gamma_{15} x^3 z. \end{aligned} \tag{4-7}$$

For r_{tu}, let $\kappa^C(r_{tu}) = g(\underline{\gamma})$. We have that $x \cdot \kappa^C(r_{tu}) = \kappa^C(r_{tu})$ and $y \cdot \kappa^C(r_{tu}) = \kappa^C(r_{tu})$ imply that $\gamma_8 = \gamma_{12}$, $\gamma_9 = \gamma_{13}$, $\gamma_{10} = \gamma_{14}$, $\gamma_{11} = \gamma_{15}$. Moreover, $z \cdot \kappa^C(r_{tu}) = \kappa^C(r_{tu})$ implies that $\gamma_1 = \gamma_5$, $\gamma_3 = \gamma_7$, $\gamma_9 = \gamma_{13}$, $\gamma_{11} = \gamma_{15}$. Therefore,

$$\begin{aligned} \kappa^C(r_{tu}) &= g(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_6, \gamma_8, \gamma_9, \gamma_{10}, \gamma_{11}) \\ &= \gamma_0 + \gamma_1(x + xy) + \gamma_2 x^2 + \gamma_3(x^3 + x^3 y) + \gamma_4 y + \gamma_6 x^2 y \\ &\quad + \gamma_8(z + yz) + \gamma_9(xz + xyz) + \gamma_{10}(x^2 z + x^2 yz) + \gamma_{11}(x^3 z + x^3 yz). \end{aligned} \tag{4-8}$$

For r_{vw}, let $\kappa^C(r_{vw}) = g(\underline{\gamma'})$. We have that $x \cdot \kappa^C(r_{vw}) = -\kappa^C(r_{vw})$ and $y \cdot \kappa^C(r_{vw}) = \kappa^C(r_{vw})$ implies that $\gamma'_0 = \cdots = \gamma'_7 = 0$, $\gamma'_8 = -\gamma'_{12}$, $\gamma'_9 = -\gamma'_{13}$, $\gamma'_{10} = -\gamma'_{14}$, and $\gamma'_{11} = -\gamma'_{15}$. Moreover, we have that $z \cdot \kappa^C(r_{vw}) = -\kappa^C(r_{vw})$. So the conditions on γ'_i in (4-7) then imply that $\gamma'_i = 0$ for $i = 0, \ldots, 7, 8, 10, 12, 14$ with $\gamma'_9 = -\gamma'_{13}$, $\gamma'_{11} = -\gamma'_{15}$. Thus,

$$\kappa^C(r_{vw}) = g(\gamma'_9, \gamma'_{11}) = \gamma'_9(xz - xyz) + \gamma'_{11}(x^3 z - x^3 yz). \tag{4-9}$$

For $r \neq r_{tu}, r_{vw}$, we have that $x \cdot \kappa^C(r) = \pm i \kappa^C(r)$ implies that $\kappa^C(r) = 0$.

We compute $\kappa^L(r)$ under condition (a) of Theorem 3.1. Fix $r \in I$ and let

$$\kappa^L(r) = t \otimes f_t + u \otimes f_u + v \otimes f_v + w \otimes f_w \in V \otimes H$$

for some $f_t, f_u, f_v, f_w \in H$. Since y is central in H and $y \cdot r = r$ for each relation r, we have that

$$\begin{aligned} \kappa^L(r) &= y \cdot \kappa^L(r) \\ &= y \cdot t \otimes yf_t S(y) + y \cdot u \otimes yf_u S(y) + y \cdot v \otimes yf_v S(y) + y \cdot w \otimes yf_w S(y) \\ &= -t \otimes f_t - u \otimes f_u - v \otimes f_v - w \otimes f_w = -\kappa^L(r). \end{aligned}$$

Thus, $\kappa^L(r) = 0$.

To finish, we apply to κ conditions (b)–(d) of Theorem 3.1. Since $\kappa^L(r) = 0$ for all $r \in I$, only condition (c) is pertinent. Namely, we only need to impose $\kappa^C \otimes \mathrm{id} = \mathrm{id} \otimes \kappa^C$ as maps on $(I \otimes V) \cap (V \otimes I)$. This intersection is a 4-dimensional k-vector space with basis

$$\begin{aligned}
s_{tuv} &:= tuv - tvu - utv + uvt + vtu - vut,\\
s_{tuw} &:= tuw + twu - utw - uwt + wtu - wut,\\
s_{tvw} &:= tvw - twv - vtw + vwt + wtv - wvt,\\
s_{uvw} &:= uvw - uwv - vuw - vwu - wuv + wvu.
\end{aligned}$$

Since $\kappa^C(r_{tv}) = \kappa^C(r_{uv}) = 0$, we get that

$$\begin{aligned}
(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)(s_{tuv}) &= \kappa^C(r_{uv})t - \kappa^C(r_{tv})u + \kappa^C(r_{tu})v\\
&\qquad - t\kappa^C(r_{uv}) + u\kappa^C(r_{tv}) - v\kappa^C(r_{tu})\\
&= \kappa^C(r_{tu})v - v\kappa^C(r_{tu}).
\end{aligned} \tag{4-10}$$

Identify $b \in V$ with $b \# 1 \in A$ and $h \in H$ with $1 \# h \in A$. Recall that in A we have $(1 \# h)(b \# 1) = \sum (h_1 \cdot b) \# h_2$. Now by using (4-8) and by setting (4-10) equal to 0, we get that

$$\kappa^C(r_{tu}) = g(\gamma_0, \gamma_2) = \gamma_0 + \gamma_2 x^2. \tag{4-11}$$

Moreover,

$$\begin{aligned}
(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)(s_{tuw}) &= -\kappa^C(r_{uw})t + \kappa^C(r_{tw})u + \kappa^C(r_{tu})w\\
&\qquad - t\kappa^C(r_{uw}) + u\kappa^C(r_{tw}) - w\kappa^C(r_{tu})\\
&= \kappa^C(r_{tu})w - w\kappa^C(r_{tu}) = 0
\end{aligned}$$

imposes no new restrictions on $\kappa^C(r_{tu})$, nor do $(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)(s_{tvw}) = 0$, $(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)(s_{uvw}) = 0$. Therefore, $\kappa^C(r_{tu})$ is given by (4-11).

To compute $\kappa^C(r_{vw})$, consider the calculation

$$\begin{aligned}
(\kappa^C \otimes \mathrm{id} - \mathrm{id} \otimes \kappa^C)(s_{tvw}) &= \kappa^C(r_{vw})t - \kappa^C(r_{tw})v + \kappa^C(r_{tv})w\\
&\qquad - t\kappa^C(r_{vw}) + v\kappa^C(r_{tw}) - w\kappa^C(r_{tv})\\
&= \kappa^C(r_{vw})t - t\kappa^C(r_{vw}).
\end{aligned} \tag{4-12}$$

Now by using (4-9) and by setting (4-12) equal to 0, we get that $\kappa^C(r_{vw}) = 0$.

Therefore, the filtered algebra $\mathcal{D}_{B,\kappa}$ is a PBW deformation of $B \# H_{a:1}$ if and only if the deformation parameter $\kappa = \kappa^C$ of $\mathcal{D}_{B,\kappa}$ is defined by (4-11) for the relation r_{tu}, and $\kappa^C(r) = 0$ for $r \neq r_{tu}$. Hence, we have a two-parameter family of PBW deformations of $B \# H_{a:1}$.

4B. ***Nonsemisimple Hopf actions.*** Here, we consider nonsemisimple Hopf actions to illustrate Theorem 3.1. We begin with an example of a Taft algebra action.

Example 4.13. Let $H = T(n)$, the n^2-dimensional nonsemisimple Taft algebra. We take $n \geq 3$ here and we consider the (slightly different) case $n = 2$ (the Sweedler algebra) in Example 4.16 below. Let $B = k[u, v]$. The Hopf algebra $T(n)$ is generated by g, x and the relations are $g^n = 1$, $x^n = 0$, and $xg = \zeta gx$, for $\zeta \in k^\times$ a primitive n-th root of unity. The rest of the structure of $T(n)$ and the left $T(n)$-action on B are given by

$$\begin{gathered}\Delta(g) = g \otimes g, \quad \Delta(x) = g \otimes x + x \otimes 1, \quad \epsilon(g) = 1, \quad \epsilon(x) = 0,\\ S(g) = g^{-1} = g^{n-1}, \quad S(x) = -g^{n-1}x,\\ g \cdot u = u, \quad g \cdot v = \zeta^{-1} v, \quad x \cdot u = 0, \quad x \cdot v = u.\end{gathered}$$

Let $r := uv - vu$ and note that $g \cdot r = \zeta^{-1} r$ and $x \cdot r = 0$. Hence, the ideal of relations $I = \langle r \rangle$ of B is H-stable.

Since $\dim_k V = 2$, only condition (a) of Theorem 3.1 applies. Now, we compute κ^C. Let $\kappa^C(r) = \sum_{i,j=0}^{n-1} \gamma_{ij} g^i x^j$. Since $h \cdot (\kappa^C(r)) = \sum h_1(\kappa^C(r))S(h_2)$ for $h \in H$, we have that the equality $g \cdot (\kappa^C(r)) = \zeta^{-1}\kappa^C(r)$ implies that all terms equal zero except when $j = 1$; hence

$$\kappa^C(r) = \gamma_0 x + \gamma_1 gx + \cdots + \gamma_{n-1} g^{n-1} x$$

for $\gamma_i \in k$. Also, the equality $x \cdot (\kappa^C(r)) = 0$ implies that all terms equal zero except for $i = n - 1$, so

$$\kappa^C(r) = \gamma g^{n-1} x \tag{4-14}$$

for $\gamma \in k$.

On the other hand, let $\kappa^L(r) = u \otimes f + v \otimes f' \in V \otimes H$ for $f = \sum_{i,j=0}^{n-1} \lambda_{ij} g^i x^j$ and $f' = \sum_{i,j=0}^{n-1} \lambda'_{ij} g^i x^j$. Notice that $h \cdot (\kappa^L(r)) = \sum h_1 \cdot u \otimes h_2 f S(h_3) + \sum h_1 \cdot v \otimes h_2 f' S(h_3)$ (see Section 1A). Since $g \cdot (\kappa^L(r)) = \zeta^{-1}\kappa^L(r)$, all terms equal zero except possibly those in the first sum for which $j = 1$ and those in the second sum for which $j = 0$. Therefore $\kappa^L(r) = u \otimes f + v \otimes f'$ for

$$f = \lambda_0 x + \lambda_1 gx + \cdots + \lambda_{n-1} g^{n-1} x \quad \text{and} \quad f' = \lambda'_0 + \lambda'_1 g + \cdots + \lambda'_{n-1} g^{n-1}$$

with $\lambda_i, \lambda'_i \in k$. Applying x, we obtain

$$\begin{aligned}0 = x \cdot \kappa^L(r) &= (g \cdot u) \otimes (gfS(x) + xfS(1)) + (x \cdot u) \otimes f\\ &\qquad\qquad + (g \cdot v) \otimes (gf'S(x) + xf'S(1)) + (x \cdot v) \otimes f'\\ &= u \otimes (-gfg^{n-1}x + xf + f') + \zeta^{-1} v \otimes (-gf'g^{n-1}x + xf').\end{aligned}$$

It follows that

$$-gfg^{n-1}x + xf + f' = 0 \quad \text{and} \quad -gf'g^{n-1}x + xf' = 0.$$

Since f' is in the group algebra $kG(T(n)) \cong k\mathbb{Z}_n$ and $g^n = 1$, the second equation implies that $xf' = f'x$, and so $f' = \lambda_0$ is constant. The first equation further implies that $f' = 0$ and that all terms of f are equal to zero except for $i = n-1$. Thus,

$$\kappa^L(r) = u \otimes \lambda g^{n-1}x \tag{4-15}$$

for $\lambda \in k$.

In summary,

$$\mathcal{D}_{B,\kappa} = \frac{k\langle u, v\rangle \# T(n)}{(uv - vu - \kappa(uv - vu))}$$

is a PBW deformation of $k[u, v] \# T(n)$ if and only if the deformation map κ equals $\kappa^C + \kappa^L$ as given in (4-14) and (4-15). So, we have a two-parameter family of PBW deformations of $k[u, v] \# T(n)$.

Example 4.16. Let H be $H_{Sw} = T(2)$, the 4-dimensional nonsemisimple Sweedler algebra, which is a Taft algebra with $n = 2$. Let $B = k[u, v]$. Retaining the notation from Example 4.13, the Hopf algebra H_{Sw} is generated by g, x and acts on B by $g \cdot u = u$, $g \cdot v = -v$, $x \cdot u = 0$, $x \cdot v = u$. Similar to Example 4.13, let $r := uv - vu$ and note that $g \cdot r = -r$ and $x \cdot r = 0$. So, $I = \langle r\rangle$ is H-stable.

Let $\kappa^C(r) = \gamma_0 + \gamma_1 g + \gamma_2 x + \gamma_3 gx$. We have that $g \cdot (\kappa^C(r)) = -\kappa^C(r)$ implies that $\gamma_0 = \gamma_1 = 0$. Moreover, $x \cdot (\kappa^C(r)) = 0$ does not yield new restrictions on $\kappa^C(r)$. Thus, for $\gamma, \gamma' \in k$, we get that $\kappa^C(r) = \gamma x + \gamma' gx$. In the same fashion as Example 4.13, we also get that $\kappa^L(r) = u \otimes (\lambda x + \lambda' gx)$ for $\lambda, \lambda' \in k$.

In summary,

$$\mathcal{D}_{B,\kappa} = \frac{k\langle u, v\rangle \# H_{Sw}}{(uv - vu - \kappa(uv - vu))}$$

is a PBW deformation of $k[u, v] \# H_{Sw}$ if and only if the deformation map κ equals $\kappa^C + \kappa^L$, where

$$\kappa^C(uv - vu) = \gamma x + \gamma' gx \quad \text{and} \quad \kappa^L(uv - vu) = u \otimes (\lambda x + \lambda' gx)$$

for $\gamma, \gamma', \lambda, \lambda' \in k$. Thus, we have a four-parameter family of PBW deformations of $k[u, v] \# H_{Sw}$.

Remark 4.17. The invariant ring resulting from the action of H_{Sw} on $k[u, v]$ is isomorphic to the polynomial ring $k[u, v^2]$, that is to say, $k[u, v]^{H_{Sw}}$ is regular. Recall that the Shephard–Todd–Chevalley theorem states that when given a finite group (G-) action on a commutative polynomial ring R that is linear and faithful, R^G is regular if and only if G is a reflection group. Our results would then suggest that H_{Sw} is a "reflection Hopf algebra". Ram and Shepler [2003] showed that there are no nontrivial PBW deformations of $k[v_1, \dots, v_n] \# kG$ for many complex reflection groups G; such deformations are referred to as *graded Hecke algebras*. Now by broadening their setting to Hopf actions on (possibly noncommutative) regular algebras, we consider new objects in representation theory: Hopf analogues of graded Hecke algebras. Nontrivial examples of these objects exist as we showed in the example above. Further examples and a general explanation of this phenomenon are worthy of further investigation.

Now we consider the well-known Hopf action of $U_q(\mathfrak{sl}_2)$ on $k\langle u, v\rangle/(uv - qvu)$, where $q \in k^\times$ with $q^2 \neq 1$. A PBW deformation of $(k\langle u, v\rangle/(uv - qvu)) \# U_q(\mathfrak{sl}_2)$ was studied by Gan and Khare [2007]; we recover their result below. Such algebras are known as *quantized symplectic oscillator algebras of rank 1*.

Example 4.18. Fix $q \in k^\times$, with $q^2 \neq 1$. Let H be the Hopf algebra $U_q(\mathfrak{sl}_2)$, and $B = k\langle u, v\rangle/(uv - qvu)$. As in [Brown and Goodearl 2002, I.6.2], we take $U_q(\mathfrak{sl}_2)$ to be generated by E, F, K, K^{-1} with relations:

$$EF - FE = (q - q^{-1})^{-1}(K - K^{-1}), \qquad KEK^{-1} = q^2 E,$$
$$KFK^{-1} = q^{-2} F, \qquad KK^{-1} = K^{-1}K = 1.$$

So, $U_q(\mathfrak{sl}_2)$ has a k-vector space basis $\{E^i F^j K^m\}_{i,j\in\mathbb{N};m\in\mathbb{Z}}$. The rest of the structure of $U_q(\mathfrak{sl}_2)$ and the left $U_q(\mathfrak{sl}_2)$-action on B is given by:

$$\Delta(E) = E \otimes 1 + K \otimes E, \quad \Delta(F) = F \otimes K^{-1} + 1 \otimes F,$$
$$\Delta(K) = K \otimes K, \quad \Delta(K^{-1}) = K^{-1} \otimes K^{-1},$$
$$\epsilon(E) = 0, \quad \epsilon(F) = 0, \quad \epsilon(K) = 1, \quad \epsilon(K^{-1}) = 1,$$
$$S(E) = -K^{-1}E, \quad S(F) = -FK, \quad S(K) = K^{-1}, \quad S(K^{-1}) = K,$$
$$E \cdot u = 0, \quad F \cdot u = v, \quad K \cdot u = qu, \qquad K^{-1} \cdot u = q^{-1}u,$$
$$E \cdot v = u, \quad F \cdot v = 0, \quad K \cdot v = q^{-1}v, \quad K^{-1} \cdot v = qv.$$

Let $r := uv - qvu$ and note that $E \cdot r = F \cdot r = 0$ and $K \cdot r = K^{-1} \cdot r = r$. Hence, the ideal of relations $I = \langle r\rangle$ of B is H-stable.

Since $\dim_k V = 2$, only condition (a) of Theorem 3.1 applies. Let us compute $\kappa^C(r)$. Since $K \cdot \kappa^C(r) = \kappa^C(K \cdot r) = \kappa^C(r)$, we have that $K\kappa^C(r)S(K) = \kappa^C(r)$ (see Section 1A). Hence, $K\kappa^C(r) = \kappa^C(r)K$. Moreover,

$$0 = \kappa^C(E \cdot r) = E \cdot \kappa^C(r) = E\kappa^C(r)S(1) + K\kappa^C(r)S(E),$$

so $E\kappa^C(r) = \kappa^C(r)E$. Likewise, $F \cdot \kappa^C(r) = 0$ implies that $F\kappa^C(r) = \kappa^C(r)F$. So, $\kappa^C(r)$ is in the center of $U_q(\mathfrak{sl}_2)$. For q not a root of unity, the center of $U_q(\mathfrak{sl}_2)$ is generated by the *quantum Casimir element* [Kassel 1995, Theorem VI.4.8],

$$C_q = EF + \frac{q^{-1}K + qK^{-1}}{(q-q^{-1})^2} = FE + \frac{qK + q^{-1}K^{-1}}{(q-q^{-1})^2},$$

whereas for q a root of unity, the elements E^e, F^e, K^e also belong to the center of $U_q(\mathfrak{sl}_2)$, where $e = \operatorname{ord}(q^2)$.

To compute $\kappa^L(r)$, let $\kappa^L(r) = u \otimes \sum \gamma_{ijm} E^i F^j K^m + v \otimes \sum \gamma'_{ijm} E^i F^j K^m$ for $\gamma_{ijm}, \gamma'_{ijm} \in k$. Then,

$$\begin{aligned}
\kappa^L(r) &= \kappa^L(K \cdot r) \\
&= \sum K \cdot u \otimes \gamma_{ijm} K(E^i F^j K^m) K^{-1} + \sum K \cdot v \otimes \gamma'_{ijm} K(E^i F^j K^m) K^{-1} \\
&= \sum qu \otimes \gamma_{ijm} q^{2(i-j)} E^i F^j K^m + \sum q^{-1} v \otimes \gamma'_{ijm} q^{2(i-j)} E^i F^j K^m \\
&= \sum q^{2(i-j)+1} u \otimes \gamma_{ijm} E^i F^j K^m + \sum q^{2(i-j)-1} v \otimes \gamma'_{ijm} E^i F^j K^m.
\end{aligned}$$

Thus, given $m \in \mathbb{Z}/n\mathbb{Z}$, define the subspace $V_m \subset U_q(\mathfrak{sl}_2)$ to be the k-span of all monomials $E^i F^j K^\ell$ such that $j - i \equiv m \mod n$. Then $\kappa^L(uv - qvu) \in u \otimes V_{2^{-1}} + v \otimes V_{-2^{-1}}$ if q is a primitive root of unity of odd order, and $\kappa^L(uv - qvu) = 0$ otherwise.

Therefore,

$$\mathfrak{D}_{B,\kappa} = \frac{k\langle u, v\rangle \# U_q(\mathfrak{sl}_2)}{(uv - qvu - \kappa(uv - qvu))}$$

is a PBW deformation of $(k\langle u, v\rangle/(uv - qvu)) \# U_q(\mathfrak{sl}_2)$ if and only if $\kappa = \kappa^C + \kappa^L$, where $\kappa^C(uv - qvu)$ is in the center of $U_q(\mathfrak{sl}_2)$ and $\kappa^L(uv - qvu)$ is given as above.

More generally, there is a standard $U_q(\mathfrak{sl}_n)$-action on a q-polynomial ring B in n variables.

Question 4.19. Are there nontrivial PBW deformations of the resulting smash product algebra $B \# U_q(\mathfrak{sl}_n)$?

These would be *quantized symplectic oscillator algebras of rank* $n-1$, and merit further investigation.

Acknowledgments

Part of this work was completed during the spring 2013 program on Noncommutative Algebraic Geometry and Representation Theory at the Mathematical Sciences Research Institute. We thank the organizers for the facilities and generous hospitality. Walton and Witherspoon are supported by the National Science Foundation grants DMS-1102548 and DMS-1101399, respectively.

References

[Berger and Taillefer 2007] R. Berger and R. Taillefer, "Poincaré–Birkhoff–Witt deformations of Calabi–Yau algebras", *J. Noncommut. Geom.* **1**:2 (2007), 241–270. MR 2008i:16011 Zbl 1161.16022

[Braverman and Gaitsgory 1996] A. Braverman and D. Gaitsgory, "Poincaré–Birkhoff–Witt theorem for quadratic algebras of Koszul type", *J. Algebra* **181**:2 (1996), 315–328. MR 96m:16012 Zbl 0860.17002

[Brown and Goodearl 2002] K. A. Brown and K. R. Goodearl, *Lectures on algebraic quantum groups*, Birkhäuser, Basel, 2002. MR 2003f:16067 Zbl 1027.17010

[Chan et al. 2012] K. Chan, E. Kirkman, C. Walton, and J. Zhang, Examples of actions of semisimple Hopf algebras on Artin–Schelter regular algebras, 2012. Private communication.

[Crawley-Boevey and Holland 1998] W. Crawley-Boevey and M. P. Holland, "Noncommutative deformations of Kleinian singularities", *Duke Math. J.* **92**:3 (1998), 605–635. MR 99f:14003 Zbl 0974.16007

[Ding and Tsymbaliuk 2013] F. Ding and A. Tsymbaliuk, "Representations of infinitesimal Cherednik algebras", *Represent. Theory* **17** (2013), 557–583. MR 3123740

[Drinfeld 1986] V. G. Drinfeld, "Degenerate affine Hecke algebras and Yangians", *Funktsional. Anal. i Prilozhen.* **20**:1 (1986), 69–70. In Russian; translated in *Funct. Anal. Appl.* **20**:1 (1986), 58–60. MR 87m:22044 Zbl 0599.20049

[Etingof and Ginzburg 2002] P. Etingof and V. Ginzburg, "Symplectic reflection algebras, Calogero–Moser space, and deformed Harish–Chandra homomorphism", *Invent. Math.* **147**:2 (2002), 243–348. MR 2003b:16021 Zbl 1061.16032

[Etingof et al. 2005] P. Etingof, W. L. Gan, and V. Ginzburg, "Continuous Hecke algebras", *Transform. Groups* **10**:3-4 (2005), 423–447. MR 2006h:20006 Zbl 1115.20005

[Gan and Khare 2007] W. L. Gan and A. Khare, "Quantized symplectic oscillator algebras of rank one", *J. Algebra* **310**:2 (2007), 671–707. MR 2008c:17008 Zbl 1194.17003

[Gerstenhaber 1964] M. Gerstenhaber, "On the deformation of rings and algebras", *Ann. of Math.* (2) **79** (1964), 59–103. MR 30 #2034 Zbl 0123.03101

[Guccione and Guccione 2002] J. A. Guccione and J. J. Guccione, "Hochschild (co)homology of Hopf crossed products", *K-Theory* **25**:2 (2002), 139–169. MR 2003h:16012 Zbl 1016.16005

[Kac and Paljutkin 1966] G. I. Kac and V. G. Paljutkin, "Finite ring groups", *Trudy Moskov. Mat. Obšč.* **15** (1966), 224–261. In Russian; translated in *Trans. Moscow Math. Soc.* **15** (1966), 251–294. MR 34 #8211

[Kashina 2000] Y. Kashina, "Classification of semisimple Hopf algebras of dimension 16", *J. Algebra* **232**:2 (2000), 617–663. MR 2001h:16041 Zbl 0969.16014

[Kassel 1995] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, New York, 1995. MR 96e:17041 Zbl 0808.17003

[Khare 2007] A. Khare, "Drinfeld–Hecke algebras over cocommutative algebras", preprint, 2007. arXiv 0705.2067

[Levandovskyy and Shepler 2014] V. Levandovskyy and A. V. Shepler, "Quantum Drinfeld Hecke algebras", *Canad. J. Math.* **66**:4 (2014), 874–901. MR 3224268 Zbl 06331501

[Losev and Tsymbaliuk 2014] I. Losev and A. Tsymbaliuk, "Infinitesimal Cherednik algebras as W-algebras", *Transform. Groups* **19**:2 (2014), 495–526. MR 3200433 Zbl 06315919

[Lusztig 1989] G. Lusztig, "Affine Hecke algebras and their graded version", *J. Amer. Math. Soc.* **2**:3 (1989), 599–635. MR 90e:16049 Zbl 0715.22020

[Masuoka 1995] A. Masuoka, "Semisimple Hopf algebras of dimension 6, 8", *Israel J. Math.* **92**:1-3 (1995), 361–373. MR 96j:16045 Zbl 0839.16036

[McConnell and Robson 2001] J. C. McConnell and J. C. Robson, *Noncommutative Noetherian rings*, Graduate Studies in Mathematics **30**, Amer. Math. Soc., Providence, RI, 2001. MR 2001i:16039 Zbl 0980.16019

[Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, Amer. Math. Soc., Providence, RI, 1993. MR 94i:16019 Zbl 0793.16029

[Naidu and Witherspoon 2014] D. Naidu and S. Witherspoon, "Hochshild cohomology and quantum Drinfeld Hecke algebras", preprint, 2014. arXiv 1111.5243

[Negron 2014] C. Negron, "Spectral sequences for the cohomology rings of a smash product", preprint, 2014. arXiv 1401.3551

[Norton 2013] E. Norton, "Symplectic reflection algebras in positive characteristic as Ore extensions", Preprint, 2013. arXiv 1302.5411

[Polishchuk and Positselski 2005] A. Polishchuk and L. Positselski, *Quadratic algebras*, University Lecture Series **37**, Amer. Math. Soc., Providence, RI, 2005. MR 2006f:16043 Zbl 1145.16009

[Radford 2012] D. E. Radford, *Hopf algebras*, Series on Knots and Everything **49**, World Scientific, Hackensack, NJ, 2012. MR 2894855 Zbl 1266.16036

[Ram and Shepler 2003] A. Ram and A. V. Shepler, "Classification of graded Hecke algebras for complex reflection groups", *Comment. Math. Helv.* **78**:2 (2003), 308–334. MR 2004d:20007 Zbl 1063.20005

[Shepler and Witherspoon 2012a] A. V. Shepler and S. Witherspoon, "Drinfeld orbifold algebras", *Pacific J. Math.* **259**:1 (2012), 161–193. MR 2988488 Zbl 1285.16026

[Shepler and Witherspoon 2012b] A. V. Shepler and S. Witherspoon, "A Poincaré–Birkoff–Witt theorem for quadratic algebras with group actions", preprint, 2012. To appear in *Trans. Amer. Math. Soc.* arXiv 1209.5660

[Shroff 2014] P. Shroff, "Quantum Drinfeld orbifold algebras", preprint, 2014. arXiv 1304.6134

[Tikaradze 2010] A. Tikaradze, "Center of infinitesimal Cherednik algebras of $\mathfrak{gl}_n$", *Represent. Theory* **14** (2010), 1–8. MR 2011b:16111 Zbl 1250.17015

[Tikaradze and Khare 2010] A. Tikaradze and A. Khare, "Center and representations of infinitesimal Hecke algebras of $\mathfrak{sl}_2$", *Comm. Algebra* **38**:2 (2010), 405–439. MR 2011g:17024 Zbl 1282.17015

[Tsymbaliuk 2014] A. Tsymbaliuk, "Infinitesimal Hecke algebras of $\mathfrak{so}_n$", preprint, 2014. arXiv 1306.1514

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

[Wu and Zhu 2013] Q. S. Wu and C. Zhu, "Poincaré–Birkhoff–Witt deformation of Koszul Calabi–Yau algebras", *Algebr. Represent. Theory* **16**:2 (2013), 405–420. MR 3035998 Zbl 1268.16030

Communicated by Georgia Benkart
Received 2013-12-18 Revised 2014-06-30 Accepted 2014-07-29

notlaw@math.mit.edu *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*

sjw@math.tamu.edu *Department of Mathematics, Texas A&M University, College Station, TX 77843, United States*

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1733

Highly biased prime number races

Daniel Fiorilli

Chebyshev observed in a letter to Fuss that there tends to be more primes of the form $4n+3$ than of the form $4n+1$. The general phenomenon, which is referred to as Chebyshev's bias, is that primes tend to be biased in their distribution among the different residue classes mod q. It is known that this phenomenon has a strong relation with the low-lying zeros of the associated L-functions, that is, if these L-functions have zeros close to the real line, then it will result in a lower bias. According to this principle one might believe that the most biased prime number race we will ever find is the $\mathrm{Li}(x)$ versus $\pi(x)$ race, since the Riemann zeta function is the L-function of rank one having the highest first zero. This race has density $0.99999973\ldots$, and we study the question of whether this is the highest possible density. We will show that it is not the case; in fact, there exist prime number races whose density can be arbitrarily close to 1. An example of a race whose density exceeds the above number is the race between quadratic residues and nonresidues modulo 4849845, for which the density is $0.999999928\ldots$. We also give fairly general criteria to decide whether a prime number race is highly biased or not. Our main result depends on the generalized Riemann hypothesis and a hypothesis on the multiplicity of the zeros of a certain Dedekind zeta function. We also derive more precise results under a linear independence hypothesis.

1. Introduction and statement of results

The study of prime number races started in 1853, when Chebyshev noted in a letter to Fuss that there seemed to be more primes of the form $4n+3$ than of the form $4n+1$. More precisely, Chebyshev claimed without proof that as $c \to 0$, we have

$$-\sum_p \left(\frac{-4}{p}\right) e^{-pc} = e^{-3c} - e^{-5c} + e^{-7c} + e^{-11c} - e^{-13c} - \cdots \to \infty.$$

However, as Hardy and Littlewood [1916] and Landau [1918a; 1918b] have shown, this statement is equivalent to the Riemann hypothesis for $L(s, \chi_{-4})$, where χ_{-4} denotes the primitive character modulo 4.

MSC2010: primary 11N13; secondary 11M26.
Keywords: prime number races, primes in arithmetic progressions.

Chebyshev's observation created a new area of research which goes under the names of either comparative prime number theory, Chebyshev's bias or, more colloquially, prime number races. This rich research area has a long history, encompassing authors such as Chebyshev, Littlewood, Wintner, Shanks, Knapowski, Turan and Kaczorowski, to name a few, and more recently Rubinstein, Sarnak, Schlage-Puchta, Ng, Martin, Ford, Konyagin and Lamzouri. For a good account of the history of the subject as well as recent developments, the reader is encouraged to consult the great expository paper [Granville and Martin 2006].

The modern way to study Chebyshev's observation is to look at the set of integers n for which $\pi(n;4,3) > \pi(n;4,1)$, which we denote by $P_{4;3,1}$. One would like to understand the size of this set; however, it is known that its natural density does not exist [Kaczorowski 1995]. To remedy this problem we define the *logarithmic density* of a set $P \subset \mathbb{N}$ by

$$\delta(P) := \lim_{N\to\infty} \frac{1}{\log N} \sum_{\substack{n\leq N\\ n\in P}} \frac{1}{n},$$

if the limit exists. In general we define $\underline{\delta}(P)$ and $\overline{\delta}(P)$ to be the lim inf and lim sup of this sequence. If $P = P_{4;3,1}$, then this last limit exists under the assumption of the generalized Riemann hypothesis (GRH) and the linear independence hypothesis (LI), and equals $0.9959\ldots$ (see [Rubinstein and Sarnak 1994]).

The *generalized Riemann hypothesis* states that for every primitive character $\chi \bmod q$, all nontrivial zeros of $L(s,\chi)$ lie on the line $\Re(s) = \frac{1}{2}$.

The *linear independence hypothesis* states that for every fixed modulus q, the set

$$\bigcup_{\substack{\chi \bmod q\\ \chi \text{ primitive}}} \{\Im(\rho_\chi) : L(\rho_\chi,\chi) = 0,\ 0 < \Re(\rho_\chi) < 1,\ \Im(\rho_\chi) \geq 0\}$$

is linearly independent over $\mathbb{Q}$.

Rubinstein and Sarnak developed a framework to study this question and more general prime number races. Assuming GRH and LI, they have shown that for any r-tuple $(a_1,\ldots,a_r)$ of admissible residue classes mod q (that is, $(a_i,q)=1$), the logarithmic density of the set

$$P_{q;a_1,\ldots,a_r} := \{n : \pi(n;q,a_1) > \pi(n;q,a_2) > \cdots > \pi(n;q,a_r)\},$$

which we denote by $\delta(q;a_1,\ldots,a_r)$, exists and is not equal to 0 or 1 (we call this an r-way prime number race). Moreover, they have shown that if r is fixed, then as $q\to\infty$,

$$\max_{\substack{1\leq a_1,\ldots,a_r\leq q\\ (a_i,q)=1}} \left|\delta(q;a_1,\ldots,a_r) - \frac{1}{r!}\right| \to 0.$$

In other words, the bias dissolves as $q \to \infty$. For $r = 2$, this phenomenon can be readily seen in [Fiorilli and Martin 2013], where the authors exhibited the list of the 117 densities which are greater than or equal to $\frac{9}{10}$. By the trivial inclusion

$$P_{q;a_1,\dots,a_r} \subset P_{q;a_1,a_2},$$

we see that the most biased r-way prime number race is the two-way race appearing on top of the list in that article, that is,

$$\delta(24; 5, 1) = 0.999988\dots.$$

Only one race is known to be more biased: it is the race between $\mathrm{Li}(x)$ and $\pi(x)$, for which the density is

$$\delta(1) := \delta(\{n : \mathrm{Li}(n) > \pi(n)\}) = 0.99999973\dots.$$

One can also combine different residue classes mod q to make prime number races. For two subsets $A, B \subset (\mathbb{Z}/q\mathbb{Z})^\times$, we consider the inequality

$$\frac{1}{|A|}\sum_{a\in A}\pi(n; q, a) > \frac{1}{|B|}\sum_{b\in B}\pi(n; q, b), \tag{1}$$

and denote by $\delta(q; A, B)$ the logarithmic density of the set of n for which it is satisfied, if the density exists. An example of such race was given by Rubinstein and Sarnak, who studied the race between

$$\pi(x; q, NR) = \#\{p \le x : p \text{ is not a quadratic residue mod } q\}$$

and

$$\pi(x; q, R) = \#\{p \le x : p \text{ is a quadratic residue mod } q\},$$

for moduli q having a primitive root. This race appears naturally in their work, since, as they have shown, it is the property of the competitors being a quadratic residue or not which determines whether a two-way prime number race is biased or not. These are good candidates for biased races, however, it can be shown that as $q \to \infty$, $\delta(q; NR, R) \to \frac{1}{2}$ (but at a much slower rate than two-way races [Fiorilli and Martin 2013]).

It is known [Bays et al. 2001; Fiorilli and Martin 2013] that under GRH and LI, low-lying zeros of $L(s, \chi)$ have a significant effect on decreasing the bias, which explains why races of high moduli are very moderately biased. Odlyzko [1990] has shown that the Dedekind zeta function $\zeta_K(s)$ having the highest first zero in the critical strip is the Riemann zeta function, which is $\rho_0 = \frac{1}{2} + i \cdot 14.134725\dots$. Subsequently, Miller [2002] generalized this result by showing that each member of a very large class of cuspidal GL_n L-functions of real archimedean type has the property of either having a zero in the interval $\left[\frac{1}{2} - 14.13472i, \frac{1}{2} + 14.13472i\right]$, or having a zero whose real part is strictly larger than $\frac{1}{2}$ (violating GRH). In particular, this class contains all Dirichlet, rational elliptic curve and modular form L-functions, and possibly also contains all Artin and rational abelian variety L-functions.[1] By these considerations, one might conjecture that the highest density one will ever find by doing prime number races with L-functions of real archimedean type is $\delta(1) = 0.99999973\ldots$.

As it turns out, this is false, and we can find races which are arbitrarily biased. This is achieved by considering races between linear combinations of prime counting functions, and we will see in Section 5 that the key to finding such biased races is to take a very large number of residue classes.

The first (and most extreme) example we give is a quadratic residue versus quadratic nonresidue race as in [Rubinstein and Sarnak 1994], but for a general modulus q. We take $A = NR := \{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \not\equiv \square \bmod q\}$ and $B = R := \{b \in (\mathbb{Z}/q\mathbb{Z})^\times : b \equiv \square \bmod q\}$ in (1). An elementary argument using the Chinese remainder theorem shows that $|B| = \phi(q)/\rho(q)$ and $|A| = \phi(q)(1-1/\rho(q))$, where for $G = (\mathbb{Z}/q\mathbb{Z})^\times$,

$$\rho(q) := [G : G^2] = \begin{cases} 2^{\omega(q)} & \text{if } 2 \nmid q, \\ 2^{\omega(q)-1} & \text{if } 2 \mid q \text{ but } 4 \nmid q, \\ 2^{\omega(q)} & \text{if } 4 \mid q \text{ but } 8 \nmid q, \\ 2^{\omega(q)+1} & \text{if } 8 \mid q, \end{cases}$$

and $\omega(q)$ denotes the number of distinct prime factors of q.

Theorem 1.1. *Assume GRH and LI. Then for any $\epsilon > 0$ there exists q such that*

$$1-\epsilon < \delta(q; NR, R) < 1. \tag{2}$$

Moreover, for any fixed $\frac{1}{2} \le \eta \le 1$ there exists a sequence of moduli $\{q_n\}$ such that

$$\lim_{n\to\infty} \delta(q_n; NR, R) = \eta. \tag{3}$$

In concise form,

$$\overline{\{\delta(q; NR, R)\}} = \left[\tfrac{1}{2}, 1\right].$$

To prove the existence of highly biased races we do not need the full strength of LI; in fact, we only need a much weaker hypothesis on the multiplicity of the elements of the multiset of all nontrivial zeros of quadratic Dirichlet L-functions modulo q, which we will denote by $Z(q)$. Note that LI implies that the elements of this set have multiplicity one.

[1] The restriction to L-functions of real archimedean type is crucial here, since Bober et al. [2014] have given an example of an L-function having a first zero whose imaginary part is $t_0 \approx 14.496$. They have also shown that under certain conditions, all L-functions have a zero in the interval $(-t_2, t_2)$, with $t_2 \approx 22.661$.

Theorem 1.2. *Assume GRH, and assume that there exists an increasing sequence of moduli q such that $\log q = o(\rho(q))$ and such that each element of $Z(q)$ has multiplicity $o(\rho(q)/\log q)$. Then for any $\epsilon > 0$ there exists q such that*

$$1-\epsilon < \underline{\delta}(q; NR, R) \le \overline{\delta}(q; NR, R) < 1. \tag{4}$$

Remark 1.3. The difference between (2) and (4) is explained by the fact that it is not known whether $\delta(q; NR, R)$ exists under GRH alone.

Remark 1.4. For a fixed modulus $q \ge 2$, write

$$q = 2^e \prod_{\substack{p \mid q \\ p \ne 2}} p^{e_p} \quad \text{and} \quad \ell := \prod_{\substack{p \mid q \\ p \ne 2}} p.$$

One can show that under GRH,[2] $\underline{\overline{\delta}}(q; NR, R) = \underline{\overline{\delta}}(2^{\min(3,e)}\ell; NR, R)$. Therefore, when studying $\underline{\overline{\delta}}(q; NR, R)$ one can assume without loss of generality that q is of the form $2^m\ell$, where ℓ is an odd squarefree integer and $m \le 3$.

Remark 1.5. We will see that what controls the bias in these races is the number of prime factors of q and the size of q. More precisely, under GRH and LI the two following statements are equivalent:

$$\sum_{p \mid q} \log p = o(2^{\omega(q)}), \tag{5}$$

$$\delta(q; NR, R) = 1 - o(1). \tag{6}$$

Using this, we can show that the set of moduli $q \le x$ such that $\delta(q; NR, R) = 1 - o(1)$ has density $(\log x)^{-\lambda + o(1)}$, where $\lambda = 1 - (1 + \log\log 2)/\log 2 = 0.086071\ldots$. It is an interesting coincidence that the integers satisfying (5) also appear in the Erdös multiplication table (see Ford's work [2008a; 2008b] on integers having a divisor in a given interval).

In terms of random variables, Theorem 1.2 can be explained by saying that the extreme examples we are considering correspond to random variables whose mean is much larger than their standard deviation. The easy way to show that this implies a very large bias is to use Chebyshev's inequality; however this approach is quite imprecise when the ratio $\mathbb{E}[X]/\sqrt{\mathrm{Var}[X]}$ is large. Instead, one should study the large deviations of $X - \mathbb{E}[X]$. The theory of large deviations of remainder terms arising from prime counting functions was initiated by Montgomery [1980], and has since

[2] First note that there are no real primitive characters modulo p^e with $p \ne 2$ and $e \ge 2$, and there are no real primitive characters modulo 2^e for $e \ge 4$. That is, the conductor of any real character modulo q divides $2^{\min(3,e)}\ell$. The claimed equality follows from Lemma 2.1, since $L(s, \chi^*)$ and $L(s, \chi)$ have the same nontrivial zeros, and thus $E_q(x) = E_{2^{\min(3,e)}\ell}(x) + o(1)$.

q	$\omega(q)$	$\rho(q)/\log q'$	$\delta(q; NR, R)$
3	1	1.82	0.999063
15	2	1.47	0.999907
105	3	1.71	0.999928
1155	4	2.26	0.999877
15015	5	3.33	0.999950
255255	6	5.14	0.9999946
4849845	7	8.31	0.999999928
111546435	8	13.81	0.999999999954

Table 1. First few values of $\delta(q; NR, R)$ for half-primorial moduli.

then been developed by Monach [1980], Montgomery and Odlyzko [1988], Rubinstein and Sarnak [1994], and more recently Ng [2004] and Lamzouri [2012]. Exploiting the results of Montgomery and Odlyzko we are able to be more precise in (2).

Theorem 1.6. *Assume GRH and LI, and define* $q' := \prod_{p \mid q} p$. *If* $\rho(q)/\log q'$ *is large enough, then we have*

$$\exp\left(-a_1 \frac{\rho(q)}{\log q'}\right) \leq 1 - \delta(q; NR, R) \leq \exp\left(-a_2 \frac{\rho(q)}{\log q'}\right),$$

where a_1 *and* a_2 *are positive absolute constants.*

This last theorem shows that the convergence in (2) can be quite fast. It is actually possible to explicitly compute a density which exceeds $\delta(1)$, namely $\delta(4849845; NR, R) = 0.999999928\ldots$. In Table 1 we list the first few values of $\delta(q; NR, R)$ for half-primorial moduli (that is, for q the product of the first k primes excluding $p = 2$). These values were computed using Myerscough's method [2013] and Rubinstein's `lcalc` package.

Remark 1.7. As remarked in [Rubinstein and Sarnak 1994], these densities can theoretically be computed to any given level of accuracy under GRH alone. Indeed, using the B^2 almost-periodicity of these races, this amounts to computing a finite number of zeros of Dirichlet L-functions to a certain level of accuracy.

Remark 1.8. One can summarize Remark 1.5, Theorem 1.1 and Theorem 1.6 by the statement

$$\delta(q; NR, R) \approx \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2^{\omega(q)-1}/\log q'}}^{\infty} e^{-\frac{x^2}{2}}\, dx.$$

Remark 1.9. Using our analysis, one can show under GRH and LI that for almost all squarefree integers q,

$$\delta(q; NR, R) - \tfrac{1}{2} = (\log q)^{\frac{\log 2-1}{2}+o(1)}.$$

That is to say, most such races have a very moderate bias.

It is possible to analyze highly biased races in a more general setting, and to determine which features are needed for this bias to appear. To do this we take $\vec{a} = (a_1, \dots, a_k)$ a vector of invertible reduced residues modulo q and $\vec{\alpha} = (\alpha_1, \dots, \alpha_k)$ a nonzero vector of real numbers such that $\sum_{i=1}^{k} \alpha_i = 0$. We will be interested in the race between positive and negative entries of $\vec{\alpha}$; that is, we define

$$\delta(q; \vec{a}, \vec{\alpha}) := \delta(\{n : \alpha_1 \pi(n; q, a_1) + \cdots + \alpha_k \pi(n; q, a_k) > 0\}).$$

Moreover, we define

$$\epsilon_i := \begin{cases} 1 & \text{if } a_i \equiv \square \bmod q, \\ 0 & \text{if } a_i \not\equiv \square \bmod q, \end{cases}$$

and we assume without loss of generality that

$$\sum_{i=1}^{k} \epsilon_i \alpha_i < 0.$$

(By Lemma 5.1, this will force $\delta(q; \vec{a}, \vec{\alpha}) > \frac{1}{2}$. If $\sum_{i=1}^{k} \epsilon_i \alpha_i = 0$, then $\delta(q; \vec{a}, \vec{\alpha}) = \frac{1}{2}$. If $\sum_{i=1}^{k} \epsilon_i \alpha_i > 0$, then we multiply $\vec{\alpha}$ by -1 and study the complementary probability $\delta(q; \vec{a}, -\vec{\alpha}) = 1 - \delta(q; \vec{a}, \vec{\alpha})$.)

There are many choices of vectors $\vec{a}$ and $\vec{\alpha}$ which yield highly biased races. We give some examples with constant coefficients, which we believe are the most natural.

Theorem 1.10. *Assume GRH and LI, and let*

$$k_R \le \frac{\phi(q)}{\rho(q)} \quad \textit{and} \quad k_N \le \left(1 - \frac{1}{\rho(q)}\right)\phi(q)$$

be two positive integers. Take $a_1, \dots, a_{k_N}$ to be any distinct quadratic nonresidues mod q, *with coefficients $\alpha_1 = \cdots = \alpha_{k_N} = k_R$, and $a_{k_N+1}, \dots, a_{k_N+k_R}$ to be any distinct quadratic residues* mod q, *with $\alpha_{k_N+1} = \cdots = \alpha_{k_N+k_R} = -k_N$. There exists an absolute constant $c > 0$ such that if for some $0 < \epsilon < 1/(2c)$ we have*

$$\frac{1}{k_N} + \frac{1}{k_R} < \epsilon \frac{\rho(q)^2}{\phi(q) \log q}, \tag{7}$$

then

$$\delta(q; \vec{a}, \vec{\alpha}) > 1 - c\epsilon.$$

Remark 1.11. Fix $0 < \epsilon < 1/(2c)$ and define $N_\epsilon(q)$ to be the number of positive integers k_N, k_R for which $k_N \leq (1-1/\rho(q))\phi(q)$, $k_R \leq \phi(q)/\rho(q)$ and

$$\frac{1}{k_N} + \frac{1}{k_R} < \epsilon \frac{\rho(q)^2}{\phi(q)\log q}.$$

Then, for values of q for which $\rho(q) \geq \epsilon^{-2}\log q$, we have that $N_\epsilon(q)$ tends to infinity as $q \to \infty$. Hence, for values of q for which $\log q = o(\rho(q))$, (7) has a large number of solutions.

Remark 1.12. Theorem 1.10 shows the existence of highly biased races with the same number of residue classes on each side of the inequality. Indeed, for moduli q with $\log q = o(\rho(q))$, taking $k_R = k_N$ with $\phi(q)\log q/\rho(q)^2 = o(k_R)$ and choosing any residue classes $a_1, \ldots, a_{k_N+k_R}$ gives a race with $\delta(q; \vec{a}, \vec{\alpha}) = 1 - o(1)$.

Remark 1.13. In Theorem 1.1, we have

$$k_N = \left(1 - \frac{1}{\rho(q)}\right)\phi(q) \quad \text{and} \quad k_R = \frac{\phi(q)}{\rho(q)},$$

which explains why we obtained a highly biased race when $\rho(q)$ was large compared to $\log q$.

Here is our most general class of highly biased races.

Theorem 1.14. *Assume GRH and LI. There exists an absolute constant $c > 0$ such that if for some $0 < \epsilon < 1/(2c)$ we have*

$$\frac{\sum_{i=1}^{k} \alpha_i^2}{\left(\sum_{i=1}^{k} \epsilon_i \alpha_i\right)^2} < \epsilon \frac{\rho(q)^2}{\phi(q)\log q}, \tag{8}$$

then

$$\delta(q; \vec{a}, \vec{\alpha}) > 1 - c\epsilon.$$

Remark 1.15. Trivially, one has

$$\frac{\sum_{i=1}^{k} \alpha_i^2}{\left(\sum_{i=1}^{k} \epsilon_i \alpha_i\right)^2} \geq \frac{1}{k_R},$$

where $k_R := \sum_{i=1}^{k} \epsilon_i$. Hence, for (8) to be satisfied, one needs k_R to be larger than

$$\epsilon^{-1}\frac{\phi(q)\log q}{\rho(q)^2}.$$

Since $k_R \le \phi(q)/\rho(q)$, this imposes the condition on q

$$\rho(q) \ge \epsilon^{-1} \log q.$$

Remark 1.16. The goal of Theorem 1.14 is to give a large class of biased races, without necessarily being precise on the value of $\delta(q;\vec{a},\vec{\alpha})$. One can use the Montgomery–Odlyzko bounds [1988] to obtain more precise estimates in some particular cases.

The previous examples of highly biased races all have the property that the number of residue classes involved is very large in terms of q (it is at least $q^{1-o(1)}$). In the next theorem we show that this condition is necessary, and that moreover highly biased races are very particular, in the sense that they must satisfy precise conditions.

Theorem 1.17. *Assume GRH and LI. There exist absolute positive constants K_1, K_2 and $0 < \eta < \frac{1}{2}$ such that if $k \le K_1\phi(q)$ and*

$$\frac{\left(\sum_{i=1}^{k} \epsilon_i \alpha_i\right)^2}{\sum_{i=1}^{k} \alpha_i^2} \le K_2 \frac{\phi(q) \log \dfrac{3\phi(q)}{k}}{\rho(q)^2}, \tag{9}$$

then

$$\delta(q;\vec{a},\vec{\alpha}) \le 1-\eta. \tag{10}$$

(*Hence this race cannot be too biased.*)

Remark 1.18. Applying the Cauchy–Schwarz inequality and using that $k_R := \sum_{i=1}^{k}\epsilon_i \le \phi(q)/\rho(q)$, one sees that if $\rho(q) \le K_2 \log(3\phi(q)/k)$, then whatever $\vec{a}$ and $\vec{\alpha}$ are, (9) holds. Moreover, in the range $\rho(q) > K_2 \log(3\phi(q)/k)$ we have that if $k_R \le K_2\phi(q)/\rho(q)^2$, then (9) holds. We conclude that a necessary condition to obtain a highly biased race is that $k_R \gg \phi(q)/\rho(q)^2$.

An interesting feature of prime number races is Skewes' number. It is by definition the smallest x_0 for which

$$\pi(x_0) > \mathrm{Li}(x_0).$$

This number has been extensively studied since Skewes' paper [1933] in which he showed under the Riemann hypothesis that

$$x_0 < 10^{10^{10^{34}}}.$$

The Riemann hypothesis has since then been removed and the upper bound greatly reduced; we refer the reader to [Bays and Hudson 2000] for the list of such improvements. The current record is due to the authors of that work, who showed that

$x_0 < 1.3983 \times 10^{316}$, and moreover this bound is believed to be close to the true size of x_0.

One could also study the generalized Skewes' number

$$x_{q;a,b} := \inf\{x : \pi(x;q,a) < \pi(x;q,b)\}.$$

However, two-way prime number races become less and less biased as q grows; that is, $\delta(q;a,b) \to \frac{1}{2}$ uniformly in a and b coprime to q. Hence, for large q we expect this generalized Skewes number to be small and uninteresting.

The situation is quite different with the highly biased race we constructed; in fact, we expect the Skewes number

$$x_q := \inf\{x : \pi(x;q,NR) < (\rho(q)-1)\pi(x;q,R)\}$$

to tend to infinity as $\rho(q)/\log q'$ tends to infinity (q' is the radical of q). Using similar arguments to those of [Montgomery 1980; Ng 2000], we can speculate on the exact growth rate of x_q.

Conjecture 1.19. *As $\rho(q)/\log q'$ tends to infinity we have*

$$\log\log x_q \asymp \frac{\rho(q)}{\log q'}.$$

2. Results without the linear independence hypothesis

The goal of this section is to prove Theorem 1.2 (from which the first part of Theorem 1.1 clearly follows). We first note that if $A = NR$ and $B = R$, then (1) is equivalent to

$$\pi(x;q,NR) > (\rho(q)-1)\pi(x;q,R).$$

Lemma 2.1. *Fix $q \geq 3$. Assuming GRH, we have that*

$$\begin{aligned} E_q(x) :&= \frac{\pi(x;q,NR)-(\rho(q)-1)\pi(x;q,R)}{\sqrt{x}/\log x} \\ &= \rho(q)-1+\sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}}\sum_{\gamma_\chi}\frac{x^{i\gamma_\chi}}{\rho_\chi}+o_{x\to\infty}(1). \end{aligned}$$

Proof. Let b be an invertible reduced residue mod q. We will use the orthogonality relation

$$\sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}} \chi(b) = \begin{cases} \rho(q)-1 & \text{if } b \equiv \square \bmod q, \\ -1 & \text{if } b \not\equiv \square \bmod q. \end{cases} \tag{11}$$

The explicit formula gives

$$\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0 \\ \chi \neq \chi_0}} \psi(x,\chi) = -\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0 \\ \chi \neq \chi_0}} \sum_{\rho_\chi} \frac{x^{\rho_\chi}}{\rho_\chi} + O_q(\log x), \tag{12}$$

where ρ_χ runs over the nontrivial zeros of $L(s,\chi)$. The left side of (12) is equal to

$$\begin{aligned}\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0 \\ \chi \neq \chi_0}} \sum_{p\le x} \chi(p)\log p + \sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0 \\ \chi \neq \chi_0}} \sum_{p^2\le x} \chi(p)^2 \log p + O_q\big(x^{\frac{1}{3}}\big) \\ = (\rho(q)-1) \sum_{\substack{p\le x \\ p\equiv\square \bmod q}} \log p - \sum_{\substack{p\le x \\ p\not\equiv\square \bmod q}} \log p + (\rho(q)-1)\sqrt{x} + o_q(\sqrt{x}),\end{aligned}$$

by (11) and the prime number theorem. Combining this with a standard summation by parts we get that

$$\frac{\pi(x;q,NR)-(\rho(q)-1)\pi(x;q,R)}{\sqrt{x}/\log x} = \rho(q)-1+\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0 \\ \chi \neq \chi_0}} \sum_{\gamma_\chi} \frac{x^{i\gamma_\chi}}{\rho_\chi} + o_{x\to\infty}(1). \qquad \square$$

Lemma 2.2. *Assuming GRH, the quantity $E_q(x)$ defined in Lemma 2.1 has a limiting logarithmic distribution; that is, there exists a Borel measure μ_q on $\mathbb{R}$ such that for any bounded Lipschitz continuous function $f:\mathbb{R}\to\mathbb{R}$ we have*

$$\lim_{Y\to\infty} \frac{1}{Y}\int_2^Y f(E_q(e^y))\,dy = \int_{\mathbb{R}} f(t)\,d\mu_q(t).$$

Proof. This follows from analysis in [Rubinstein and Sarnak 1994; Akbary et al. 2013]. □

Remark 2.3. By the Portmanteau theorem, the Lipschitz assumption in the last lemma can be removed.

Remark 2.4. As Schlage-Puchta has pointed out to me, it is possible to show under GRH that for all but a countable set of values of c, the density

$$F_q(c) := \lim_{Y\to\infty} \frac{1}{Y} \operatorname{meas}\{y \le Y : E_q(e^y) \le c\}$$

exists. Moreover, one can show that in the domain where F is defined,

$$\sup_{x<c} F_q(x) \le \liminf_{Y\to\infty} \frac{1}{Y} \operatorname{meas}\{y \le Y : E_q(e^y) \le c\}$$
$$\le \limsup_{Y\to\infty} \frac{1}{Y} \operatorname{meas}\{y \le Y : E_q(e^y) \le c\} \le \inf_{x>c} F_q(x),$$

and so in particular if $F_q(x)$ is continuous at $x=c$, then the set $\{y \le Y : E_q(e^y) \le c\}$ has a density.

Let X_q be the random variable associated to μ_q. We will show that X_q can be very biased, in the sense that $\operatorname{Prob}[X_q > 0]$ can be very close to 1. To do so we will compute the first two moments of $E_q(e^y)$, which we relate to the random variable X_q.

Lemma 2.5. *Under GRH, we have that*

$$\lim_{Y\to\infty} \frac{1}{Y} \int_2^Y E_q(e^y)\,dy = \int_{\mathbb{R}} t\,d\mu_q(t),$$
$$\lim_{Y\to\infty} \frac{1}{Y} \int_2^Y E_q(e^y)^2\,dy = \int_{\mathbb{R}} t^2\,d\mu_q(t).$$

Proof. We will only prove the second statement, as the first follows along the same lines. Similarly as in [Schlage-Puchta 2000], we can compute that

$$\lim_{Y\to\infty} \frac{1}{Y} \int_0^Y |E_q(e^y)|^4\,dy = \sum_{\rho_1+\rho_2+\rho_3+\rho_4=0} \frac{1}{\rho_1\rho_2\rho_3\rho_4} < \infty,$$

where the last sum runs over quadruples of nontrivial zeros of quadratic Dirichlet L-functions modulo q. This implies that as $M \to \infty$,

$$\limsup_{Y\to\infty} \frac{1}{Y} \int\limits_{\substack{0\le y\le Y\\ |E_q(e^y)|>M}} |E_q(e^y)|^2\,dy \to 0. \tag{13}$$

Indeed, if this was not the case then we would have that for all $M > M_0$,

$$\limsup_{Y\to\infty} \frac{1}{Y} \int\limits_{\substack{0\le y\le Y\\ |E_q(e^y)|>M}} |E_q(e^y)|^2\,dy \ge \eta > 0,$$

and so

$$\limsup_{Y\to\infty} \frac{1}{Y} \int\limits_{\substack{0\le y\le Y\\ |E_q(e^y)|>M}} |E_q(e^y)|^4\,dy \ge \eta M^2,$$

which would contradict the fact that the fourth moment is finite. We now define the bounded Lipschitz function

$$H_M(t) := \begin{cases} t^2 & \text{if } |t| \le M, \\ M^2(M+1-|t|) & \text{if } M < |t| \le M+1, \\ 0 & \text{if } |t| \ge M+1. \end{cases}$$

We then have

$$\frac{1}{Y}\int_2^Y E_q(e^y)^2\,dy = \frac{1}{Y}\int_{2\le y\le Y} H_M(E_q(e^y))\,dy - \frac{1}{Y}\int_{\substack{2\le y\le Y \\ M<|E_q(e^y)|\le M+1}} H_M(E_q(e^y))\,dy + \frac{1}{Y}\int_{\substack{2\le y\le Y \\ |E_q(e^y)|>M}} E_q(e^y)^2\,dy;$$

therefore by (13) and Lemma 2.2 we get that

$$\limsup_{Y\to\infty}\frac{1}{Y}\int_2^Y E_q(e^y)^2\,dy = \int_{\mathbb{R}} H_M(t)\,d\mu_q(t) + \epsilon_M,$$

where ϵ_M tends to zero as $M \to \infty$. Using the bound

$$\mu_q((-\infty,-M]\cup[M,\infty)) \ll \exp(-c_2\sqrt{M})$$

(see [Rubinstein and Sarnak 1994, Theorem 1.2]), we get by taking $M\to\infty$ that

$$\limsup_{Y\to\infty}\frac{1}{Y}\int_2^Y E_q(e^y)^2\,dy = \int_{\mathbb{R}} t^2\,d\mu_q(t).$$

The same reasoning applies to the lim inf, and thus the proof is finished. □

The following calculation is similar to that of Schlage-Puchta [2000], who computed the moments of $e^{-t/2}\psi(e^t;\chi)$.

Lemma 2.6. *Assume GRH. Then,*

$$\mathbb{E}[X_q] = \rho(q) - 1 + z(q) \quad \textit{and} \quad \operatorname{Var}[X_q] = \sum\nolimits^*_{\gamma\ne 0}\frac{m_\gamma^2}{\frac14+\gamma^2},$$

where the last sum runs over the imaginary parts of the nontrivial zeros of

$$Z_q(s) := \prod_{\substack{\chi^2=\chi_0 \\ \chi\ne\chi_0}} L(s,\chi),$$

m_γ denotes the multiplicity of the zero $\frac{1}{2}+i\gamma$, the star means that we count the zeros without multiplicity, and $z(q)$ denotes the multiplicity of the (possible) real zero $\gamma=0$.

Proof. By Lemma 2.1 we have that

$$\begin{aligned}\int_2^Y E_q(e^y)\,dy &= (\rho(q)-1+z(q))(Y-2)\\ &\quad+\sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}}\sum_{\gamma_\chi\neq 0}\frac{1}{\frac{1}{2}+i\gamma_\chi}\int_2^Y e^{i\gamma_\chi y}\,dy+o_{Y\to\infty}(Y)\\ &= (\rho(q)-1+z(q))(Y-2)+O_q(1)+o_{Y\to\infty}(Y),\end{aligned}$$

by absolute convergence. Taking $Y\to\infty$ and applying Lemma 2.5 gives that

$$\mathbb{E}[X_q]=\lim_{Y\to\infty}\frac{1}{Y}\int_2^Y E_q(e^y)\,dy=\rho(q)-1+z(q).$$

The calculation of the variance follows from Lemma 2.1 and from Parseval's identity for B^2 almost-periodic functions [Besicovitch 1926]. (An alternative way to compute the variance is to argue as in [Schlage-Puchta 2000].) ☐

Remark 2.7. It is a general fact that Besicovitch almost-periodic functions [1955] always have a mean value. Moreover, Parseval's identity [Besicovitch 1955; 1926] shows that Besicovitch B^2 almost-periodic functions $f(y)$ have a second moment, given by

$$\lim_{Y\to\infty}\frac{1}{Y}\int_0^Y f(y)^2\,dy=\sum_{n\geq 1}A_n^2,$$

where the A_n are the Fourier coefficients of f.

Lemma 2.8. *Let χ mod q be a Dirichlet character. We have for $k\geq 1$ that*

$$b_k(\chi):=\sum_{\gamma_\chi}\frac{1}{\left(\frac{1}{4}+\gamma_\chi^2\right)^k}\asymp_k \log q^*,$$

where the sum is counted with multiplicity.

Remark 2.9. One has an exact formula for $b_k(\chi)$, in terms of the values of the derivatives of $\log L(s,\chi)$ evaluated at $s=1$ [Fiorilli and Martin 2013, Lemma 3.15].

Proof. This follows from applying the Riemann–von Mangoldt formula and summation by parts:

$$\begin{aligned} b_k(\chi) &= \sum_{|\gamma_\chi|<1} \frac{1}{\left(\frac14+\gamma_\chi^2\right)^k} + \int_1^\infty \frac{dN(t,\chi)}{\left(\frac14+t^2\right)^k} \\ &\asymp_k N(1,\chi) + 2k\int_1^\infty \frac{tN(t,\chi)}{\left(\frac14+t^2\right)^{k+1}}\,dt \asymp_k \log q^*. \end{aligned}$$ □

Lemma 2.10. *Assume GRH. If*

$$B(q) := \frac{\mathbb{E}[X_q]}{\sqrt{\operatorname{Var}[X_q]}}$$

is large enough, then

$$\underline{\delta}(q; NR, R) \geq 1 - 2\frac{\operatorname{Var}[X_q]}{\mathbb{E}[X_q]^2}.$$

Proof. It is clear from Lemmas 2.6 and 2.8 that $\operatorname{Var}[X_q] \gg \log q'$, and therefore our assumption that $B(q)$ is large enough implies that $\mathbb{E}[X_q]$ is also large enough, say at least 4. Now let

$$H(x) := \begin{cases} 0 & \text{if } x<0, \\ 1 & \text{if } x \geq 0, \end{cases} \qquad f(x) := \begin{cases} 0 & \text{if } x<0, \\ x & \text{if } 0\leq x<1, \\ 1 & \text{if } x\geq 1. \end{cases}$$

Clearly, $f(x)$ is bounded Lipschitz continuous and $f(x) \leq H(x)$. Therefore,

$$\underline{\delta}(q; NR, R) = \liminf_{Y\to\infty} \frac{1}{Y}\int_2^Y H(E_q(e^y))\,dy \geq \liminf_{Y\to\infty}\frac{1}{Y}\int_2^Y f(E_q(e^y))\,dy,$$

which by Lemma 2.2 is equal to

$$\begin{aligned} \int_{\mathbb{R}} f(t)\,d\mu_q(t) &= 1 - \int_{\mathbb{R}} (1-f(t))\,d\mu_q(t) \\ &= 1 - \int_{-\infty}^1 (1-f(t))\,d\mu_q(t) \geq 1 - \mu_q(-\infty,1]. \end{aligned}$$

We now apply Chebyshev's inequality:

$$\begin{aligned} \mu_q(-\infty,1] &= \operatorname{Prob}[X_q \leq 1] = \operatorname{Prob}[X_q - \mathbb{E}[X_q] \leq 1 - \mathbb{E}[X_q]] \\ &\leq \operatorname{Prob}[|X_q - \mathbb{E}[X_q]| \geq \mathbb{E}[X_q]-1] \leq \frac{\operatorname{Var}[X_q]}{(\mathbb{E}[X_q]-1)^2} \leq 2\frac{\operatorname{Var}[X_q]}{\mathbb{E}[X_q]^2}, \end{aligned}$$

since $\mathbb{E}[X_q] \geq 4$, and therefore

$$\underline{\delta}(q; NR, R) \geq 1 - 2\frac{\operatorname{Var}[X_q]}{\mathbb{E}[X_q]^2}.$$ □

Proof of Theorem 1.2. By Lemma 2.6, our hypothesis implies that for the sequence of moduli q under consideration,

$$\mathrm{Var}[X_q] \leq \max_{\gamma}(m_\gamma) \sum_{\gamma}^{*} \frac{m_\gamma}{\frac{1}{4}+\gamma_\chi^2} = o\left(\frac{\rho(q)}{\log q}\rho(q)\log q\right) = o\big(\rho(q)^2\big),$$

by Lemma 2.8. Lemma 2.6 also implies that $\mathbb{E}[X_q] \gg \rho(q)$, and hence Lemma 2.10 implies that

$$\underline{\delta}(q; NR, R) \geq 1 - o(1).$$

The last inequality to show, that is, $\bar{\delta}(q; NR, R) < 1$, follows from a lower bound on $\mu_E(-\infty, -1]$ similar to that in [Rubinstein and Sarnak 1994, Theorem 1.2], which holds in greater generality [Akbary et al. 2013]. Using this lower bound, one does an analysis similar to that in the proof of Lemma 2.10, replacing the function $f(x)$ with

$$g(x) := \begin{cases} 1 & \text{if } x < -1, \\ -x & \text{if } -1 \leq x < 0, \\ 0 & \text{if } x \geq 0, \end{cases}$$

in order to obtain a lower bound for

$$1 - \bar{\delta}(q; NR, R) = \limsup_{Y\to\infty} \frac{1}{Y}\int_2^\infty (1 - H(E_q(e^y)))\, dy. \qquad \square$$

3. A central limit theorem

The goal of this section is to show a central limit theorem under GRH and LI, from which the second part of Theorem 1.1 will follow. We first translate our problem to questions on sums of independent random variables, which can be done thanks to LI. Recall that we are interested in the set of n such that

$$\pi(n; q, NR) > (\rho(q) - 1)\pi(n; q, R).$$

Lemma 3.1. *Assume GRH and LI. Then the logarithmic density of the set of n for which $\pi(n; q, NR) > (\rho(q) - 1)\pi(n; q, R)$ exists and is equal to*

$$\mathrm{Prob}[X_q > 0],$$

where X_q is the random variable defined in Section 2. Moreover, we have

$$X_q \sim \rho(q) - 1 + \sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0 \\ \chi \neq \chi_0}} \sum_{\gamma_\chi > 0} \frac{2\Re(Z_{\gamma_\chi})}{\sqrt{\frac{1}{4}+\gamma_\chi^2}}, \tag{14}$$

where the Z_{γ_χ} are independent identically distributed random variables following a uniform distribution on the unit circle in $\mathbb{C}$.

Proof. By Lemma 2.1, we have that

$$\frac{\pi(x;q,NR)-(\rho(q)-1)\pi(x;q,R)}{\sqrt{x}/\log x}=\rho(q)-1+\sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}}\sum_{\gamma_\chi}\frac{x^{i\gamma_\chi}}{\rho_\chi}+o(1),$$

since LI implies that there are no real zeros. It follows by the work of Rubinstein and Sarnak that $\delta(q;NR,R)$ exists and equals $\mathrm{Prob}[X_q>0]$ (their analysis shows that the distribution function of X_q is continuous). Moreover, an argument similar to the proof of [Fiorilli and Martin 2013, Proposition 2.3] shows that (14) holds. □

One can show that the random variables in (14) have variance $\mathrm{Var}[\Re(Z_{\gamma_\chi})]=\frac{1}{2}$ and have mean $\mathbb{E}[Z_{\gamma_\chi}]=0$. Using this and the fact that they are mutually independent, we recover Lemma 2.6:

$$\mathbb{E}[X_q]=\rho(q)-1,\quad \mathrm{Var}[X_q]=\sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}}\sum_{\gamma_\chi}\frac{1}{\frac{1}{4}+\gamma_\chi^2},\tag{15}$$

since the zeros come in conjugate pairs (χ is real). We will see in the following lemma that $\mathrm{Var}[X_q]\asymp\rho(q)\log q'$ (recall that $q':=\prod_{p\mid q}p$), and this is a crucial fact in our analysis.

Lemma 3.2. *Assume GRH and LI, and let X_q be the random variable defined in (14). We have that*

$$\mathrm{Var}[X_q]=2^{\omega(q)-1-\epsilon_q}\log q'\left[1+O\left(\frac{\log\log q'}{\log q'}\right)\right],$$

where $\epsilon_q=1$ if $2\mid q$, and $\epsilon_q=0$ otherwise. In particular,

$$\mathrm{Var}[X_q]\asymp\rho(q)\log q'.$$

Proof. By Remark 1.4, we have that

$$\mathrm{Var}[X_q]=\mathrm{Var}[X_{2^e\ell}],$$

where $e\le 3$, $2^e\parallel q$ and $\ell:=\prod_{p\mid q,\,p\neq 2}p$. Therefore we assume from now on (without loss of generality) that $q=2^e\ell$, with $e\le 3$ and ℓ an odd squarefree integer.

Lemma 3.5 of [Fiorilli and Martin 2013] gives that

$$\begin{aligned}\sum_{\gamma_\chi}\frac{1}{\frac{1}{4}+\gamma_\chi^2}&=\log q^*-\log\pi-\gamma-(1+\chi(-1))\log 2+2\Re\frac{L'}{L}(1,\chi^*)\\&=\log q^*+O(\log\log q^*),\end{aligned}\tag{16}$$

by Littlewood's GRH bound on $(L'/L)(1,\chi)$. Plugging this into (15), we get

$$\operatorname{Var}[X_q] = \sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0}} \log q^* + O\big(2^{\omega(q)} \log\log q\big).$$

If q is odd, then there is exactly one primitive real character mod d for every $d \mid q$, hence

$$\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0}} \log q^* = \sum_{d\mid q} \log d = \sum_{d\mid q}\sum_{p\mid d} \log p = \sum_{p\mid q} (\log p) 2^{\omega(q)-1} = 2^{\omega(q)-1}\log q.$$

If $2 \parallel q$, then there are no primitive characters modulo even divisors of q, so

$$\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0}} \log q^* = \sum_{d\mid \frac{q}{2}} \log d = 2^{\omega(q)-2} \log \frac{q}{2}.$$

If $4 \parallel q$, then there is exactly one primitive real character modulo divisors which are multiples of 4, so

$$\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0}} \log q^* = \sum_{d\mid \frac{q}{4}} \log d + \sum_{4\mid d\mid q} \log d = 2^{\omega(q)-2} \log(2q).$$

If $8 \parallel q$, then there are exactly two primitive real characters modulo divisors which are multiples of 8, so

$$\sum_{\substack{\chi \bmod q \\ \chi^2=\chi_0}} \log q^* = \sum_{d\mid \frac{q}{8}} \log d + \sum_{\substack{4\mid d\mid q \\ 8\nmid d}} \log d + 2\sum_{8\mid d\mid q} \log d = 2^{\omega(q)-2} \log(8q). \qquad \square$$

Let X_q be the random variable defined in (14), and define

$$B(q) := \frac{\mathbb{E}[X_q]}{\sqrt{\operatorname{Var}[X_q]}}.$$

It is $B(q)$ which dictates the behavior of the race we are considering: if $B(q)$ is small, then the race will not be very biased, whereas if $B(q)$ is large, then the race will have a significant bias. By Lemma 3.2, we have under GRH and LI the estimate

$$B(q) = \sqrt{\frac{2^{\omega(q)+1+\epsilon_q}}{\log q'}}\left[1 + O\left(2^{-\omega(q)} + \frac{\log\log q'}{\log q'}\right)\right]. \tag{17}$$

To prove the second part of Theorem 1.1 we will need a sequence of moduli for which $B(q)$ is very regular.

Lemma 3.3. *For any fixed $0 < c < \infty$, there exists an increasing sequence of squarefree odd integers $\{q_n\}$ such that*

$$2^{\omega(q_n)+1} = (c + o(1)) \log q_n.$$

Proof. Fix $0 < c < \infty$, and define $e_c := \min\{e \geq 1 : 2^{-e} c < 2/\log 4\}$ and $c_1 := 2^{-e_c} c < 2/\log 4$. Define for $\ell = 1, 2, \dots$ the intervals

$$I_\ell := \big(\exp(c_1^{-1} 2^\ell),\, 2\exp(c_1^{-1} 2^\ell)\big), \quad J_\ell := \big(2\exp(c_1^{-1} 2^\ell),\, 4\exp(c_1^{-1} 2^\ell)\big).$$

Since $c_1 < 2/\log 4$, we have that for all $\ell \geq 1$,

$$4\exp(c_1^{-1} 2^\ell) < \exp(c_1^{-1} 2^{\ell+1});$$

hence our intervals are all disjoint. We define p_ℓ to be any prime in the interval I_ℓ, and similarly for $p'_\ell \in J_\ell$. The existence of such primes is granted by Bertrand's postulate (note that $\exp(c_1^{-1} 2^1) > 4$). Now, the sequence of moduli we are looking for is

$$q_n := \prod_{1 \leq \ell \leq e_c} p'_\ell \prod_{1 \leq \ell \leq n} p_\ell,$$

since

$$\begin{aligned}
\frac{2^{\omega(q_n)+1}}{\log q_n} &= \frac{2^{n+e_c+1}}{O_c(1) + \sum_{1 \leq \ell \leq n} \big(c_1^{-1} 2^\ell + O(1)\big)} = \frac{2^{n+e_c+1}}{c_1^{-1} 2^{n+1} + O_c(n)} \\
&= 2^{e_c} c_1 \left(1 + O_c\left(\frac{n}{2^n}\right)\right) = c(1 + o(1)),
\end{aligned}$$

by definition of c_1. □

Before proving the second part of Theorem 1.1, we give some information about the characteristic function of the random variables we are interested in. The following lemma implies a central limit theorem.

Lemma 3.4. *Assume GRH and LI. Let X_q be the random variable defined in (14), and define*

$$Y_q := \frac{X_q - \mathbb{E}[X_q]}{\sqrt{\operatorname{Var}[X_q]}} = \frac{1}{\sqrt{\operatorname{Var}[X_q]}} \sum_{\substack{\chi \bmod q \\ \chi^2 = \chi_0 \\ \chi \neq \chi_0}} \sum_{\gamma_\chi > 0} \frac{2\Re(Z_{\gamma_\chi})}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}.$$

The characteristic function of Y_q satisfies, for $|\xi| \leq \frac{3}{5}\sqrt{\operatorname{Var}[X_q]}$,

$$\hat{Y}_q(\xi) = -\frac{\xi^2}{2} + O\left(\frac{\xi^4}{\rho(q) \log q'}\right).$$

Moreover, in the same range, we have

$$\widehat{Y}_q(\xi) \le -\frac{\xi^2}{2}. \tag{18}$$

Proof. The proof is very similar to that of [Fiorilli and Martin 2013, Theorem 3.22]. Using the additivity of the cumulant-generating function of X_q, one can show that

$$\log \widehat{X}_q(\xi) = i\,\mathbb{E}[X_q]\xi + \sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}} \sum_{\gamma_\chi>0} \log\left(J_0\left(\frac{2\xi}{\sqrt{\frac14+\gamma_\chi^2}}\right)\right), \tag{19}$$

where $J_0(x)$ is the Bessel function of the first kind:

$$J_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n (x/2)^{2n}}{n!^2}.$$

We will use the following Taylor expansion, which is valid for $|\xi| \le \frac{12}{5}$ (see [Fiorilli and Martin 2013, Section 2.2]):

$$\log J_0(\xi) = -\frac{\xi^2}{4} + O(\xi^4). \tag{20}$$

Plugging this estimate into (19), we get that for $|\xi| \le \frac35$,

$$\log \widehat{X}_q(\xi) = i\,\mathbb{E}[X_q]\xi - \xi^2 \sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}} \sum_{\gamma_\chi>0} \frac{1}{\frac14+\gamma_\chi^2} + O\left(\xi^4 \sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}} \sum_{\gamma_\chi>0} \frac{1}{\left(\frac14+\gamma_\chi^2\right)^2}\right).$$

Applying Lemma 2.8 gives

$$\sum_{\substack{\chi \bmod q\\ \chi^2=\chi_0\\ \chi\neq\chi_0}} \sum_{\gamma_\chi>0} \frac{1}{\left(\frac14+\gamma_\chi^2\right)^2} \ll \rho(q)\log q'.$$

Moreover, by Lemma 3.2 we have $\mathrm{Var}[X_q] \asymp \rho(q)\log q'$. Putting these together and using (15), we get that

$$\log \widehat{Y}_q(\xi) = \log \widehat{X}_q\left(\frac{\xi}{\sqrt{\mathrm{Var}[X_q]}}\right) - i\,\mathbb{E}[X_q]\frac{\xi}{\sqrt{\mathrm{Var}[X_q]}} = -\frac{\xi^2}{2} + O\left(\frac{\xi^4}{\rho(q)\log q'}\right),$$

showing the first assertion. For the second we use the same argument, but we replace the estimate (20) with the following inequality, valid in the range $|\xi| \le \frac{12}{5}$:

$$\log J_0(\xi) \le -\frac{\xi^2}{4}. \qquad \square$$

Lemma 3.5 (Berry–Esseen inequality). *Assume GRH and LI. Denote by F_q the distribution function of*

$$Y_q := \frac{X_q - \mathbb{E}[X_q]}{\sqrt{\mathrm{Var}[X_q]}},$$

and by F that of the Gaussian distribution. We have that

$$\sup_{x\in\mathbb{R}} |F_q(x) - F(x)| \ll \frac{1}{\rho(q)\log q'}.$$

Remark 3.6. One could get a more precise estimate using the Feuerverger–Martin formula [2000]. However, the estimate of Lemma 3.5 is sufficient for our purposes.

Proof. Since the statement is trivial if $\rho(q)\log q'$ is bounded, we can assume without loss of generality that $\mathrm{Var}[X_q] \ge 1$ (by Lemma 3.2).

The Berry–Esseen inequality in the form of [Esseen 1945, Theorem 2a] gives that for any $T > 0$,

$$\sup_{x\in\mathbb{R}} |F_q(x) - F(x)| \ll \int_{-T}^{T} \frac{\widehat{Y}_q(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\, d\xi + \frac{1}{T}. \tag{21}$$

We take $T := \mathrm{Var}[X_q]$. By Lemma 3.4, the part of the integral with $|\xi| \le \frac{3}{5}\mathrm{Var}[X_q]^{\frac{1}{4}}$ is at most

$$\int_{-\frac{3}{5}\mathrm{Var}[X_q]^{\frac{1}{4}}}^{\frac{3}{5}\mathrm{Var}[X_q]^{\frac{1}{4}}} \frac{e^{-\frac{\xi^2}{2}}\left(e^{O\left(\frac{\xi^4}{\rho(q)\log q'}\right)} - 1\right)}{\xi}\, d\xi \ll \frac{1}{\rho(q)\log q'} \int_{\mathbb{R}} \xi^3 e^{-\frac{\xi^2}{2}}\, d\xi \ll \frac{1}{\rho(q)\log q'}.$$

We now bound the remaining part of the integral using an argument analogous to [Fiorilli and Martin 2013, Proposition 2.14]. Fix $0 \le \lambda \le \frac{5}{6}$. By the properties of the Bessel function $J_0(x)$, we have that if $|\xi| > \lambda/2$, then whatever $\gamma_\chi \in \mathbb{R}$ is,

$$\left| J_0\left(\frac{2\xi}{\sqrt{\frac{1}{4}+\gamma_\chi^2}}\right)\right| \le J_0\left(\frac{\lambda}{\sqrt{\frac{1}{4}+\gamma_\chi^2}}\right).$$

By (19), this shows that in the range $|\xi| > \frac{5}{12}\operatorname{Var}[X_q]^{\frac14}$ we have $|\widehat{X}_q(\xi)| \le |\widehat{X}_q(\frac{5}{12}\operatorname{Var}[X_q]^{\frac14})|$ (since $\operatorname{Var}[X_q] \ge 1$), and so

$$\int_{\frac35 \operatorname{Var}[X_q]^{\frac14} < |\xi| \le \operatorname{Var}[X_q]} \frac{\widehat{Y}_q(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\, d\xi$$
$$\ll \widehat{Y}_q\big(\tfrac{5}{12}\operatorname{Var}[X_q]^{\frac14}\big) \log \operatorname{Var}[X_q] + \int_{|\xi| > \frac35 \operatorname{Var}[X_q]^{\frac14}} \frac{e^{-\frac{\xi^2}{2}}}{\xi}\, d\xi$$
$$\ll \exp(-\tfrac{25}{577}\operatorname{Var}[X_q]^{\frac12}) + \exp(-\tfrac{9}{51}\operatorname{Var}[X_q]^{\frac12}),$$

by (18). Applying Lemma 3.2, we conclude that the right-hand side of (21) is at most a constant times $(\rho(q)\log q')^{-1}$. $\square$

Proof of Theorem 1.1, second part. Fix $\eta \in \left[\frac12, 1\right]$. We wish to find a sequence of moduli $\{q_n\}$ such that $\delta(q_n, NR, R) \to \eta$. The case $\eta = 1$ was already covered in part (1), and the case $\eta = \frac12$ follows from taking prime values of q, by the central limit theorem [Rubinstein and Sarnak 1994]. Therefore we can assume that $\frac12 < \eta < 1$.

Let $\kappa > 0$ be the unique real solution of the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\kappa}^{\infty} e^{-\frac{t^2}{2}}\, dt = \eta.$$

Moreover, let $\{q_n\}$ be the sequence of squarefree odd integers from Lemma 3.3 for which

$$2^{\omega(q_n)+1} = \log q_n' (\kappa^2 + o(1)).$$

By (17), this gives that as $n \to \infty$,

$$B(q_n) := \frac{\mathbb{E}[X_{q_n}]}{\sqrt{\operatorname{Var}[X_{q_n}]}} \to \kappa.$$

Define

$$Y_{q_n} := \frac{X_{q_n} - \mathbb{E}[X_{q_n}]}{\sqrt{\operatorname{Var}[X_{q_n}]}} = \frac{X_{q_n}}{\sqrt{\operatorname{Var}[X_{q_n}]}} - B(q_n).$$

We will use the central limit theorem of Lemma 3.4, as well as the Berry–Esseen inequality (21). Denoting by F_{q_n} the distribution function of Y_{q_n} and by F that of the Gaussian distribution, we have that

$$\begin{aligned}|\delta(q_n, NR, R) - \eta| &= |\operatorname{Prob}[X_{q_n} > 0] - \eta| = |\operatorname{Prob}[X_{q_n} \le 0] - (1-\eta)| \\ &= |F_{q_n}(-B(q_n)) - F(-\kappa)| \\ &\le |F_{q_n}(-B(q_n)) - F(-B(q_n))| + |F(-B(q_n)) - F(-\kappa)| \\ &\ll \frac{1}{\rho(q_n)\log q_n'} + |\kappa - B(q_n)|,\end{aligned}$$

by Lemma 3.5 and by the fact that the probability density function of the Gaussian is bounded on $\mathbb{R}$. Looking at the proof of Lemma 3.3, we see that $\rho(q_n) \to \infty$, hence this last quantity tends to zero as $n \to \infty$, concluding the proof. □

4. A more precise estimation of the bias using the theory of large deviations

To give a more precise estimate for the bias we are interested in under LI, we use the theory of large deviations of independent random variables. The fundamental estimate of this section is given in the following theorem.

Theorem 4.1 [Montgomery and Odlyzko 1988, Theorem 2]. *For $n = 1, 2, \ldots,$ let Y_n be independent real-valued random variables such that $\mathbb{E}[Y_n] = 0$ and $|Y_n| \le 1$. Suppose that there is a constant $c > 0$ such that $\mathbb{E}[Y_n^2] \ge c$ for all n. Put $Y = \sum r_n Y_n$ where $\sum r_n^2 < \infty$.*

If $\sum_{|r_n| \ge \alpha} |r_n| \le V/2$ then

$$\operatorname{Prob}[Y \ge V] \le \exp\biggl(-\frac{1}{16}V^2\Bigl(\sum_{|r_n|<\alpha} r_n^2\Bigr)^{-1}\biggr).$$

If $\sum_{|r_n| \ge \alpha} |r_n| \ge 2V$ then

$$\operatorname{Prob}[Y \ge V] \ge a_1 \exp\biggl(-a_2 V^2\Bigl(\sum_{|r_n|<\alpha} r_n^2\Bigr)^{-1}\biggr).$$

Here $a_1 > 0$ and $a_2 > 0$ depend only on c.

To make use of these bounds we need to give estimates on sums over zeros.

Lemma 4.2. *For $T \ge 1$, we have*

$$\sum_{|\gamma_\chi|<T} \frac{1}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} = \frac{1}{\pi}\log(q^*\sqrt{T})\log T + O(\log(q^* T)).$$

Proof. We start from the Riemann–von Mangoldt formula,

$$N(T, \chi) = \frac{T}{\pi}\log\frac{q^* T}{2\pi e} + O(\log q^* T).$$

With a summation by parts we get

$$\begin{aligned}\sum_{|\gamma_\chi|<T} \frac{1}{\sqrt{\frac14+\gamma_\chi^2}} &= O(\log q^*) + \int_1^T \frac{dN(t,\chi)}{\sqrt{\frac14+t^2}} \\ &= \frac{N(T,\chi)}{\sqrt{\frac14+T^2}} + \int_1^T \frac{tN(t,\chi)}{\left(\frac14+t^2\right)^{\frac32}}\,dt + O(\log q^*) \\ &= \int_1^T \frac{\frac{t^2}{\pi}\log\frac{q^* t}{2\pi e}}{\left(\frac14+t^2\right)^{\frac32}}\,dt + O(\log(q^*T)) \\ &= \frac{1}{\pi}\log\bigl(q^*\sqrt{T}\bigr)\log T + O(\log q^* T). \qquad \square\end{aligned}$$

Lemma 4.3. *Assume LI, and let $\mathcal{F}(q)$ be a subset of the group of Dirichlet characters* mod q *such that $\chi \in \mathcal{F}(q) \Rightarrow \bar\chi \in \mathcal{F}(q)$. Define the random variable*

$$Y := \sum_{\chi\in\mathcal{F}(q)} \sum_{\gamma_\chi>0} \frac{2\Re(Z_{\gamma_\chi})}{\sqrt{\frac14+\gamma_\chi^2}},$$

where the Z_{γ_χ} are i.i.d. uniformly distributed on the unit circle. Then, we have for q large enough that

$$a_1 \exp\left(-a_2\frac{|\mathcal{F}(q)|}{L(q)}\right) \le \mathrm{Prob}[Y \ge |\mathcal{F}(q)|] \le \exp\left(-a_3\frac{|\mathcal{F}(q)|}{L(q)}\right),$$

where the a_i are absolute constants and

$$L(q) := \frac{\sum_{\chi\in\mathcal{F}(q)} \log q^*}{|\mathcal{F}(q)|} \ge \frac{\log 2}{2}.$$

Proof. This is a direct application of Theorem 4.1. Taking the sequence $\{r_i\}$ to be the $2/\sqrt{\frac14+\gamma_\chi^2}$ ordered by size, and denoting by C the constant $\sqrt{4/\alpha^2-1/4}$, we have for $0<\alpha\le 4$ that

$$\sum_{|r_n|\ge\alpha} |r_n| = \sum_{\chi\in\mathcal{F}(q)} \sum_{0<\gamma_\chi\le C} \frac{2}{\sqrt{\frac14+\gamma_\chi^2}},$$

$$\sum_{|r_n|>\alpha} |r_n|^2 = \sum_{\chi\in\mathcal{F}(q)} \sum_{\gamma_\chi>C} \frac{4}{\frac14+\gamma_\chi^2}.$$

For the upper bound we take $\alpha=4$, then we trivially have $\sum_{|r_n|\ge\alpha} |r_n| \le |\mathcal{F}(q)|/2$, so

$$\mathrm{Prob}[Y \ge |\mathcal{F}(q)|] \le \exp\left(-\tfrac{1}{16}|\mathcal{F}(q)|^2\left(c_1\sum_{\chi\in\mathcal{F}(q)}\log q^*\right)^{-1}\right)$$

for some absolute constant c_1. For the lower bound we take $\alpha = 2/\sqrt{\frac{1}{4}+T_0^2}$, where $T_0 > 1$ is a fixed large real number (independent of q and $\mathscr{F}(q)$) such that

$$\sum_{\chi\in\mathscr{F}(q)}\sum_{|\gamma_\chi|\le T_0}\frac{1}{\sqrt{\frac{1}{4}+\gamma_\chi^2}} \ge \frac{4}{\log 2}L(q)|\mathscr{F}(q)| \ge 2|\mathscr{F}(q)|,$$

whose existence is granted by Lemma 4.2 (we grouped together conjugate characters). Then Theorem 4.1 gives the bound

$$\begin{aligned}\mathrm{Prob}[Y \ge |\mathscr{F}(q)|] &\ge c_2\exp\biggl(-c_3|\mathscr{F}(q)|^2\biggl(\sum_{\chi\in\mathscr{F}(q)}\sum_{\gamma_\chi>T_0}\frac{4}{\frac{1}{4}+\gamma_\chi^2}\biggr)^{-1}\biggr)\\ &\ge c_2\exp\biggl(-c_3|\mathscr{F}(q)|^2\biggl(c_4\sum_{\chi\in\mathscr{F}(q)}\log q^*\biggr)^{-1}\biggr)\end{aligned}$$

for q large enough and some absolute constants c_2, c_3 and c_4, since if we choose $T_1 > T_0$ independent of χ and large enough such that $N(2T_1,\chi)-N(T_1,\chi) \gg \log q^*$ (this is possible by the Riemann–von Mangoldt formula), then we have

$$\begin{aligned}\sum_{\chi\in\mathscr{F}(q)}\sum_{\gamma_\chi>T_0}\frac{4}{\frac{1}{4}+\gamma_\chi^2} &\ge \sum_{\chi\in\mathscr{F}(q)}\sum_{T_1<\gamma_\chi<2T_1}\frac{4}{\frac{1}{4}+\gamma_\chi^2}\\ &\ge \sum_{\chi\in\mathscr{F}(q)}\frac{4}{\frac{1}{4}+(2T_1)^2}(N(2T_1,\chi)-N(T_1,\chi))\\ &\gg \sum_{\chi\in\mathscr{F}(q)}\log q^*.\end{aligned}$$ □

Proof of Theorem 1.6. Let X_q be the random variable in (14) and define the symmetric random variable

$$Y_q := X_q - \mathbb{E}[X_q].$$

By Lemma 3.1,

$$\begin{aligned}\delta(q;NR,R) &= \mathrm{Prob}[X_q>0] = \mathrm{Prob}[Y_q > -\mathbb{E}[X_q]]\\ &= \mathrm{Prob}[Y_q < \mathbb{E}[X_q]] = 1-\mathrm{Prob}[Y_q \ge \mathbb{E}[X_q]].\end{aligned}$$

The proof follows by taking $\mathscr{F}(q) := \{\chi \bmod q : \chi^2=\chi_0, \chi\ne\chi_0\}$ in Lemma 4.3 and by estimating $L(q)$ as in the proof of Lemma 3.2. □

5. A more general analysis

In this section we do a more general analysis by studying arbitrary linear combinations of prime counting functions.

Throughout the section, $\vec{a}=(a_1,\dots,a_k)$ will be a vector of invertible reduced residues mod q and $\vec{\alpha}=(\alpha_1,\dots,\alpha_k)$ will be a nonzero vector of real numbers such that $\sum_{i=1}^k \alpha_i=0$. Recall that

$$\epsilon_i=\begin{cases}1 & \text{if } a_i\equiv\square \bmod q,\\ 0 & \text{if } a_i\not\equiv\square \bmod q,\end{cases}$$

and we assume without loss of generality that

$$\sum_{i=1}^k \epsilon_i\alpha_i<0.$$

To prove Theorems 1.10, 1.14 and 1.17, we need a few lemmas.

Lemma 5.1. *Assume GRH and LI. Then the quantity*

$$E(y;q,\vec{a};\vec{\alpha}):=\phi(q)\frac{\alpha_1\pi(e^y;q,a_1)+\cdots+\alpha_k\pi(e^y;q,a_k)}{e^{y/2}/y}$$

has the same distribution as the random variable

$$X_{q;\vec{a},\vec{\alpha}}:=-\rho(q)\sum_{i=1}^k\epsilon_i\alpha_i+\sum_{\chi\neq\chi_0}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|\sum_{\gamma_\chi>0}\frac{2\Re(Z_{\gamma_\chi})}{\sqrt{\frac14+\gamma_\chi^2}},\qquad(22)$$

where the Z_{γ_χ} are independent random variables following a uniform distribution on the unit circle in $\mathbb{C}$.

Remark 5.2. If we take $a_1,\dots,a_{\phi(q)(1-\rho(q)^{-1})}$ to be the set of all quadratic nonresidues mod q with $\alpha_1=\cdots=\alpha_{\phi(q)(1-\rho(q)^{-1})}=1/\phi(q)$, and we take $a_{\phi(q)(1-\rho(q)^{-1})+1},\dots,a_{\phi(q)}$ to be the set of all quadratic residues mod q with $\alpha_{\phi(q)(1-\rho(q)^{-1})+1}=\cdots=\alpha_{\phi(q)}=(1-\rho(q))/\phi(q)$, then we recover formula (14).

Proof. In the same way as in the proof of Lemma 3.1, we get by the explicit formula and by applying GRH that

$$\begin{aligned}F(y;q,\vec{a},\vec{\alpha})&:=\phi(q)\frac{\alpha_1\psi(e^y;q,a_1)+\cdots+\alpha_k\psi(e^y;q,a_k)}{e^{y/2}}\\&=-\sum_{\chi\neq\chi_0}(\alpha_1\bar\chi(a_1)+\cdots+\alpha_k\bar\chi(a_k))\sum_{\gamma_\chi}\frac{e^{i\gamma_\chi y}}{\rho_\chi}+o_q(1)\end{aligned}$$

(the main terms are canceled since $\sum_{i=1}^k\alpha_i=0$). By [Rubinstein and Sarnak 1994], $F(y;q,\vec{a},\vec{\alpha})$ has the same distribution as $X_{q;\vec{a},\vec{\alpha}}-\mathbb{E}[X_{q;\vec{a},\vec{\alpha}}]$, since LI implies that there are no real zeros. The second step is to use summation by parts and to remove squares and other prime powers; this gives that

$$E(y;q,\vec{a},\vec{\alpha}) + \rho(q)\sum_{i=1}^{k}\epsilon_i\alpha_i + o(1) = F(y;q,\vec{a},\vec{\alpha}),$$

completing the proof. □

Before we give a bound on the variance of this distribution, we prove a lemma about conductors.

Lemma 5.3. *Let* $1 \le L \le \phi(q)$. *Then,*

$$\#\{\chi \bmod q : q^* \le L\} \le \min\{L\tau(q), L^2\}.$$

Proof. Denoting by $\phi^*(d)$ the number of primitive characters modulo d, we have

$$\sum_{\substack{d\,|\,q\\ d\le L}} \phi^*(d) \le \min\Big\{\sum_{d\le L} d,\ L\sum_{d\,|\,q} 1\Big\}.$$ □

Lemma 5.4. *Assume LI. Let* $V(q;\vec{a},\vec{\alpha}) := \mathrm{Var}[X_{q;\vec{a},\vec{\alpha}}]$, *where* $X_{q;\vec{a},\vec{\alpha}}$ *is the random variable defined in* (22). *Then,*

$$\phi(q)\|\vec{a}\|_2^2 \log\frac{3\phi(q)}{k} \ll V(q;\vec{a},\vec{\alpha}) \ll \phi(q)\|\vec{a}\|_2^2\log q, \tag{23}$$

where

$$\|\vec{a}\|_2^2 := \sum_{i=1}^{k}\alpha_i^2.$$

Remark 5.5. The upper bound in (23) is attained when q is prime by Lemma 5.8. As for the lower bound, if we take moduli q with a fixed set of distinct prime factors (for instance powers of a fixed prime) and consider the race between residues and non-residues with the weights of Remark 5.2, we obtain by Lemma 3.2 that $V(q;\vec{a},\vec{\alpha}) = O(1)$, and this is of the same order of magnitude as the lower bound in (23).

Proof. Since the Z_{γ_χ} in (22) are independent and have variance $\frac{1}{2}$, we have that

$$\mathrm{Var}[X_{q;\vec{a},\vec{\alpha}}] = \sum_{\chi\ne\chi_0} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2 \sum_{\gamma_\chi}\frac{1}{\frac{1}{4}+\gamma_\chi^2} \tag{24}$$

(LI implies there are no real zeros). Combining this with Lemma 2.8 gives

$$V(q;\vec{a},\vec{\alpha}) \asymp \sum_{\chi\ne\chi_0} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2\log q^*. \tag{25}$$

Now, $\alpha_1\chi_0(a_1)+\cdots+\alpha_k\chi_0(a_k)=0$, so

$$\begin{aligned}\sum_{\chi\neq\chi_0}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2 &= \sum_{\chi \bmod q}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2\\ &= \sum_{1\le i,j\le k}\alpha_i\alpha_j\sum_{\chi \bmod q}\chi(a_ia_j^{-1})\\ &= \phi(q)\sum_{i=1}^k\alpha_i^2. \end{aligned} \tag{26}$$

Using this and (25), the upper bound follows from the fact that $\log q^*\le\log q$. This also gives the lower bound $V(q;\vec{a},\vec{\alpha})\ge(\log 3)\phi(q)\|\vec{\alpha}\|_2^2$, which proves the claim for bounded values of $\phi(q)/k$. Hence we assume from now on that $\phi(q)/k\ge 576$. We fix a parameter $1<L<\phi(q)$ and discard the characters of conductor at most L:

$$\begin{aligned}V(q;\vec{a},\vec{\alpha}) &\gg \log L\sum_{\substack{\chi \bmod q:\\ q^*>L}}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2\\ &= \log L\sum_{1\le i,j\le k}\alpha_i\alpha_j\sum_{\substack{\chi \bmod q:\\ q^*>L}}\chi(a_ia_j^{-1})\\ &= \log L\Bigg[\sum_{i=1}^k\alpha_i^2\sum_{\substack{\chi \bmod q\\ q^*>L}}1+\sum_{1\le i\neq j\le k}\alpha_i\alpha_j\sum_{\substack{\chi \bmod q:\\ q^*>L}}\chi(a_ia_j^{-1})\Bigg],\end{aligned}$$

which by Lemma 5.3 and the orthogonality relations is

$$\begin{aligned}&\ge \log L\Bigg[\sum_{i=1}^k\alpha_i^2(\phi(q)-\min\{L\tau(q),L^2\})-\sum_{1\le i\neq j\le k}|\alpha_i\alpha_j|\min\{L\tau(q),L^2\}\Bigg]\\ &\ge \log L\|\vec{\alpha}\|_2^2[\phi(q)-(k+1)\min\{L\tau(q),L^2\}]\end{aligned}$$

by the Cauchy–Schwarz inequality. Taking $L:=(3\phi(q)/k)^{\frac{1}{3}}$ gives the result, since then $\phi(q)/k\ge 576$ implies that $(k+1)L^2\le\phi(q)/2$. □

Remark 5.6. In the last proof, we did not lose a lot by discarding the characters of conductor at most $(3\phi(q)/k)^{\frac{1}{3}}$, since by (26) their contribution is

$$\ll \phi(q)\|\vec{\alpha}\|_2^2\log\frac{3\phi(q)}{k}.$$

Proof of Theorem 1.14. We have by Lemma 5.4 that there exists an absolute constant $c>0$ such that

$$B(q;\vec{a},\vec{\alpha}) := \frac{\mathbb{E}[X_{q;\vec{a},\vec{a}}]}{\sqrt{\mathrm{Var}[X_{q;\vec{a},\vec{a}}]}} \geq \frac{\rho(q)\left|\sum_{i=1}^{k}\epsilon_i\alpha_i\right|}{\sqrt{c\phi(q)\log q\sum_{i=1}^{k}\alpha_i^2}},$$

a quantity which is greater or equal to $(c\epsilon)^{-\frac{1}{2}}$ by the condition of the theorem. We conclude that $1-\delta(q;\vec{a},\vec{\alpha})\leq c\epsilon$ by using Chebyshev's bound in the same way as in the proof of Theorem 1.1. ☐

Proof of Theorem 1.10. It is a particular case of Theorem 1.14. ☐

We now prove our negative results. To do so, we need to provide a central limit theorem, analogous to Lemma 3.4.

Lemma 5.7. *Assume LI, and let*

$$Y_{q;\vec{a},\vec{\alpha}} := \frac{X_{q;\vec{a},\vec{\alpha}}-\mathbb{E}[X_{q;\vec{a},\vec{\alpha}}]}{\sqrt{\mathrm{Var}[X_{q;\vec{a},\vec{\alpha}}]}}.$$

The characteristic function of $Y_{q;\vec{a},\vec{\alpha}}$ satisfies

$$\log\widehat{Y}_{q;\vec{a},\vec{\alpha}}(\xi) = -\frac{\xi^2}{2}+O\left(\frac{\xi^4}{\log(3\phi(q)/k)}\min\left\{1,\ \frac{k^2\log q}{\phi(q)\log(3\phi(q)/k)}\right\}\right)$$

in the range $|\xi|\leq 3/(5\|\vec{\alpha}\|_1)$, where $\|\vec{\alpha}\|_1 := \sum_{i=1}^{k}|\alpha_i|$.

Proof. As in Lemma 3.4, we compute

$$\log\widehat{X}_{q;\vec{a},\vec{\alpha}}(\xi) = i\mathbb{E}[X_{q;\vec{a},\vec{a}}]\xi+\sum_{\chi\neq\chi_0}\sum_{\gamma_\chi>0}\log\left(J_0\left(\frac{2|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|\xi}{\sqrt{\frac{1}{4}+\gamma_\chi^2}}\right)\right).$$

We now use the Taylor expansion (20), which is valid as soon as $|\xi|\leq 3/(5\|\vec{\alpha}\|_1)$, since under this condition we have

$$\frac{2|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|\,|\xi|}{\sqrt{\frac{1}{4}+\gamma_\chi^2}} \leq \frac{2\|\vec{\alpha}\|_1}{\frac{1}{2}}\frac{3}{5\|\vec{\alpha}\|_1} = \frac{12}{5}.$$

Applying Lemma 2.8, we obtain that

$$\log \widehat{Y}_{q;\vec{a},\vec{\alpha}}(\xi) = -\frac{\xi^2}{2} + O\left(\xi^4 \frac{\sum_{\chi\neq\chi_0} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^4 \log q^*}{\left(\sum_{\chi\neq\chi_0} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2 \log q^*\right)^2}\right). \tag{27}$$

If $\phi(q)/k$ is bounded, then the statement trivially follows from the bound $\sum_i a_i^4 \leq \left(\sum_i a_i^2\right)^2$. Therefore we assume from now on that $\phi(q)/k \geq 576$.

We now use two different approaches to bound the error term. The first idea is to "factor out $\sqrt{\log q^*}$" before applying the trivial inequality $\sum_i a_i^4 \leq \left(\sum_i a_i^2\right)^2$. We have seen in Remark 5.6 that the main contribution to the variance is that of the characters with $q^* \geq L := (3\phi(q)/k)^{\frac{1}{3}}$. We use the same idea here. Setting $\Theta_\chi := |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2$, we have

$$\begin{aligned}
\sum_{\chi\neq\chi_0} \Theta_\chi \log q^* &\geq \sum_{\substack{\chi\neq\chi_0 \\ q^*>L}} \Theta_\chi \log q^* \\
&\geq \sqrt{\log L} \sum_{\substack{\chi\neq\chi_0 \\ q^*>L}} \Theta_\chi \sqrt{\log q^*} \\
&\geq \sqrt{\log L}\left(\sum_{\chi\neq\chi_0} \Theta_\chi \sqrt{\log q^*} - kL^2\sqrt{\log L}\|\vec{\alpha}\|_2^2\right)
\end{aligned} \tag{28}$$

by Lemma 5.3 and the Cauchy–Schwarz inequality. Now, by our choice of L, the fact that $\phi(q)/k \geq 576$ and the equality $\sum_{\chi\neq\chi_0}\Theta_\chi = \phi(q)\|\vec{\alpha}\|_2^2$ (see (26)), we have

$$\begin{aligned}
kL^2\sqrt{\log L}\|\vec{\alpha}\|_2^2 &\leq \frac{1}{2}\sqrt{\log L}\left[\phi(q)\|\vec{\alpha}\|_2^2 - kL^2\|\vec{\alpha}\|_2^2\right] \\
&\leq \frac{1}{2}\sum_{\substack{\chi\neq\chi_0 \\ q^*\geq L}} \Theta_\chi \sqrt{\log q^*} \leq \frac{1}{2}\sum_{\chi\neq\chi_0} \Theta_\chi \sqrt{\log q^*},
\end{aligned}$$

hence (28) gives that

$$\sum_{\chi\neq\chi_0} \Theta_\chi \log q^* \gg \sqrt{\log L} \sum_{\chi\neq\chi_0} \Theta_\chi \sqrt{\log q^*}.$$

Plugging this into (27) and using the trivial bound

$$\sum_{\chi\neq\chi_0} \Theta_\chi^2 \log q^* \leq \left(\sum_{\chi\neq\chi_0} \Theta_\chi \sqrt{\log q^*}\right)^2,$$

we get that the error term is $\ll \xi^4/\log L$.

For the second upper bound we use Lemma 5.4 and the Cauchy–Schwarz inequality:

$$\frac{\sum_{\chi\neq\chi_0}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^4\log q^*}{\left(\sum_{\chi\neq\chi_0}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2\log q^*\right)^2}$$
$$\ll \frac{\log q}{\log(3\phi(q)/k)^2}\frac{\sum_{\chi\neq\chi_0}|\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^4}{(\phi(q)\|\vec{\alpha}\|_2^2)^2}$$
$$= \frac{\log q}{\log(3\phi(q)/k)^2}\frac{\sum_{\substack{i,j,i',j'\\ a_ia_{i'}\equiv a_ja_{j'} \bmod q}}\alpha_i\alpha_{i'}\alpha_j\alpha_{j'}}{\phi(q)\|\vec{\alpha}\|_2^4}$$
$$\le \frac{\log q}{\log(3\phi(q)/k)^2}\frac{\left(\sqrt{\sum_{i=1}^k\alpha_i^2}\sqrt{\sum_{j=1}^k 1}\right)^4}{\phi(q)\|\vec{\alpha}\|_2^4},$$

which gives the claimed bound. □

Proof of Theorem 1.17. Let $K\ge 1$ and define $c>0$ to be the constant implied in the lower bound in Lemma 5.4. Assume that $k\le e^{-e^{4K}}\phi(q)$ and that (9) holds with $K_2=cK$. Define the vector $\vec{\beta}:=\left(e^{-K}/\|\vec{\alpha}\|_1\right)\vec{\alpha}$, so that $\|\vec{\beta}\|_1=e^{-K}$, which will allow us to apply Lemma 5.7. Clearly,

$$\delta(q;\vec{a},\vec{\alpha})=\delta(q;\vec{a},\vec{\beta}),$$

since multiplying $\vec{\alpha}$ by a positive constant does not affect the inequality

$$\alpha_1\pi(n;q,a_1)+\cdots+\alpha_k\pi(n;q,a_k)>0.$$

We have by Lemma 5.4 and by the definition of c that

$$B(q;\vec{a},\vec{\beta}):=\frac{\mathbb{E}[X_{q;\vec{a},\vec{\beta}}]}{\sqrt{\mathrm{Var}[X_{q;\vec{a},\vec{\beta}}]}}\le\frac{\rho(q)\left|\sum_{i=1}^k\epsilon_i\beta_i\right|}{\sqrt{c\phi(q)\log(3\phi(q)/k)\sum_{i=1}^k\beta_i^2}}$$
$$=c^{-\frac{1}{2}}\frac{\rho(q)\left|\sum_{i=1}^k\epsilon_i\alpha_i\right|}{\sqrt{\phi(q)\log(3\phi(q)/k)\sum_{i=1}^k\alpha_i^2}},$$

a quantity which is at most $\sqrt{K}$ by (9). Defining

$$Y_{q;\vec{a},\vec{\beta}}:=\frac{X_{q;\vec{a},\vec{\beta}}-\mathbb{E}[X_{q;\vec{a},\vec{\beta}}]}{\sqrt{\mathrm{Var}[X_{q;\vec{a},\vec{\beta}}]}},$$

we have by Lemma 5.7 and by our condition on k that in the range $|\xi| \le \frac{3}{5}e^{K}$,

$$\log \widehat{Y}_{q;\vec{a},\vec{\beta}}(\xi) = -\frac{\xi^2}{2} + O\left(\frac{\xi^4}{e^{4K}}\right).$$

Combining this with the Berry–Esseen inequality (21) and taking W to be a standard Gaussian random variable with mean 0 and variance 1, we get

$$\begin{aligned}
\mathrm{Prob}[Y_{q;\vec{a},\vec{\beta}} > -B(q;\vec{a},\vec{\beta})] - \mathrm{Prob}[W > -B(q;\vec{a},\vec{\beta})] \\
\ll \int_{-\frac{3}{5}e^K}^{\frac{3}{5}e^K} \frac{\widehat{Y}_{q;\vec{a},\vec{\beta}}(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\, d\xi + \tfrac{5}{3}e^{-K} \\
\ll \int_{-\frac{3}{5}e^K}^{\frac{3}{5}e^K} \frac{\xi^3 e^{-\frac{\xi^2}{2}}}{e^{4K}}\, d\xi + e^{-K} \\
\ll e^{-K}. \qquad (29)
\end{aligned}$$

However, since $B(q;\vec{a},\vec{\beta}) \le \sqrt{K}$, we have that

$$\mathrm{Prob}[W \le -B(q;\vec{a},\vec{\beta})] \ge c_1 \frac{e^{-\frac{K}{2}}}{K}$$

for some absolute constant c_1. Therefore, applying (29) gives

$$\begin{aligned}
\delta(q;\vec{a},\vec{\beta}) &= \mathrm{Prob}[Y_{q;\vec{a},\vec{\beta}} > -B(q;\vec{a},\vec{\beta})] \\
&= \mathrm{Prob}[W > -B(q;\vec{a},\vec{\beta})] + O(e^{-K}) \\
&\le 1 - c_1 e^{-\frac{K}{2}}/K + c_2 e^{-K},
\end{aligned}$$

a quantity which is less than the right-hand side of (10) for K large enough. The proof is finished since $\delta(q;\vec{a},\vec{\alpha}) = \delta\big(q;\vec{a},\vec{\beta}\big)$. □

To end this section we give an estimate for the variance $V(q;\vec{a},\vec{\alpha})$. While we have not explicitly made use of this expression, we include it for its intrinsic interest, and for its ability to give a precise evaluation of the variance $V(q;\vec{a},\vec{\alpha})$ for values of q having prescribed prime factors.

Lemma 5.8. *Assuming GRH and LI, we have that*

$$\begin{aligned}
&V(q;\vec{a},\vec{\alpha}) \\
&\quad = \phi(q)\|\vec{\alpha}\|_2(\log q + O(\log\log q)) - \phi(q)\sum_{i\ne j}\alpha_i\alpha_j \frac{\Lambda\left(\frac{q}{(q,a_i a_j^{-1}-1)}\right)}{\phi\left(\frac{q}{(q,a_i a_j^{-1}-1)}\right)}. \qquad (30)
\end{aligned}$$

Proof. Using [Fiorilli and Martin 2013, Proposition 3.3], we obtain that

$$\begin{aligned}\sum_{\chi \bmod q} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2 \log q^* \\ &= \sum_{1\le i,j\le k} \alpha_i\alpha_j \sum_{\chi \bmod q} \chi(a_i a_j^{-1}) \log q^* \\ &= \phi(q)\Big(\log q - \sum_{p\mid q} \frac{\log p}{p-1}\Big)\sum_{i=1}^k \alpha_i^2 - \phi(q)\sum_{i\ne j}\alpha_i\alpha_j \frac{\Lambda\Big(\frac{q}{(q,a_ia_j^{-1}-1)}\Big)}{\phi\Big(\frac{q}{(q,a_ia_j^{-1}-1)}\Big)}.\end{aligned}$$

We have

$$\sum_{p\mid q}\frac{\log p}{p-1} \le \sum_{i=1}^{\omega(q)} \frac{\log p_i}{p_i-1} \ll \log\log q,$$

where p_i denotes the i-th prime. The claimed estimate then follows by combining this with the formula

$$V(q;\vec{a},\vec{\alpha}) = \sum_{\chi \bmod q} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2 \sum_{\gamma_\chi} \frac{1}{\frac14+\gamma_\chi^2}$$

(see (24)) and with (16). Note that by the Littlewood bound $(L'/L)(1,\chi) \ll \log\log q^*$, the implied error term is

$$\ll \sum_{\chi \bmod q} |\alpha_1\chi(a_1)+\cdots+\alpha_k\chi(a_k)|^2 \log\log q = \phi(q)\|\vec{\alpha}\|_2 \log\log q. \qquad \square$$

It might seem like the second term of (30) is an error term; however, this is not necessarily true for large values of k (see Lemma 3.2). Nevertheless, we expect many cancellations to occur since

$$\sum_{i\ne j}\alpha_i\alpha_j = \Big(\sum_{i=1}^k \alpha_i\Big)^2 - \sum_{i=1}^k \alpha_i^2 = -\sum_{i=1}^k \alpha_i^2.$$

Acknowledgements

I would like to thank my former advisor Andrew Granville for his very interesting question which motivated this work, and for his comments and encouragement. I would also like to thank Enrico Bombieri and Peter Sarnak for their advice and their encouragement, and the Institute for Advanced Study for providing excellent research conditions. I thank Kevin Ford, Youness Lamzouri and the people at the University of Illinois at Urbana–Champaign for their hospitality and for very fruitful conversations which led me to consider general linear combinations of prime counting functions. I thank Jan-Christoph Schlage-Puchta for Remark 2.4

and for suggesting the current proof of Lemma 2.5. I also thank Carl Pomerance for suggesting Remark 1.9, and Greg Martin for suggesting Remark 1.12. I thank Barry Mazur for motivating me to weaken the linear independence hypothesis. I thank Byungchul Cha, Ke Gong, Peter Humphries, Chen Meiri, Nathan Ng and Anders Södergren for helpful conversations and for their comments. Finally I thank the anonymous referees for their careful reading of the manuscript and their very useful comments. This work was supported by an NSERC Postdoctoral Fellowship, as well as NSF grant DMS-0635607.

References

[Akbary et al. 2013] A. Akbary, N. Ng, and M. Shahabi, "Limiting distributions of the classical error terms of prime number theory", preprint, 2013. arXiv 1306.1657

[Bays and Hudson 2000] C. Bays and R. H. Hudson, "A new bound for the smallest x with $\pi(x) > \mathrm{li}(x)$", *Math. Comp.* **69**:231 (2000), 1285–1296. MR 2001c:11138 Zbl 1042.11001

[Bays et al. 2001] C. Bays, K. Ford, R. H. Hudson, and M. Rubinstein, "Zeros of Dirichlet L-functions near the real axis and Chebyshev's bias", *J. Number Theory* **87**:1 (2001), 54–76. MR 2001m:11148 Zbl 1009.11057

[Besicovitch 1926] A. S. Besicovitch, "On generalized almost periodic functions", *Proc. London Math. Soc.* **S2-25**:1 (1926), 495–512. MR 1575297 JFM 52.0263.01

[Besicovitch 1955] A. S. Besicovitch, *Almost periodic functions*, Dover, New York, 1955. MR 16,817a Zbl 0065.07102

[Bober et al. 2014] J. Bober, J. B. Conrey, D. W. Farmer, A. Fujii, S. Koutsoliotas, S. Lemurell, M. Rubinstein, and H. Yoshida, "The highest lowest zero of general L-functions", preprint, 2014. arXiv 1211.5996

[Esseen 1945] C.-G. Esseen, "Fourier analysis of distribution functions: A mathematical study of the Laplace–Gaussian law", *Acta Math.* **77** (1945), 1–125. MR 7,312a Zbl 0060.28705

[Feuerverger and Martin 2000] A. Feuerverger and G. Martin, "Biases in the Shanks–Rényi prime number race", *Experiment. Math.* **9**:4 (2000), 535–570. MR 2002d:11111 Zbl 0976.11041

[Fiorilli and Martin 2013] D. Fiorilli and G. Martin, "Inequities in the Shanks–Rényi prime number race: An asymptotic formula for the densities", *J. Reine Angew. Math.* **676** (2013), 121–212. MR 3028758 Zbl 1276.11150

[Ford 2008a] K. Ford, "The distribution of integers with a divisor in a given interval", *Ann. of Math.* (2) **168**:2 (2008), 367–433. MR 2009m:11152 Zbl 1181.11058

[Ford 2008b] K. Ford, "Integers with a divisor in $(y, 2y]$", pp. 65–80 in *Anatomy of integers*, edited by J.-M. De Koninck et al., CRM Proc. Lecture Notes **46**, Amer. Math. Soc., Providence, RI, 2008. MR 2009i:11113 Zbl 1175.11053

[Granville and Martin 2006] A. Granville and G. Martin, "Prime number races", *Amer. Math. Monthly* **113**:1 (2006), 1–33. MR 2006h:11112 Zbl 1139.11037

[Hardy and Littlewood 1916] G. H. Hardy and J. E. Littlewood, "Contributions to the theory of the Riemann Zeta-function and the theory of the distribution of primes", *Acta Math.* **41**:1 (1916), 119–196. MR 1555148 JFM 46.0498.01

[Kaczorowski 1995] J. Kaczorowski, "On the distribution of primes (mod 4)", *Analysis* **15**:2 (1995), 159–171. MR 96h:11095 Zbl 0826.11042

[Lamzouri 2012] Y. Lamzouri, "Large deviations of the limiting distribution in the Shanks–Rényi prime number race", *Math. Proc. Cambridge Philos. Soc.* **153**:1 (2012), 147–166. MR 2943671 Zbl 1286.11143

[Landau 1918a] E. Landau, "Über einige ältere Vermutungen und Behauptungen in der Primzahltheorie, I", *Math. Z.* **1** (1918), 1–24. JFM 46.0263.02

[Landau 1918b] E. Landau, "Über einige ältere Vermutungen und Behauptungen in der Primzahltheorie, II", *Math. Z.* **1**:2-3 (1918), 213–219. MR 1544293 JFM 46.0263.02

[Miller 2002] S. D. Miller, "The highest lowest zero and other applications of positivity", *Duke Math. J.* **112**:1 (2002), 83–116. MR 2003b:11047 Zbl 1014.11036

[Monach 1980] W. R. Monach, *Numerical investigation of several problems in number theory*, Ph.D. thesis, University of Michigan, 1980, Available at http://search.proquest.com/docview/302978421.

[Montgomery 1980] H. L. Montgomery, "The zeta function and prime numbers", pp. 1–31 in *Proceedings of the Queen's Number Theory Conference, 1979* (Kingston, ON, 1979), edited by P. Ribenboim, Queen's Papers in Pure and Appl. Math. **54**, Queen's Univ., Kingston, ON, 1980. MR 82k:10047 Zbl 0451.10023

[Montgomery and Odlyzko 1988] H. L. Montgomery and A. M. Odlyzko, "Large deviations of sums of independent random variables", *Acta Arith.* **49**:4 (1988), 427–434. MR 89m:11075 Zbl 0641.60032

[Myerscough 2013] C. Myerscough, "Application of an accurate remainder term in the calculation of residue class distributions", preprint, 2013. arXiv 1301.1434

[Ng 2000] N. Ng, *Limiting distributions and zeros of Artin L-functions*, Ph.D. thesis, University of British Columbia, 2000, Available at http://www.cs.uleth.ca/∼nathanng/RESEARCH/phd.thesis.pdf.

[Ng 2004] N. Ng, "The distribution of the summatory function of the Möbius function", *Proc. London Math. Soc.* (3) **89**:2 (2004), 361–389. MR 2005f:11215 Zbl 1138.11341

[Odlyzko 1990] A. M. Odlyzko, "Bounds for discriminants and related estimates for class numbers, regulators and zeros of zeta functions: A survey of recent results", *Sém. Théor. Nombres Bordeaux* (2) **2**:1 (1990), 119–141. MR 91i:11154 Zbl 0722.11054

[Rubinstein and Sarnak 1994] M. Rubinstein and P. Sarnak, "Chebyshev's bias", *Experiment. Math.* **3**:3 (1994), 173–197. MR 96d:11099 Zbl 0823.11050

[Schlage-Puchta 2000] J.-C. Puchta, "On large oscillations of the remainder of the prime number theorems", *Acta Math. Hungar.* **87**:3 (2000), 213–227. MR 2001g:11146 Zbl 0963.11051

[Skewes 1933] S. Skewes, "On the difference $\pi(x) - \mathrm{li}(x)$, (I)", *J. Lond. Math. Soc.* **S8-4**:4 (1933), 277–283. Zbl 0007.34003

Communicated by Andrew Granville
Received 2014-03-26 Revised 2014-03-26 Accepted 2014-05-23

daniel.fiorilli@uottawa.ca *Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa, Ontario, K1N 6N5, Canada*

ALGEBRA AND NUMBER THEORY 8:7 (2014)
dx.doi.org/10.2140/ant.2014.8.1769

Bounded gaps between primes with a given primitive root

Paul Pollack

Fix an integer $g \neq -1$ that is not a perfect square. In 1927, Artin conjectured that there are infinitely many primes for which g is a primitive root. Forty years later, Hooley showed that Artin's conjecture follows from the generalized Riemann hypothesis (GRH). We inject Hooley's analysis into the Maynard–Tao work on bounded gaps between primes. This leads to the following GRH-conditional result: *Fix an integer $m \geq 2$. If $q_1 < q_2 < q_3 < \cdots$ is the sequence of primes possessing g as a primitive root, then $\liminf_{n\to\infty}(q_{n+(m-1)} - q_n) \leq C_m$, where C_m is a finite constant that depends on m but not on g.* We also show that the primes $q_n, q_{n+1}, \ldots, q_{n+m-1}$ in this result may be taken to be consecutive.

1. Introduction

The following conjecture was proposed by Emil Artin in the course of a September 1927 conversation with Helmut Hasse:

Artin's primitive root conjecture. *Fix an integer $g \neq -1$ that is not a square. There are infinitely many primes p for which g is a primitive root modulo p. In fact, the number of such $p \leq x$ is (as $x \to \infty$) asymptotically $c_g \pi(x)$ for a certain $c_g > 0$.*

While there is a substantial literature surrounding Artin's conjecture (lovingly catalogued in the survey [Moree 2012]), we still know infuriatingly little. In particular, there is no specific value of g which is known to occur as a primitive root for infinitely many primes. However, thanks to work of Heath-Brown [1986] (refining earlier results of Gupta and Murty [1984]), we know that at least one of 2, 3, and 5 has this property. In fact, one can replace "2, 3, and 5" with any three multiplicatively independent integers satisfying mild conditions.

In a seminal paper, Hooley [1967] (see also his exposition in [Hooley 1976, Chapter 3]) showed that the Chebotarev density theorem with a sufficiently sharp error term would imply the quantitative form of Artin's conjecture. Moreover, he showed that such a variant of Chebotarev's density theorem — at least for the cases relevant for this application — follows from the generalized Riemann hypothesis

MSC2010: primary 11A07; secondary 11N05.
Keywords: primitive root, Artin's conjecture, bounded gaps, Maynard–Tao theorem.

(GRH) for Dedekind zeta functions. Thus, under GRH, we have a complete proof of Artin's conjecture.

In this paper, we combine Hooley's work on Artin's conjecture with recent methods used to study gaps between primes. In sensational work of Maynard [2013] and Tao, it is shown that $\liminf_{n\to\infty}(p_{n+m-1}-p_n)<\infty$ for every m. Here $p_1<p_2<p_3<\cdots$ is the sequence of all primes, in the usual order. Our main theorem is an analogous bounded gaps result for primes possessing a prescribed primitive root.

Theorem 1.1 (conditional on GRH). *Fix an integer $g\neq -1$ and not a square. Let $q_1<q_2<q_3<\cdots$ denote the sequence of primes for which g is a primitive root. Then, for each m,*

$$\liminf_{n\to\infty}(q_{n+m-1}-q_n)\le C_m,$$

where C_m is a finite constant depending on m but not on g.

In the last section of the paper, we show how to modify the proof of Theorem 1.1 to impose the additional restriction that the m primes $q_n, q_{n+1}, \ldots, q_{n+m-1}$ are in fact *consecutive* (Theorem 4.1).

We remark that other recent work producing bounded gaps between primes in special sets has been done by Thorner [2014], who handles primes restricted by Chebotarev conditions, and by Li and Pan [2014], who work with primes p for which $p+2$ is an "almost prime".

Notation. The letters p and q always denote primes. We use the Bachmann–Landau O and o notations, as well as the associated Vinogradov symbols $\ll$ and $\gg$, with their usual meanings.

2. Technical preparation

Configurations of quadratic residues and nonresidues. We will use that certain configurations of residues and nonresidues are guaranteed to appear for all large enough primes. This is a fairly standard consequence of the Riemann Hypothesis for curves, as proved by Weil, but we give the argument for completeness. The following lemma is a special case of [Wan 1997, Corollary 2.3].

Lemma 2.1. *Let p be a prime. Suppose that $f(T)$ is a monic polynomial in $\mathbb{F}_p[T]$ of degree d and that $f(T)$ is not a square in $\mathbb{F}_p[T]$. Then*

$$\left|\sum_{a \bmod p}\left(\frac{f(a)}{p}\right)\right|\le (d-1)\sqrt{p}.$$

Lemma 2.2. *Let p be a prime, and let k be a positive integer. Suppose that $h_1,\ldots,h_k$ are integers, no two of which are congruent modulo p. Suppose*

$\epsilon_1, \ldots, \epsilon_k \in \{\pm 1\}$. *The number of mod p solutions n to the system of equations*

$$\left(\frac{n+h_i}{p}\right) = \epsilon_i \quad \textit{for all} \;\; 1 \le i \le k \tag{2-1}$$

is at least $p/2^k - (k-1)\sqrt{p} - k$.

Proof. For each n, let $\iota(n) = (1/2^k) \prod_{i=1}^{k} \bigl(1 + \epsilon_i \bigl(\frac{n+h_i}{p}\bigr)\bigr)$. If we suppose that $n \not\equiv -h_1, \ldots, -h_k \pmod p$, then $\iota(n)$ equals 1 when (2-1) holds, and 0 otherwise. Since $|\iota(n)| \le 1$ for all n, the number of solutions to (2-1) is at least $-k + \sum_{n \bmod p} \iota(n)$. For each subset $S \subset \{1, 2, 3, \ldots, k\}$, put $f_S(T) = \prod_{i \in S}(T + h_i) \in \mathbb{F}_p[T]$. Then

$$\sum_{n \bmod p} \iota(n) = \frac{1}{2^k} \sum_{S \subset \{1,2,\ldots,k\}} \left(\prod_{i \in S} \epsilon_i\right) \sum_{n \bmod p} \left(\frac{f_S(n)}{p}\right).$$

If $S = \varnothing$, then $f_S = 1$, and we get a contribution of $p/2^k$. In all other cases, f_S is a nonsquare polynomial of degree at most k. By Lemma 2.1, the total contribution from all nonempty subsets of $\{1, 2, \ldots, k\}$ is bounded in absolute value by $((2^k - 1)/2^k)(k-1)\sqrt{p} \le (k-1)\sqrt{p}$. Thus, $\sum_{n \bmod p} \iota(n) \ge p/2^k - (k-1)\sqrt{p}$, and the lemma follows. □

Effective Chebotarev. The next result is due in essence to Lagarias and Odlyzko [1977], although the precise formulation we give is due to Serre [1981, §2.4]:

Theorem 2.3 (conditional on GRH). *Let L be a finite Galois extension of $\mathbb{Q}$ with Galois group G, and let C be a conjugacy class of G. The number of unramified primes $p \le x$ whose Frobenius conjugacy class $(p, L/\mathbb{Q})$ is C is given by*

$$\frac{\#C}{\#G}\mathrm{Li}(x) + O\Bigl(\frac{\#C}{\#G} x^{1/2} \bigl(\log|\Delta_L| + [L:\mathbb{Q}] \log x\bigr)\Bigr)$$

for all $x \ge 2$. Here Δ_L denotes the discriminant of L and the O-constant is absolute.

To apply Theorem 2.3, we require an upper bound for the term $\log|\Delta_L|$. The following result, which is contained in [Serre 1981, Proposition 6], suffices for our applications.

Lemma 2.4. *For every Galois extension $L/\mathbb{Q}$, we have*

$$\log|\Delta_L| \le ([L:\mathbb{Q}] - 1) \sum_{p \mid \Delta_L} \log p + [L:\mathbb{Q}] \log [L:\mathbb{Q}].$$

3. Proof of Theorem 1.1

The Maynard–Tao strategy. We begin by recalling the strategy of [Maynard 2013] for producing bounded gaps between primes. Let $k \ge 2$ be a fixed positive integer, and let $\mathcal{H} = \{h_1 < h_2 < \cdots < h_k\}$ denote a fixed *admissible k-tuple*, i.e., a set of

k distinct integers that does not occupy all of the residue classes modulo p for any prime p. With N a large positive integer, we seek values of n belonging to the dyadic interval $[N, 2N)$ for which the shifted tuple $n+h_1, n+h_2, \dots, n+h_k$ contains several primes.

Let $W := \prod_{p \le \log\log\log N} p$. Choose an integer ν so that $\gcd(\nu + h_i, W) = 1$ for all $1 \le i \le k$; the existence of such a ν is implied by the admissibility of $\mathcal{H}$. We restrict attention to integers $n \equiv \nu \pmod{W}$. This has the effect of pre-sieving the values of n to ensure that none of the $n+h_i$ have any small prime factors. Let $w(n)$ denote nonnegative weights (to be chosen momentarily), and let $\chi_{\mathcal{P}}$ denote the characteristic function of the set $\mathcal{P}$ of prime numbers. One studies the sums

$$S_1 := \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod{W}}} w(n) \quad \text{and} \quad S_2 := \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod{W}}} \left(\sum_{i=1}^{k} \chi_{\mathcal{P}}(n+h_i) \right) w(n).$$

The ratio S_2/S_1 is a weighted average of the number of primes among $n+h_1, \dots,$ $n+h_k$, as n ranges over $[N, 2N)$. Consequently, if $S_2 > (m-1)S_1$ for the positive integer m, then at least m of the numbers $n+h_1, \dots, n+h_k$ are primes. So, if the inequality $S_2 > (m-1)S_1$ is achieved for a sequence of n tending to infinity, then $\liminf(p_{n+m-1} - p_n) \le h_k - h_1 < \infty$.

As we have described it so far, this strategy goes back to Goldston, Pintz, and Yıldırım. The key innovation in the approach of Maynard and Tao is the choice of congenial weights $w(n)$. The following result, which is a restatement of [Maynard 2013, Proposition 4.1], is crucial.

Proposition 3.1. *Let θ be a real number, $0 < \theta < \frac{1}{4}$. Let F be a piecewise differentiable function supported on the simplex $\{(x_1, \dots, x_k) : \text{each } x_i \ge 0, \sum_{i=1}^{k} x_i \le 1\}$. With $R := N^\theta$, put*

$$\lambda_{d_1,\dots,d_k} := \left(\prod_{i=1}^{k} \mu(d_i) d_i \right) \sum_{\substack{r_1,\dots,r_k \\ d_i \mid r_i \,\forall i \\ (r_i, W)=1 \,\forall i}} \frac{\mu\left(\prod_{i=1}^{k} r_i\right)^2}{\prod_{i=1}^{k} \varphi(r_i)} F\left(\frac{\log r_1}{\log R}, \dots, \frac{\log r_k}{\log R} \right)$$

whenever $\gcd\left(\prod_{i=1}^{k} d_i, W\right) = 1$, *and let* $\lambda_{d_1,\dots,d_k} = 0$ *otherwise. Let*

$$w(n) := \left(\sum_{d_i \mid n+h_i \,\forall i} \lambda_{d_1,\dots,d_k} \right)^2.$$

Then, as $N \to \infty$,

$$S_1 \sim \frac{\varphi(W)^k}{W^{k+1}} N (\log R)^k I_k(F) \quad \textit{and}$$

$$S_2 \sim \frac{\varphi(W)^k}{W^{k+1}} \frac{N}{\log N} (\log R)^{k+1} \sum_{m=1}^{k} J_k^{(m)}(F),$$

provided that $I_k(F) \neq 0$ *and* $J_k^{(m)}(F) \neq 0$ *for each* m*, where*

$$I_k(F) := \int \cdots \int_{[0,1]^k} F(t_1, \ldots, t_k)^2 \, \mathrm{d}t_1 \, \mathrm{d}t_2 \cdots \mathrm{d}t_k,$$

$$J_k^{(m)}(F) := \int \cdots \int_{[0,1]^{k-1}} \left(\int_0^1 F(t_1, \ldots, t_k) \, \mathrm{d}t_m \right)^2 \mathrm{d}t_1 \cdots \mathrm{d}t_{m-1} \, \mathrm{d}t_{m+1} \cdots \mathrm{d}t_k.$$

From our interpretation of S_2/S_1 as a weighted average, we know that there is an $n \in [N, 2N)$ for which at least S_2/S_1 of the numbers $n+h_1, \ldots, n+h_k$ are prime. Proposition 3.1 shows that $S_2/S_1 \to (\theta/I_k(F)) \sum_{m=1}^{k} J_k^{(m)}(F)$ as $N \to \infty$. For each F satisfying the conditions of Proposition 3.1, put

$$M_k(F) := \frac{1}{I_k(F)} \sum_{m=1}^{k} J_k^{(m)}(F), \quad \text{and set } M_k := \sup_F M_k(F). \tag{3-1}$$

Upon choosing θ close to $\frac{1}{4}$ and F so that $M_k(F)$ is close to M_k, we find that, infinitely often, at least $\lceil \frac{1}{4} M_k \rceil$ of the numbers $n+h_1, \ldots, n+h_k$ are prime. The following lower bound on M_k is due to Maynard [2013, Proposition 4.3].

Proposition 3.2. $M_k \to \infty$ *as* $k \to \infty$*. In fact, for all sufficiently large values of* k,

$$M_k > \log k - 2 \log\log k - 2.$$

Consequently, once k is a little larger than e^{4m}, we have $\lceil \frac{1}{4} M_k \rceil > m-1$. From the above discussion, $\liminf_{n\to\infty}(p_{n+m-1} - p_n) \le h_k - h_1 < \infty$ for every admissible k-tuple $\mathcal{H}$. Choosing $\mathcal{H}$ carefully, this argument gives $\liminf_{n\to\infty}(p_{n+m-1} - p_n) \ll m^3 e^{4m}$; see the proof of [Maynard 2013, Theorem 1.1] for details.

Modifying Maynard–Tao. For the rest of the paper, we fix an integer $g \neq -1$ that is not a square. Let $\widetilde{\mathcal{P}}$ denote the set of primes having g as a primitive root. Fix an integer $k \ge 2$, and let

$$K := 9k^2 \cdot 4^k.$$

We fix $\mathcal{H}$ as the admissible k-tuple having $h_i = (i-1)K!$ for all $1 \le i \le k$; that is,

$$\mathcal{H} := \{0, K!, 2K!, \ldots, (k-1)K!\}. \tag{3-2}$$

We work below with a fixed function F satisfying the conditions of Proposition 3.1. For the rest of the argument, implied constants may depend on g, k, and F without further mention.

In what follows, we think of N as very large, in particular much larger than g. We use the Maynard–Tao strategy to detect integers $n \in [N, 2N)$ for which the list $n+h_1, \ldots, n+h_k$ contains several primes belonging to $\widetilde{\mathcal{P}}$. Let g_0 denote the discriminant of the quadratic field $\mathbb{Q}(\sqrt{g})$. Set

$$W := \operatorname{lcm}\Big[g_0, \prod_{p \le \log\log\log N} p\Big].$$

Once again, we pre-sieve values of n by putting n in an appropriate residue class $\nu \bmod W$. Whereas Maynard could choose any ν with $\gcd(\nu + h_i, W) = 1$ for all $1 \le i \le k$, we must tread more carefully. We choose ν so that the primes detected by the sieve are heavily biased towards having g as a primitive root.

Lemma 3.3. *We can choose an integer ν with all of the following properties*:

(i) $\nu + h_i$ *is coprime to W for all* $1 \le i \le k$.

(ii) $\nu + h_i - 1$ *is coprime to* $\prod_{2<p\le\log\log\log N} p$ *for all* $1 \le i \le k$.

(iii) *The Kronecker symbol* $\left(\frac{g_0}{\nu+h_i}\right)$ *equals* -1 *for all* $1 \le i \le k$.

Proof. Factor g_0 as a product $D_1 D_2 \dots D_\ell$ of coprime prime discriminants, where the *prime discriminants* are the numbers $-4, -8, 8$, and $(-1)^{(p-1)/2} p$ for odd primes p. Reordering the factorization if necessary, we can assume all of the following:

- If all $|D_i| \le K$ and g_0 is even, then $D_1 \in \{-4, -8, 8\}$.
- If all $|D_i| \le K$, g_0 is odd, and $\ell > 1$, then $|D_1| \ge 5$.
- If some $|D_i| > K$, then $|D_1| > K$.

We start by choosing any odd integer ν_1 that avoids the residue classes $-h_1, \dots, -h_k$, $1 - h_1, \dots, 1 - h_k$ modulo p for each odd prime $p \le \log\log\log N$ not dividing D_1. Note that when $p \le K$ the only requirement on ν_1 is that it avoids the residue classes 0 and 1 mod p, while when $p > K$ we are to avoid at most $2k$ of the $p > K > 2k$ residue classes modulo p. So such a choice of ν_1 certainly exists by the Chinese remainder theorem. We choose ν to satisfy

$$\nu \equiv \nu_1 \pmod{[W/D_1, 2]}.$$

To ensure (i), (ii), and (iii), it suffices to impose a further condition on ν guaranteeing

(i′) $\nu + h_i$ is coprime to all odd p dividing D_1 for all $1 \le i \le k$,

(ii′) $\nu + h_i - 1$ is coprime to all odd p dividing D_1 for all $1 \le i \le k$,

(iii′) $\left(\frac{D_1}{\nu+h_i}\right) = -\left(\frac{D_2 \cdots D_\ell}{\nu_1+h_i}\right)$ for all $1 \le i \le k$.

Notice that for all $1 \le i \le k$ we have $\left(\frac{D_2 \cdots D_\ell}{\nu_1+h_i}\right) \ne 0$, by the choice of ν_1.

Case I: all $|D_i| \le K$. In this case, (i′) and (ii′) are satisfied as long as $\nu \not\equiv 0$ or 1 (mod p) for any odd p dividing D_1, while (iii′) is satisfied as long as

$$\left(\frac{D_1}{\nu}\right) = -\left(\frac{D_2 \cdots D_\ell}{\nu_1}\right).$$

Assume first that g_0 is even. Then $D_1 \in \{-4, -8, 8\}$ and (i′) and (ii′) hold vacuously. Choose ν_2 so that $\left(\frac{D_1}{\nu_2}\right) = -\left(\frac{D_2 \cdots D_\ell}{\nu_1}\right)$. We ensure (iii′) by selecting ν as any solution to the simultaneous congruences

$$\nu \equiv \nu_1 \pmod{[W/D_1, 2]} \quad \text{and} \quad \nu \equiv \nu_2 \pmod{D_1}. \tag{3-3}$$

While the moduli here share a factor of 2, it is clear that these congruences still admit a simultaneous solution, since the only 2-adic information encoded by the first congruence is that ν is odd, which is certainly compatible with the second!

Now assume instead that g_0 is odd, so that $|D_1|$ is an odd prime. Either $|D_1| = 3$ and $\ell = 1$, or $|D_1| \ge 5$. If the former, then (i′), (ii′), and (iii′) hold upon selecting $\nu_2 = 2$ and choosing ν to satisfy (3-3). If the latter, choose $\nu_2 \not\equiv 1 \pmod{D_1}$ with $\left(\frac{D_1}{\nu_2}\right) = -\left(\frac{D_2 \cdots D_\ell}{\nu_1}\right)$ (possible since that equality of Kronecker symbols holds for a total of $\frac{1}{2}(|D_1| - 1) > 1$ residue classes ν_2 mod D_1). Once again, choosing ν to satisfy (3-3) completes the proof.

Case II: some $|D_i| > K$. In this case, $|D_1| > K$. Since $K > 8$, we see that $|D_1|$ is an odd prime. To satisfy (i′), (ii′), and (iii′), it suffices to show that there is an integer $\nu_2 \not\equiv 1 - h_1, \ldots, 1 - h_k \pmod{D_1}$ with

$$\left(\frac{\nu_2 + h_i}{|D_1|}\right) = -\left(\frac{D_2 \cdots D_\ell}{\nu_1 + h_i}\right) \quad \text{for all } 1 \le i \le k, \tag{3-4}$$

for then we can choose as ν any solution to (3-3). (We used here that $\left(\frac{D_1}{\nu + h_i}\right) = \left(\frac{\nu + h_i}{|D_1|}\right)$.) The integers $h_1, \ldots, h_k$ are incongruent modulo D_1, as each nonzero difference $h_j - h_i = (j - i)K!$ has only prime factors smaller than K. So Lemma 2.2 gives that the number of ν_2 mod D_1 satisfying (3-4) is at least $|D_1|/2^k - (k-1)\sqrt{|D_1|} - k$. Since $|D_1| > K = 9k^2 \cdot 4^k$, this count of solutions exceeds k. In particular, we can satisfy (3-4) with $\nu_2 \not\equiv 1 - h_1, \ldots, 1 - h_k \pmod{D_1}$. □

Assume that ν has been chosen to satisfy the conditions of Lemma 3.3. We let $R = N^\theta$, with θ to be specified momentarily, and we define the weights $w(n)$ exactly as in the statement of Proposition 3.1. We let

$$\tilde{S}_1 := \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod{W}}} w(n) \quad \text{and} \quad \tilde{S}_2 := \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod{W}}} \left(\sum_{i=1}^{k} \chi_{\tilde{\mathscr{P}}}(n + h_i)\right) w(n).$$

Theorem 1.1 is a consequence of the following result, established in the next section.

Proposition 3.4 (assuming GRH). *Fix a positive real number $\theta < \frac{1}{4}$. As $N \to \infty$, we have the same asymptotic estimates for $\tilde{S}_1$ and $\tilde{S}_2$ as those for S_1 and S_2 given in Proposition 3.1.*

Once Proposition 3.4 has been established, the earlier analysis we applied to Maynard's Proposition 3.1 applies, and we immediately obtain Theorem 1.1.

Proof of Proposition 3.4. The $\tilde{S}_1$ estimate is established in precisely the same way as Maynard's S_1 estimate in Proposition 3.1; see the proofs of Lemmas 5.1 and 6.2 in [Maynard 2013]. So we describe only the estimation of $\tilde{S}_2$. We write $\tilde{S}_2 = \sum_{m=1}^{k} \tilde{S}_2^{(m)}$, where

$$\tilde{S}_2^{(m)} := \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod W}} \chi_{\tilde{\mathscr{P}}}(n + h_m) w(n).$$

This is precisely analogous to Maynard's decomposition of S_2 as $\sum_{m=1}^{k} S_2^{(m)}$, where

$$S_2^{(m)} := \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod W}} \chi_{\mathscr{P}}(n + h_m) w(n).$$

Maynard's proof of Proposition 3.1 gives that each

$$S_2^{(m)} \sim \frac{\varphi(W)^k}{W^{k+1}} \frac{N}{\log N} (\log R)^{k+1} \cdot J_k^{(m)}(F).$$

So, to prove Proposition 3.4, it suffices to show that for each m we have

$$S_2^{(m)} - \tilde{S}_2^{(m)} = o\left(\frac{\varphi(W)^k}{W^{k+1}} N (\log N)^k \right) \tag{3-5}$$

as $N \to \infty$. From now on, we think of m as fixed, and we focus our energies on proving (3-5).

To prepare for the proof of (3-5), for each prime q we let $\mathscr{P}_q^{(0)}$ denote the set of all primes p satisfying

$$p \equiv 1 \pmod q \quad \text{and} \quad g^{(p-1)/q} \equiv 1 \pmod p. \tag{3-6}$$

Let

$$\mathscr{P}_q := \mathscr{P}_q^{(0)} \setminus \bigcup_{q' < q} \mathscr{P}_{q'}^{(0)}.$$

Provided that the argument is not a prime divisor of g,

$$0 \le \chi_{\mathscr{P}} - \chi_{\tilde{\mathscr{P}}} \le \sum_q \chi_{\mathscr{P}_q}. \tag{3-7}$$

Indeed, if p is a prime not dividing g, then either g is a primitive root mod p or g is a q-th power residue mod p for some prime q dividing $p-1$. From (3-7), it follows immediately that

$$0 \le S_2^{(m)} - \tilde{S}_2^{(m)} \le \sum_q \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod W}} \chi_{\mathscr{P}_q}(n+h_m) w(n). \tag{3-8}$$

We claim that the primes $q \le \log\log\log N$ make no contribution to the right-hand side of (3-8). Indeed, suppose $p := n + h_m$ is prime with $N \le n < 2N$ and $n \equiv \nu \pmod W$. By Lemma 3.3(ii), the number $p-1$ has no odd prime factors up to $\log\log\log N$; it follows trivially that $\chi_{\mathscr{P}_q}(p) = 0$ for odd $q \le \log\log\log N$. By Lemma 3.3(iii), $\chi_{\mathscr{P}_2}(p) = 0$, since, modulo p,

$$g^{(p-1)/2} \equiv \left(\frac{g}{p}\right) = \left(\frac{g}{n+h_m}\right) = \left(\frac{g_0}{n+h_m}\right) = -1.$$

Thus, the right-hand side of (3-8) can be rewritten as $\Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4$, where each Σ_i represents a partial sum of (3-8) over values of q in the following ranges:

Σ_1: $\log\log\log N < q \le (\log N)^{100k}$,

Σ_2: $(\log N)^{100k} < q \le N^{1/2}(\log N)^{-100k}$,

Σ_3: $N^{1/2}(\log N)^{-100k} < q \le N^{1/2}(\log N)^{100k}$,

Σ_4: $q > N^{1/2}(\log N)^{100k}$.

We treat these ranges of q separately.

Estimation of Σ_2 *and* Σ_4. We need the following lemma, which facilitates later applications of Cauchy–Schwarz.

Lemma 3.5.
$$\sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod W}} w(n)^2 \ll \frac{N}{W} (\log R)^{19k}.$$

Proof. Let $\boldsymbol{d} = (d_1, \dots, d_k)$, $\boldsymbol{e} = (e_1, \dots, e_k)$, $\boldsymbol{f} = (f_1, \dots, f_k)$, and $\boldsymbol{g} = (g_1, \dots, g_k)$ represent k-tuples of positive integers. Expanding the sum using the definition of $w(n)$ gives

$$\sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod W}} \sum_{\substack{\boldsymbol{d},\boldsymbol{e},\boldsymbol{f},\boldsymbol{g} \\ [d_i,e_i,f_i,g_i] \mid n+h_i \ \forall i}} \lambda_{\boldsymbol{d}} \lambda_{\boldsymbol{e}} \lambda_{\boldsymbol{f}} \lambda_{\boldsymbol{g}} = \sum_{\boldsymbol{d},\boldsymbol{e},\boldsymbol{f},\boldsymbol{g}} \lambda_{\boldsymbol{d}} \lambda_{\boldsymbol{e}} \lambda_{\boldsymbol{f}} \lambda_{\boldsymbol{g}} \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod W \\ [d_i,e_i,f_i,g_i] \mid n+h_i \ \forall i}} 1.$$

Remembering that $\lambda_{d_1,\dots,d_k}$ vanishes unless $d_1 \cdots d_k$ is prime to W, we see that a quadruple $\boldsymbol{d}$, $\boldsymbol{e}$, $\boldsymbol{f}$, $\boldsymbol{g}$ makes no contribution to the right-hand side unless the numbers $[d_i, e_i, f_i, g_i]$, for $1 \le i \le k$, are pairwise coprime and all coprime to W. In that case, the conditions on n in the inner sum put n in a uniquely determined congruence class modulo $W \prod_{i=1}^k [d_i, e_i, f_i, g_i]$. It follows that our sum is bounded above by

$$\sum_{\boldsymbol{d},\boldsymbol{e},\boldsymbol{f},\boldsymbol{g}} |\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}\lambda_{\boldsymbol{f}}\lambda_{\boldsymbol{g}}| \left(\frac{N}{W\prod_{i=1}^{k}[d_i, e_i, f_i, g_i]} + 1 \right).$$

Let

$$r := \prod_{i=1}^{k} [d_i, e_i, f_i, g_i]. \tag{3-9}$$

Since $\lambda_{d_1,\dots,d_k}$ vanishes unless $d_1 \cdots d_k$ is a squarefree integer smaller than R, we may restrict attention to squarefree $r < R^4$. Given r, there are $\tau_{15k}(r)$ choices of $\boldsymbol{d}$, $\boldsymbol{e}$, $\boldsymbol{f}$, and $\boldsymbol{g}$ giving (3-9). Hence, writing $\lambda_{\max} = \max_{d_1,\dots,d_k} |\lambda_{d_1,\dots,d_k}|$, we find that

$$\sum_{\boldsymbol{d},\boldsymbol{e},\boldsymbol{f},\boldsymbol{g}} |\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}\lambda_{\boldsymbol{f}}\lambda_{\boldsymbol{g}}| \left(\frac{N}{W\prod_{i=1}^{k}[d_i, e_i, f_i, g_i]} + 1 \right)$$
$$\le \lambda_{\max}^4 \sum_{r<R^4} \mu^2(r)\tau_{15k}(r) \left(\frac{N}{Wr} + 1 \right) \le \lambda_{\max}^4 \left(\frac{N}{W} + R^4 \right) \sum_{r<R^4} \frac{\mu^2(r)\tau_{15k}(r)}{r}. \tag{3-10}$$

The remaining sum on r is bounded above by $\prod_{p<R^4}(1 + 15k/p) \ll (\log R)^{15k}$. Since $R = N^{\theta}$ with $\theta < \frac{1}{4}$ fixed, we get that $R^4 \ll N/W$. Finally, we note that $\lambda_{\max} \ll (\log R)^k$ (see [Maynard 2013, equations (5.9) and (6.3)], and recall that our implied constants may depend on F). Inserting these estimates into (3-10) gives the lemma. □

Proof that $\Sigma_2 = o\big((\varphi(W)^k/W^{k+1})N(\log N)^k\big)$. Let $\mathscr{Q}$ be the union of the sets $\mathscr{P}_q$ for $(\log N)^{100k} < q \le N^{1/2}(\log N)^{-100k}$. Then

$$\Sigma_2 = \sum_{\substack{N \le n < 2N \\ n \equiv \nu \ (\mathrm{mod}\ W)}} \chi_{\mathscr{Q}}(n + h_m) w(n).$$

Applying Cauchy–Schwarz and Lemma 3.5, we see that

$$\Sigma_2 \ll W^{-1/2} N^{1/2} (\log R)^{9.5k} \Bigg(\sum_{\substack{N \le n < 2N \\ n \equiv \nu \ (\mathrm{mod}\ W)}} \chi_{\mathscr{Q}}(n + h_m) \Bigg)^{1/2}. \tag{3-11}$$

The remaining sum on n is certainly bounded above by the total number of primes $p \in [N, 3N]$ belonging to $\mathscr{Q}$. For each such p, we may select a q with $(\log N)^{100k} < q \le N^{1/2}(\log N)^{-100k}$ for which (3-6) holds. Given q, we count the number of corresponding p using effective Chebotarev.

Since g is fixed and q is large, we see that $g \notin (\mathbb{Q}^{\times})^q$. So, by a theorem of Capelli on irreducible binomials, the extension $\mathbb{Q}(\sqrt[q]{g})/\mathbb{Q}$ has degree q. For later use, we note that the discriminant of $\mathbb{Q}(\sqrt[q]{g})$ divides $(gq)^q$ — so the only ramified primes divide gq. By a theorem of Dedekind and Kummer, a prime $p \in [N, 3N]$ satisfies (3-6) precisely when p splits completely in $L := \mathbb{Q}(\zeta_q, \sqrt[q]{g})$. To continue, we need to know the degree of $L/\mathbb{Q}$. Now $\sqrt[q]{g}$ is not contained in $\mathbb{Q}(\zeta_q)$ — otherwise, $\sqrt[q]{g}$ would generate a Galois extension of $\mathbb{Q}$, contradicting that $\mathbb{Q}(\sqrt[q]{g})$ contains only a single q-th root of unity (since it can be viewed as a subfield of $\mathbb{R}$). So, by another application of Capelli's theorem,

$$[L : \mathbb{Q}] = [L : \mathbb{Q}(\zeta_q)] \cdot [\mathbb{Q}(\zeta_q) : \mathbb{Q}] = q(q-1).$$

Moreover, since q is the only ramified prime in $\mathbb{Q}(\zeta_q)/\mathbb{Q}$, the only primes that may ramify in $L/\mathbb{Q}$ all divide gq. By Lemma 2.4, $\log|\Delta_L| \ll q^2 \log(|g|q) \ll q^2 \log N$. We plug this estimate into Theorem 2.3, taking C as the conjugacy class of the identity. We find that the number of $p \in [N, 3N]$ for which (3-6) holds for a given q is

$$\frac{1}{q(q-1)} \int_N^{3N} \frac{dt}{\log t} + O(N^{1/2} \log N).$$

Summing this upper bound over primes q with $(\log N)^{100k} < q \le N^{1/2}(\log N)^{-100k}$, we get that the total number of these p is $O(N(\log N)^{-100k})$.

Now, referring back to (3-11), we see that $\Sigma_2 \ll W^{-1/2} N(\log N)^{-40k}$. But this is $o(N)$, and so certainly also $o\big((\varphi(W)^k/W^{k+1})N(\log N)^k\big)$. □

Proof that $\Sigma_4 = o\big((\varphi(W)^k/W^{k+1})N(\log N)^k\big)$. We proceed as above, but now with $\mathscr{Q}$ equal to the union of the sets $\mathscr{P}_q$ for $q > N^{1/2}(\log N)^{100k}$. We will show that $\#\mathscr{Q} \cap [N, 3N] \ll N(\log N)^{-200k}$. By the previous Cauchy–Schwarz argument, this is (more than) enough. If $p \in \mathscr{Q} \cap [N, 3N]$, then the order of g modulo p, call it ℓ, divides $(p-1)/q$ for some $q > N^{1/2}(\log N)^{100k}$. In particular, $\ell < 3N^{1/2}(\log N)^{-100k}$. Since $g^\ell - 1$ has only $O(\ell)$ prime factors, summing on $\ell < 3N^{1/2}(\log N)^{-100k}$ shows there are $O(N(\log N)^{-200k})$ possibilities for p. □

Estimation of Σ_3. For each prime q, we let $\mathscr{A}_q$ denote the set of natural numbers $n \equiv 1 \pmod q$. We estimate Σ_3 using the trivial bound $\chi_{\mathscr{P}_q} \le \chi_{\mathscr{A}_q}$. To save space, write $\mathscr{I} := (N^{1/2}(\log N)^{-100k}, N^{1/2}(\log N)^{100k}]$. Then

$$\Sigma_3 \le \sum_{q \in \mathscr{I}} \sum_{\substack{N \le n < 2N \\ n \equiv \nu \ (\mathrm{mod}\ W)}} \chi_{\mathscr{A}_q}(n + h_m) w(n).$$

Expanding out the right-hand side yields

$$\sum_{q \in \mathscr{I}} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k} \sum_{\substack{N \le n < 2N \\ n \equiv \nu \ (\mathrm{mod}\ W) \\ [d_i, e_i] \mid n + h_i \ \forall i}} \chi_{\mathscr{A}_q}(n + h_m). \tag{3-12}$$

We can assume $d_1 \cdots d_k$ is a squarefree integer coprime to W and not exceeding R, since otherwise $\lambda_{d_1,\dots,d_k} = 0$. A similar assumption can be made for $e_1 \cdots e_k$. Since

$q \in \mathcal{J}$, it follows that q is coprime to each d_i and each e_i, and W. Now the innermost sum in (3-12) vanishes unless $[d_1, e_1], [d_2, e_2], \ldots, [d_k, e_k]$, and W are pairwise coprime. Using a $'$ to denote this restriction on the d_i and e_i, we get that

$$\sum_{q\in\mathcal{J}} \sum_{\substack{d_1,\ldots,d_k\\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k} \sum_{\substack{N\le n<2N\\ n\equiv\nu \pmod W\\ [d_i,e_i]\mid n+h_i\ \forall i}} \chi_{\mathscr{A}_q}(n+h_m)$$
$$= \sum_{q\in\mathcal{J}} \sum_{\substack{d_1,\ldots,d_k\\ e_1,\ldots,e_k}}{}' \lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}\left(\frac{N}{qW\prod_{i=1}^k[d_i,e_i]} + O(1)\right).$$

The error here is

$$\ll \left(\sum_{q\in\mathcal{J}} 1\right)\left(\sum_{d_1,\ldots,d_k} |\lambda_{d_1,\ldots,d_k}|\right)^2 \ll N^{1/2}(\log N)^{100k}\, \lambda_{\max}^2 \left(\sum_{r<R}\mu^2(r)\tau_k(r)\right)^2.$$

Recalling that $\lambda_{\max} \ll (\log R)^k$ and that $\sum_{r<R}\tau_k(r) \ll R(\log R)^{k-1}$, our final O error term is $O(N^{1/2}R^2\cdot(\log N)^{104k})$. Since $R = N^\theta$ with $\theta < \frac{1}{4}$, this error is $o(N)$ and so is negligible for us.

We now turn to the main term, which has the form

$$\left(\sum_{q\in\mathcal{J}} \frac{1}{q}\right)\left(\frac{N}{W}\sum_{\substack{d_1,\ldots,d_k\\ e_1,\ldots,e_k}}{}' \frac{\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}}{\prod_{i=1}^k[d_i,e_i]}\right).$$

The first factor here is $O(\log\log N/\log N)$, and so in particular is $o(1)$. Maynard's analysis (see the proofs of [Maynard 2013, Lemmas 5.1, 6.2]) shows that the second factor here satisfies the asymptotic formula asserted for S_1 in Proposition 3.1. Hence, $\Sigma_3 = o\big((\varphi(W)^k/W^{k+1})N(\log N)^k\big)$, as desired.

Estimation of Σ_1. For this case, let $\mathcal{J} := (\log\log\log N, (\log N)^{100k}]$. Using the bound $\chi_{\mathscr{P}_q} \le \chi_{\mathscr{P}_q^{(0)}}$, we get that

$$\Sigma_1 \le \sum_{q\in\mathcal{J}} \sum_{N\le n<2N} \chi_{\mathscr{P}_q^{(0)}}(n+h_m)w(n).$$

Expanding out the right-hand side gives

$$\sum_{q\in\mathcal{J}} \sum_{\substack{d_1,\ldots,d_k\\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k} \sum_{\substack{N\le n<2N\\ n\equiv\nu \pmod W\\ [d_i,e_i]\mid n+h_i\ \forall i}} \chi_{\mathscr{P}_q^{(0)}}(n+h_m). \tag{3-13}$$

The inner sum can be written as a sum over a single residue class modulo $f := W\prod_{i=1}^k[d_i,e_i]$, provided that $W, [d_1,e_1], \ldots, [d_k,e_k]$ are pairwise coprime; otherwise we get no contribution. We also need that $n+h_m$ lies in a residue class coprime to f, which happens precisely when $d_m = e_m = 1$. Also, $\chi_{\mathscr{P}_q^{(0)}}(n+h_m)$

vanishes unless $q \mid n+h_m-1$, and this implies that the inner sum in (3-13) vanishes unless q is coprime to each d_i and e_i. Indeed, if q divides d_i or e_i without the inner sum vanishing, then $q \mid h_m - h_i - 1$. But that divisibility cannot hold for $q \in \mathcal{I}$, since $0 < |h_m - h_i - 1| < k \cdot K!$.

Thus, we only see a contribution to (3-13) if $[d_1, e_1], [d_2, e_2], \ldots, [d_k, e_k], W$, and q are pairwise coprime. Under these conditions, we claim that

$$\sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod{W} \\ [d_i, e_i] \mid n+h_i \ \forall i}} \chi_{\mathscr{P}_q^{(0)}}(n+h_m) \\ = \frac{1}{q(q-1)\varphi(W)\prod_{i=1}^{k}\varphi([d_i, e_i])} \int_{N+h_m}^{2N+h_m} \frac{dt}{\log t} + O(N^{1/2}\log N). \quad (3\text{-}14)$$

To see this, let $p := n + h_m$. Then the prime $p \in [N + h_m, 2N + h_m)$ makes a contribution to the left-hand sum precisely when Frob_p is a certain element of $\mathrm{Gal}(\mathbb{Q}(\zeta_f)/\mathbb{Q})$ — determined by the congruence conditions modulo the $[d_i, e_i]$ and W — *and* when p splits completely in $\mathbb{Q}(\zeta_q, \sqrt[q]{g})$. Now $\mathbb{Q}(\sqrt[q]{g}) \not\subset \mathbb{Q}(\zeta_{qf})$, since $\mathbb{Q}(\sqrt[q]{g})$ is not a Galois extension of $\mathbb{Q}$. Thus, letting $L := \mathbb{Q}(\zeta_{qf}, \sqrt[q]{g})$, we find that

$$[L : \mathbb{Q}] = [L : \mathbb{Q}(\zeta_{qf})][\mathbb{Q}(\zeta_{qf}) : \mathbb{Q}] = q \cdot \varphi(qf) = q(q-1)\varphi(W)\prod_{i=1}^{k}\varphi([d_i, e_i]).$$

Hence, $\mathbb{Q}(\zeta_f)$ and $\mathbb{Q}(\zeta_q, \sqrt[q]{g})$ are linearly disjoint extensions of $\mathbb{Q}$ with compositum L. Our conditions on p amount to placing Frob_p in a certain uniquely determined conjugacy class of size 1 in $\mathrm{Gal}(L/\mathbb{Q})$. Since the only primes that ramify in L divide qfg, Lemma 2.4 gives that

$$\log|\Delta_L| \ll [L : \mathbb{Q}](\log(qfg) + \log[L : \mathbb{Q}]) \ll [L : \mathbb{Q}]\log N.$$

Inserting this estimate into Theorem 2.3 yields (3-14).

Returning now to (3-13), we see that the error term in (3-14) yields a total error of size

$$\ll N^{1/2}\log N\Bigl(\sum_{q \in \mathcal{I}} 1\Bigr)\Bigl(\sum_{d_1,\ldots,d_k} |\lambda_{d_1,\ldots,d_k}|\Bigr)^2 \\ \ll N^{1/2}(\log N)^{100k+1}\lambda_{\max}^2\Bigl(\sum_{r<R}\tau_k(r)\Bigr)^2 \ll N^{1/2}R^2(\log N)^{104k+1}.$$

This is $o(N)$ and so is again negligible for us. Letting

$$X_N := \int_{N+h_m}^{2N+h_m} \frac{dt}{\log t},$$

the main term has the shape

$$\sum_{q\in\mathcal{I}}\frac{1}{q(q-1)}\left(\frac{X_N}{\varphi(W)}\sideset{}{'}\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ d_m=e_m=1}}\frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k\varphi([d_i,e_i])}\right). \tag{3-15}$$

Here the $'$ on the sum indicates that W, $[d_1,e_1],\dots,[d_k,e_k]$, and q are pairwise coprime. Owing to the support of the λ's, this restriction on the sum has the same effect as requiring that $(d_i,e_j)=1$ for all $i\neq j$ and that $(d_i,q)=(e_j,q)=1$ for all $1\le i,j\le k$. We incorporate the restrictions that $(d_i,e_j)=1$ by multiplying through by $\sum_{s_{i,j}\mid d_i,e_j}\mu(s_{i,j})$ for $i\neq j$. Similarly, we incorporate the restrictions that $(d_i,q)=(e_j,q)=1$ by multiplying through by $\sum_{\delta_i\mid d_i,q}\mu(\delta_i)$ and $\sum_{\epsilon_j\mid e_j,q}\mu(\epsilon_j)$, for all pairs of i and j.

Let g be the completely multiplicative function defined by $g(p)=p-2$ for all primes p, and note that

$$\frac{1}{\varphi([d_i,e_i])}=\frac{1}{\varphi(d_i)\varphi(e_i)}\sum_{u_i\mid d_i,e_i}g(u_i)$$

for squarefree d_i and e_i. This allows us to rewrite the parenthesized portion of (3-15) as

$$\frac{X_N}{\varphi(W)}\sum_{u_1,\dots,u_k}\left(\prod_{i=1}^k g(u_i)\right)\sideset{}{^*}\sum_{s_{1,2},\dots,s_{k,k-1}}\left(\prod_{\substack{1\le i,j\le k\\ i\neq j}}\mu(s_{i,j})\right)\sum_{\substack{\delta_1,\dots,\delta_k\mid q\\ \epsilon_1,\dots,\epsilon_k\mid q}}\left(\prod_{i=1}^k\mu(\delta_i)\prod_{j=1}^k\mu(\epsilon_j)\right)$$
$$\times\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ u_i\mid d_i,e_i\ \forall i\\ s_{i,j}\mid d_i,e_j\ \forall i\neq j\\ \delta_i\mid d_i,\epsilon_j\mid e_j\ \forall i,j\\ d_m=e_m=1}}\frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k\varphi(d_i)\varphi(e_i)}, \tag{3-16}$$

where the $*$ on the sum indicates that $s_{i,j}$ is restricted to be coprime to u_i, u_j, $s_{i,a}$, and $s_{b,j}$ for all $a\neq j$ and $b\neq i$. (The other values of $s_{i,j}$ make no contribution.) Introducing the new variables

$$y^{(m)}_{r_1,\dots,r_k}:=\left(\prod_{i=1}^k\mu(r_i)g(r_i)\right)\sum_{\substack{d_1,\dots,d_k\\ r_i\mid d_i\ \forall i\\ d_m=1}}\frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k\varphi(d_i)},$$

we may rewrite (3-16) as

$$\frac{X_N}{\varphi(W)}\sum_{u_1,\dots,u_k}\Bigl(\prod_{i=1}^{k}g(u_i)\Bigr)\sideset{}{^*}\sum_{s_{1,2},\dots,s_{k,k-1}}\Bigl(\prod_{\substack{1\le i,j\le k\\ i\ne j}}\mu(s_{i,j})\Bigr)\sum_{\substack{\delta_1,\dots,\delta_k\mid q\\ \epsilon_1,\dots,\epsilon_k\mid q}}\Bigl(\prod_{i=1}^{k}\mu(\delta_i)\prod_{j=1}^{k}\mu(\epsilon_j)\Bigr)$$
$$\times\Bigl(\prod_{i=1}^{k}\frac{\mu(a_i)}{g(a_i)}\Bigr)\Bigl(\prod_{j=1}^{k}\frac{\mu(b_j)}{g(b_j)}\Bigr)y^{(m)}_{a_1,\dots,a_k}y^{(m)}_{b_1,\dots,b_k},$$

where $a_i=\operatorname{lcm}\bigl[u_i\prod_{j\ne i}s_{i,j},\delta_i\bigr]$ and $b_j=\operatorname{lcm}\bigl[u_j\prod_{i\ne j}s_{i,j},\epsilon_j\bigr]$. Define $\delta_i'\in\{1,q\}$ and $\epsilon_j'\in\{1,q\}$ by the equations

$$a_i=\Bigl(u_i\prod_{j\ne i}s_{i,j}\Bigr)\delta_i',\qquad b_j=\Bigl(u_j\prod_{i\ne j}s_{i,j}\Bigr)\epsilon_j'.$$

Exploiting coprimality, we can write $\mu(a_i)=\bigl(\mu(u_i)\prod_{j\ne i}\mu(s_{i,j})\bigr)\mu(\delta_i')$, and similarly for $\mu(b_j)$, $g(a_i)$, and $g(b_j)$. This transforms (3-16) into

$$\frac{X_N}{\varphi(W)}\sum_{u_1,\dots,u_k}\Bigl(\prod_{i=1}^{k}\frac{\mu(u_i)^2}{g(u_i)}\Bigr)\sideset{}{^*}\sum_{s_{1,2},\dots,s_{k,k-1}}\Bigl(\prod_{\substack{1\le i,j\le k\\ i\ne j}}\frac{\mu(s_{i,j})}{g(s_{i,j})^2}\Bigr)$$
$$\times\sum_{\substack{\delta_1,\dots,\delta_k\mid q\\ \epsilon_1,\dots,\epsilon_k\mid q}}\Bigl(\prod_{i=1}^{k}\frac{\mu(\delta_i)\mu(\delta_i')}{g(\delta_i')}\prod_{j=1}^{k}\frac{\mu(\epsilon_j)\mu(\epsilon_j')}{g(\epsilon_j')}\Bigr)y^{(m)}_{a_1,\dots,a_k}y^{(m)}_{b_1,\dots,b_k}.$$

Let $y^{(m)}_{\max}=\max_{r_1,\dots,r_k}\bigl|y^{(m)}_{r_1,\dots,r_k}\bigr|$. From [Maynard 2013, equation (6.10)], we have $y^{(m)}_{\max}\ll(\varphi(W)/W)\log R$. Inserting these bounds into the previous display, we find that (3-16) is

$$\ll\frac{X_N}{\varphi(W)}\Bigl(\sum_{\substack{u<R\\ \gcd(u,W)=1}}\frac{\mu(u)^2}{g(u)}\Bigr)^{k-1}\Bigl(\sum_{s}\frac{\mu(s)^2}{g(s)^2}\Bigr)^{k(k-1)}y^{(m)\,2}_{\max}$$
$$\ll\frac{X_N}{\varphi(W)}\Bigl(\frac{\varphi(W)}{W}\Bigr)^{k+1}(\log R)^{k+1}\ll\Bigl(\frac{\varphi(W)^k}{W^{k+1}}\Bigr)N(\log N)^k.$$

We used here that there are only $O(1)$ possibilities for the δ_i and ϵ_j, and that for each of these, $\prod_i(1/g(\delta_i'))\prod_j(1/g(\epsilon_j'))\le 1$. Referring back to (3-15), we see that our original main term contributes

$$\ll\Bigl(\frac{\varphi(W)^k}{W^{k+1}}\Bigr)N(\log N)^k\sum_{q\in\mathcal{J}}\frac{1}{q(q-1)}=o\Bigl(\frac{\varphi(W)^k}{W^{k+1}}N(\log N)^k\Bigr),$$

as desired.

Remark. The truth of Theorem 1.1 could also have been predicted on heuristic grounds. Indeed, there are well-known heuristics for Artin's primitive root conjecture, suggesting even the "correct" value of c_g (see [Moree 2012, §§2–5]), as well as heuristics for the prime k-tuples conjecture (see, for instance, [Crandall and Pomerance 2005, pp. 14–15]), and these can be fitted together. As an example, this combined heuristic suggests that the count of twin prime pairs p, $p+2$ with $p \le x$ and with 2 a primitive root of both p and $p+2$ should be approximately

$$\mathfrak{S} \int_2^x \frac{dt}{(\log t)^2}, \quad \text{where } \mathfrak{S} := \frac{1}{4} \prod_{p>3} \left(1 - \frac{3}{(p-1)^2}\right).$$

Quantitative conjectures of this kind, but in the context of primes represented by a single irreducible polynomial rather than primes produced by linear forms, appear in recent work of Moree [2007] and of Akbary and Scholten [2013].

4. Concluding remarks

We conclude with a proof of the following result, which seems of independent interest:

Theorem 4.1 (conditional on GRH). *Fix an integer $g \neq -1$ and not a square. For every positive integer m, there are m consecutive primes all of which possess g as a primitive root.*

Theorem 4.1 might be compared with Shiu's celebrated result [2000] that each coprime residue class $a \bmod q$ contains arbitrarily long runs of consecutive primes. Our proof of Theorem 4.1 is similar in spirit to a short proof of Shiu's theorem recently given by Banks, Freiberg, and Turnage-Butterbaugh [Banks et al. 2013].

It will be useful to first translate the proof of Theorem 1.1 into probabilistic terms. Let k be a fixed positive integer, and let $h_1, \dots, h_k$ be given by (3-2). We view the set of $n \in [N, 2N)$ with $n \equiv \nu \pmod{W}$ as a finite probability space where the probability mass at each n_0 is given by

$$w(n_0) \Big/ \sum_{\substack{N \le n < 2N \\ n \equiv \nu \pmod{W}}} w(n).$$

Here the weights $w(n)$ are assumed to be of the form specified in Proposition 3.1. Introduce the random variables

$$X := \sum_{i=1}^{k} \chi_{\mathscr{P}}(n+h_i) \quad \text{and} \quad Y := \sum_{i=1}^{k} \chi_{\mathscr{P} \setminus \tilde{\mathscr{P}}}(n+h_i).$$

Then $\mathbb{E}[X] = S_2/S_1$. Given suitable parameters F and θ, Proposition 3.1 gives us the limiting value of $\mathbb{E}[X]$ as $N \to \infty$. Combining Propositions 3.1 and 3.2, we see that for k large enough in terms of m, we can choose parameters so this limiting value exceeds $m-1$. On the other hand, it was shown in Section 3 that (with the same choice of parameters) $\mathbb{E}[Y] = o(1)$ as $N \to \infty$. Thus, $\mathbb{E}[X - Y] > m - 1$ for all large N. But $X - Y = \sum_{i=1}^{m} \chi_{\widetilde{\mathscr{P}}}(n + h_i)$. Hence, for some $n \in [N, 2N)$, the list $n + h_1, \dots, n + h_k$ contains at least m primes having g as a primitive root. Theorem 1.1 follows, with $C_m = h_k - h_1$.

We now present the minor variation of this argument needed to establish Theorem 4.1.

Proof of Theorem 4.1. Given m, we fix a large enough value of k (and parameters F and θ) so that the limiting value of $\mathbb{E}[X]$ exceeds $m-1$. Then, for all large N,

$$\Pr(X \geq m) \geq \mathbb{E}\left[\frac{X-(m-1)}{k}\right] = \frac{1}{k}(\mathbb{E}[X] - (m-1)) \gg 1.$$

Note that $\Pr(Y > 0) \leq \mathbb{E}[Y] = o(1)$, as $N \to \infty$. So, for large N, there is a positive probability that both $X \geq m$ and $Y = 0$. This allows us to select $n \in [N, 2N)$ with $n \equiv \nu \pmod{W}$ satisfying

(i) at least m of $n + h_1, \dots, n + h_k$ are prime,

(ii) all of the primes among $n + h_1, \dots, n + h_k$ possess g as a primitive root.

We will argue momentarily that we can also assume

(iii) the only primes in the interval $[n + h_1, n + h_k]$ are the primes in the list $n + h_1, \dots, n + h_k$.

From (i), (ii), and (iii), we see that the set of primes in $[n + h_1, n + h_k]$ contains at least m elements, all of which have g as a primitive root. Theorem 4.1 follows.

In order to show we may assume (iii), we tweak the choice of the residue class ν mod W from which n is sampled. In the proof of Lemma 3.3, we chose ν_1 as any odd integer avoiding $-h_1, \dots, -h_k, 1 - h_1, \dots, 1 - h_k$ modulo p, for all odd $p \leq \log\log\log N$ not dividing D_1. We now add an extra condition on ν_1. Choose distinct primes $p^{(h)} \in [\frac{1}{2}\log\log\log N, \log\log\log N)$ for all even $h \in [h_1, h_k] \setminus \mathscr{H}$. We add the requirement that $\nu_1 \equiv -h \pmod{p^{(h)}}$ for each such h. This is consistent with our earlier restrictions, since h is not congruent modulo $p^{(h)}$ to any of $h_1, \dots, h_k$ (since $h \notin \mathscr{H}$) or to any of $h_1 - 1, \dots, h_k - 1$ (since h and the h_i are all even). Using the resulting value of ν from Lemma 3.3, we see that for even $h \in [h_1, h_k] \setminus \mathscr{H}$, we have $p_h \mid n + h$ whenever $n \equiv \nu \pmod{W}$. For all odd $h \in [h_1, h_k]$, we have trivially that $2 \mid n + h$ whenever $n \equiv \nu \pmod{W}$. Thus, $n + h$ is composite if $h \in [h_1, h_k] \setminus \mathscr{H}$, and so (iii) holds. $\square$

Acknowledgement

The author thanks the referee for a careful reading of the manuscript.

References

[Akbary and Scholten 2013] A. Akbary and K. Scholten, "Artin prime producing polynomials", preprint, 2013. To appear in *Math. Comp.* arXiv 1310.5198

[Banks et al. 2013] W. D. Banks, T. Freiberg, and C. L. Turnage-Butterbaugh, "Consecutive primes in tuples", preprint, 2013. To appear in *Acta Arith.* arXiv 1311.7003

[Crandall and Pomerance 2005] R. Crandall and C. Pomerance, *Prime numbers: A computational perspective*, 2nd ed., Springer, New York, 2005. MR 2006a:11005 Zbl 1088.11001

[Gupta and Murty 1984] R. Gupta and M. R. Murty, "A remark on Artin's conjecture", *Invent. Math.* **78**:1 (1984), 127–130. MR 86d:11003 Zbl 0549.10037

[Heath-Brown 1986] D. R. Heath-Brown, "Artin's conjecture for primitive roots", *Quart. J. Math. Oxford Ser.* (2) **37**:145 (1986), 27–38. MR 88a:11004 Zbl 0586.10025

[Hooley 1967] C. Hooley, "On Artin's conjecture", *J. Reine Angew. Math.* **225** (1967), 209–220. MR 34 #7445 Zbl 0221.10048

[Hooley 1976] C. Hooley, *Applications of sieve methods to the theory of numbers*, Cambridge Tracts in Mathematics **70**, Cambridge University Press, 1976. MR 53 #7976 Zbl 0327.10044

[Lagarias and Odlyzko 1977] J. C. Lagarias and A. M. Odlyzko, "Effective versions of the Chebotarev density theorem", pp. 409–464 in *Algebraic number fields: L-functions and Galois properties* (Durham, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR 56 #5506 Zbl 0362.12011

[Li and Pan 2014] H. Li and H. Pan, "Bounded gaps between primes of the special form", preprint, 2014. arXiv 1403.4527

[Maynard 2013] J. Maynard, "Small gaps between primes", preprint, 2013. To appear in *Ann. Math.* arXiv 1311.4600v2

[Moree 2007] P. Moree, "Artin prime producing quadratics", *Abh. Math. Sem. Univ. Hamburg* **77** (2007), 109–127. MR 2008m:11194 Zbl 1214.11107

[Moree 2012] P. Moree, "Artin's primitive root conjecture—a survey", *Integers* **12**:6 (2012), 1305–1416. MR 3011564 Zbl 1271.11002

[Serre 1981] J.-P. Serre, "Quelques applications du théorème de densité de Chebotarev", *Inst. Hautes Études Sci. Publ. Math.* 54 (1981), 323–401. MR 83k:12011 Zbl 0496.12011

[Shiu 2000] D. K. L. Shiu, "Strings of congruent primes", *J. London Math. Soc.* (2) **61**:2 (2000), 359–373. MR 2001f:11155 Zbl 0973.11083

[Thorner 2014] J. Thorner, "Bounded gaps between primes in Chebotarev sets", *Res. Math. Sci.* **1**:4 (2014), 1–16.

[Wan 1997] D. Wan, "Generators and irreducible polynomials over finite fields", *Math. Comp.* **66**:219 (1997), 1195–1212. MR 97j:11060 Zbl 0879.11072

Communicated by Andrew Granville
Received 2014-04-27 Revised 2014-06-21 Accepted 2014-07-19

pollack@uga.edu *Department of Mathematics, University of Georgia, Boyd Graduate Studies Building, Athens, GA 30602, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 8 No. 7 2014