Algebra & Number Theory

Volume 9
2015
No. 9

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2015 is US $255/year for the electronic version, and $440/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 9:9 (2015)
dx.doi.org/10.2140/ant.2015.9.1955

Families of nearly ordinary Eisenstein series on unitary groups

Xin Wan

With an appendix by Kai-Wen Lan

We use the doubling method to construct p-adic L-functions and families of nearly ordinary Klingen Eisenstein series from nearly ordinary cusp forms on unitary groups of signature (r, s) and Hecke characters, and prove the constant terms of these Eisenstein series are divisible by the p-adic L-function, following earlier constructions of Eischen, Harris, Li, Skinner and Urban. We also make preliminary computations for the Fourier–Jacobi coefficients of the Eisenstein series. This provides a framework to do Iwasawa theory for cusp forms on unitary groups.

A list of symbols can be found on page 2034.

1. Introduction

Let p be an odd prime. Let $\mathcal{K}$ be a CM field with the maximal totally real subfield F such that $[F : \mathbb{Q}] = d$. Suppose p is totally split at $\mathcal{K}$. We fix an isomorphism $\iota_p := \mathbb{C}_p \simeq \mathbb{C}$ and a CM type Σ_∞, which means a set of d different embeddings $\mathcal{K} \to \mathbb{C}$ such that $\Sigma_\infty \cup \Sigma_\infty^c$, where c means complex conjugation, is the set of all embeddings of $\mathcal{K}$ into $\mathbb{C}$. This determines a set of embeddings $\mathcal{K} \hookrightarrow \mathbb{C}_p$ using ι_p,

MSC2010: 11R23.
Keywords: Iwasawa theory, ordinary, Klingen Eisenstein series, unitary groups, p-adic L-function,

which we denote by Σ_p. Let $r \geq s \geq 0$ be integers. We often write $a = r - s$ and $b = s$. Let $\mathrm{U}(r, s)$ be the unitary group associated to the skew-Hermitian matrix

$$\begin{pmatrix} & & 1_b \\ & \zeta & \\ -1_b & & \end{pmatrix},$$

where ζ is a diagonal matrix such that $i^{-1}\zeta$ is positive definite.

Eischen et al. [≥ 2015] constructed the p-adic L-function for an irreducible cuspidal automorphic representation of $\mathrm{U}(r, s)$ that is nearly ordinary at all primes dividing p, which interpolates (the algebraic part of) critical values of the standard L-function of the representation twisted by general CM characters at far-from-center critical points. The main tool used in [loc. cit.] is the doubling method of Piatetski-Shapiro and Rallis. This paper can be thought of as a continuation of their work, but instead using a more general pullback formula of Shimura (which is actually due to Garrett [1984; 1989] and is called the "Garrett map") to construct p-adic families of Klingen Eisenstein series on $\mathrm{U}(r + 1, s + 1)$ from the original automorphic representation.

The motivation for doing this is to provide a framework to generalize the important work of Skinner and Urban [2014] on the Iwasawa main conjectures for GL_2 to forms on general unitary groups. The general strategy is, starting with a family of cuspforms on the unitary group $\mathrm{U}(r, s)$ and a family of CM characters, we construct a family of Klingen Eisenstein series on the bigger group $\mathrm{U}(r + 1, s + 1)$. One tries to prove the constant terms of the Klingen Eisenstein family are divisible by the standard p-adic L-function of the cuspforms on $\mathrm{U}(r, s)$ and, therefore, the Eisenstein family is congruent to cuspidal families modulo this p-adic L-function. Passing to the Galois side, such congruences enable us to construct elements in the Selmer groups, proving one divisibility of the corresponding Iwasawa main conjecture.

We have been able to use it to prove one divisibility of the Iwasawa main conjectures for Hilbert modular forms and some kinds of Rankin–Selberg p-adic L-functions; see [Wan 2013; 2015]. C. Skinner has recently been able to use the result of [Wan 2015] to prove a converse of a theorem of Gross, Zagier and Kolyvagin that states that, if the rank of the Selmer group of an elliptic curve is one and the p-part of the Shafarevich–Tate group is finite, then the Heegner point is nontorsion and the central L-value vanishes at order exactly one [Skinner 2014]. The first step towards the plan outlined above is to construct the family of Klingen–Eisenstein series and study the p-adic properties of its Fourier–Jacobi coefficients, which is the main task of the present paper.

In [Eischen et al. ≥ 2015] the interpolation formulas are proved at all arithmetic points. However, in this paper we are only able to understand the pullback Eisenstein sections in the "generic case" (to be defined in Definition 4.42; basically this puts restrictions on the ramification of the form at primes dividing p). The reason is that it seems difficult in general to describe the nearly ordinary Klingen Eisenstein sections. Fortunately, since along a Hida family the set of forms that are "generic" is Zariski-dense, these computations are enough to construct the whole Hida family of Klingen Eisenstein series (similar to the [Skinner and Urban 2014] case). Thus, we only work with a Hida family of forms instead of a single cusp form, due to this "generic" condition. We remark that when $s=0$, by working with forms of general vector-valued weights, we are able to construct a class of the p-adic L-function and Klingen Eisenstein family for a single form unramified at p (not necessarily ordinary; see [Eischen and Wan 2014]).

Now we state the main results. Let $\mathcal{K}_\infty$ be the maximal abelian pro-p-extension of $\mathcal{K}$ unramified outside p. We write $\Gamma_{\mathcal{K}}=\mathrm{Gal}(\mathcal{K}_\infty/\mathcal{K})$. This is a free $\mathbb{Z}_p$-module whose rank should be $d+1$, assuming the Leopoldt conjecture. Take a finite extension L over $\mathbb{Q}_p$. Let $\mathcal{O}_L$ be the integer ring of L. Let $\mathcal{O}_L^{\mathrm{ur}}$ be the completion of the integer ring of the maximal unramified extension of L. We define $\Lambda_{\mathcal{K}}=\mathcal{O}_L[\![\Gamma_{\mathcal{K}}]\!]$. Let $\kappa>4$ be an integer and τ_0 a Hecke character of $\mathcal{K}^\times\backslash\mathbb{A}_{\mathcal{K}}^\times$ whose infinite types are $\left(-\frac{1}{2}\kappa,\frac{1}{2}\kappa\right)$ at all infinite places. We have a $\Lambda_{\mathcal{K}}$-valued family of Hecke characters of $\mathcal{K}^\times\backslash\mathbb{A}_{\mathcal{K}}^\times$ containing τ_0 as a specialization (to be made precise later). Let Λ be the weight algebra for $\mathrm{U}(r,s)$, defined later, and $\mathbb{I}$ a normal domain containing Λ which is finite over Λ. Let $\mathbb{I}^{\mathrm{ur}}$ be the normalization of an irreducible component of $\mathbb{I}\hat{\otimes}_{\mathcal{O}_L}\mathcal{O}_L^{\mathrm{ur}}$. (In fact, for each such irreducible component we can make the following construction.) Let $\Omega_\infty\in\mathbb{C}^{\Sigma_\infty}$ be the CM period of the CM field $\mathcal{K}$ and $\Omega_p\in(\mathbb{Z}_p^{\mathrm{ur}})^{\Sigma_\infty}$ be the p-adic period (we refer to [Hida 2004a] for the definition). We write $\Omega_\infty^{\Sigma_\infty}$ for the product of the d elements of Ω_∞ and define $\Omega_p^{\Sigma_\infty}$ similarly. Throughout this paper, we write

$$z_\kappa=\tfrac{1}{2}(\kappa-r-s-1),$$
$$z'_\kappa=\tfrac{1}{2}(\kappa-r-s).$$

Theorem 1.1. *Let $\boldsymbol{f}$ be an $\mathbb{I}$-coefficient, nearly ordinary, cuspidal eigenform on* $\mathrm{GU}(r,s)$ *such that the specialization $\boldsymbol{f}_\phi$ at a Zariski-dense set of "generic" arithmetic points ϕ is classical and generates an irreducible automorphic representation of* $\mathrm{U}(r,s)$. *Let Σ be a finite set of primes containing all primes dividing any entry of ζ, or the conductor of $\boldsymbol{f}$, or $\mathcal{K}$.*

In the case when $s\neq 0$, we make the assumptions TEMPERED, $\mathrm{Proj}_{\boldsymbol{f}^\vee}$ *and* DUAL, *or assumptions* TEMPERED, $\mathrm{Proj}_{\boldsymbol{f}}$ *and* $\mathrm{Proj}_{\boldsymbol{f}^\vee}$ (*to be defined in Section 5A*).

Then:

(i) *There is an element $\mathcal{L}^\Sigma_{\boldsymbol{f},\tau_0}\in\mathbb{I}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]\otimes_{\mathbb{I}^{\mathrm{ur}}}F_{\mathbb{I}^{\mathrm{ur}}}$ such that, for a Zariski-dense subset of arithmetic points $\phi\in$* $\mathrm{Spec}\,\mathbb{I}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]$ (*to be specified in Definition 4.42*), *we have* *that, if $s=0$, then $\mathscr{L}^{\Sigma}_{\boldsymbol{f},\tau_0}\in \mathbb{I}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]$ and*

$$\phi(\mathscr{L}^{\Sigma}_{\boldsymbol{f},\tau_0})=c'_{\kappa}(z'_{\kappa_\phi})\left(\frac{(-2)^{-d(a+2b)}(2\pi i)^{d(a+2b)\kappa_\phi}(2/\pi)^{d(a+2b)(a+2b-1)/2}}{\prod_{j=0}^{a+2b-1}(\kappa_\phi-j-1)^d}\right)^{-1}\cdot C^p_{\boldsymbol{f}_\phi}$$
$$\times\prod_{v\mid p}\left(|p^{t_1+\cdots+t_r}|^{-\kappa_\phi/2}\times p^{-((r+1)/2)\sum_{j=1}^r t_j}\prod_{j=1}^r\mathfrak{g}(\chi_j\tau_1^{-1})\chi_j^{-1}\tau_1(p^{t_j})\right)$$
$$\times\frac{L^{\Sigma}(\tilde{\pi}_{\boldsymbol{f}_\phi},\bar{\tau}^c_\phi,\kappa_\phi-r)\Omega_p^{r\kappa_\phi\Sigma_\infty}}{\Omega_\infty^{r\kappa_\phi\Sigma_\infty}}.$$

If $s\neq 0$, then

$$\phi(\mathscr{L}^{\Sigma}_{\boldsymbol{f},\tau_0})$$
$$=c'_{\kappa}(z'_{\kappa_\phi})\left(\frac{(-2)^{-d(a+2b)}(2\pi i)^{d(a+2b)\kappa_\phi}(2/\pi)^{d(a+2b)(a+2b-1)/2}}{\prod_{j=0}^{a+2b-1}(\kappa_\phi-j-1)^d}\right)^{-1}\cdot C^p_{\boldsymbol{f}_\phi}$$
$$\times\prod_{v\mid p}\left(\frac{p^{(r+s)(r+s-1)/2}\cdot(p-1)^{r+s}}{\left(\prod_{i=1}^r p^{t_i\cdot(r+s-i)}\right)\cdot\left(\prod_{i=1}^s p^{t_{r+i}(r+s-i)}\right)\cdot\prod_{j=1}^{r+s}(p^j-1)}\right.$$
$$\times p^{-ss_2(1+a+2b)/2}p^{-\sum_{j=1}^r t_j(a+1)/2}p^{\sum_{i=1}^s t_{r+i}(a+b)}|p^{t_1+\cdots+t_r+s\cdot s_2}|^{-\kappa_\phi/2}$$
$$\left.\times\prod_{i=r+1}^{r+s}\mathfrak{g}(\chi_i^{-1}\tau_2)\chi_i\tau_2^{-1}(p^{s_2})\prod_{j=1}^r\mathfrak{g}(\chi_j\tau_1^{-1})\chi_j^{-1}\tau_1(p^{t_j})\right)$$
$$\times\frac{L^{\Sigma}(\tilde{\pi}_{\boldsymbol{f}_\phi},\bar{\tau}^c_\phi,\kappa_\phi-r-s)}{\langle\tilde{\varphi}^{\mathrm{ord}}_\phi,\varphi_\phi\rangle},$$

where the χ_i are defined in Definition 4.42 and $\tau_{\phi,p}=(\tau_1,\tau_2^{-1})$ such that τ_i has conductor p^{s_i} with $s_2>s_1$. Also,

$$C^p_{\boldsymbol{f}_\phi}=\prod_{v\nmid p,\,v\in\Sigma}\tau(y_v\bar{y}_vx_v)|(y_v\bar{y}_v)^2x_v\bar{x}_v|_v^{-z_{\kappa'_\phi}-(a+2b)/2}\,\mathrm{Vol}(\mathfrak{Y}_v)$$

(*the x_v and y_v are the x and y in Section 4C1 and $\mathfrak{Y}_v$ is defined in Definition 4.11.*) *The $c_\kappa(z)$ and $c'_\kappa(z)$ are defined in Lemma 4.3 and κ_ϕ is the weight associated to the arithmetic point ϕ. The φ_ϕ and $\tilde{\varphi}^{\mathrm{ord}}_\phi$ are the specialization of $\boldsymbol{f}$ and the $\boldsymbol{f}^\vee$ provided by the assumption* $\mathrm{Proj}_{\boldsymbol{f}^\vee}$ (*notice that they are ordinary vectors with respect to different Borel groups, e.g., when $s=0$, the level group for φ_ϕ at p is with respect to the upper-triangular Borel subgroup, while that for $\tilde{\varphi}^{\mathrm{ord}}_\phi$ is with respect to the lower-triangular Borel subgroup*). *The factor*

$$\frac{p^{(r+s)(r+s-1)/2}\cdot(p-1)^{r+s}}{\left(\prod_{i=1}^r p^{t_i\cdot(r+s-i)}\right)\cdot\left(\prod_{i=1}^s p^{t_{r+i}(r+s-i)}\right)\cdot\prod_{j=1}^{r+s}(p^j-1)}$$

is the volume of a set $\tilde{K}'$ *defined in Definition 4.34* (*this is smaller than the level group for* $\tilde{\varphi}_\phi^{\mathrm{ord}}$). *The* $F_{\mathbb{I}^{\mathrm{ur}}}$ *is the fraction field of* $\mathbb{I}^{\mathrm{ur}}$. *The* τ_ϕ *are specializations of the family of CM characters containing* τ_0. *The* p^{t_i} *are conductors of some characters defined in Definition 4.21. The* $\tilde{\lambda}_{\beta,v}$ *is defined in* (17), *whose* p*-order is* $\sum_{i=1}^{b} t_{a+b+i}(a+b-\kappa)$.

(ii) *There is a set of formal* q*-expansions* $\boldsymbol{E}_{\boldsymbol{f},\tau_0} := \{\sum_\beta a^h_{[g]}(\beta)q^\beta\}_{([g],h)}$ *for* $\sum_\beta a^h_{[g]}(\beta)q^\beta \in (\mathbb{I}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]\hat{\otimes}_{\mathbb{Z}_p}\mathcal{R}_{[g],\infty}) \otimes_{\mathbb{I}^{\mathrm{ur}}} F_{\mathbb{I}^{\mathrm{ur}}}$, *where* $\mathcal{R}_{[g],\infty}$ *is some ring to be defined later*, *in Section 5*, *and* $([g],h)$ *are* p*-adic cusp labels* (*Definition 2.6*) *such that*, *for a Zariski-dense set of arithmetic points* $\phi \in \operatorname{Spec}\mathbb{I}[\![\Gamma_{\mathcal{K}}]\!]$, $\phi(\boldsymbol{E}_{\boldsymbol{f},\tau_0})$ *is the Fourier–Jacobi expansion of the holomorphic*, *nearly ordinary Klingen Eisenstein series* $E(f_{\mathrm{Kling},\phi}, z_{\kappa_\phi}, -)$ *we construct in Section 5C* (*see the interpolation formula in Proposition 5.8*). *Here*, f_{Kling} *is a certain "Klingen section" to be defined there.*

(iii) *The terms* $a^t_{[g]}(0)$ *are divisible by* $\mathcal{L}^\Sigma_{\boldsymbol{f},\tau_0}\cdot\mathcal{L}^\Sigma_{\bar{\tau}_0'}$, *where* $\mathcal{L}^\Sigma_{\bar{\tau}_0'}$ *is the* p*-adic* L*-function of a Dirichlet character to be defined in the text.*

The assumption "TEMPERED" is included so that we can easily write down the explicit range of absolute convergence for pullback formulas. It is not serious and may be relaxed using ideas of [Harris 1984]. Besides the theorem, we also make some preliminary computations for the Fourier–Jacobi coefficients for Siegel Eisenstein series. This is crucial for analyzing the p-adic properties of the Klingen Eisenstein series we construct. When doing arithmetic application we need to prove that certain Fourier–Jacobi coefficient of this Eisenstein family is prime to the p-adic L-function.

This paper is organized as follows. In Section 2 we recall various backgrounds. In Section 3 we recall the notion of p-adic automorphic forms on unitary groups and Fourier–Jacobi expansion. In Section 4 we recall the notion of Klingen and Siegel Eisenstein series, the pullback formulas relating them and their Fourier–Jacobi coefficients, and then do the local calculations. (This is the most technical part of this paper.) We manage to take the Siegel sections so that, when we are moving our Eisenstein datum p-adically, these Siegel Eisenstein series also move p-adic analytically. The hard part is to choose the sections at p-adic places. At non-Archimedean cases prime to p the choice is more flexible. (We might change this choice whenever doing arithmetic applications; see [Wan 2013; 2015].) At the Archimedean places we restrict ourselves to the parallel scalar weight case, which is enough for doing Hida theory. In Section 5 we make the global calculations and construct the nearly ordinary Klingen Eisenstein series by the pullbacks of a Siegel Eisenstein series from a larger unitary group. Finally, we include an Appendix by Kai-Wen Lan for detailed proofs of some facts used for the p-adic q-expansion principle. (This is not strictly needed in our construction. But we think it is good to include it for completeness and for the convenience of readers.)

2. Background

In this section we recall notations for holomorphic automorphic forms on unitary groups, Eisenstein series and Fourier–Jacobi expansions.

2A. ***Notation.*** Suppose F is a totally real field such that $[F:\mathbb{Q}]=d$ and $\mathcal{K}$ is a totally imaginary quadratic extension of F. For a finite place v of F or $\mathcal{K}$, we usually write ϖ_v for a uniformizer and q_v for the cardinality of its residue field. Let c be the nontrivial element of $\mathrm{Gal}(\mathcal{K}/F)$. Let r and s be two integers with $r\geq s\geq 0$. We fix an odd prime p that splits completely in $\mathcal{K}/\mathbb{Q}$. We fix $i_\infty:\overline{\mathbb{Q}}\hookrightarrow\mathbb{C}$ and $\iota:\mathbb{C}\simeq\mathbb{C}_p$ and write i_p for $\iota\circ i_\infty$. Let Σ_∞ be the set of Archimedean places of F. We take a CM type of $\mathcal{K}$, still denoted by Σ_∞ (thus $\Sigma_\infty\sqcup\Sigma_\infty^c$ are all embeddings $\mathcal{K}\to\mathbb{C}$, where $\Sigma_\infty^c=\{\tau\circ c\mid\tau\in\Sigma_\infty\}$).

We use ϵ to denote the cyclotomic character and ω the Teichmüller character. We will often adopt the following notation: for an idele class character $\chi=\bigotimes_v\chi_v$, we write $\chi_p(x)=\prod_{v|p}\chi_v(x_v)$. For a character ψ of $\mathcal{K}_v^\times$ or $\mathbb{A}_\mathcal{K}^\times$, we often write ψ' for the restriction to $F_v^\times$ or $\mathbb{A}_F^\times$. For a character τ of $\mathcal{K}^\times$ or $\mathbb{A}_\mathcal{K}^\times$, we define τ^c by $\tau^c(x)=\tau(x^c)$. (Note: we will write $\bar{\tau}(x)$ for the complex conjugation of $\tau(x)$ while the "c" means taking complex conjugation for the source.)

If v is a prime of F with characteristic ℓ and $\mathfrak{d}_v\mathcal{O}_{F,v}=(d_v)$, $d_v\in F_v^\times$ is the different of $F/\mathbb{Q}$ at v and, if ψ_v is a character of $F_v^\times$ and $(c_{\psi,v})\subset\mathcal{O}_{F,v}$ is the conductor, then we define the local *Gauss sums*

$$\mathfrak{g}(\psi_v,c_{\psi,v}d_v):=\sum_{a\in(\mathcal{O}_{F,v}/c_{\psi,v})^\times}\psi_v(a)e\left(\mathrm{Tr}_{F_v/\mathbb{Q}_\ell}\left(\frac{a}{c_{\psi,v}d_v}\right)\right),$$

where ℓ is the rational prime above v. If $\bigotimes\psi_v$ is an idele class character of $\mathbb{A}_F^\times$ then we set the global Gauss sum,

$$\mathfrak{g}\left(\bigotimes\psi_v\right):=\prod_v\psi_v^{-1}(c_{\psi,v}d_v)\mathfrak{g}(\psi,c_{\psi,v}d_v).$$

This is independent of all the choices of d_v and $C_{\psi,v}$. Also, if $F_v\simeq\mathbb{Q}_p$ and (p^t) is the conductor for ψ_v, then we write $\mathfrak{g}(\psi_v):=\mathfrak{g}(\psi_v,p^t)$. We define the Gauss sums for $\mathcal{K}$ similarly.

Let $\mathcal{K}_\infty$ be the maximal abelian $\mathbb{Z}_p$-extension of $\mathcal{K}$ unramified outside p. Write $\Gamma_\mathcal{K}:=\mathrm{Gal}(\mathcal{K}_\infty/\mathcal{K})$ and $G_\mathcal{K}$ the absolute Galois group of $\mathcal{K}$. Define $\Lambda_\mathcal{K}:=\mathbb{Z}_p[[\Gamma_\mathcal{K}]]$. For any finite extension A of $\mathbb{Z}_p$ define $\Lambda_{\mathcal{K},A}:=A[[\Gamma_\mathcal{K}]]$. Let $\boldsymbol{\varepsilon}_\mathcal{K}:G_\mathcal{K}\to\Gamma_\mathcal{K}\hookrightarrow\Lambda_\mathcal{K}^\times$ be the canonical character. We define $\Psi_\mathcal{K}$ to be the composition of $\boldsymbol{\varepsilon}_\mathcal{K}$ with the reciprocity map of global class field theory, which we denote as $\mathrm{rec}_\mathcal{K}:\mathcal{K}^\times\backslash\mathbb{A}_\mathcal{K}^\times\to G_\mathcal{K}^{\mathrm{ab}}$. Here we used the geometric normalization of class field theory. We make the corresponding definitions for F as well.

Let $S_m(R)$ be the set of matrices $S \in M_m(R \otimes_{\mathcal{O}_F} \mathcal{O}_{\mathcal{K}})$ such that $S = {}^t\bar{S}$, where conjugation is with respect to the second variable of $R \otimes_{\mathcal{O}_F} \mathcal{O}_{\mathcal{K}}$. We write $B = B_n$ and $N = N_n$ for the upper-triangular Borel subgroup and unipotent radical of the group GL_n. Let N^{opp} or N^- be the opposite unipotent radical of N. We define the function $e_{\mathbb{A}_{\mathbb{Q}}} = \prod_v e_v$ with e_v the function on $\mathbb{Q}_v^\times$ such that $e_v(x_v) = e^{2\pi i \cdot \{x_v\}}$ for $\{x_v\}$ the fractional part of x_v and $e_\infty(x) = e^{-2\pi i x}$. We will usually write $\eta = \begin{pmatrix} & 1_m \\ -1_m & \end{pmatrix}$ if m is clear from the context.

2B. ***Unitary Groups.*** We define

$$\theta_{r,s} = \begin{pmatrix} & & 1_s \\ & \zeta & \\ -1_s & & \end{pmatrix},$$

where ζ is a fixed diagonal matrix such that $i^{-1}\zeta$ is totally positive. Let $V = V(r, s)$ be the skew-Hermitian space over $\mathcal{K}$ with respect to this metric, i.e., $\mathcal{K}^{r+s}$ equipped with the metric given by $\langle u, v\rangle := u\theta_{r,s}{}^t\bar{v}$. We define algebraic groups $\mathrm{GU}(r, s)$ and $\mathrm{U}(r, s)$ as follows: for any $\mathcal{O}_F$-algebra R, the R points are

$$\mathrm{GU}(r, s)(R) := \{g \in \mathrm{GL}_{r+s}(\mathcal{O}_{\mathcal{K}} \otimes_{\mathcal{O}_F} R) \mid g\theta_{r,s}g^* = \mu(g)\theta_{r,s},\ \mu(g) \in R^\times\}$$

(where $g^* = {}^t\bar{g}$ and $\mu : \mathrm{GU}(r, s) \to \mathbb{G}_{m,F}$ is called the similitude character) and

$$\mathrm{U}(r, s)(R) := \{g \in \mathrm{GU}(r, s)(R) \mid \mu(g) = 1\}.$$

So the unitary group $\mathrm{U}(r, s)$ in this paper really means the unitary group with respect to our fixed metric $\theta_{r,s}$. Sometimes we write GU_n and U_n for $\mathrm{GU}(n, n)$ and $\mathrm{U}(n, n)$. For two forms φ_1, φ_2 on $\mathrm{U}(r, s)(\mathbb{A}_F)$, we define the inner product by

$$\langle\varphi_1, \varphi_2\rangle := \int_{\mathrm{U}(r,s)(F)\backslash \mathrm{U}(r,s)(\mathbb{A}_F)} \varphi_1(g)\varphi_2(g)\, dg,$$

where the measure is chosen so that $\mathrm{U}(r, s)(\mathcal{O}_{F_v}) = 1$ for all finite v and we take the measure at Archimedean places as in [Shimura 1997, (7.14.5)].

We have the embedding

$$\mathrm{GU}(r, s) \times \mathrm{Res}_{\mathcal{O}_{\mathcal{K}}/\mathcal{O}_F} \mathbb{G}_m \to \mathrm{GU}(r + 1, s + 1),$$

$$g \times x = \begin{pmatrix} a & b & c \\ d & e & f \\ h & l & k \end{pmatrix} \times x \mapsto \begin{pmatrix} a & & b & c & \\ & \mu(g)\bar{x}^{-1} & & & \\ d & & e & f & \\ h & & l & k & \\ & & & & x \end{pmatrix}.$$

We write $m(g, x)$ for the right-hand side. The image of the above map is the Levi subgroup of the Klingen parabolic subgroup P of $\mathrm{GU}(r + 1, s + 1)$, which consists of matrices in $\mathrm{GU}(r+1, s+1)$ such that the off-diagonal entries of the $(s+1)$-st column and the last row are 0. We denote this Levi subgroup by M_P. We also write N_P for the unipotent radical of P. We also define $B = B(r, s)$ to be the standard Borel subgroup, consisting of matrices

$$g = \begin{pmatrix} A_g & B_g \\ & D_g \end{pmatrix},$$

where the blocks are with respect to the partition $r+s$ and we require that A_g is lower-triangular and D_g is upper-triangular.

We write $-V(r, s) = V(s, r)$ for the Hermitian space whose metric is $-\theta_{r,s}$. We define some embeddings of $\mathrm{GU}(r+1, s+1) \times \mathrm{GU}(-V(r, s))$ into some larger unitary groups. These will be used in the doubling method. Recall we wrote $a = r - s$ and $b = s$ at the beginning of the introduction; we define $\mathrm{GU}(r+s+1, r+s+1)'$ to be the unitary similitude group associated to

$$\begin{pmatrix} & & & & 1_b & & & \\ & & & & & 1 & & \\ & & \zeta & & & & & \\ & & & & & & & -1_b \\ -1_b & & & & & & & \\ & -1 & & & & & & \\ & & & & & & -\zeta & \\ & & & 1_b & & & & \end{pmatrix}$$

and $G(r+s, r+s)'$ to be associated to

$$\begin{pmatrix} & & & 1_b & & \\ & \zeta & & & & \\ & & & & & -1_b \\ -1_b & & & & & \\ & & & & -\zeta & \\ & & 1_b & & & \end{pmatrix}.$$

We define an embedding

$$\alpha : \{g_1 \times g_2 \in \mathrm{GU}(r+1, s+1) \times \mathrm{GU}(-V(r, s)) \mid \mu(g_1) = \mu(g_2)\} \\ \to \mathrm{GU}(r+s+1, r+s+1)'$$

by viewing g_1 as a block matrix with respect to the partition $s+1+(r-s)+s+1$ (this means we use this partition to divide both the rows and the columns into blocks) and g_2 as a block matrix with respect to $s+(r-s)+s$, then we define α by requiring the 1, 2, 3, 4, 5-th (blockwise) rows and columns of $\mathrm{GU}(r+1, s+1)$ embed to the 1, 2, 3, 5, 6-th (blockwise) rows and columns of $\mathrm{GU}(r+s+1, r+s+1)'$ and the

1, 2, 3-rd (blockwise) rows and columns of $\mathrm{GU}(V(s,r))$ embed to the 8, 7, 4-th rows and columns (blockwise) of $\mathrm{GU}(r+s+1, r+s+1)'$.

We also define an embedding

$$\alpha' : \{g_1 \times g_2 \in \mathrm{GU}(r,s) \times \mathrm{GU}(-V(r,s)) \mid \mu(g_1) = \mu(g_2)\} \to \mathrm{GU}(r+s, r+s)'$$

in a similar way to above: Consider $\mathrm{GU}(r,s)$ and $\mathrm{GU}(-V(r,s))$ as block matrices with respect to the partition $s + (r-s) + s$. Put the 1, 2, 3-rd (blockwise) rows and columns of the first $\mathrm{GU}(r,s)$ into the 1, 2, 4-th (blockwise) rows and columns of $\mathrm{GU}(r+s, r+s)'$ and put the 1, 2, 3-rd (blockwise) rows and columns of the second $\mathrm{GU}(r,s)$ into the 6, 5, 4-th rows and columns of $\mathrm{GU}(r+s, r+s)'$.

We also define isomorphisms

$$\beta : \mathrm{GU}(r+s+1, r+s+1)' \xrightarrow{\sim} \mathrm{GU}(r+s+1, r+s+1), \quad g \mapsto S^{-1} g S,$$

and

$$\beta' : \mathrm{GU}(r+s, r+s)' \xrightarrow{\sim} \mathrm{GU}(r+s, r+s), \quad g \mapsto S'^{-1} g S',$$

where

$$S = \begin{pmatrix} 1_b &&&&&&& -\frac{1}{2}\cdot 1_b \\ & 1 &&&&&& \\ && 1_a &&&& -\frac{1}{2}\zeta & \\ &&& -1_b & \frac{1}{2}\cdot 1_b &&& \\ &&& 1_b & \frac{1}{2}\cdot 1_b &&& \\ &&&&& 1 && \\ && -1_a &&&& -\frac{1}{2}\zeta & \\ -1_b &&&&&&& -\frac{1}{2}\cdot 1_b \end{pmatrix} \tag{1}$$

and

$$S' = \begin{pmatrix} 1_b &&&&& -\frac{1}{2}\cdot 1_b \\ & 1_a &&& -\frac{1}{2}\zeta & \\ && -1_b & \frac{1}{2}\cdot 1_b && \\ && 1_b & \frac{1}{2}\cdot 1_b && \\ & -1_a &&& -\frac{1}{2}\zeta & \\ -1_b &&&&& -\frac{1}{2}\cdot 1_b \end{pmatrix}. \tag{2}$$

Remark 2.1 (about unitary groups). In order to have Shimura varieties for doing p-adic modular forms and Galois representations, we need to use a unitary group defined over $\mathbb{Q}$. More precisely, consider V as a skew-Hermitian space over $\mathbb{Q}$ and still write $\theta_{r,s}$ for the metric on it. Let T be a $\mathcal{O}_{\mathcal{K}}$ lattice that we use to define $\mathrm{GU}(r,s)$. Then the correct unitary similitude group should be

$$\mathrm{GU}^0(r,s)(A) := \{g \in \mathrm{GL}_{\mathcal{O}_{\mathcal{K}} \otimes_{\mathbb{Z}} A}(T \otimes_{\mathbb{Z}} A) \mid g\theta_{r,s}g^* = \mu(g)\theta_{r,s},\ \mu(g) \in A\}$$

for any commutative ring A. This group is smaller than the one we defined before. However, this group is not convenient for local computations, since we cannot treat the primes of F independently. So what we do (implicitly) is: For the analytic construction, we write down forms on the larger unitary similitude group defined above and then restrict to the smaller one. For the algebraic construction, we only do the pullbacks for unitary (instead of similitude) groups.

We are going to fix some bases of the various Hermitian spaces. We let

$$y^1, \dots, y^s, w^1, \dots, w^{r-s}, x^1, \dots, x^s$$

be the standard basis of V such that the Hermitian forms is given above. Let W be the span over $\mathcal{K}$ of $w^1, \dots, w^{r-s}$. Let $X^\vee = \mathcal{O}_{\mathcal{K}}x^1 \oplus \cdots \oplus \mathcal{O}_{\mathcal{K}}x^s$ and $Y = \mathcal{O}_{\mathcal{K}}y^1 \oplus \cdots \oplus \mathcal{O}_{\mathcal{K}}y^s$. Let L be an $\mathcal{O}_{\mathcal{K}}$-maximal lattice such that $L_p := L \otimes_{\mathbb{Z}} \mathbb{Z}_p = \sum_{i=1}^{r-s}(\mathcal{O}_{\mathcal{K}} \otimes_{\mathbb{Z}} \mathbb{Z}_p)w^i$. We define a $\mathcal{O}_{\mathcal{K}}$-lattice M of V by

$$M := Y \oplus L \oplus X^\vee.$$

Let $M_p = M \otimes_{\mathbb{Z}} \mathbb{Z}_p$. A pair of sublattices $\mathrm{Pol}_p = \{N^{-1}, N^0\}$ of M_p is called an ordered polarization of M_p if N^{-1} and N^0 are maximal isotropic direct summands in M_p and they are dual to each other with respect to the Hermitian pairing. Moreover, we require that, for each $v = ww^c$ with $w \in \Sigma_p$, $\operatorname{rank} N_w^{-1} = \operatorname{rank} N_{w^c}^0 = r$ and $\operatorname{rank} N_{w^c}^{-1} = \operatorname{rank} N_w^0 = s$. The standard polarization of M_p is given by $M_v^{-1} = Y_w \oplus L_w \oplus Y_{w^c}$ and $M_v^0 = X_{w^c} \oplus L_{w^c} \oplus X_w$. We let $-V$ be the Hermitian space V with the metric given by the negative of V. We let $\tilde{y}^1, \dots, \tilde{y}^s$, $\tilde{w}^1, \dots, \tilde{w}^{r-s}$, $\tilde{x}^1, \dots, \tilde{x}^s$ be the corresponding basis. Let $\mathcal{K}y^{s+1} \oplus \mathcal{K}x^{s+1}$ be a 2-dimensional Hermitian space with metric $\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$. We define

$$\boldsymbol{W} := V \oplus \mathcal{K}y^{s+1} \oplus \mathcal{K}x^{s+1} \oplus (-V).$$

Let $\Upsilon \in \mathrm{U}(n+1, n+1)(F_p)$ be such that, for each $v|p$ with $v = ww^c$, where w is in our p-adic CM type Σ_p, $\Upsilon_w = S_w^{-1}$. We define another basis of $\boldsymbol{W}$ by

$$\begin{aligned}{}^t(y^1, \dots, y^{s+1}, w^1, \dots, w^{r-s}, x^1, \dots, x^{s+1}, y^1, \dots, y^s, w^1, \dots, w^{r-s}, x^1, \dots, x^s)\Upsilon \\ = {}^t(\boldsymbol{y}^1, \dots, \boldsymbol{y}^{r+s+1}, \boldsymbol{x}^1, \dots, \boldsymbol{x}^{r+s+1}).\end{aligned}$$

Then $\boldsymbol{Y} := \bigoplus_{i=1}^{r+s+1}(\mathcal{O}_{\mathcal{K}} \otimes_{\mathbb{Z}} \mathbb{Z}_p)\boldsymbol{y}^i$ and $\boldsymbol{X} := \bigoplus_{i=1}^{r+s+1}(\mathcal{O}_{\mathcal{K}} \otimes_{\mathbb{Z}} \mathbb{Z}_p)\boldsymbol{x}^i$ gives another polarization $(\boldsymbol{Y}, \boldsymbol{X})$ of $\boldsymbol{L}_p := M_p \oplus (-M_p) \oplus \mathcal{O}_{\mathcal{K}}y^{s+1} \oplus \mathcal{O}_{\mathcal{K}}x^{s+1}$.

2C. *Automorphic forms.*

2C1. *Hermitian symmetric domain.* Suppose $r \geq s > 0$. Then the Hermitian symmetric domain for $G := \mathrm{GU}(r, s)$ is

$$X^+ = X_{r,s} = \left\{\tau = \begin{pmatrix} x \\ y \end{pmatrix} \,\middle|\, x \in M_s(\mathbb{C}^\Sigma),\ y \in M_{(r-s)\times s}(\mathbb{C}^\Sigma),\ i(x^* - x) > -iy^*\theta^{-1}y\right\}.$$

For $\alpha \in \mathrm{GU}(r,s)(F_\infty)$, where $F_\infty := F \otimes_{\mathbb{Q}} \mathbb{R}$, we write

$$\alpha = \begin{pmatrix} a & b & c \\ d & e & f \\ h & l & d \end{pmatrix}$$

according to the standard basis of V together with the block decomposition with respect to $s + (r-s) + s$. There is an action of $\alpha \in G(F_\infty)^+$ (here the superscript $+$ means the component with positive similitude factor at all Archimedean places) on $X_{r,s}$, defined by

$$\alpha \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by+c \\ gx+ey+f \end{pmatrix} (hx+ly+d)^{-1}.$$

If $rs=0$, $X_{r,s}$ consists of a single point, written $\boldsymbol{x}_0$, with the trivial action of $G(F_\infty)^+$. For an open compact subgroup U of $G(\mathbb{A}_{F,f})$, put

$$M_G(X^+, U) := G(F)^+ \backslash X^+ \times G(\mathbb{A}_{F,f})/U,$$

where U is an open compact subgroup of $G(\mathbb{A}_{F,f})$. We let

$$\mathbb{C}^{r,s} = \mathbb{C}(\Sigma^c)^s \otimes \mathbb{C}(\Sigma^c)^{r-s} \otimes \mathbb{C}(\Sigma)^s$$

and define a map $c_{r,s}$ on it by $(u_1, u_2, u_3)c_{r,s} = (\bar{u}_1, \bar{u}_2, u_3)$. We define the map $p(\tau) : V \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{C}^{r,s}$ by $p(\tau)v = vB(\tau)c_{r,s}$. Let

$$B(\tau) = \begin{pmatrix} x^* & y^* & x \\ 0 & \zeta & y \\ 1_s & 0 & 1_s \end{pmatrix}.$$

We define the automorphic factors $\kappa(\alpha, \tau)$ and $\mu(\alpha, \tau)$ by

$$\alpha B(\tau) = B(\alpha\tau)(\overline{\kappa(\alpha, \tau)}, \mu(\alpha, \tau))$$

for $\alpha \in G(\mathbb{R})$ and $\tau \in X^+$. We sometimes write $\kappa_{r,s}(\alpha, \tau)$, $\mu_{r,s}(\alpha, \tau)$ to emphasize the group $\mathrm{U}(r,s)$. We define $j(g,z) := \det(\mu(g,z))$. For $z \in X_{r+1,s+1}$, we define $\wp(z) \in X_{r,s}$ to be the lower-right $r \times s$ submatrix. For $z_1 = \binom{x_1}{y_1}$ and $z \in \binom{x}{y}$, we define $\eta(z_1, z) = i(x_1^* - x) - y_1^*(i\zeta^{-1})y$ and $\delta(z_1, z) = 2^{-s}\det(\eta(z_1, z))$.

2C2. *Automorphic forms.* We will mainly follow [Hsieh 2014] to define the space of automorphic forms, with slight modifications. We define a cocycle

$$J : R_{F/\mathbb{Q}}G(\mathbb{R})^+ \times X^+ \to \mathrm{GL}_r(\mathbb{C}^\Sigma) \times \mathrm{GL}_s(\mathbb{C}^\Sigma) := H(\mathbb{C}),$$
$$(\alpha, \tau) \mapsto (\kappa(\alpha, \tau), \mu(\alpha, \tau)),$$

where

$$\kappa(\alpha,\tau)=\begin{pmatrix}\bar{h}\,{}^t x+\bar{d} & \bar{h}\,{}^t y+l\bar{\theta}\\ -\bar{\theta}^{-1}(\bar{g}\,{}^t x+\bar{f}) & -\bar{\theta}^{-1}\bar{g}\,{}^t y+\bar{\theta}^{-1}\bar{e}\bar{\theta}\end{pmatrix}\quad\text{and}\quad \mu(\alpha,\tau)=hx+ly+d$$

$$\text{for }\tau=\begin{pmatrix}x\\ y\end{pmatrix}\text{ and }\alpha=\begin{pmatrix}a&b&c\\ g&e&f\\ h&l&d\end{pmatrix}.$$

Let $\boldsymbol{i}$ be the point $\binom{i1_s}{0}$ on the Hermitian symmetric domain for $\mathrm{GU}(r,s)$ (here 0 means the $(r-s)\times s$ matrix 0). Let $\mathrm{GU}(r,s)(\mathbb{R})^+$ be the subgroup of $\mathrm{GU}(r,s)(\mathbb{R})$ whose similitude factor is totally positive. Let K_∞^+ be the compact subgroup of $\mathrm{U}(r,s)(\mathbb{R})$ stabilizing $\boldsymbol{i}$ and K_∞ be the group generated by K_∞^+ and $\mathrm{diag}(1_{r+s},-1_s)$. Then $J:K_\infty^+\to H(\mathbb{C})$, $k_\infty\mapsto J(k_\infty,\boldsymbol{i})$, defines an algebraic representation of K_∞^+.

Definition 2.2. A weight $\underline{k}$ is defined by a set $\{\underline{k}_\sigma\}_{\sigma\in\Sigma_\infty}$ where

$$\underline{k}_\sigma=(c_{r+s,\sigma},\ldots,c_{s+1,\sigma};c_{1,\sigma},\ldots,c_{s,\sigma})$$

with $c_{1,\sigma}\ge\cdots\ge c_{s,\sigma}\ge c_{s+1,\sigma}+r+s\ge\cdots\ge c_{s+r,\sigma}+r+s$ for the $c_{i,\sigma}$ in $\mathbb{Z}$.

Remark 2.3. Our convention is different from others in the literature. For example, in [Hsieh 2014] the a_{r+1-i} there for $1\le i\le r$ is our $-c_{s+i}$, and b_{s+1-j} there for $1\le j\le s$ is our c_j. We let $\underline{k}':=(a_1,\ldots,a_r;b_1,\ldots,b_s)$. We also note that if each $\underline{k}_\sigma=(0,\ldots,0;\kappa,\ldots,\kappa)$ then $L^{\underline{k}}(\mathbb{C})$ is 1-dimensional with $\rho^{\underline{k}}(h)=\det\mu(h,i)^\kappa$.

For a weight $\underline{k}=(c_{r+s},\ldots,c_{s+1};c_1,\ldots,c_s)$, we define the representation of $\mathrm{GL}_r\times\mathrm{GL}_s$ with minimal weight $-\underline{k}$ by

$$L_{\underline{k}}=\{f\in\mathbb{O}_{\mathrm{GL}_r\times\mathrm{GL}_s}\mid f(tn_+g)=\underline{k}'^{-1}(t)f(g),\ t\in T_r\times T_s,\ n_+\in N_r\times{}^tN_s\},$$

where $\mathbb{O}_{\mathrm{GL}_r\times\mathrm{GL}_s}$ is the structure sheaf of the algebraic group $\mathrm{GL}_r\times\mathrm{GL}_s$; see [Hsieh 2014, Section 3]. The group action is denoted by $\rho_{\underline{k}}$. We define the functional/$l_{\underline{k}}$ on $L_{\underline{k}}$ by evaluating at the identity and define a model $L^{\underline{k}}(\mathbb{C})$ of the representation $H(\mathbb{C})$ with the highest weight $\underline{k}$ as follows. The underlying space of $L^{\underline{k}}(\mathbb{C})$ is $L_{\underline{k}}(\mathbb{C})$ and the group action is defined by

$$\rho^{\underline{k}}(h)=\rho_{\underline{k}}({}^th^{-1}),\quad h\in H(\mathbb{C}).$$

For a weight $\underline{k}$, define $\|\underline{k}\|=\{\|\underline{k}\|_\sigma\}_{\sigma\in\Sigma}\in\mathbb{Z}^\Sigma$ by

$$\|\underline{k}\|_\sigma:=-c_{s+1,\sigma}-\cdots-c_{s+r,\sigma}+c_{1\sigma}+\cdots+c_{s,\sigma}$$

and $|\underline{k}|\in\mathbb{Z}^{\Sigma\sqcup\Sigma^c}$ by

$$|\underline{k}|=\sum_{\sigma\in\Sigma}(c_{1,\sigma}+\cdots+c_{s,\sigma})\cdot\sigma-(c_{s+1,\sigma}+\cdots+c_{s+r,\sigma})\cdot\sigma^c.$$

Let χ be a Hecke character of $\mathcal{K}$ with infinite type $|\underline{k}|$, i.e., the Archimedean part of χ is given by

$$\chi(z_\infty) = \left(\prod_\sigma z_\sigma^{(c_{1,\sigma}+\cdots+c_{s,\sigma})} \cdot z_{\sigma^c}^{-(c_{s+1,\sigma}+\cdots+c_{s+r,\sigma})}\right).$$

Definition 2.4. Let U be an open compact subgroup in $G(\mathbb{A}_{F,f})$. We denote by $M_{\underline{k}}(U, \mathbb{C})$ the space of holomorphic $L^{\underline{k}}(\mathbb{C})$-valued functions f on $X^+ \times G(\mathbb{A}_{F,f})$ such that, for $\tau \in X^+$, $\alpha \in G(F)^+$ and $u \in U$, we have

$$f(\alpha\tau, \alpha g u) = \mu(\alpha)^{-\|\underline{k}\|} \rho^{\underline{k}}(J(\alpha, \tau)) f(\tau, g).$$

Now we consider automorphic forms on unitary groups in the adelic language. Let $\boldsymbol{i} \in X^+$ and $K_\infty^+ \subset \mathrm{U}(r, s)(F_\infty)$ be the stabilizer of $\boldsymbol{i}$. The space of automorphic forms of weight $\underline{k}$ and level U with central character χ consists of smooth and slowly increasing functions $F : G(\mathbb{A}_F) \to L_{\underline{k}}(\mathbb{C})$ such that, for every $(\alpha, k_\infty, u, z) \in G(F) \times K_\infty^+ \times U \times Z(\mathbb{A}_F)$,

$$F(z\alpha g k_\infty u) = \rho^{\underline{k}}(J(k_\infty, \boldsymbol{i})^{-1}) F(g) \chi^{-1}(z).$$

2C3. *The group* $\mathrm{GU}(s, r)$. Now we consider the unitary group $\mathrm{GU}(s, r)$ which has the same Hermitian space as $\mathrm{GU}(r, s)$ but with the metric $\langle\,,\,\rangle_{s,r} := -\langle\,,\,\rangle_{r,s}$. We define the symmetric domain $X_{s,r} = X_{r,s}$ but with the complex structure such that a function is holomorphic on $X_{s,r}$ if and only if it is holomorphic on $X_{r,s}$ after composition with the map

$$X_{r,s} \to X_{s,r}, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -\bar{x} \\ -\bar{y} \end{pmatrix}.$$

We let $\mathbb{C}^{s,r} = \mathbb{C}(\Sigma)^s \otimes \mathbb{C}(\Sigma)^{r-s} \otimes \mathbb{C}(\Sigma^c)^s$ and define $c_{s,r}$ by $(u_1, u_2, u_3)c_{s,r} = (u_1, u_2, \bar{u}_3)$. For $\mathrm{GU}(s, r)$, we define $p(\tau) : V \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{C}^{s,r}$ by $p(\tau)v = vB(\tau)c_{s,r}$. We define the automorphic factors $\kappa_{s,r}(\alpha, \tau)$ and $\mu_{s,r}(\alpha, \tau)$ by

$$\alpha B(\tau) = B(\alpha\tau)(\overline{\mu_{s,r}(\alpha, \tau)}, \kappa_{s,r}(\alpha, \tau)).$$

We define a weight $\underline{k}$ of $\mathrm{U}(r, s)$ such that $\underline{k} = (c_{r+1,\sigma}, \ldots, c_{r+s,\sigma}; c_{1,\sigma}, \ldots, c_{r,\sigma})_\sigma$ with $c_{1,\sigma} \geq \cdots \geq c_{r,\sigma} \geq c_{r+1,\sigma} + r + s \geq \cdots \geq c_{r+s,\sigma} + r + s$. Using these we can develop the theory of holomorphic automorphic forms on $\mathrm{GU}(s, r)$ similar to the $\mathrm{GU}(r, s)$ case.

2C4. *Embeddings of symmetric domains.* We still follow [Shimura 1997]. Pick one Archimedean place. Write $z = \binom{x}{y} \in X_{r+1,s+1}$, $X_{r,s}$, and $w = \binom{u}{v} \in X_{s,r}$. We define the embeddings ι from $X_{r+1,s+1} \times X_{s,r}$ or $X_{r,s} \times X_{s,r}$ to $X_{r+s+1,r+s+1}$ or $X_{r+s,r+s}$

by

$$\iota(z,w)\to\begin{pmatrix} x & 0 & 0\\ y & \frac{1}{2}\zeta & 0\\ -\zeta^{-1}v^*y & -v^* & -u^*\end{pmatrix}.$$

(The ζ really means the image of ζ at this Archimedean place.) Let $U=RTQ$, for

$$Q=\begin{pmatrix} 1 &&&&&&& -\frac{1}{2}\\ & 1 &&&&&&\\ && 1 &&&& \frac{1}{2}\zeta &\\ &&& 1 & \frac{1}{2} &&&\\ -1 &&&&&&& -\frac{1}{2}\\ &&&&& 1 &&\\ && -1 &&&& -\frac{1}{2}\zeta &\\ &&& -1 & \frac{1}{2} &&& \end{pmatrix},\qquad RT=\begin{pmatrix} 1_{1+s} &&&&&\\ & 2^{-1}1_{r-s} &&& -2^{-1}1_{r-s} &\\ &&& 1_s &&\\ && A &&&\\ & -\zeta^{-1} &&& -\zeta^{-1} &\\ &&&&& 1_s \end{pmatrix},$$

where $A=\begin{pmatrix} & 1\\ 1_s & \end{pmatrix}$. (The U here is the U_v defined in [Shimura 1997, Section 22] and other notations are slightly different.) We also define Q' to be Q with the second and sixth rows and columns (blockwise) deleted. Let

$$R'T'=\begin{pmatrix} 1_s &&&&&\\ & 2^{-1}1_{r-s} &&& -2^{-1}1_t &\\ &&& 1_s &&\\ && A' &&&\\ & -\zeta^{-1} &&& -\zeta^{-1} &\\ &&&&& 1_s \end{pmatrix}$$

with $A'=1_s$. Define $U'=R'T'Q'$. Let $\wp(z)$ be the lower-right $r\times s$ block for $z\in X_{r+1,s+1}$ and $\iota_U(z,w)=(U^{-1}\iota(z,w)$ as in [Shimura 1997, (22.2.1)]. If $z=\binom{x}{y}$ and $z_1=\binom{x_1}{y_1}$, let $\delta(z_1,z)=2^{s-r}\det[i(x_1^*-x)-y_1^*\theta^{-1}y]$. If we write $[h]_S$ for $S^{-1}hS$ then we have $[\mathrm{diag}(g,g_1)]_S\iota_U(z,w)=\iota_U(gz,g_1w)$, $[\mathrm{diag}(g,g_1)]_{S'}\iota_{U'}(z,w)=\iota_{U'}(gz,g_1w)$ and

$$j\big([\mathrm{diag}(g,g_1)]_S,\iota_U(z,w)\big)=\delta(w,\wp(z))^{-1}\delta(gw,\wp(g_1z))\det(\gamma)\overline{j_g(w)}\,j_{g_1}(z).\quad(3)$$

For a function g on $X_{r+s+1,r+s+1}$ or $X_{r+s,r+s}$, we define the pullback g° to be the function on $X_{r+1,s+1}\times X_{s,r}$ or $X_{r,s}\times X_{s,r}$ given by

$$g^\circ(z,w)=\delta(w,\wp(z))^{-k}g(\iota_U(z,w)).$$

Definition 2.5. We define a scalar weight κ of $\mathrm{U}(s,r)$ to be the weight

$$(\underbrace{-\kappa,\dots,-\kappa}_{s};\underbrace{0,\dots,0}_{r}).$$

2D. ***Shimura varieties and Igusa varieties.*** Fix a neat open compact subgroup K of $\mathrm{GU}^0(r,s)(\mathbb{A}_f)$ whose p-component is $\mathrm{GU}^0(r,s)(\mathbb{Z}_p)$; we refer to [Hsieh 2014] for the definitions and arithmetic models of Shimura varieties over the reflex field E, which we denote by $S_G(K)$. It parameterizes isomorphism classes of the quadruples $(A,\lambda,\iota,\bar{\eta}^{(\Box)})_{/S}$, where $\Box$ is a finite set of primes, (A,λ) is a polarized abelian variety over some base ring S, λ is an orbit (see [Hsieh 2014, Definition 2.1]) of prime-to-$\Box$ polarizations of A, ι is an embedding of $\mathcal{O}_{\mathcal{K}}$ into the endomorphism ring of A, and $\bar{\eta}^{(\Box)}$ is some prime-to-$\Box$ level structure of A. To each point $(\tau,g)\in X^+\times G(\mathbb{A}_{F,f})$, we attach a quadruple as follows:

- The abelian variety $\mathcal{A}_g(\tau):=V\otimes_{\mathbb{Q}}\mathbb{R}/M_{[g]}(M_{[g]}:=H_1(\mathcal{A}_g(\tau),\hat{\mathbb{Z}}^p))$.
- The polarization of $\mathcal{A}$ is given by the pullback of $-\langle\,,\rangle_{r,s}$ on $\mathbb{C}^{r,s}$ to $V\otimes_{\mathbb{Q}}\mathbb{R}$ via $p(\tau)$.
- The complex multiplication ι is the $\mathcal{O}_{\mathcal{K}}$-action induced by the action on V.
- The prime-to-p level structure $\eta_g^{(p)}:M\otimes\hat{\mathbb{Z}}^p\simeq M_{[g]}$ is defined by $\eta_g^{(p)}(x)=g*x$ for $x\in M$.

We have a similar theory for Shimura varieties for $\mathrm{GU}(s,r)$ as well.

There is also a theory of compactifications of $S_G(K)$, developed in [Lan 2008]. We let $\bar{S}_G(K)$ be a fixed choice of a toroidal compactification and $S^*_G(K)$ the minimal compactification.

We define some level groups at p, as in [Hsieh 2014, Section 1.10]. Recall that $G(\mathbb{A}_f)\supseteq K=\prod_V K_v$ is an open compact subgroup such that $K_p=G(\mathbb{Z}_p)$ and let Σ be a finite set of primes including all primes above p such that K_v is spherical for all $v\notin\Sigma$. If we write $g_p=\left(\begin{smallmatrix}A&B\\C&D\end{smallmatrix}\right)$ for the p-component of g, then define

$$K^n=\left\{g\in K\;\middle|\;g_p\equiv\begin{pmatrix}1_r&*\\0&1_s\end{pmatrix}\bmod p^n\right\},$$
$$K_1^n=\{g\in K\mid A\in N_r(\mathbb{Z}_p)\bmod p^n,\ D\in N_s^-(\mathbb{Z}_p)\bmod p^n,\ C=0\},$$
$$K_0^n=\{g\in K\mid A\in B_r(\mathbb{Z}_p)\bmod p^n,\ D\in B_s^-(\mathbb{Z}_p)\bmod p^n,\ C=0\}.$$

Now we recall briefly the notion of Igusa schemes over $\mathcal{O}_{v_0}$ (the localization of the integer ring of the reflex field at the p-adic place v_0 determined by $\iota_p:\mathbb{C}\simeq\mathbb{C}_p$) in [Hsieh 2014, Section 2]. Let V be the Hermitian space for $\mathrm{U}(r,s)$, M be a standard lattice of V and $M_p=M\otimes_{\mathbb{Z}}\mathbb{Z}_p$. Let $\mathrm{Pol}_p=\{N^{-1},N^0\}$ be a polarization of M_p. The Igusa variety $I_G(K^n)$ of level p^n is the scheme representing the usual quadruple for a Shimura variety together with

$$j:\mu_{p^n}\otimes_{\mathbb{Z}}N^0\hookrightarrow A[p^n],$$

where A is the abelian variety in the quadruple. Note that the existence of j implies that if p is nilpotent in the base ring then A must be ordinary. For any integer $m > 0$, let $\mathbb{O}_m := \mathbb{O}_{v_0}/p^m$.

Igusa schemes over $\bar{S}_G(K)$. To define p-adic automorphic forms one needs Igusa schemes over $\bar{S}_G(K)$. We fix such a toroidal compactification and refer to [Hsieh 2014, Section 2.7.6] for the construction. We still denote it by $I_G(K^n)$. Then, over $\mathbb{O}_m$, $I_G(K^n)$ is a Galois covering of the ordinary locus of the Shimura variety with Galois group $\prod_{v|p} \mathrm{GL}_r(\mathbb{O}_{F,v}/p^n) \times \mathrm{GL}_s(\mathbb{O}_{F,v}/p^n)$. We write $I_G(K_0^n) = I_G(K^n)^{K_0^n}$ and $I_G(K_1^n) = I_G(K^n)^{K_1^n}$ over $\mathbb{O}_m$.

Cusps. Let $1 \le t \le s$. We let P_t be the maximal parabolic subgroup of $\mathrm{GU}(r, s)$ consisting of matrices which, in the block form with respect to $t + (r + s - 2t) + t$, are of the form

$$\begin{pmatrix} \times & \times & \times \\ & \times & \times \\ & & \times \end{pmatrix}.$$

Let G_{P_t} be the unitary similitude group with respect to the skew-Hermitian space for ζ. Let Y_t be the $\mathbb{O}_{\mathcal{K}}$ span of $\{y^1, \dots, y^t\}$. We define the set of cusp labels by

$$C_t(K) := (\mathrm{GL}(Y_t) \times G_{P_t}(\mathbb{A}_f))N_{P_t}(\mathbb{A}_f)\backslash G(\mathbb{A}_f)/K.$$

This is a finite set. We write $[g]$ for the class represented by $g \in G(\mathbb{A}_f)$. For each such g whose p-component is 1, we define $K^g_{P_t} = G_{P_t}(\mathbb{A}_f) \cap gKg^{-1}$ and write $S_{[g]} := S_{G_{P_t}}(K^g_{P_t})$ the corresponding Shimura variety for the group G_P with level group $K^g_{P_t}$). By the strong approximation we can choose a set $\underline{C}_t(K)$ of representatives of $C_t(K)$ consisting of elements $g = pk^0$ for $p \in P_t(\mathbb{A}_f^\Sigma)$ and $k^0 \in K^0$ for K^0 the spherical compact subgroup.

Definition 2.6 (p-adic cusps). As in [Hsieh 2014], each pair $(g, h) \in C_t(K) \times H(\mathbb{Z}_p)$ can be regarded as a p-adic cusp, i.e., cusps of the Igusa tower.

Igusa schemes for unitary groups. We refer to [Hsieh 2014, Section 2.5] for the notion of Igusa schemes for the unitary groups $\mathrm{U}(r, s)$ (not the similitude group). It parameterizes quintuples $(A, \lambda, \iota, \bar{\eta}^{(p)}, j)_{/S}$ similar to the Igusa schemes for unitary similitude groups but requiring λ to be a prime-to-p polarization of A (instead of an orbit). In order to use the pullback formula algebraically we need a map of Igusa schemes, given by

$$\begin{aligned} i\big([(A_!, \lambda_!, \iota_1, \eta_!^p K_!, j_1)], [(A_2, \lambda_2, \iota_2, \eta_2^p K_2, j_2)]\big) \\ = [(A_1 \times A_2, \lambda_1 \times \lambda_2, \iota_1, \iota_2, (\eta_1^p \times \eta_2^p)K_3, j_1 \times j_2)]. \end{aligned}$$

Similar to [Hsieh 2014], we know that, taking the change of polarization into consideration,

$$i([z,g],[w,h]) = [\iota(z,w),(g,h)\Upsilon]$$

(Υ is as defined at the end of Section 2B).

2D1. *Geometric modular forms.* Let $H = \prod_{v|p}(\mathrm{GL}_r \times \mathrm{GL}_s)$ and let $N \subset H$ be $\prod_{v|p}(N_r \times {}^tN_s)$. To save notation we also write $H = \prod_{v|p} \mathrm{GL}_r(\mathcal{O}_{F,v}) \times \mathrm{GL}_s(\mathcal{O}_{F,v})$ and let $N \subset H$ be $\prod_{v|p} N_r(\mathcal{O}_{F,v}) \times {}^tN_s(\mathcal{O}_{F,v})$. We define $\underline{\omega} = e^*\Omega_{\mathcal{G}/\bar{S}_G(K)}$ for Ω the sheaf of differentials on the universal semiabelian scheme $\mathcal{G}$ over the toroidal compactification (see [Hsieh 2014, Section 2.7.2] for a brief discussion). Recall that for $v|p$ we have $v = w\bar{w}$ in $\mathcal{K}$ with $w \in \Sigma_p$. Let e_w and $e_{\bar{w}}$ be the corresponding projections for $\mathcal{K}_v \simeq \mathcal{K}_w \times \mathcal{K}_{\bar{w}}$; then $\underline{\omega} = e_w\underline{\omega} \oplus e_{\bar{w}}\underline{\omega}$. We also define

$$\begin{aligned}
\mathcal{E}^+ &:= \underline{\mathrm{Isom}}(\mathcal{O}^r_{\bar{S}_G(K)}, e_w\underline{\omega}),\\
\mathcal{E}^- &:= \underline{\mathrm{Isom}}(\mathcal{O}^s_{\bar{S}_G(K)}, e_{\bar{w}}\underline{\omega}),\\
\mathcal{E} &:= \mathcal{E}^+ \oplus \mathcal{E}^-.
\end{aligned}$$

This is an H-torsor over $\bar{S}_G(K)$. We can define the automorphic sheaf $\omega_{\underline{k}} = \mathcal{E} \times^H L_{\underline{k}}$. A section f of $\underline{\omega_{\underline{k}}}$ is a morphism $f : \mathcal{E} \to L_{\underline{k}}$ such that

$$f(x, h\boldsymbol{w}) = \rho_{\underline{k}}(h) f(x, \boldsymbol{\omega}), \quad h \in H,\ x \in \bar{S}_G(K).$$

2E. ***p-adic automorphic forms on unitary groups.*** Let R be a p-adic $\mathbb{Z}_p$-algebra and let $R_m := R/p^m$. Let $T_{n,m} := I_G(K^n)_{/R_m}$. Define

$$\begin{aligned}
V_{n,m} &= H^0(T_{n,m}, \mathcal{O}_{T_{n,m}}),\\
V_{\underline{k}}(K^n_\bullet, R_m) &= H^0(T_{n,m/R_m}, \omega_{\underline{k}})^{K^n_\bullet}.
\end{aligned}$$

Let $V_{\infty,m} = \varinjlim_n V_{n,m}$ and $V_{\infty,\infty} = \varprojlim_m V_{\infty,m}$. Define $V_p(G,K) := V^N_{\infty,\infty}$, the space of p-adic modular forms. Let $\boldsymbol{T} = T(\mathbb{Z}_p) \subset H$ and let $\Lambda_{\boldsymbol{T}} := \mathbb{Z}_p[\![\boldsymbol{T}]\!]$. The Galois action of $\boldsymbol{T}$ on $V^N_{\infty,m}$ makes the space of p-adic modular forms a discrete $\Lambda_{\boldsymbol{T}}$-module.

Suppose $n \geq m$. To each R_m-quintuple $(\underline{A}, j)$ of level K^n, we can attach a canonical basis $\boldsymbol{\omega}(j)$ of $H^0(A, \Omega_A)$. Therefore, we have a canonical isomorphism

$$H^0(T_{n,m/R_m}, \omega_{\underline{k}}) \simeq V_{n,m} \otimes L_{\underline{k}}(R_m)$$

given by

$$f \mapsto \hat{f}(\underline{A}, j) = f(\underline{A}, j, \boldsymbol{\omega}(j)).$$

We call $\hat{f}$ the p-adic avatar of f.

Similarly, we can define an embedding of geometric modular forms into p-adic modular forms by

$$f \mapsto \hat{f}(\underline{A}, j) = f(\underline{A}, \boldsymbol{\omega}(j)).$$

We also define the morphism

$$\beta_{\underline{k}} : V_{\underline{k}}(K_1^n, R_m) \to V_{n,m}^N, \quad f \mapsto \beta_{\underline{k}}(f) := l_{\underline{k}}(\hat{f}).$$

We can also pass to the limit for $m \to \infty$ to get the embedding of $V_{\underline{k}}(K_1^n, R)$ into $V_{\infty,\infty}^N$. We refer to [Hsieh 2014, Sections 3.8–3.9] for the definition of a U_p Hecke operator and define Hida's ordinary projector

$$e := \lim_n U_p^{n!}.$$

2F. ***Algebraic theory for Fourier–Jacobi expansions.*** We suppose $s > 0$ in this subsection. Let $X_t^\vee = \mathrm{span}_{\mathcal{O}_\mathcal{K}}\{x^1, \cdots, x^t\}$ and $Y_t = \mathrm{span}_{\mathcal{O}_\mathcal{K}}\{y^1, \ldots, y^t\}$. Let W_t be the skew-Hermitian space $\mathrm{span}_{\mathcal{O}_\mathcal{K}}\{y^{t+1}, \ldots, y^s, w_1, \ldots, x^{t+1}, \ldots, x^s\}$. Let G_t^0 be the unitary similitude group of W_t. Let $[g] \in C_t(K)$ and $K_{G_{P_t}} = G_{P_t}(\mathbb{A}_f) \cap gKg^{-1}$ (we suppress the subscript $[g]$ so as to not make the notation too cumbersome). Let $\mathcal{A}_t$ be the universal abelian scheme over the Shimura variety $S_{G_{P_t}}(K_{G_{P_t}})$. Write $g^\vee = kg_i^\vee\gamma$ for $\gamma \in G(F)^+$ and $k \in K$. Define $X_g^\vee = X_t^\vee g_i^\vee \gamma$ and $Y_g = Y_t g_i^\vee \gamma$. Let $X_g = \{y \in (Y_t \otimes_\mathbb{Q} \mathbb{Z}) \cdot \gamma \mid \langle y, X_g^\vee \rangle \in \mathbb{Z}\}$. Then we have

$$i : Y_g \hookrightarrow X_g.$$

Let $\mathcal{Z}_{[g]}$ be

$$\underline{\mathrm{Hom}}_{\mathcal{O}_\mathcal{K}}(X_g, \mathcal{A}_t^\vee) \times_{\underline{\mathrm{Hom}}_{\mathcal{O}_\mathcal{K}}(Y_g, \mathcal{A}_t^\vee)} \underline{\mathrm{Hom}}_{\mathcal{O}_\mathcal{K}}(Y_g, \mathcal{A}_t) \\ := \{(c, c^t) \mid c(i(y)) = \lambda(c^t(y)),\ y \in Y_g\}.$$

Here the $\underline{\mathrm{Hom}}$ are the obvious sheaves over the big étale site of $S_{G_{P_t}}$, represented by abelian schemes. Let $\boldsymbol{c}$ and $\boldsymbol{c}^\vee$ be the universal morphisms over $\underline{\mathrm{Hom}}_{\mathcal{O}_\mathcal{K}}(X_g, \mathcal{A}_t^\vee)$ and $\underline{\mathrm{Hom}}_{\mathcal{O}_\mathcal{K}}(Y_g, \mathcal{A}_t)$. Let N_{P_t} be the unipotent radical of P_t and $Z(N_{P_t})$ be its center. Let $H_{[g]} := Z(N_{P_t}(F)) \cap g_i K g_i^{-1}$. Note that if we replace the components of K at $v|p$ by K_1^n then the set $H_{[g]}$ remains unchanged. Let $\Gamma_{[g]} := \mathrm{GL}_\mathcal{K}(Y_t) \cap g_i K g_i^{-1}$. Let $\mathcal{P}_{\mathcal{A}_t}$ be the Poincaré sheaf over $\mathcal{A}_t^\vee \times \mathcal{A}_t/_{\mathcal{Z}_{[g]}}$ and $\mathcal{P}_{\mathcal{A}_t}^\times$ its associated $\mathbb{G}_m$-torsor. Let $S_{[g]} := \mathrm{Hom}(H_{[g]}, \mathbb{Z})$. For any $h \in S_{[g]}$, let $c(h)$ be the tautological map $\mathcal{Z}_{[g]} \to \mathcal{A}_t^\vee \times \mathcal{A}_t$ and $\mathcal{L}(h) := c(h)^*\mathcal{P}_{\mathcal{A}_t}^\times$ its associated $\mathbb{G}_m$ torsor over $\mathcal{Z}_{[g]}$.

It is well known (see [Lan 2008, Chapter 7], for example) that the minimal compactification $S_G^*(K)$ is the disjoint union of boundary components corresponding to each $t = 1, \ldots, s$. Let $\mathcal{O}_{\mathbb{C}_p}$ be the valuation ring for $\mathbb{C}_p$. The following proposition is proved in [Lan 2008, Proposition 7.2.3.16]. Let $[g] \in C_t(K)$ and $\bar{x}$ be a $\mathcal{O}_{\mathbb{C}_p}$-point of the t-stratum of $S_G^*(K)(1/E)$ corresponding to $[g]$.

Proposition 2.7. *Let $[g]$ and $\bar{x}$ be as above. We write the subscript $\bar{x}$ to mean formal completion along $\bar{x}$. Let π be the map $\bar{S}_G(K) \to S_G^*(K)$. Then $\pi_*(\mathcal{O}_{\bar{S}_G(K)})_{\bar{x}}$ is isomorphic to*

$$\left\{ \sum_{h \in S_{[g]}^+} H^0(\mathfrak{X}_{[g]}, \mathcal{L}(h))_{\bar{x}} q^h \right\}^{\Gamma_{[g]}}.$$

Here $S_{[g]}^+$ means the totally nonnegative elements in $S_{[g]}$. The q^h is just regarded as a formal symbol and $\Gamma_{[g]}$ acts on the set by a certain formula, which we omit.

For each $[g] \in C_t(K)$ we fix an $\bar{x}$ corresponding to it as above. Now we consider the diagram

$$\begin{array}{ccc} T_{n,m} & \xrightarrow{\pi_{n,m}} & T_{n,m}^* \\ \downarrow & & \downarrow \\ \bar{S}_G(K)[1/E]_{\mathcal{O}_m} & \xrightarrow{\pi} & S_G^*(K)[1/E]_{\mathcal{O}_m}, \end{array}$$

where $T_{n,m} \to T_{n,m}^* \to S_G^*(K)[1/E]_{\mathcal{O}_m}$ is the Stein factorization. By [Lan 2013b, Corollary 6.2.2.8], $T_{n,m}^*$ is finite étale over $S_G^*(K)[1/E]_{\mathcal{O}_m}$. Take a preimage of $\bar{x}$ in $T_{n,m}^*$, which we still denote by $\bar{x}$. (To do this, we have to extend the field of definition to include the maximal unramified extension of L.) Then the formal completion of the structure sheaf of $T_{n,m}^*$ and $S_G^*(K)[1/E]_{\mathcal{O}_m}$ at $\bar{x}$ are isomorphic. So, for any p-adic automorphic form $f \in \varprojlim_m \varinjlim_n H^0(T_{n,m}, \mathcal{O}_{n,m})$ (with trivial coefficients), we have a Fourier–Jacobi coefficient

$$\mathrm{FJ}(f) \in \left\{ \prod_{h \in S_{[g]}^+} \varprojlim_m \varinjlim_n H^0(\mathfrak{X}_{[g]}, \mathcal{L}(h))_{\bar{x}} \cdot q^h \right\}_{[g]} \tag{4}$$

by considering f as a global section of $\pi_{n,m}^*(\mathcal{O}_{T_{n,m}}) = \mathcal{O}_{T_{n,m}^*}$ and pullback at the $\bar{x}$. Note that if $t = s = 1$ then there is no need to choose the $\bar{x}$ and pullback, since the Shimura variety for G_t is 0-dimensional (see [Hsieh 2014, (2.18)]). In application, when we construct families of Klingen Eisenstein series in terms of Fourier–Jacobi coefficients, we will take $t = 1$ and define

$$\mathcal{R}_{[g],\infty} := \prod_{h \in S_{[g]}^+} \varprojlim_m \varinjlim_n H^0(\mathfrak{X}_{[g]}, \mathcal{L}(h))_{\bar{x}} \cdot q^h. \tag{5}$$

We remark that the map FJ is injective on the space of forms with prescribed nebentypus at p (this is not needed for our result though). This can be seen using the discussion in [Skinner and Urban 2014] right before Section 6.2 (which in turn uses result of Hida [2011] about the irreducibility of Igusa towers for the group $\mathrm{SU}(r, s) \subset \mathrm{U}(r, s)$, the kernel of the determinant map). In particular, to see this injectivity we need the fact that there is a bijection between the irreducible components of the generic and special fibers of $S_G^*(K)$ (see [Lan 2008, Section 6.4.1]) and that there is at least one cusp of any given genus on the ordinary locus of each irreducible component. (Note that the signature is (r, s) for $r \geq s > 0$ at all Archimedean places, so there is at least one cusp in $C_t(K)$ at each irreducible component. Since p splits completely in $\mathcal{K}$, the cusps of minimal genus must be in the ordinary locus. On the other hand by the construction of minimal compactification the closure of the stratum of any genus r is the union of all stratums of genus less than or equal to r. Note also that, since the geometric fibers of the minimal compactification are normal, their irreducible components are also connected components. This implies the existence of such a cusp on the ordinary locus.) See the Appendix for more details.

3. Eisenstein series and Fourier–Jacobi coefficients

The materials of this section are straightforward generalizations of parts of [Skinner and Urban 2014, Sections 9 and 11] and we use the same notations as there, so everything in this section should work out the same as [Skinner and Urban 2014] when specialized to the group $\mathrm{GU}(2, 2)_{/\mathbb{Q}}$.

3A. ***Klingen Eisenstein series.*** Let $\mathfrak{gu}(\mathbb{R})$ be the Lie algebra of $\mathrm{GU}(r, s)(\mathbb{R})$. Let δ be a character of the Klingen parabolic subgroup P such that $\delta^{a+2b+1} = \delta_P$ (the modulus character of P).

3A1. *Archimedean picture.* Let v be an infinite place of F, so that $F_v \simeq \mathbb{R}$. Let $\boldsymbol{i}'$ and $\boldsymbol{i}$ be the points on the Hermitian symmetric domain for $\mathrm{GU}(r, s)$ and $\mathrm{GU}(r+1, s+1)$ which are $\binom{i1_s}{0}$ and $\binom{i1_{s+1}}{0}$, respectively (here 0 means the $(r-s) \times s$ or $(r-s) \times (s+1)$ matrix 0). Let $\mathrm{GU}(r, s)(\mathbb{R})^+$ be the subgroup of $\mathrm{GU}(r, s)(\mathbb{R})$ whose similitude factor is positive. Let K_∞^+ and $K_\infty^{+,\prime}$ be the compact subgroups of $\mathrm{U}(r+1, s+1)(\mathbb{R})$ and $\mathrm{U}(r, s)(\mathbb{R})$ stabilizing $\boldsymbol{i}$ or $\boldsymbol{i}'$, and let K_∞ (resp. K'_∞) be the group generated by K_∞^+ (resp. $K_\infty^{+,\prime}$) and $\mathrm{diag}(1_{r+s+1}, -1_{s+1})$ (resp. $\mathrm{diag}(1_{r+s}, -1_s)$).

Now let (π, H) be a unitary tempered Hilbert representation of $\mathrm{GU}(r, s)(\mathbb{R})$ with H_∞ the space of smooth vectors. We define a representation of $P(\mathbb{R})$ on H_∞ as follows: for $p = mn$, where $n \in N_P(\mathbb{R})$ and $m = m(g, a) \in M_P(\mathbb{R})$ with $a \in \mathbb{C}^\times$, $g \in \mathrm{GU}(r+1, s+1)(\mathbb{R})$, put

$$\rho(p)v := \tau(a)\pi(g)v, \quad v \in H_\infty.$$

We define a representation by smooth induction, $I(H_\infty) := \mathrm{Ind}_{P(\mathbb{R})}^{\mathrm{GU}(r+1,s+1)(\mathbb{R})} \rho$ and write $I(\rho)$ for the space of K_∞-finite vectors in $I(H_\infty)$. For $f \in I(\rho)$ we also define, for each $z \in \mathbb{C}$, a function

$$f_z(g) := \delta(m)^{(a+2b+1)/2+z} \rho(m) f(k), \quad g = mk \in P(\mathbb{R})K_\infty,$$

and an action of $\mathrm{GU}(r+1, s+1)(\mathbb{R})$ on it by

$$(\sigma(\rho, z)(g) f)(k) := f_z(kg).$$

Let $(\pi^\vee, V)$ be the irreducible $(\mathfrak{gu}(\mathbb{R}), K'_\infty)$-module given by $\pi^\vee(x) = \pi(\eta^{-1} x \eta)$ for

$$\eta = \begin{pmatrix} & & 1_b \\ & 1_a & \\ -1_b & & \end{pmatrix}$$

and x in $\mathfrak{gu}(\mathbb{R})$ or K'_∞ (this does not mean the contragradient representation!). Let $\rho^\vee$, $I(\rho^\vee)$, $I^\vee(H_\infty)$, $\sigma(\rho^\vee, z)$ and $I(\rho^\vee)$ be the representations and spaces defined as above but with π and τ replaced by $\pi^\vee \otimes (\tau \circ \det)$ and $\bar{\tau}^c$. We are going to define an intertwining operator. Let

$$w = \begin{pmatrix} & & 1_{b+1} \\ & 1_a & \\ -1_{b+1} & & \end{pmatrix}.$$

For any $z \in \mathbb{C}$, $f \in I(H_\infty)$ and $k \in K_\infty$, consider the integral

$$A(\rho, z, f)(k) := \int_{N_P(\mathbb{R})} f_z(wnk)\, dn. \tag{6}$$

This is absolutely convergent when $\mathrm{Re}(z) > \frac{1}{2}(a + 2b + 1)$ and $A(\rho, z, -)$ in $\mathrm{Hom}_{\mathbb{C}}(I(H_\infty), I^\vee(H_\infty))$ intertwines the actions of $\sigma(\rho, z)$ and $\sigma(\rho^\vee, -z)$.

Suppose π is the holomorphic discrete series representation associated to the (scalar) weight $(0, \dots, 0; \kappa, \dots, \kappa)$; then it is well known that there is a unique (up to scalar) vector $v \in \pi$ such that $k \cdot v = \det \mu(k, i)^{-\kappa}$ (here μ means the second component of the automorphic factor J instead of the similitude character) for any $k \in K^{+,\prime}_\infty$. Then, by the Frobenius reciprocity law, there is a unique (up to scalar) vector $\tilde{v} \in I(\rho)$ such that $k \cdot \tilde{v} = \det \mu(k, i)^{-\kappa} \tilde{v}$ for any $k \in K^+_\infty$. We fix v and multiply $\tilde{v}$ by a constant so that $\tilde{v}(1) = v$. In $\pi^\vee$, $\pi(w)v$ has the action of K^+_∞ given by multiplying by $\det \mu(k, i)^{-\kappa}$. We define $w' \in \mathrm{U}(r+1, s+1)$ by

$$w' = \begin{pmatrix} 1_b & & & & \\ & & & & 1 \\ & & 1_a & & \\ & & & 1_b & \\ & -1 & & & \end{pmatrix}.$$

There is a unique vector $\tilde{v}^\vee \in I(\rho^\vee)$ such that the action of K^+_∞ is given by $\det \mu(k, i)^{-\kappa}$ and $\tilde{v}^\vee(w') = \pi(w)v$. Then, by uniqueness, there is a constant $c(\rho, z)$ such that $A(\rho, z, \tilde{v}) = c(\rho, z)\tilde{v}^\vee$.

Definition 3.1. We define $F_\kappa \in I(\rho)$ to be the $\tilde{v}$ as above.

3A2. *Prime-to-p picture.* Our discussion here follows [Skinner and Urban 2014, §9.1.2]. Let (π, V) be an irreducible, admissible representation of $\mathrm{GU}(r,s)(F_v)$ which is unitary and tempered. Let ψ and τ be unitary characters of $\mathcal{K}_v^\times$ such that ψ is the central character for π. We define a representation ρ of $P(F_v)$ as follows. For $p = mn$, where $n \in N_P(F_v)$ and $m = m(g,a) \in M_P(F_v)$ with $a \in K_v^\times$ and $g \in \mathrm{GU}(F_v)$, let

$$\rho(p)v := \tau(a)\pi(g)v, \quad v \in V.$$

Let $I(\rho)$ be the representation defined by admissible induction, that is, $I(\rho) = \mathrm{Ind}_{P(F_v)}^{\mathrm{GU}(r+1,s+1)(F_v)} \rho$. As in the Archimedean case, for each $f \in I(\rho)$ and each $z \in \mathbb{C}$ we define a function f_z on $\mathrm{GU}(r+1, s+1)(F_v)$ by

$$f_z(g) := \delta(m)^{(a+2b+1)/2+z} \rho(m) f(k) \quad \text{for } g = mk \in P(F_v)K_v$$

and a representation $\sigma(\rho, z)$ of $\mathrm{GU}(r+1, s+1)(F_v)$ on $I(\rho)$ by

$$(\sigma(\rho, z)(g) f)(k) := f_z(kg).$$

Let $(\pi^\vee, V)$ be given by $\pi^\vee(g) = \pi(\eta^{-1} g\eta)$. This representation is also tempered and unitary. We denote by $\rho^\vee$, $I(\rho^\vee)$ and $(\sigma(\rho^\vee, z), I(\rho^\vee))$ the representations and spaces defined as above but with π and τ replaced by $\pi^\vee \otimes (\tau \circ \det)$ and $\bar{\tau}^c$, respectively.

For $f \in I(\rho)$, $k \in K_v$ and $z \in \mathbb{C}$, consider the integral

$$A(\rho, z, v)(k) := \int_{N_P(F_v)} f_z(wnk)\, dn. \tag{7}$$

As a consequence of our hypotheses on π this integral converges absolutely and uniformly for z and k in compact subsets of $\{z \mid \mathrm{Re}(z) > \frac{1}{2}(a+2b+1)\} \times K_v$. Moreover, for such z, $A(\rho, z, f) \in I(\rho^\vee)$, and the operator $A(\rho, z, -) \in \mathrm{Hom}_{\mathbb{C}}(I(\rho), I(\rho^\vee))$ intertwines the actions of $\sigma(\rho, z)$ and $\sigma(\rho^\vee, -z)$.

For any open subgroup $U \subseteq K_v$, let $I(\rho)^U \subseteq I(\rho)$ be the finite-dimensional subspace consisting of functions satisfying $f(ku) = f(k)$ for all $u \in U$. Then the function

$$\{z \in \mathbb{C} \mid \mathrm{Re}(z) > \tfrac{1}{2}(a+2b+1)\} \to \mathrm{Hom}_{\mathbb{C}}(I(\rho)^U, I(\rho^\vee)^U), \quad z \mapsto A(\rho, z, -)$$

is holomorphic. This map has a meromorphic continuation to all of $\mathbb{C}$.

We finally remark that, when π and τ are unramified, there is a unique (up to scalar) unramified vector $F_{\rho_v} \in I(\rho)$.

3A3. *Global picture.* We follow [Skinner and Urban 2014, §9.1.4]. Let (π, V) be an irreducible, cuspidal, tempered automorphic representation of $\mathrm{GU}(r,s)(\mathbb{A}_F)$. This is an admissible $(\mathfrak{gu}(\mathbb{R}), K'_\infty)_{v|\infty} \times \mathrm{GU}(r,s)(\mathbb{A}_f)$-module which is a restricted tensor product of local irreducible admissible representations. Let $\psi, \tau : \mathbb{A}_{\mathcal{K}}^\times \to \mathbb{C}^\times$

be Hecke characters such that ψ is the central character of π. Let $\tau = \bigotimes \tau_w$ and $\psi = \bigotimes \psi_w$ be their local decompositions, w running over places of F. Define a representation of $(P(F_\infty) \cap K_\infty) \times P(\mathbb{A}_{F,f})$ by putting

$$\rho(p)v := \bigotimes(\rho_w(p_w)v_w),$$

Let $I(\rho)$ be the restricted product $\bigotimes I(\rho_w)$ with respect to the F_{ρ_w} at those w at which τ_w, ψ_w and pi_w are unramified. As before, for each $z \in \mathbb{C}$ and $f \in I(\rho)$, we define a function f_z on $\mathrm{GU}(r+1, s+1)(\mathbb{A}_F)$ as

$$f_z(g) := \bigotimes f_{w,z}(g_w),$$

where $f_{w,z}$ are defined as before, and an action $\sigma(\rho, z)$ of

$$(\mathfrak{gu}, K_\infty) \otimes \mathrm{GU}(r+1, s+1)(\mathbb{A}_f)$$

on $I(\rho)$ by $\sigma(\rho, z) := \bigotimes \sigma(\rho_w, z)$. Similarly, we define $\rho^\vee$, $I(\rho^\vee)$ and $\sigma(\rho^\vee, z)$, but with the corresponding things replaced by their $\vee$, and we have global versions of the intertwining operators $A(\rho, f, z)$.

Definition 3.2. Let Σ be a finite set of primes of F containing all the infinite places, primes dividing p, and places where π or τ is ramified. Then we call the triple $\mathfrak{D} = (\pi, \tau, \Sigma)$ an Eisenstein datum.

3A4. *Klingen-type Eisenstein series on G.* We follow [Skinner and Urban 2014, §9.1.5] in this subsubsection. Let π, ψ and τ be as above. For $f \in I(\rho)$ and $z \in \mathbb{C}$, there are maps from $I(\rho)$ and $I(\rho^\vee)$ to spaces of automorphic forms on $P(\mathbb{A}_F)$ given by

$$f \longmapsto (g \mapsto f_z(g)(1)).$$

In the following we often write f_z for the automorphic form on $P(\mathbb{A}_F)$ given by this recipe.

If $g \in \mathrm{GU}(r+1, s+1)(\mathbb{A}_F)$, it is well known that

$$E(f, z, g) := \sum_{\gamma \in P(F)\backslash G(F)} f_z(\gamma g) \tag{8}$$

converges absolutely and uniformly for (z, g) in compact subsets of

$$\left\{z \in \mathbb{C} \,\middle|\, \mathrm{Re}(z) > \tfrac{1}{2}(a+2b+1)\right\} \times \mathrm{GU}(r+1, s+1)(\mathbb{A}_F).$$

Therefore, we get some automorphic forms which are called Klingen Eisenstein series.

Definition 3.3. For any parabolic subgroup R of $\mathrm{GU}(r+1, s+1)$ and an automorphic form φ, we define φ_R to be the constant term of φ along R, defined by

$$\varphi_R(g)=\int_{n\in N_R(F)\backslash N_R(\mathbb{A}_F)}\varphi(ng)\,dn.$$

The following lemma is well known (see [Skinner and Urban 2014, Lemma 9.2]).

Lemma 3.4. *Let R be a standard F-parabolic subgroup of* $\mathrm{GU}(r+1,s+1)$ (*i.e*, $R\supseteq B$, *where B is the standard Borel subgroup*). *Suppose* $\mathrm{Re}(z)>\frac{1}{2}(a+2b+1)$. *Then*:

(i) *If $R\neq P$ then $E(f,z,g)_R=0$.*

(ii) $E(f,z,-)_P=f_z+A(\rho,f,z)_{-z}$.

3B. *Siegel Eisenstein series on G_n.*

3B1. *Local picture.* Our discussion in this subsection follows [Skinner and Urban 2014, §§11.1–11.3] closely. Let $Q=Q_n$ be the Siegel parabolic subgroup of GU_n consisting of matrices $\left(\begin{smallmatrix}A_q & B_q\\ 0 & D_q\end{smallmatrix}\right)$. It consists of matrices whose lower-left $n\times n$ block is zero.

For a finite place v of F and a character χ of $\mathcal{K}_v^\times$, we let $I_n(\chi)$ be the space of smooth $K_{n,v}$-finite functions (here $K_{n,v}$ means the open compact group $G_n(\mathcal{O}_{F,v})$) $f:K_{n,v}\to\mathbb{C}$ such that $f(qk)=\chi(\det D_q)f(k)$ for all $q\in Q_n(F_v)\cap K_{n,v}$, where we write q as a block matrix $q=\left(\begin{smallmatrix}A_q & B_q\\ 0 & D_q\end{smallmatrix}\right)$. For $z\in\mathbb{C}$ and $f\in I(\chi)$, we also define a function $f(z,-):G_n(F_v)\to\mathbb{C}$ by $f(z,qk):=\chi(\det D_q)|\det A_qD_q^{-1}|_v^{z+n/2}f(k)$ for $q\in Q_n(F_v)$ and $k\in K_{n,v}$.

For $f\in I_n(\chi)$, $z\in\mathbb{C}$ and $k\in K_{n,v}$, the intertwining integral is defined by

$$M(z,f)(k):=\bar{\chi}^n(\mu_n(k))\int_{N_{Q_n}(F_v)}f(z,w_nrk)\,dr.$$

For z in compact subsets of $\left\{\mathrm{Re}(z)>\frac{1}{2}n\right\}$, this integral converges absolutely and uniformly, with the convergence being uniform in k. In this case it is easy to see that $M(z,f)\in I_n(\bar{\chi}^c)$. A standard fact from the theory of Eisenstein series says that this has a continuation to a meromorphic section on all of $\mathbb{C}$.

Let $\mathcal{U}\subseteq\mathbb{C}$ be an open set. By a meromorphic section of $I_n(\chi)$ on $\mathcal{U}$ we mean a function $\varphi:\mathcal{U}\to I_n(\chi)$ taking values in a finite-dimensional subspace $V\subset I_n(\chi)$ and such that $\varphi:\mathcal{U}\to V$ is meromorphic.

For Archimedean places there is a similar picture (see [loc. cit.]).

3B2. *Global picture.* For an idele class character $\chi=\bigotimes\chi_v$ of $\mathbb{A}_{\mathcal{K}}^\times$, we define a space $I_n(\chi)$ to be the restricted tensor product defined using the spherical vectors $f_v^{\mathrm{sph}}\in I_n(\chi_v)$, $f_v^{\mathrm{sph}}(K_{n,v})=1$, at the finite places v where χ_v is unramified.

For $f\in I_n(\chi)$ we consider the Eisenstein series

$$E(f;z,g):=\sum_{\gamma\in Q_n(F)\backslash G_n(F)}f(z,\gamma g).$$

This series converges absolutely and uniformly for (z, g) in compact subsets of $\left\{\mathrm{Re}(z) > \frac{1}{2}n\right\} \times G_n(\mathbb{A}_F)$. The automorphic form defined is called Siegel Eisenstein series.

Let $\varphi : \mathcal{U} \to I_n(\chi)$ be a meromorphic section; then we put $E(\varphi; z, g) = E(\varphi(z); z, g)$. This is defined at least on the region of absolute convergence and it is well known that it can be meromorphically continued to all $z \in \mathbb{C}$.

Now, for $f \in I_n(\chi)$, $z \in \mathbb{C}$ and $k \in \prod_{v\nmid\infty} K_{n,v} \prod_{v|\infty} K_\infty$, there is a similar intertwining integral $M(z, f)(k)$ as above, but with the integration being over $N_{Q_n}(\mathbb{A}_F)$. This again converges absolutely and uniformly for z in compact subsets of $\left\{\mathrm{Re}(z) > \frac{1}{2}n\right\} \times K_n$. Thus, $z \mapsto M(z, f)$ defines a holomorphic section $\left\{\mathrm{Re}(z) > \frac{1}{2}n\right\} \to I_n(\bar{\chi}^c)$. This has a continuation to a meromorphic section on $\mathbb{C}$. For $\mathrm{Re}(z) > \frac{1}{2}n$, we have

$$M(z, f) = \bigotimes\nolimits_v M(z, f_v), \quad f = \bigotimes f_v.$$

The functional equation for Siegel Eisenstein series is

$$E(f, z, g) = \chi^n(\mu(g))E(M(z, f); -z, g),$$

in the sense that both sides can be meromorphically continued to all $z \in \mathbb{C}$ and the equality is understood as of meromorphic functions of $z \in \mathbb{C}$.

3B3. *The pullback formulas.* Let χ be a unitary idele class character of $\mathbb{A}_{\mathcal{K}}^\times$. Given a unitary, tempered, cuspidal eigenform φ on $\mathrm{GU}(r, s)$ which is a pure tensor, we formally define the integral

$$F_\varphi(f; z, g) := \int_{\mathrm{U}(r,s)(\mathbb{A}_F)} f(z, S^{-1}\alpha(g, g_1h)S)\bar{\chi}(\det g_1g)\varphi(g_1h)\,dg_1,$$
$$f \in I_{r+s+1}(\chi),\ g \in \mathrm{GU}(r+1, s+1)(\mathbb{A}_F),\ h \in \mathrm{GU}(r, s)(\mathbb{A}_F),\ \mu(g) = \mu(h).$$

This is independent of h. (We suppress the χ in the notation for F_φ since its choice is implicitly given by f.) We also formally define

$$F'_\varphi(f; z, g) := \int_{\mathrm{U}(r,s)(\mathbb{A}_F)} f(z, S'^{-1}\alpha(g, g_1h)S')\bar{\chi}(\det g_1g)\varphi(g_1h)\,dg_1,$$
$$f \in I_{r+s}(\chi),\ g \in \mathrm{GU}(r, s)(\mathbb{A}_F),\ h \in \mathrm{GU}(r, s)(\mathbb{A}_F),\ \mu(g) = \mu(h).$$

The pullback formulas are the identities in the following proposition.

Proposition 3.5. *Let χ be a unitary idele class character of $\mathbb{A}_{\mathcal{K}}^\times$.*

(i) *If $f \in I_{r+s}(\chi)$, then $F_\varphi(f; z, g)$ converges absolutely and uniformly for (z, g) in compact sets of $\{\mathrm{Re}(z) > r+s\} \times \mathrm{GU}(r, s)(\mathbb{A}_F)$ and, for any $h \in \mathrm{GU}(r, s)(\mathbb{A}_F)$*

such that $\mu(h)=\mu(g)$,

$$\int_{\mathrm{U}(r,s)(F)\backslash \mathrm{U}(r,s)(\mathbb{A}_F)} E(f;z,S'^{-1}\alpha(g,g_1h)S')\bar{\chi}(\det g_1h)\varphi(g_1h)\,dg_1 = F'_\varphi(f;z,g). \quad (9)$$

(ii) *If* $f\in I_{r+s+1}(\chi)$, *then* $F_\varphi(f;z,g)$ *converges absolutely and uniformly for* (z,g) *in compact sets of* $\left\{\mathrm{Re}(z)>r+s+\frac{1}{2}\right\}\times \mathrm{GU}(r+1,s+1)(\mathbb{A}_F)$ *and, for any* $h\in \mathrm{GU}(r,s)(\mathbb{A}_F)$ *such that* $\mu(h)=\mu(g)$,

$$\int_{\mathrm{U}(r,s)(F)\backslash \mathrm{U}(r,s)(\mathbb{A}_F)} E(f;z,S^{-1}\alpha(g,g_1h)S)\bar{\chi}(\det g_1h)\varphi(g_1h)\,dg_1 = \sum_{\gamma\in P(F)\backslash G(r+1,s+1)(F)} F_\varphi(f;z,\gamma g), \quad (10)$$

with the series converging absolutely and uniformly for (z,g) *in compact subsets of* $\left\{\mathrm{Re}(z)>r+s+\frac{1}{2}\right\}\times \mathrm{GU}(r+1,s+1)(\mathbb{A}_F)$.

Proof. The global integral F_φ and F'_φ can be written as a product of local integrals. The absolute convergence of local integrals for F'_φ is proved in [Lapid and Rallis 2005, Lemma 2]. The absolute convergence for the global integral F'_φ follows from this and the explicit computations in [Lapid and Rallis 2005] at all unramified places, together with the assumption that φ is tempered. The absolute convergence for F_φ is proved in the same way. Then part (i) is proved by Piatetski-Shapiro and Rallis [Gelbart et al. 1987] and (ii) is a straightforward generalization by Shimura [1997], which is in turn due to earlier works of Garrett [1984; 1989]. Both are straightforward consequences of the double coset decompositions in [Shimura 1997, Propositions 2.4 and 2.7]. □

3C. ***Fourier–Jacobi expansion.***

3C1. *Fourier–Jacobi expansion.* We will usually write $e_{\mathbb{A}}(x)=e_{\mathbb{A}_{\mathbb{Q}}}(\mathrm{Tr}_{\mathbb{A}_F/\mathbb{A}_{\mathbb{Q}}}x)$ for $x\in\mathbb{A}_F$. For any automorphic form φ on $\mathrm{GU}(r,s)(\mathbb{A}_F)$, $\beta\in S_m(F)$ for $m\le s$. We define the Fourier–Jacobi coefficient at $g\in \mathrm{GU}(r,s)(\mathbb{A}_F)$ as

$$\varphi_\beta(g)=\int_{S_m(F)\backslash S_m(\mathbb{A}_F)}\varphi\left(\begin{pmatrix}1_s & 0 & \begin{matrix} S & 0\\ 0 & 0\end{matrix}\\ 0 & 1_{r-s} & 0\\ 0 & 0 & 1_s\end{pmatrix} g\right)e_{\mathbb{A}}(-\mathrm{Tr}(\beta S))\,dS.$$

In fact, we are mainly interested in two cases: $m=s$, or $r=s$ and arbitrary $m\le s$. In particular, suppose $G=G_n=\mathrm{GU}(n,n)$, $0\le m\le n$ are integers, and $\beta\in S_m(F)$. Let φ be a function on $G(F)\backslash G(\mathbb{A})$. The β-th Fourier–Jacobi coefficient φ_β of φ

at g is defined by

$$\varphi_\beta(g) := \int \varphi\left(\begin{pmatrix} 1_n & \begin{matrix} S & 0 \\ 0 & 0 \end{matrix} \\ & 1_n \end{pmatrix} g\right) e_{\mathbb{A}}(-\operatorname{Tr}\beta S)\, dS.$$

Now we prove a useful formula on the Fourier–Jacobi coefficients for Siegel Eisenstein series.

Definition 3.6. Put

$$Z := \left\{ \begin{pmatrix} 1_n & \begin{matrix} z & 0 \\ 0 & 0 \end{matrix} \\ 0_n & 1_n \end{pmatrix} \;\middle|\; z \in S_m(\mathcal{K}) \right\},$$

$$V := \left\{ \begin{pmatrix} 1_m & x & z & y \\ & 1_{n-m} & y^* & 0 \\ & & 1_m & \\ 0_n & & -x^* & 1_{n-m} \end{pmatrix} \;\middle|\; x, y \in M_{m(n-m)}(\mathcal{K}),\ z - xy^* \in S_m(\mathcal{K}) \right\},$$

$$X := \left\{ \begin{pmatrix} 1_m & x & & \\ & 1_{n-m} & & 0_n \\ & & 1_m & \\ 0_n & & -x^* & 1_{n-m} \end{pmatrix} \;\middle|\; x \in M_{m(n-m)}(\mathcal{K}) \right\},$$

$$Y := \left\{ \begin{pmatrix} 1_n & \begin{matrix} z & y \\ y^* & 0 \end{matrix} \\ 0_n & 1_n \end{pmatrix} \;\middle|\; y \in M_{m(n-m)}(\mathcal{K}) \right\}.$$

From now on we will usually write w_n for $\left(\begin{smallmatrix} & 1_n \\ -1_n & \end{smallmatrix}\right)$.

Proposition 3.7. *Let f be in $I_n(\tau)$ and suppose $\beta \in S_m(F)$ is totally positive. If $E(f; z, g)$ is the Siegel Eisenstein series on* GU_n *defined by f for some* $\mathrm{Re}(z)$ *sufficiently large, then the β-th Fourier–Jacobi coefficient $E_\beta(f; z, g)$ satisfies*

$$E_\beta(f; z, g) = \sum_{\gamma \in Q_{n-m}(F)\backslash \mathrm{GU}_{n-m}(F)} \sum_{y \in Y} \int_{S_m(\mathbb{A})} f\left(w_n \begin{pmatrix} 1_n & \begin{matrix} S & y \\ {}^t\bar{y} & 0 \end{matrix} \\ & 1_n \end{pmatrix} \alpha_{n-m}(1, \gamma) g\right) e_{\mathbb{A}}(-\operatorname{Tr}\beta S)\, dS,$$

where

$$\alpha_{n-m}(\gamma) = \begin{pmatrix} 1 & & & \\ & D & & C \\ & & 1 & \\ & B & & A \end{pmatrix}$$

if $g_1 = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, where A, B, C and D are $(n-m) \times (n-m)$ matrices.

Proof. We follow [Ikeda 1994, Section 3]. Let H be the normalizer of V in G. Then

$$G_n(F)=\bigsqcup_{i=1}^{m} Q_n(F)\xi_i H(F)$$

for

$$\xi_i := \begin{pmatrix} 0_{m-i} & 0 & -1_{m-i} & 0 \\ 0 & 1_{n-m+i} & 0 & 0 \\ 1_{m-i} & 0 & 0_{m-i} & 0 \\ 0 & 0_{n-m+i} & 0 & 1_{n-m+i} \end{pmatrix}.$$

Unfolding the Eisenstein series, we get

$$E_\beta(f;z,g)=\sum_{i>0}\sum_{\gamma\in Q_n(F)\backslash Q_n(F)\xi_i H(F)}\int f\left(\gamma\begin{pmatrix}1_n & \begin{matrix} S & 0\\ 0 & 0\end{matrix}\\ & 1_n\end{pmatrix}g\right)e_{\mathbb{A}}(-\mathrm{Tr}(\beta S))\,dS$$
$$+\sum_{\gamma\in Q_n(F)\backslash Q_n(F)\xi_0 H(F)}\int f\left(\gamma\begin{pmatrix}1_n & \begin{matrix} S & 0\\ 0 & 0\end{matrix}\\ & 1_n\end{pmatrix}g\right)e_{\mathbb{A}}(-\mathrm{Tr}(\beta S))\,dS.$$

By [Ikeda 1994, Lemma (3.1)] (see [ibid., p. 628]), the first term vanishes. Also, we have [loc. cit.]

$$\begin{aligned} Q_n(F)\backslash Q_n(F)\xi_0 H(F) &= \xi_0 Z(F)X(F)Q_{n-m}(F)\backslash G_{n-m}(F)\\ &= \xi_0 X(F)\cdot Q_{n-m}(F)\backslash G_{n-m}(F)\cdot Z(F)\\ &= w_n Y(F)S_m(F)w_{n-m}Q_{n-m}(F)\backslash G_{n-m}(F)\end{aligned}$$

(note that S_m commutes with X and G_{n-m}). So

$$E_\beta(f;z,g)=\sum_{\gamma\in Q_{n-m}(F)\backslash G_{n-m}(F)}\sum_{y\in Y(F)}\int_{S_m(\mathbb{A})} f\left(w_n\begin{pmatrix}1_n & \begin{matrix} S & y\\ {}^t\bar{y} & 0\end{matrix}\\ & 1_n\end{pmatrix}\alpha_{n-m}(1,\gamma)g\right)$$
$$\times e_{\mathbb{A}}(-\mathrm{Tr}(\beta S))\,dS$$

Note that the final integral is essentially a product of local ones. □

Now we record some useful formulas:

Definition 3.8. If $g_v\in \mathrm{U}_{n-m}(F_v)$ and $x\in \mathrm{GL}_m(\mathcal{K}_v)$, then define

$\mathrm{FJ}_\beta(f_v;z,y,g,x)$

$$=\int_{S_m(F_v)} f\left(w_n\begin{pmatrix}1_n & \begin{matrix} S & y\\ {}^t\bar{y} & 0\end{matrix}\\ & 1_n\end{pmatrix}\alpha(\mathrm{diag}(x,{}^t\bar{x}^{-1}),g)\right)e_{F_v}(-\mathrm{Tr}\,\beta S)\,dS,$$

where

$$\alpha(g_1, g_2) = \begin{pmatrix} A & & B & \\ & D' & & C' \\ C & & D & \\ & B' & & A' \end{pmatrix} \quad \text{if } g_1 = \begin{pmatrix} A & B \\ C & D \end{pmatrix},\ g_2 = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}.$$

We also define

$$f_{v,\beta,z}(g) := f\left(z, w_n \begin{pmatrix} 1_n & S \\ & 1_n \end{pmatrix} g\right) e_v(-\operatorname{Tr} \beta S)\, dS.$$

Since

$$\begin{pmatrix} 1_n & S & X \\ & {}^t\bar{X} & \\ & 1_n & \end{pmatrix} \begin{pmatrix} 1_m & & & \\ & \bar{A}^{-1} & & \\ & & 1_m & \\ & B\bar{A}^{-1} & & A \end{pmatrix} = \begin{pmatrix} 1_m & XB\bar{A}^{-1} & & \\ & \bar{A}^{-1} & & \\ & & 1_m & \\ & B\bar{A}^{-1} & & A \end{pmatrix} \begin{pmatrix} 1_n & S - XB\,{}^t\bar{X} & XA \\ & \bar{A}\,{}^t\bar{X} & \\ & & \\ & 1_n & \end{pmatrix},$$

it follows that:

$$\mathrm{FJ}_\beta\left(f; z, X, \begin{pmatrix} A & B\bar{A}^{-1} \\ & \bar{A}^{-1} \end{pmatrix} g, Y\right)$$
$$= \tau_v^c(\det A)^{-1} |\det A\bar{A}|_v^{z+n/2} e_v(-\operatorname{Tr}({}^t\bar{X}\beta X B))\, \mathrm{FJ}_\beta(f; z, XA, g, Y).$$

Also, we have

$$\mathrm{FJ}_\beta(f; z, y, g, x) = \tau_v(\det x) |\det x\bar{x}|_{\mathbb{A}}^{-(z+n/2-m)}\, \mathrm{FJ}_{{}^t\bar{x}\beta x}(f; z, x^{-1}y, g, 1).$$

3C2. *Weil representations.* We define the Weil representations which will be used in calculating local Fourier–Jacobi coefficients in the next section.

The local set-up. Let v be a place of F. Let $h \in S_m(F_v)$, $\det h \neq 0$. Let U_h be the unitary group of this metric and denote by V_v the corresponding Hermitian space. Let $V_{n-m} := \mathcal{K}_v^{(n-m)} \oplus \mathcal{K}_v^{(n-m)} := X_v \oplus Y_v$ be the skew-Hermitian space associated to $\mathrm{U}(n-m, n-m)$. Let $W := V_v \otimes_{\mathcal{K}_v} V_{n-m,v}$. Then $(-,-) := \operatorname{Tr}_{\mathcal{K}_v/F_v}(\langle -,-\rangle_h \otimes_{\mathcal{K}_v} \langle -,-\rangle_{n-m})$ is a F_v linear pairing on W that makes W into a $4m(n-m)$-dimensional symplectic space over F_v. The canonical embedding of $\mathrm{U}_h \times \mathrm{U}_{n-m}$ into $\mathrm{Sp}(W)$ realizes the pair $(\mathrm{U}_h, \mathrm{U}_{n-m})$ as a dual pair in $\mathrm{Sp}(W)$. Let λ_v be a character of $\mathcal{K}_v^\times$ such that $\lambda_v|_{F_v^\times} = \chi_{\mathcal{K}/F,v}^m$. It is well known (see [Kudla 1994]) that there is a splitting $\mathrm{U}_h(F_v) \times \mathrm{U}_{n-m}(F_v) \hookrightarrow \mathrm{Mp}(W, F_v)$ of the metaplectic cover $\mathrm{Mp}(W, F_v) \to \mathrm{Sp}(W, F_v)$ determined by the character λ_v. This gives the Weil representation $\omega_{h,v}(u, g)$ of $\mathrm{U}_h(F_v) \times \mathrm{U}_{n-m}(F_v)$, where $u \in \mathrm{U}_h(F_v)$ and $g \in \mathrm{U}_{n-m}(F_v)$, via the Weil representation of $\mathrm{Mp}(W, F_v)$ on the space of Schwartz functions $\mathcal{S}(V_v \otimes_{\mathcal{K}_v} X_v)$. Moreover, we write $\omega_{h,v}(g)$ to mean $\omega_{h,v}(1, g)$. For $X \in M_{m\times(n-m)}(\mathcal{K}_v)$, we define $\langle X, X\rangle_h := {}^t\bar{X}\beta X$ (note this is an $(n-m)\times(n-m)$

matrix). We record here some useful formulas for $\omega_{h,v}$ which are generalizations of the formulas in [Skinner and Urban 2014, Section 10]:

- $\omega_{h,v}(u,g)\Phi(X)=\omega_{h,v}(1,g)\Phi(u^{-1}X)$.
- $\omega_{h,v}(\mathrm{diag}(A,{}^t\bar{A}^{-1}))\Phi(X)=\lambda(\det A)|\det A|_{\mathcal{K}}\Phi(XA)$.
- $\omega_{h,v}(r(S))\Phi(x)=\Phi(x)e_v(\mathrm{Tr}\langle X,X\rangle_h S)$.
- $\omega_{h,v}(\eta)\Phi(x)=|\det h|_v\int\Phi(Y)e_v(\mathrm{Tr}_{\mathcal{K}_v/F_v}(\mathrm{Tr}\langle Y,X\rangle_h))\,dY$.

Global setup. Let $h\in S_m(F)$ be positive definite. We can define global versions of U_h, GU_h, X, Y, W, and $(-,-)$, analogously to the local case. Fixing an idele class character $\lambda=\bigotimes\lambda_v$ of $\mathbb{A}_{\mathcal{K}}^\times/\mathcal{K}^\times$ such that $\lambda|_{F^\times}=\chi^m_{\mathcal{K}/F}$, the associated local splitting described above then determines a global splitting $\mathrm{U}_h(\mathbb{A}_F)\times\mathrm{U}_1(\mathbb{A}_F)\hookrightarrow\mathrm{Mp}(W,\mathbb{A}_F)$ and hence an action $\omega_h:=\otimes\omega_{h,v}$ of $\mathrm{U}_h(\mathbb{A}_F)\times\mathrm{U}_1(\mathbb{A}_F)$ on the Schwartz space $\mathcal{S}(V_{\mathbb{A}_{\mathcal{K}}}\otimes_{\mathcal{K}}X)$.

4. Local computations

In this section we do the local computations for Klingen Eisenstein sections realized as the pullbacks of Siegel Eisenstein sections. We will compute the Fourier and Fourier–Jacobi coefficients for the Siegel sections and the pullback Klingen Eisenstein sections.

4A. ***Archimedean computations.*** Let v be an Archimedean place of F.

4A1. *Fourier coefficients.*

Definition 4.1. $f_{\kappa,n}(z,g)=J_n(g,i1_n)^{-\kappa}|J_n(g,i1_n)|^{\kappa-2z-n}$.

Now we recall [Skinner and Urban 2014, Lemma 11.4]. Let $J_n(g,i1_n):=\det(C_g i1_n+D_g)$ for $g=\left(\begin{smallmatrix}A_g&B_g\\C_g&D_g\end{smallmatrix}\right)$.

Lemma 4.2. *Suppose $\beta\in S_n(\mathbb{R})$. Then the function $z\mapsto f_{\kappa,\beta}(z,g)$ has a meromorphic continuation to all of $\mathbb{C}$. Furthermore, if $\kappa\geq n$ then $f_{\kappa,n,\beta}(z,g)$ is holomorphic at $z_\kappa:=\frac{1}{2}(\kappa-n)$ and, for $y\in\mathrm{GL}_n(\mathbb{C})$, $f_{\kappa,n,\beta}(z_\kappa,\mathrm{diag}(y,{}^t\bar{y}^{-1}))=0$ if $\det\beta\leq 0$, while, if $\det\beta>0$, then*

$$f_{\kappa,n,\beta}(z_\kappa,\mathrm{diag}(y,{}^t\bar{y}^{-1}))=\frac{(-2)^{-n}(2\pi i)^{n\kappa}(2/\pi)^{n(n-1)/2}}{\prod_{j=0}^{n-1}(\kappa-j-1)!}e_v(i\,\mathrm{Tr}(\beta y\,{}^t\bar{y}))\det(\beta)^{\kappa-n}\det\bar{y}^\kappa.$$

4A2. *Pullback sections.* Now we assume that our π is the holomorphic discrete series representation associated to the (scalar) weight $(0,\dots,0;\kappa,\dots,\kappa)$ and let φ be the unique (up to scalar) vector such that the action of $K_\infty^{+,\prime}$ (see Section 3A) is given by $\det\mu(k,i)^{-\kappa}$. Recall also that in Section 3A we defined the Klingen section $F_\kappa(z, g)$ (denoted F_κ). Recall we have defined S and S' in equations (1) and (2). Let

$$\boldsymbol{i} := \begin{pmatrix} \frac{1}{2}i\,1_b & & & \\ & i & & \\ & & \frac{1}{2}\zeta & \\ & & & \frac{1}{2}i\,1_b \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \frac{1}{2}i\,1_b & & \\ & \frac{1}{2}\zeta & \\ & & \frac{1}{2}i\,1_b \end{pmatrix}$$

be the distinguished point in the symmetric domain for $\mathrm{GU}(n, n)$ or $\mathrm{GU}(n+1, n+1)$, for $n = a + 2b$. We define Archimedean sections to be

$$f_\kappa(g) = J_{n+1}(g, \boldsymbol{i})^{-\kappa}|J_{n+1}(g, \mathbf{i})|^{\kappa-2z-n-1}$$

and

$$f'_\kappa(g) = J_n(g, \boldsymbol{i})^{-\kappa}|J_n(g, \mathbf{i})|^{\kappa-2z-n}$$

and the pullback sections on $\mathrm{GU}(a+b+1, b+1)$ and $\mathrm{GU}(a+b, a)$ to be

$$F_\kappa(z, g) := \int_{\mathrm{U}(a+b,b)(\mathbb{R})} f_\kappa(z, S^{-1}\alpha(g, g_1)S)\bar{\tau}(\det g_1)\pi(g_1)\varphi\, dg_1$$

and

$$F'_\kappa(z, g) := \int_{\mathrm{U}(a+b,b)(\mathbb{R})} f'_\kappa(z, S'^{-1}\alpha(g, g_1)S')\bar{\tau}(\det g_1)\pi(g_1)\varphi\, dg_1.$$

Lemma 4.3. *The integrals F_κ and F'_κ are absolutely convergent for* $\mathrm{Re}(z)$ *sufficiently large and, for such z, we have*

(i) $F_\kappa(z, g) = c_\kappa(z)F_{\kappa,z}(g)$;

(ii) $F'_\kappa(z, g) = c'_\kappa(z)\pi(g)\varphi$;

where

$$c'_\kappa(z, g) = 2^\nu|\det\zeta|_v^b \begin{cases} \pi^{(a_v+b_v)b_v}\Gamma_{b_v}\big(z+\frac{1}{2}(n+\kappa)-a_v-b_v\big) \times \Gamma_{b_v}\big(z+\frac{1}{2}(n+k)\big)^{-1} & \text{if } b > 0, \\ 1 & \text{otherwise}. \end{cases}$$

and $c_\kappa(z, g) = c'_\kappa\big(z+\frac{1}{2}, g\big)$. *Here* $\Gamma_m(s) := \pi^{m(m+1)/2}\prod_{k=0}^{m-1}\Gamma(s-k)$ *and* $\nu := (a+2b)db$ *(recall that* $d = [F:\mathbb{Q}]$*).*

Proof. See [Shimura 1997, Propositions 22.2 and A2.9]. Note that the action of $(\beta, \gamma) \in \mathrm{U}(r, s) \times \mathrm{U}(s, r)$ is given by (β', γ'), defined there. Taking this into consideration, our conjugation matrix S is Shimura's S times Σ^{-1} (with notation as there), which is defined in (22.1.2) in [Shimura 1997]. Also our result differs from [Skinner and Urban 2014, Lemma 11.6] by a power of 2, since we are using a different S here. $\square$

4A3. *Fourier–Jacobi coefficients.* We write $\mathrm{FJ}_{\beta,\kappa}$ for the Fourier–Jacobi coefficient defined in Definition 3.8 with f_v chosen as $f_{\kappa,n}$.

Lemma 4.4. *Let* $z_\kappa = \frac{1}{2}(\kappa - n)$, $\beta \in S_m(\mathbb{R})$, $m < n$ *and* $\det\beta > 0$. *Then*:

(i) $\mathrm{FJ}_{\beta,\kappa}(z_\kappa, x, \eta, 1) = f_{\kappa,m,\beta}\big(z_\kappa + \frac{1}{2}(n-m), 1\big) e(i\,\mathrm{Tr}({}^t\overline{X}\beta X))$.

(ii) *If* $g \in U_{n-m}(\mathbb{R})$, *then*

$$\mathrm{FJ}_{\beta,\kappa}(z_\kappa, X, g, 1) = e(i\,\mathrm{Tr}\,\beta)c_m(\beta,\kappa) f_{\kappa-m,n-m}(z_\kappa, g')w_\beta(g')\Phi_{\beta,\infty}(x),$$

where

$$g' = \begin{pmatrix} 1_n & \\ & -1_n \end{pmatrix} g \begin{pmatrix} & 1_n \\ 1_n & \end{pmatrix},$$

$$c_t(\beta,\kappa) = \frac{(-2)^{-t}(2\pi i)^{t\kappa}(2/\pi)^{t(t-1)/2}}{\prod_{j=0}^{t-1}(\kappa - j - 1)} \det\beta^{\kappa - t}$$

and $\Phi_{\beta,\infty}(x) = e^{-2\pi\,\mathrm{Tr}(\langle x,x\rangle_\beta)}$.

Proof. Our proof is similar to [Skinner and Urban 2014, Lemma 11.5]. For (i) we first assume that $m \le \frac{1}{2}n$; then there is a matrix $U \in \mathrm{U}_{n-m}$ such that $XU = (0, A)$ for A an $m \times m$ positive semidefinite Hermitian matrix. It follows that $\mathrm{FJ}_{\beta,\kappa}(z, X, \eta, 1) = \mathrm{FJ}_{\beta,\kappa}(z, (0, A), \eta, 1)$ and $e(iTr({}^t\overline{X}\beta X)) = e(i\,\mathrm{Tr}(U^{-1}\,{}^t\overline{X}\beta XU))$, so we are reduced to the case when $X = (0, A)$.

Let C be an $m \times m$ positive definite Hermitian matrix defined by $C = \sqrt{A^2 + 1}$. (Since A is positive semidefinite Hermitian, this C exists by linear algebra.) We have

$$\begin{pmatrix} & & & A \\ 1_n & & & \\ & A & & \\ & & 1_n & \end{pmatrix} = \begin{pmatrix} C & & & & & \\ & 1 & & & & \\ & & C & & & \\ & & AC^{-1} & C^{-1} & & \\ & & & & 1 & \\ AC^{-1} & & & & & C^{-1} \end{pmatrix} \begin{pmatrix} C^{-1} & & & & & C^{-1}A \\ & 1 & & & & \\ & & C^{-1} & C^{-1}A & & \\ & & -C^{-1}AC^{-1} & & & \\ & & & & 1 & \\ -C^{-1}A & & & & & C^{-1} \end{pmatrix}.$$

Write $k(A)$ for the second matrix in the right of the above, which belongs to $K^+_{n,\infty}$; then, as in [Skinner and Urban 2014, Lemma 11.5],

$$w_n \begin{pmatrix} & S & X \\ 1_n & {}^t\overline{X} & \\ & & 1_n \end{pmatrix} = \begin{pmatrix} C^{-1} & & & \times & \times & \times \\ & 1 & & \times & \times & \times \\ \times & \times & C^{-1} & \times & \times & \times \\ & & & C & & \\ & & & & 1 & \\ & & & \times & \times & C \end{pmatrix} w_n \begin{pmatrix} & U^{-1}SU^{-1} \\ 1_n & \\ & 1_n \end{pmatrix} k(A).$$

Thus,

$$\begin{aligned}\mathrm{FJ}_{\beta,\kappa}(z_\kappa,(0,A),\eta,1)&=(\det C)^{2m-2\kappa}\,\mathrm{FJ}_{\beta',\kappa}(z_\kappa,0,\eta,1),\ \beta'=C\beta C\\&=(\det C)^{2m-2\kappa}\,f_{\kappa,m,\beta'}\big(z_\kappa+\tfrac{1}{2}(n-m),1\big)\\&=f_{\kappa,m,\beta}\big(z+\tfrac{1}{2}(n-m),1\big)e(i\,\mathrm{Tr}(C\beta C-\beta)).\end{aligned}$$

But

$$\begin{aligned}e(i\,\mathrm{Tr}(C\beta C-\beta))=e(i\,\mathrm{Tr}(C^2\beta-\beta))&=e\big(i\,\mathrm{Tr}((C^2-1)\beta)\big)\\&=e(i\,\mathrm{Tr}(A^2\beta))=e(i\,\mathrm{Tr}(A\beta A)).\end{aligned}$$

This proves part (i).

Part (ii) is proved completely the same as [Skinner and Urban 2014, Lemma 11.5].

In the case when $m>\frac{1}{2}n$, we proceed similarly as in [Skinner and Urban 2014, Lemma 11.5], replacing a and u there by corresponding block matrices just as above. We omit the details. □

4B. *Finite primes, unramified case.*

4B1. *Pullback integrals.*

Lemma 4.5. *Suppose π, ψ and τ are unramified and $\varphi\in\pi$ is a new vector. If $\mathrm{Re}(z)>\frac{1}{2}(a+b)$ then the pullback integral converges and*

$$F_\varphi(f_v^{\mathrm{sph}};z,g)=\frac{L(\tilde\pi,\bar\tau^c,z+1)}{\prod_{i=0}^{a+2b-1}L(2z+a+2b+1-i,\bar\tau'\chi_{\mathcal{K}}^i)}F_{\rho,z}(g),$$

where $F_{\rho,z}$ is the spherical section taking value φ at the identity and

$$F_\varphi(f_v^{\mathrm{sph}};z,g)=\frac{L\big(\tilde\pi,\bar\tau^c,z+\frac{1}{2}\big)}{\prod_{i=0}^{a+2b-1}L(2z+a+2b-i,\bar\tau'\chi_{\mathcal{K}}^i)}\pi(g)\varphi.$$

This is computed in [Lapid and Rallis 2005, Proposition 3.3].

4B2. *Fourier–Jacobi coefficients.* Let v be a prime of F not dividing p and τ a character of $\mathcal{K}_v^\times$. For $f\in I_n(\tau)$ and $\beta\in S_m(F_v)$, $0\le m\le n$, we define the local Fourier–Jacobi coefficient to be

$$f_\beta(z;g):=\int_{S_m(F_v)}f\left(z,w_n\begin{pmatrix}1_n&\begin{matrix}S&0\\0&0\end{matrix}\\&1_n\end{pmatrix}g\right)e_v(-\mathrm{Tr}\,\beta S)\,dS.$$

We first record straightforward generalizations of [Skinner and Urban 2014, Lemmas 11.7 and 11.8] to any fields [Shimura 1997, Propositions 18.14 and 19.2].

Lemma 4.6. *Let $\beta \in S_n(F_v)$ and let $r := \mathrm{rank}(\beta)$. Then, for $y \in \mathrm{GL}_n(\mathcal{K}_v)$,*

$$f^{\mathrm{sph}}_{v,\beta}(z, \mathrm{diag}(y, {}^t\bar{y}^{-1})) = \tau(\det y)|\det y\bar{y}|_v^{-z+n/2} D_v^{-n(n-1)/4} \times \frac{\prod_{i=r}^{n-1} L(2z+i-n+1, \bar{\tau}'\chi^i_{\mathcal{K}})}{\prod_{i=0}^{n-1} L(2z+n-i, \bar{\tau}'\chi^i_{\mathcal{K}})} h_{v,{}^t\bar{y}\beta y}(\bar{\tau}'(\varpi)q_v^{-2z-n}),$$

where $h_{v,{}^t\bar{y}\beta y} \in \mathbb{Z}[X]$ is a monic polynomial depending on v and ${}^t\bar{y}\beta y$ but not on τ. If $\beta \in S_n(\mathcal{O}_{F,v})$ and $\det\beta \in \mathcal{O}^\times_{F,v}$, then we say that β is v-primitive and, in this case, $h_{v,\beta} = 1$.

Lemma 4.7. *Suppose v is unramified in $\mathcal{K}$. Let $\beta \in S_m(F_v)$ with $\det\beta \neq 0$ and let $\beta \in S_m(\mathcal{O}_{F_v})$ and let λ be an unramified character of $\mathcal{K}_v^\times$ such that $\lambda|_{F_v^\times} = 1$. If $\beta \in \mathrm{GL}_m(\mathcal{O}_{F_v})$ then, for $u \in \mathrm{U}_\beta(F_v)$,*

$$\mathrm{FJ}_\beta(f^{\mathrm{sph}}_n; z, x, g, u) = \tau(\det u)|\det u\bar{u}|_v^{-z+1/2} \frac{f^{\mathrm{sph}}_{n-m}(z, g)\omega_\beta(u, g)\Phi_0(x)}{\prod_{i=0}^{m-1} L(2z+n-i, \bar{\tau}'\chi^i_{\mathcal{K}})}.$$

4C. *Prime-to-p ramified case.*

4C1. *Pullback integrals.* Again let v be a prime of F not dividing p. We fix some x and y in $\mathcal{K}$ which are divisible by some high power of ϖ_v (which can be made precise from the proof of the following two lemmas). (When we are moving things p-adically, the x and y are not going to change.) We define $f^\dagger \in I_{n+1}(\tau)$ to be the Siegel section supported on the cell $Q(F_v)w_{a+2b+1}N_Q(\mathcal{O}_{F,v})$, where $w_{a+2b+1} = \begin{pmatrix} & 1_{a+2b+1} \\ -1_{a+2b+1} & \end{pmatrix}$ and the value at $N_Q(\mathcal{O}_{F,v})$ equals 1. Similarly, we define $f^{\dagger,\prime} \in I_n(\tau)$ to be the section supported in $Q(F_v)w_{a+2b}N_Q(\mathcal{O}_{F,v})$ that takes value 1 on $N_Q(\mathcal{O}_{F,v})$.

Definition 4.8. $$f_{v,\mathrm{sieg}}(g) := f^\dagger(g\tilde{S}_v^{-1}\tilde{\gamma}_v) \in I_{n+1}(\tau),$$

where $\tilde{\gamma}_v$ is defined to be

$$\begin{pmatrix}
1_b & & & & & & & (1/x)1_b \\
& 1 & & & & & & \\
& & 1_a & & & (1/(y\bar{y}))1_a & & \\
& & & 1_b & (1/\bar{x})1_b & & & \\
& & & & 1_b & & & \\
& & & & & 1 & & \\
& & & & & & 1_a & \\
& & & & & & & 1_b
\end{pmatrix}$$

and

$$\tilde{S}_v = \begin{pmatrix} 1_b & & & & & & & -\frac{1}{2}1_b \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & -1_b & \frac{1}{2}1_b & & & \\ & & & 1_b & \frac{1}{2}1_b & & & \\ & & & & & 1 & & \\ & & & & & & 1_a & \\ -1_b & & & & & & & -\frac{1}{2}1_b \end{pmatrix}.$$

Similarly, we define $f'_{v,\mathrm{sieg}}(g) := f^{\dagger,\prime}(g\tilde{S}_v^{-1}\tilde{\gamma}'_v)$ for

$$\tilde{S}'_v := \begin{pmatrix} 1_b & & & & & -\frac{1}{2}1_b \\ & 1_a & & & & \\ & & -1_b & \frac{1}{2}1_b & & \\ & & 1_b & \frac{1}{2}1_b & & \\ & & & & 1_a & \\ -1_b & & & & & -\frac{1}{2}1_b \end{pmatrix}$$

and

$$\tilde{\gamma}_v = \begin{pmatrix} 1_b & & & & & (1/x)1_b \\ & 1_a & & & (1/(y\bar{y}))1_a & \\ & & 1_b & (1/\bar{x})1_b & & \\ & & & 1_b & & \\ & & & & 1_a & \\ & & & & & 1_b \end{pmatrix}.$$

Lemma 4.9. *Let $K_v^{(2)}$ be the subgroup of $G(F_v)$ of matrices of the form*

$$\begin{pmatrix} 1_b & & & & d \\ a & 1 & f & b & c \\ & & 1_a & & g \\ & & & 1_b & e \\ & & & & 1 \end{pmatrix},$$

where $e = -{}^t\bar{a}$, $b = {}^t\bar{d}$, $g = -\zeta\,{}^t\bar{f}$, $b \in M(\mathbb{O}_v)$, $c - f\zeta\,{}^t\bar{f} \in \mathbb{O}_{F,v}$, $a \in (x)$, $e \in (\bar{x})$, $f \in (y\bar{y})$ and $g \in (2\zeta y\bar{y})$. Then $F_\varphi(z; g, f)$ is supported in $PwK_v^{(2)}$ and is invariant under the action of $K_v^{(2)}$.

Proof. Let $S_{x,y}$ consist of matrices

$$S := \begin{pmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{pmatrix}$$

in the space of Hermitian $(a+2b+1)\times(a+2b+1)$ matrices (the blocks are with respect to the partition $b+1+a+b$) such that the entries of S_{13} and S_{23} (resp. S_{14} and S_{24}, S_{31} and S_{32}, S_{41} and S_{42}) are divisible by y (resp. x, $\bar{y}$, $\bar{x}$), while the entries of S_{33} (resp. S_{34}, S_{43}, S_{44}) are divisible by $y\bar{y}$ (resp. $x\bar{y}$, $\bar{x}y$, $x\bar{x}$). Let $Q_{x,y} := Q(F_v)\cdot\left(\begin{smallmatrix} 1 & \\ S_{x,y} & 1\end{smallmatrix}\right)$.

Write

$$\eta = \begin{pmatrix} & & 1_b \\ & 1_a & \\ -1_b & & \end{pmatrix}.$$

As in [Skinner and Urban 2014, Proposition 11.16], for

$$g = \begin{pmatrix} a_1 & a_2 & a_3 & b_1 & b_2 \\ a_4 & a_5 & a_6 & b_3 & b_4 \\ a_7 & a_8 & a_9 & b_5 & b_6 \\ c_1 & c_2 & c_3 & d_1 & d_2 \\ c_4 & c_5 & c_6 & d_3 & d_4 \end{pmatrix},$$

we have

$$\begin{aligned} \gamma(g,1)\in \operatorname{supp} f_{v,\mathrm{sieg}} &\iff S_v^{-1}\alpha(g,1)Aw_{a+2b+1}d_{x,y}\tilde{\gamma}^{-1}\in Q_{x,y} \\ &\iff S_v^{-1}\alpha(gw,\eta\,\mathrm{diag}(\bar{x}^{-1},1,x))Aw'd_y\tilde{\gamma}^{-1}\in Q_{x,y}. \end{aligned}$$

Here,

$$A = \begin{pmatrix} 1_b &&&&&&& \\ & 1 &&&&&& \\ && 1_a &&&-\frac{1}{2}\zeta && \\ &&& 1_b &&&& \\ &&&& 1_b &&& \\ &&&&& 1 && \\ && -1_a &&& -\frac{1}{2}\zeta && \\ &&&&&&& 1_b \end{pmatrix},$$

$$d_{x,y} = \mathrm{diag}(1,1,y,x,1,1,\bar{y}^{-1},\bar{x}^{-1}),$$
$$d_y = \mathrm{diag}(1,1,y,1,1,1,\bar{y}^{-1},1),$$

$$w' = \begin{pmatrix} 1_b &&&&&&& \\ & 1 &&&&&& \\ &&&&&& 1_a & \\ &&& 1_b &&&& \\ &&&& 1_b &&& \\ &&&&& 1 && \\ && -1_a &&&&& \\ &&&&&&& 1_b \end{pmatrix} \quad \text{and} \quad \tilde{\gamma} = \begin{pmatrix} 1_b &&&&&&& \\ & 1 &&&&&& \\ && 1_a &&&&& \\ &&& 1_b &&&& \\ &&& 1_b & 1_b &&& \\ &&&&& 1 && \\ && 1_a &&&& 1_a & \\ 1_b &&&&&&& 1_b \end{pmatrix}.$$

Here, x and y stand for the corresponding block matrices of the corresponding size. Recall that $\gamma(m(g_1, 1), g_1) \in Q$; by multiplying this to the left for $g_1 = \operatorname{diag}(\bar{x}, 1, x^{-1})\eta^{-1}$, we are reduced to proving that, if $\gamma(g, 1)w'd_y\tilde{\gamma}^{-1} \in Q_{x,y}$, then $g \in PwK_v^{(2)}w^{-1}$. A computation tells us that $\gamma(g, 1)w'd_y\tilde{\gamma}^{-1}$ equals the product

$$\begin{pmatrix} 1_b &&&&&&& -\frac{1}{2}1_b \\ & 1 &&&&&& \\ && 1_a &&&&& \\ &&& -1_b & \frac{1}{2}1_b &&& \\ &&&& 1_b &&& \\ &&&&& 1 && \\ &&&&&& 1_a & \\ &&&&&&& -1_b \end{pmatrix}$$
$$\times \begin{pmatrix} a_1 & a_2 & \frac{1}{2}\zeta a_3 y - a_3\bar{y}^{-1} & -b_1 & b_1 & b_2 & a_3\bar{y}^{-1} & \\ a_4 & a_5 & \frac{1}{2}\zeta a_6 y - a_6\bar{y}^{-1} & -b_3 & b_3 & b_4 & a_6\bar{y}^{-1} & \\ \frac{1}{2}a_7 & \frac{1}{2}a_8 & \frac{1}{4}\zeta y(a_9-1) - \frac{1}{2}(a_9+1)\bar{y}^{-1} & -\frac{1}{2}b_5 & \frac{1}{2}b_5 & \frac{1}{2}b_6 & \frac{1}{2}(a_9+1)\bar{y}^{-1} & \\ &&& 1 &&&& \\ c_1 & c_2 & \frac{1}{2}\zeta c_3 y - c_3\bar{y}^{-1} & 1-d_1 & d_1 & d_2 & c_3\bar{y}^{-1} & \\ c_4 & c_5 & \frac{1}{2}\zeta c_6 y - c_6\bar{y}^{-1} & -d_3 & d_3 & d_4 & c_6\bar{y}^{-1} & \\ -\mathfrak{z}a_7 & -\mathfrak{z}a_8 & -\frac{1}{2}(a_9+1)y + \mathfrak{z}(a_9-1)\bar{y}^{-1} & \mathfrak{z}b_5 & -\mathfrak{z}b_5 & -\mathfrak{z}b_6 & \mathfrak{z}(1-a_9)\bar{y}^{-1} & \\ a_1-1 & a_2 & \frac{1}{2}\zeta a_3 y - a_3\bar{y}^{-1} & -b_1 & b_1 & b_2 & a_3\bar{y}^{-1} & 1 \end{pmatrix},$$

where $\mathfrak{z} = \zeta^{-1}$.

One first proves that $d_4 \neq 0$ by looking at the second row of the lower left of the above matrix, so by left-multiplying g by some matrix in N_P, we may assume that $d_2 = b_2 = b_4 = b_6 = 0$, then the result follows by an argument similarly to the proof of Lemma 4.36 later on. $\square$

Now recall that

$$g = \begin{pmatrix} a_5 & a_6 & a_4 \\ a_8 & a_9 & a_7 \\ a_2 & a_3 & a_1 \end{pmatrix}.$$

Let $\mathfrak{Y}$ be the set of g such that the entries of a_2 are integers, the entries of a_3 (resp. $a_1-1,\ 1-a_5,\ a_6,\ a_4,\ a_8,\ a_7$) are divisible by $y\bar{y}$ (resp. $\bar{x}$, x, $\bar{x}y$, $x\bar{x}$, $\frac{1}{2}\bar{y}y\zeta$, $\bar{y}yx\zeta$), and $1-a_9=y\bar{y}\zeta(1+y\bar{y}N)$ for some N with integral entries.

Lemma 4.10. *Let $\varphi_x=\pi(\mathrm{diag}(\bar{x},1,x^{-1})\eta^{-1})\varphi$, where φ is invariant under the action of $\mathfrak{Y}$ defined above*; *then*:

(i) $F_{\varphi_x}(f_{v,\mathrm{sieg}};z,w)=\tau(y\bar{y}x)|(y\bar{y})^2x\bar{x}|_v^{-z-(a+2b+1)/2}\,\mathrm{Vol}(\mathfrak{Y})\cdot\varphi.$

(ii) $F'_{\varphi_x}(f'_{v,\mathrm{sieg}};z,w)=\tau(y\bar{y}x)|(y\bar{y})^2x\bar{x}|_v^{-z-(a+2b)/2}\,\mathrm{Vol}(\mathfrak{Y})\cdot\varphi.$

Proof. First, one computes

$$\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ &&\frac{1}{2}&&&&-\frac{1}{2}&\\ &&&1&&&&\\ &&&1&1&&&\\ &&&&&1&&\\ &&-\zeta^{-1}&&&&-\zeta^{-1}&\\ 1&&&&&&&1 \end{pmatrix}\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ &&1&&&&&\\ &&&a_1&&&a_3&a_2\\ &&&&1&&&\\ &&&&&1&&\\ &&&a_7&&&a_9&a_8\\ &&&a_4&&&a_6&a_5 \end{pmatrix}\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ &&1&&&&-\frac{1}{2}\zeta&\\ &&&1&&&&\\ &&&&1&&&\\ &&&&&1&&\\ &&-1&&&&-\frac{1}{2}\zeta&\\ &&&&&&&1 \end{pmatrix}$$

$$\times\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ &&&&&&1&\\ &&&1&&&&\\ &&&&1&&&\\ &&&&&1&&\\ &&-1&&&&&\\ &&&&&&&1 \end{pmatrix}\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ &&y&&&&&\\ &&&1&&&&\\ &&&&1&&&\\ &&&&&1&&\\ &&&&&&\bar{y}^{-1}&\\ &&&&&&&1 \end{pmatrix}\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ &&1&&&&&\\ &&&1&&&&\\ &&&-1&1&&&\\ &&&&&1&&\\ &&-1&&&&1&\\ -1&&&&&&&1 \end{pmatrix}$$

$$=\begin{pmatrix} 1&&&&&&&\\ &1&&&&&&\\ \frac{1}{2}a_8&&\frac{1}{4}\zeta y(1-a_9)-\frac{1}{2}\bar{y}^{-1}(1+a_9)&-\frac{1}{2}a_7&&&\frac{1}{2}\bar{y}^{-1}(1+a_9)&-\frac{1}{2}a_8\\ -a_2&&\frac{1}{2}a_3\zeta y+a_3\bar{y}^{-1}&a_1&&&-a_3\bar{y}^{-1}&a_2\\ -a_2&&\frac{1}{2}a_3\zeta y+a_3\bar{y}^{-1}&a_1-1&1&&-a_3\bar{y}^{-1}&a_2\\ &&&&&1&&\\ \zeta^{-1}a_8&&-\frac{1}{2}y(1+\zeta^{-1}a_9\zeta)+\zeta^{-1}\bar{y}^{-1}(1-a_9)&-\zeta^{-1}a_7&&&-\zeta^{-1}\bar{y}^{-1}(1-a_9)&-\zeta^{-1}a_8\\ 1-a_5&&\frac{1}{2}a_6\zeta y+a_6\bar{y}^{-1}&a_4&&&-a_6\bar{y}^{-1}&a_5 \end{pmatrix}.$$

One checks the above matrix belongs to $Q_{x,y}$ if and only if the a_i satisfy the conditions required by the definition of $\mathfrak{Y}$. The lemma follows by a similar argument to Lemma 4.38 below. $\square$

Definition 4.11. We will sometimes write $\mathfrak{Y}_v$ for the $\mathfrak{Y}$ above to emphasize the dependence on v.

4C2. *Fourier–Jacobi coefficient.* We first give a formula for the Fourier coefficients for $\tilde{f}_{v,\mathrm{sieg}} := \rho(\tilde{\gamma}_v) f^{\dagger}_{v,\mathrm{sieg}}$ and $\tilde{f}'_{v,\mathrm{sieg}} := \rho(\tilde{\gamma}'_v) f^{\dagger,\prime}_{v,\mathrm{sieg}}$.

Lemma 4.12. (i) *Let $\beta = (\beta_{ij}) \in S_{n+1}(F_v)$; then for all $z \in \mathbb{C}$ we have*

$$\tilde{f}_{v,\mathrm{sieg},\beta}(z,1) = \mathrm{Vol}(S_{n+1}(\mathbb{O}_{F,v}))e_v\biggl(\mathrm{Tr}_{\mathcal{K}_v/F_v}\biggl(\frac{\beta_{a+b+2,1}+\cdots+\beta_{a+2b+1,b}}{x}\biggr) + \frac{\beta_{b+2,b+2}+\cdots+\beta_{b+1+a,b+1+a}}{y\bar{y}}\biggr).$$

(ii) *Let $\beta = (\beta_{ij}) \in S_v(F_v)$. Then*

$$\tilde{f}'_{v,\mathrm{sieg},\beta}(z,1) = \mathrm{Vol}(S_n(\mathbb{O}_{F,v}))e_v\biggl(\mathrm{Tr}_{\mathcal{K}_v/F_v}\biggl(\frac{\beta_{a+b+1,1}+\cdots+\beta_{a+2b,b}}{x}\biggr) + \frac{\beta_{b+1,b+1}+\cdots+\beta_{b+a,b+a}}{y\bar{y}}\biggr).$$

The proof is straightforward.

Here we record a lemma on the Fourier–Jacobi coefficient for $f^{\dagger}_v \in I_n(\tau_v)$ and $\beta \in S_m(F_v)$.

Lemma 4.13. *If $\beta \notin S_m(\mathbb{O}_{F_v})^*$ then $\mathrm{FJ}_\beta(f^{\dagger}; z, u, g, hy) = 0$. If $\beta \in S_n(\mathbb{O}_{F_v})^*$ then*

$$\mathrm{FJ}_\beta(f^{\dagger}; z, u, g, h) = f^{\dagger}(z, g'\eta)\omega_\beta(h, g'\eta)\Phi_{0,y}(u)\cdot \mathrm{Vol}(S_m(\mathbb{O}_{F_v})),$$

where $g' = \begin{pmatrix} 1_{n-m} & \\ & -1_{n-m}\end{pmatrix} g \begin{pmatrix} 1_{n-m} & \\ & -1_{n-m}\end{pmatrix}$.

The proof is similar to [Skinner and Urban 2014, Lemma 11.15].

4D. ***p-adic computations.*** In this subsection we first prove that, under some "generic conditions", the unique (up to scalar) nearly ordinary vector in $I(\rho)$ is just the unique (up to scalar) vector with certain prescribed action of level subgroup. Then we construct a section $F^{\dagger}$ in $I(\rho^{\vee})$ which is the pullback of a Siegel section $f^{\dagger}$ supported in the big cell. We can understand the action of the level group of this section. Then we define F^0 to be the image of $F^{\dagger}$ under the intertwining operator. By checking the action of the level subgroup on F^0, we can prove that it is just the nearly ordinary vector.

In our calculations we will usually use the projection to the first component of $\mathcal{K}_v \simeq \mathcal{K}_w \times \mathcal{K}_{\bar{w}} \simeq \mathbb{Q}_p \times \mathbb{Q}_p$.

4D1. *Nearly ordinary sections.* Let $\lambda_1, \ldots, \lambda_n$ be n characters of $F_v^{\times}$, which we identify with $\mathbb{Q}_p^{\times}$, and $\pi = \mathrm{Ind}_B^{\mathrm{GL}_n}(\lambda_1, \ldots, \lambda_n)$.

Definition 4.14. Let $n = r + s$ and $\underline{k} = (c_{r+s}, \dots, c_{s+1}; c_1, \dots, c_s)$ be a weight. We say $(\lambda_1, \dots \lambda_n)$ is nearly ordinary with respect to $\underline{k}$ if

$$\{\mathrm{val}_p\lambda_1(p),\dots,\mathrm{val}_p\lambda_n(p)\} = \bigl\{c_1+s-1-\tfrac{1}{2}n+\tfrac{1}{2}, c_2+s-2-\tfrac{1}{2}n+\tfrac{1}{2},\dots,c_s-\tfrac{1}{2}n+\tfrac{1}{2},\\ c_{s+1}+r+s-1-\tfrac{1}{2}n+\tfrac{1}{2},\dots,c_{r+s}+s-\tfrac{1}{2}n+\tfrac{1}{2}\bigr\}.$$

We write the elements of the right side in order as $\kappa_1, \dots, \kappa_{r+s}$, so $\kappa_1 > \dots > \kappa_{r+s}$.

Let $\mathcal{A}_p := \mathbb{Z}_p[t_1, t_2, \dots, t_n, t_n^{-1}]$ be the Atkin–Lehner ring of $G(\mathbb{Q}_p)$, where t_i is defined by $t_i = N(\mathbb{Z}_p)\alpha_i N(\mathbb{Z}_p)$, $\alpha_i = \begin{pmatrix} 1_{n-i} & \\ & p1_i \end{pmatrix}$. Then t_i acts on $\pi^{N(\mathbb{Z}_p)}$ by

$$v|t_i = \sum_{x \in N|\alpha_i^{-1}N\alpha_i} x_i\alpha_i^{-1}v.$$

We also define a normalized action with respect to the weight $\underline{k}$, following [Hida 2004b]:

$$v\|t_i := \delta(\alpha_i)^{-1/2}p^{\kappa_1+\dots+\kappa_i}v|t_i$$

Definition 4.15. A vector $v \in \pi$ is called nearly ordinary if it is an eigenvector for all $\|t_i$ with eigenvalues that are p-adic units.

We identify π as a set of smooth functions on $\mathrm{GL}_n(\mathbb{Q}_p)$:

$$\pi = \{f : \mathrm{GL}_n(\mathbb{Q}_p) \to \mathbb{C} \mid f(bx) = \lambda(b)\delta_B(b)^{1/2}f(x)\}.$$

Here, $\lambda(b) := \prod_{i=1}^n \lambda_i(b_i)$ for

$$b = \begin{pmatrix} b_1 & \times & \times \\ & \ddots & \times \\ & & b_n \end{pmatrix}$$

and δ_B is the modulus function for the upper-triangular Borel subgroup. Let w_ℓ be the longest Weyl element,

$$\begin{pmatrix} & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & \end{pmatrix},$$

and let f^ℓ be the element in π (which is unique up to scalar) that is supported in $Bw_\ell N(\mathbb{Z}_p)$ and invariant under $N(\mathbb{Z}_p)$. We have:

Lemma 4.16. *f^ℓ is an eigenvector for all t_i.*

Proof. Note that, for any i, $f^\ell|t_i$ is invariant under $N(\mathbb{Z}_p)$. By looking at the definition of $v|t_i$ for the above model of π, it is not hard to see that $f^\ell|t_i$ is supported in $B(\mathbb{Q}_p)w_\ell B(\mathbb{Z}_p)$. So $f^\ell|t_i$ must be a multiple of f^ℓ. □

Lemma 4.17. *Suppose that* $(\lambda_1, \dots, \lambda_n)$ *is nearly ordinary with respect to* $\underline{k}$ *and suppose*

$$\nu_p(\lambda_1(p)) > \nu_p(\lambda_2(p)) > \cdots > \nu_p(\lambda_n(p));$$

then the eigenvalues of $\|t_i$ *acting on* f^ℓ *are* p*-adic units. In other words,* f^ℓ *is an ordinary vector.*

Proof. A straightforward computation gives that

$$f^\ell \| t_i = \lambda_1 \cdots \lambda_i (p^{-1}) p^{\kappa_1 + \cdots + \kappa_i} f^\ell,$$

which is clearly a p-adic unit by the definition of $(\lambda_1, \dots, \lambda_n)$ being nearly ordinary with respect to $\underline{k}$. $\square$

Remark 4.18. Hida proved [2004b, Theorem 5.3] that the nearly ordinary vector is unique up to scalar.

Lemma 4.19. *Let* $\lambda_1, \dots, \lambda_{a+2b}$ *be characters of* $\mathbb{Q}_p^\times$ *such that* $\mathrm{cond}(\lambda_{a+2b}) > \cdots > \mathrm{cond}(\lambda_{b+1}) > \mathrm{cond}(\lambda_1) > \cdots > \mathrm{cond}(\lambda_b)$. *We define the subgroup* K_λ *of* $\mathrm{GL}_{a+2b}(\mathbb{Z}_p)$ *to be those matrices whose below-diagonal entries of the* i*-th column are divisible by cond*$(\lambda_{a+2b+1-i})$ *for* $1 \le i \le a+b$, *and the left-to-diagonal entries of the* j*-th row are divisible by cond*$(\lambda_{a+2b+1-j})$ *for* $a+b+2 \le j \le a+2b$. *Let* λ^{op} *be the character of* K_λ *defined by*

$$\lambda_{a+2b}(g_{11})\lambda_{a+2b-1}(g_{22}) \cdots \lambda_1(g_{a+2b\ a+2b}).$$

Then f^ℓ *is the unique* (*up to scalar*) *vector in* π *such that the action of* K_λ *is given by multiplying by* λ^{op}.

Proof. We only need to prove the uniqueness. We use the model of induced representation as above. Let $n = a+2b$ and let $e_1, \dots, e_n$ be the standard basis of the standard representation of GL_n. Let p^{t_i} be the conductor of λ_i. So $t_{a+2b} = \max\{t_i\}_i$. Write $K_0(p) \subset \mathrm{GL}_n(\mathbb{Z}_p)$ for the subgroup consisting of elements in $B(\mathbb{Z}_p)$ modulo p. Suppose f is any vector satisfying the requirement of the lemma. Let w be a Weyl element of GL_n such that f is not identically 0 on $wK_0(p)$. Then we see that $w \cdot e_1 = e_{a+2b}$ by considering right-multiplication by $\mathrm{diag}(1 + p^{t_{a+2b}-1}, 1, \dots, 1)$. Continuing this argument, we see that $w \cdot e_2 = e_{a+2b-1}, \dots$. Finally, we have $w = w^\ell$ and the lemma is clear by Bruhat decomposition. $\square$

We let

$$w_1 := \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & & & 1 \\ & & & & \iddots & \\ & & & 1 & & \end{pmatrix}.$$

Now let $\tilde{B} = B^{w_1}$ and $\tilde{K}_\lambda = K_\lambda^{w_1}$.

Corollary 4.20. *Denote $a_i := \nu_p(\lambda_i(p))$. Suppose $\lambda_1, \dots, \lambda_{a+2b}$ are such that $\mathrm{cond}(\lambda_1) > \cdots > \mathrm{cond}(\lambda_{a+2b})$ and $a_1 < \cdots < a_{a+b} < a_{a+2b} < \cdots < a_{a+b+1}$. Then the unique (up to scalar) ordinary section with respect to $\tilde{B}$ is*

$$f^{\mathrm{ord}}(x) = \begin{cases} \lambda_1(g_{11}) \cdots \lambda_{a+2b}(g_{a+2b,a+2b}) & \text{if } g \in \tilde{K}_\lambda, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We only need to prove that $\pi(w_1) f^{\mathrm{ord}}(x)$ is ordinary with respect to $\tilde{B}^{w_1} = B$. Let $\lambda'_1 = \lambda_{a+b+1}, \dots, \lambda'_b = \lambda_{a+2b}, \lambda'_{b+1} = \lambda_{a+b}, \dots, \lambda'_{a+2b} = \lambda_1$. Then λ' satisfies Lemma 4.17 and thus the ordinary section for B (up to scalar) is $f^\ell_{\lambda'}$. Since λ' also satisfies the assumptions of Lemma 4.19, $f^\ell_{\lambda'}$ is the unique section such that the action of K_λ is given by $\lambda'_{a+2b}(g_{11}) \cdots \lambda'_1(g_{a+2b,a+2b})$. But λ is clearly regular, so $\mathrm{Ind}_B^{\mathrm{GL}_{a+2b}}(\lambda) \simeq \mathrm{Ind}_B^{\mathrm{GL}_{a+2b}}(\lambda')$. So the ordinary section of $\mathrm{Ind}_B^{\mathrm{GL}_{a+2b}}(\lambda)$ for B also has the action of K_λ given by this character. It is easy to check that $\pi(w_1) f^{\mathrm{ord}}$ has this property and the uniqueness (up to scalar) gives the result. □

4D2. *Pullback sections.* In this subsubsection we construct a Siegel section on $\mathrm{U}(a+2b+1, a+2b+1)$ which pulls back to the nearly ordinary Klingen sections on $\mathrm{U}(a+b+1, b+1)$. We need to rearrange the basis since we are going to study large block matrices and the new basis will simplify the explanation. One can check that the Klingen Eisenstein series we construct in this subsection, when going back to our previous basis, is nearly ordinary with respect to the Borel subgroup

$$B_1 := \begin{pmatrix} * & * & & * & * \\ & * & & * & * \\ & * & * & * & * \\ & & & * & \\ & & & * & * \end{pmatrix},$$

where the first four blocks are upper-triangular and the fifth is lower-triangular. But the one we need is nearly ordinary with respect to the Borel subgroup

$$B_2 := \begin{pmatrix} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & \\ & & & * & * \end{pmatrix}$$

(it is for this one that we can use the Λ-adic Fourier–Jacobi expansions). (Here the blocks are with respect to the partition $b+1+a+b+1$.) There is a Weyl element w_{Borel} of GL_{a+2b+2} such that $w_{\mathrm{Borel}}^{-1} B_2 w_{\mathrm{Borel}} = B_1$. This w_{Borel} is in fact in the Weyl group of GL_{b+1+a} embedded as the upper-left minor. In the case of the doubling method $(\mathrm{U}(r,s) \times \mathrm{U}(s,r) \hookrightarrow \mathrm{U}(r+s, r+s))$ we have a corresponding change of index and we write w'_{Borel} for the corresponding Weyl element. In Section 4D4 we will come back to the original basis.

Now we explain the new basis. Let $V_{a,b}$ and $V_{a,b+1}$ be the Hermitian space with respective matrices

$$\begin{pmatrix} \zeta 1_a & & \\ & & 1_b \\ & -1_b & \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \zeta & & \\ & & 1_{b+1} \\ & -1_{b+1} & \end{pmatrix}.$$

(These are our skew-Hermitian spaces for $\mathrm{U}(r, s)$ and $\mathrm{U}(r+1, s+1)$ under the new basis.) The matrix S for the embedding $\mathrm{U}(V_{a,b}) \times \mathrm{U}(V_{a,b+1}) \hookrightarrow \mathrm{U}(V_{a+2b+1})$ becomes

$$\begin{pmatrix} 1 & & & & -\frac{1}{2}\zeta & & & \\ & 1 & & & & & & \frac{1}{2} \\ & & 1 & & & & & \\ & & & -1 & & \frac{1}{2} & & \\ -1 & & & & -\frac{1}{2}\zeta & & & \\ & & & 1 & & \frac{1}{2} & & \\ & & & & & & 1 & \\ & -1 & & & & & & -\frac{1}{2} \end{pmatrix}.$$

Godement sections at p. Let $v|p$ be a prime of F and $\mathcal{K}_v \simeq \mathbb{Q}_p \times \mathbb{Q}_p$. Let τ be a character of $\mathbb{Q}_p^\times \times \mathbb{Q}_p^\times$. Suppose $\tau = (\tau_1, \tau_2^{-1})$ and let p^{s_i} be the conductor of τ_i, $i = 1, 2$. Let $\chi_1, \dots \chi_a, \chi_{a+1}, \dots \chi_{a+2b}$ be characters of $\mathbb{Q}_p^\times$ whose conductors are $p^{t_1}, \dots, p^{t_{a+2b}}$. Suppose we are in the generic case:

Definition 4.21 (generic case).

$$t_1 > t_2 > \cdots > t_{a+b} > s_1 > t_{a+b+1} > \cdots > t_{a+2b} > s_2.$$

Also, let $\xi_i = \chi_i \tau_1^{-1}$ for $1 \le i \le a+b$ and $\xi_j = \chi_j^{-1} \tau_2$ for $a+b+2 \le j \le a+2b+1$. Let $\xi_{a+b+1} = 1$.

Let Φ_1 be the following Schwartz function on $M_{a+2b+1}(\mathbb{Q}_p)$: let Γ be the subgroup of $\mathrm{GL}_{a+2b+1}(\mathbb{Z}_p)$ consisting of matrices $\gamma = (\gamma_{ij})$ such that p^{t_k} divides the below-diagonal entries (i.e., $i > j$) of the k-th column for $1 \le k \le a+b$ and p^{s_1} divides γ_{ij} when $a+b+2 \le i \le a+2b+1$ and $j = a+b+1$; while $p^{t_{j-1}}$ divides γ_{ij} when $a+b+2 \le j \le a+2b+1$ and either $i \le a+b+1$ or $i > j$.

Let $\xi'_i = \chi_i \tau_2^{-1}$, $1 \le i \le a+b$, $\xi'_j = \chi_j^{-1} \tau_1$, $a+b+2 \le j \le a+2b+1$, and $\xi'_{a+b+1} = \tau_1 \tau_2^{-1}$. (Thus, $\xi'_k = \xi_k \tau_1 \tau_2^{-1}$ for any k.)

Definition 4.22. $$\Phi_1(x) = \begin{cases} 0 & \text{if } x \notin \Gamma, \\ \prod_{k=1}^{a+b+1} \xi'_k(x_{kk}) & \text{if } x \in \Gamma. \end{cases}$$

Now we define another Schwartz function Φ_2 on $M_{a+2b+1}(\mathbb{Q}_p)$.

Let $\mathfrak{X}$ be the following set: if

$$x = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix} \in \mathfrak{X}$$

is in block matrix form with respect to the partition $a+2b+1 = a+b+1+b$, then

- x has entries in $\mathbb{Z}_p$;
- $\left(\begin{smallmatrix} A_{11} & A_{14} \\ A_{21} & A_{24} \end{smallmatrix}\right)$ has i-th upper-left minors A_i such that $\det A_i \in \mathbb{Z}_p^\times$ for $i = 1, \ldots, a+b$; and
- A_{42} has i-th upper-left minors B_i such that $\det B_i \in \mathbb{Z}_p^\times$ for $i = 1, \ldots, b$.

We define

$$\Phi_\xi(x) = \begin{cases} 0 & \text{if } x \notin \mathfrak{X}, \\ \dfrac{\xi_1}{\xi_2}(\det A_1) \cdots \dfrac{\xi_{a+b-1}}{\xi_{a+b}}(\det A_{a+b-1})\xi_{a+b}(\det A_{a+b}) & \\ \quad \times \dfrac{\xi_{a+b+2}}{\xi_{a+b+3}}(\det B_1) \cdots \dfrac{\xi_{a+2b}}{\xi_{a+2b+1}}(\det B_{b-1})\xi_{a+2b+1}(\det B_b) & \text{if } x \in \mathfrak{X}. \end{cases} \tag{11}$$

This is a locally constant function with compact support. Let

$$\Phi_2(x) := \tilde{\Phi}_\xi(x) = \int_{M_{a+2b+1}(\mathbb{Q}_p)} \Phi_\xi(y) e_p(\mathrm{Tr}\, y\, {}^t x)\, dy$$

(where tilde stands for Fourier transform). Let Φ be the Schwartz function on $M_{a+2b+1,2(a+2b+1)}(\mathbb{Q}_p)$ defined by

$$\Phi(X, Y) := \Phi_1(X)\Phi_2(Y)$$

and define a Godement section (terminology of Jacquet) by

$$f^\Phi(g) = \tau_2(\det g)|\det g|_p^{-s+(a+2b+1)/2} \times \int_{\mathrm{GL}_{a+2b+1}(\mathbb{Q}_p)} \Phi((0, X)g)\tau_1^{-1}\tau_2(\det X)|\det X|_p^{-2s+a+2b+1}\, d^\times X.$$

Lemma 4.23. *If $\gamma \in \Gamma$, then*

$$\Phi_\xi({}^t\gamma X) = \prod_{k=1}^{a+2b+1} (\xi_k(\gamma_{kk}))\Phi_\xi(X).$$

Proof. This is straightforward. For example, to see that the A_{42} block of ${}^t\gamma X$ has invertible upper-left minors (i.e., has determinants in $\mathbb{Z}_p^\times$) for $\gamma \in \Gamma$, $X \in \mathfrak{X}$, one notes that all entries of the upper-right block of γ are zero modulo p, and that multiplying by invertible matrices which are lower-triangular modulo p does not change the property that all upper-left minors are invertible. □

Fourier coefficients. For z in the absolutely convergent range and $\beta \in S_{a+2b+1}(\mathbb{Q}_p)$ (which is isomorphic to $M_{a+2b+1}(\mathbb{Q}_p)$ through the first projection), the Fourier coefficient is defined by

$$\begin{aligned} f_\beta^\Phi(1,z) &= \int_{M_{a+2b+1}(\mathbb{Q}_p)} f^\Phi\left(\begin{pmatrix} & 1_{a+2b+1} \\ -1_{a+2b+1} & \end{pmatrix}\begin{pmatrix} 1 & N \\ & 1\end{pmatrix}\right) e_p(-\operatorname{Tr}\beta N)\,dN \\ &= \int_{M_{a+2b+1}(\mathbb{Q}_p)} \int_{\mathrm{GL}_{a+2b+1}(\mathbb{Q}_p)} \Phi\left((0,X)\begin{pmatrix} & 1_{a+2b+1} \\ -1_{a+2b+1} & -N\end{pmatrix}\right) \\ &\qquad\qquad \times \tau_1^{-1}\tau_2(\det X)|\det X|_p^{-2z+a+2b+1} e_p(-\operatorname{Tr}\beta N)\,dN\,d^\times X \\ &= \int_{\mathrm{GL}_{a+2b+1}(\mathbb{Q}_p)} \Phi_1(-X)\Phi_\xi(-{}^tX^{-1}\,{}^t\beta)\tau_1^{-1}\tau_2(\det X)|\det X|_p^{-2z}\,d^\times X \\ &= \tau_1^{-1}\tau_2(-1)\operatorname{Vol}(\Gamma)\Phi_\xi({}^t\beta). \end{aligned} \tag{12}$$

Definition 4.24. Let $\tilde{f}^\dagger = \tilde{f}^\dagger_{a+2b+1}$ be the Siegel section supported on

$$Q(\mathbb{Q}_p)w_{a+2b+1}\begin{pmatrix} 1 & M_{a+2b+1}(\mathbb{Z}_p) \\ & 1\end{pmatrix}$$

and $\tilde{f}^\dagger\big(w_{a+2b+1}\left(\begin{smallmatrix} 1 & X \\ & 1\end{smallmatrix}\right)\big) = 1$ for $X \in M_{a+2b+1}(\mathbb{Z}_p)$.

Lemma 4.25.
$$\tilde{f}^\dagger_\beta(1) = \begin{cases} 1 & \text{if } \beta \in M_{a+2b+1}(\mathbb{Z}_p), \\ 0 & \text{if } \beta \notin M_{a+2b+1}(\mathbb{Z}_p), \end{cases}$$

(*here we used the projection of* β *onto its first component in* $\mathcal{K}_v = F_v \times F_v$), *where the first component corresponds to the element inside our CM-type* Σ_∞ *under* $\iota := \mathbb{C} \simeq \mathbb{C}_p$ (*see Section 2A*).

Definition 4.26.
$$f^\dagger := \frac{f^\Phi}{\tau_1^{-1}\tau_2(-1)\operatorname{Vol}(\Gamma)}.$$

Thus, $f^\dagger_\beta = \Phi_\xi({}^t\beta)$.

We define

$$c_n(\tau', z) := \begin{cases} \tau'(p^{nt})p^{2ntz-tn(n+1)/2} & \text{if } t > 0, \\ p^{2nz-n(n+1)/2} & \text{if } t = 0. \end{cases} \tag{13}$$

Now we recall a lemma from Skinner and Urban [2014, Lemma 11.12], which will be useful later.

Lemma 4.27. *Suppose* $v|p$ *and* $\beta \in S_n(\mathbb{Q}_v)$, $\det\beta \neq 0$. *Then*:

(i) *If* $\beta \notin S_n(\mathbb{Z}_v)$ *then* $M(z, \tilde{f}^\dagger_n)_\beta(-z, 1) = 0$.

(ii) *Suppose $\beta \in S_n(\mathbb{Z}_v)$. Let $t := \mathrm{ord}_v(\mathrm{cond}(\tau'))$. Then*

$$M(z, \tilde{f}_n^{\dagger})_\beta(-z, 1) = \tau'(\det\beta)|\det\beta|_v^{-2z}\mathfrak{g}(\bar{\tau}')^n c_n(\tau', z).$$

Note that our $\tilde{f}^{\dagger}$ is the $f^{\dagger}$ in [Skinner and Urban 2014] and our τ is their χ.

Now we want to write down our Godement section f^{Φ} in terms of $\tilde{f}^{\dagger}$. First we prove the following:

Lemma 4.28. *Suppose $\Phi_{\xi,n}$ is the function on $M_n(\mathbb{Q}_p)$ defined as follows: If $\mathrm{cond}(\xi_i) = (p^{t_i})$ for $t_1 > \cdots > t_n$ and ξ_i are characters of $\mathbb{Q}_p^\times$ with conductor p^{t_i}. Let $\mathfrak{X}_n$ be the subset of $M_n(\mathbb{Z}_p)$ such that the i-th upper-left minor M_i has determinant in $\mathbb{Z}_p^\times$. Define $\Phi_{\xi,n}$ to be*

$$\frac{\xi_1}{\xi_2}(\det M_1)\cdots\frac{\xi_{n-1}}{\xi_n}(\det M_{n-1})\xi_n(\det M_n)$$

on $\mathfrak{X}_n$ and 0 otherwise. Let

$$\tilde{\mathfrak{X}}_{\xi,n} = \tilde{\mathfrak{X}}_\xi := N(\mathbb{Z}_p)\begin{pmatrix} p^{-t_1}\mathbb{Z}_p^\times & & \\ & \ddots & \\ & & p^{-t_n}\mathbb{Z}_p^\times\end{pmatrix} N^{\mathrm{opp}}(\mathbb{Z}_p).$$

Then the Fourier transform $\hat{\Phi}_\xi$ of Φ_ξ is the function

$$\tilde{\Phi}_\xi(x) = \begin{cases} 0 & \text{if } x \notin \tilde{\mathfrak{X}}_\xi, \\ \prod_{i=1}^n \mathfrak{g}(\xi_i)\prod_{i=1}^n \bar{\xi}_i(x_i p^{t_i}) & \text{if } x \in \tilde{\mathfrak{X}}_\xi,\end{cases}$$

$$\text{where} \quad x = \begin{pmatrix} 1 & \times & \times \\ & \ddots & \times \\ & & 1\end{pmatrix}\begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_n\end{pmatrix}\begin{pmatrix} 1 & & \\ \times & \ddots & \\ \times & \times & 1\end{pmatrix}.$$

Proof. First suppose x is in the "big cell" $N(\mathbb{Q}_p)T(\mathbb{Q}_p)N^{\mathrm{opp}}(\mathbb{Q}_p)$. It is easily seen that we can write x in terms of block matrices,

$$x = \begin{pmatrix} 1_{n-1} & u \\ & 1\end{pmatrix}\begin{pmatrix} z & \\ & w\end{pmatrix}\begin{pmatrix} 1_{n-1} & \\ v & 1\end{pmatrix},$$

where $z \in \mathrm{GL}_{n-1}(\mathbb{Q}_p)$, $w \in \mathbb{Q}_p^\times$, $u \in M_{n-1,1}(\mathbb{Q}_p)$ and $v \in M_{1,n-1}(\mathbb{Q}_p)$. A first observation is that $\tilde{\Phi}_\xi$ is invariant under right-multiplication by $N^{\mathrm{opp}}(\mathbb{Z}_p)$ and left-multiplication by $N(\mathbb{Z}_p)$. We show that $v \in M_{1\times(n-1)}(\mathbb{Z}_p)$ if $\tilde{\Phi}_\xi(x) \neq 0$. By definition,

$$\tilde{\Phi}_\xi(x) = \int_{M_n(Q_p)} \Phi_\xi(y)e_p(\mathrm{Tr}\, y\, {}^t x)\, dy,$$

so, writing

$$y = \begin{pmatrix} 1_{n-1} & \\ \ell & 1\end{pmatrix}\begin{pmatrix} a & \\ & b\end{pmatrix}\begin{pmatrix} 1 & m \\ & 1\end{pmatrix} = \begin{pmatrix} a & am \\ \ell a & \ell am + b\end{pmatrix},$$

we have

$$\begin{aligned}\tilde{\Phi}_\xi(x) &= \int_{a\in\mathfrak{X}_{\xi,n-1},\, m\in M(\mathbb{Z}_p),\, \ell\in M(\mathbb{Z}_p),\, b\in\mathbb{Z}_p^\times} \Phi_\xi\left(\begin{pmatrix}1 & \\ \ell & 1\end{pmatrix}\begin{pmatrix}a & \\ & b\end{pmatrix}\begin{pmatrix}1 & m \\ & 1\end{pmatrix}\right)\\ &\qquad\times e_p\left(\mathrm{Tr}\left(\begin{pmatrix}1 & \\ {}^tm & 1\end{pmatrix}\begin{pmatrix}{}^ta & \\ & b\end{pmatrix}\begin{pmatrix}1 & {}^t\ell \\ & 1\end{pmatrix}\begin{pmatrix}1 & u \\ & 1\end{pmatrix}\begin{pmatrix}z & \\ & w\end{pmatrix}\begin{pmatrix}1 & \\ v & 1\end{pmatrix}\right)\right)dy\\ &= \int \Phi_\xi\left(\begin{pmatrix}a & \\ & b\end{pmatrix}\right)e_p\left(\mathrm{Tr}\begin{pmatrix}1 & \\ {}^tm+v & 1\end{pmatrix}\begin{pmatrix}{}^ta & \\ & b\end{pmatrix}\begin{pmatrix}1 & {}^t\ell+u \\ & 1\end{pmatrix}\begin{pmatrix}z & \\ & w\end{pmatrix}\right)dy\\ &= \int \Phi_\xi\left(\begin{pmatrix}a & \\ & b\end{pmatrix}\right)\\ &\qquad\times e_p\left(\mathrm{Tr}\left(\begin{pmatrix}{}^ta & {}^ta({}^t\ell+u) \\ ({}^tm+v)\,{}^ta & ({}^tm+v)\,{}^ta({}^t\ell+u)+b\end{pmatrix}\begin{pmatrix}z & \\ & w\end{pmatrix}\right)\right)dy\\ &= \int \Phi_\xi\left(\begin{pmatrix}a & \\ & b\end{pmatrix}\right)e_p\big(\mathrm{Tr}({}^taz+(({}^tm+v)\,{}^ta({}^t\ell+u)+b)w)\big)\,dy.\end{aligned}$$

(Note that Φ_ξ is invariant under transpose.)

If $\tilde{\Phi}_\xi(x)\neq 0$, then it follows from the last expression that $w\in p^{-t_n}\mathbb{Z}_p^\times$. Suppose $v\notin M_{1\times(n-1)}(\mathbb{Z}_p)$; then ${}^tm+v\notin M_{1\times(n-1)}(\mathbb{Z}_p)$. We let a, m and b be fixed and let ℓ vary in $M_{1\times(n-1)}(\mathbb{Z}_p)$; we find that this integral must be 0. (Notice that $a\in\mathfrak{X}_{\xi,n-1}$ and $w\in p^{-t_n}\mathbb{Z}_p^\times$, thus $({}^tm+v)\,{}^taw\notin M_{1\times n-1}(\mathbb{Z}_p)$). Thus, a contradiction. Therefore, $v\in M_{1\times n-1}(\mathbb{Z}_p)$, and similarly $u\in M_{n-1,1}(\mathbb{Z}_p)$. Thus, by the observation at the beginning of the proof, we may assume $u=0$ and $v=0$ without loss of generality.

Thus, if we write $\Phi_{\xi,n-1}$ as the restriction of Φ_ξ to the upper-left $(n-1)\times(n-1)$ minor,

$$\begin{aligned}\tilde{\Phi}_\xi(x) &= \int \Phi_\xi\left(\begin{pmatrix}a & \\ & b\end{pmatrix}\right)e_p\big(\mathrm{Tr}({}^taz+({}^tm\,{}^ta\,{}^t\ell+b)w)\big)\,dy\\ &= p^{-nt_n}\mathfrak{g}(\xi_n)\bar{\xi}_n(wp^{t_n})\int_{a\in\mathfrak{X}_{\xi,n-1}}\Phi_{\xi,n-1}(a)e_p({}^taz)\,dy.\end{aligned}$$

By an induction procedure one gets

$$\tilde{\Phi}_\xi(x)=\begin{cases}0 & \text{if } x\notin\tilde{\mathfrak{X}}_{\xi,n},\\ p^{-\sum_{i=1}^n it_i}\prod_{i=1}^n\mathfrak{g}(\xi_i)\prod_{i=1}^n\bar{\xi}_i(x_ip^{t_i}) & \text{if } x\in\tilde{\mathfrak{X}}_{\xi,n}.\end{cases}$$

We have thus proved that $\tilde{\Phi}_{\xi,n}$, when restricted to the "big cell", has support in $\tilde{\mathfrak{X}}_{\xi,n}$. Since $\tilde{\mathfrak{X}}_{\xi,n}$ is compact, $\tilde{\Phi}_{\xi,n}$ itself must be supported in $\tilde{\mathfrak{X}}_{\xi,n}$. □

Lemma 4.29. *Let $\tilde{\mathfrak{X}}_\xi$ be the support of $\Phi_2 = \hat{\Phi}_\xi$; then a complete set of representatives of $\tilde{\mathfrak{X}}_\xi$ $\bmod M_{a+2b+1}(\mathbb{Z}_p)$ is given by the elements*

$$\begin{pmatrix} A & & B \\ C & & D \\ & & \\ & E & \end{pmatrix},$$

where the blocks are with respect to the partition $a+b+1+b$ and $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ runs over the set

$$\begin{pmatrix} 1 & m_{12} & \cdots & m_{1,a+b} \\ & \ddots & \ddots & \vdots \\ & & \ddots & m_{a+b-1,a+b} \\ & & & 1 \end{pmatrix} \begin{pmatrix} x_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & x_{a+b} \end{pmatrix} \begin{pmatrix} 1 & & & \\ n_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ n_{a+b,1} & \cdots & n_{a+b,a+b-1} & 1 \end{pmatrix},$$

where x_i runs over $p^{-t_i}\mathbb{Z}_p^\times \bmod \mathbb{Z}_p$, m_{ij} runs over $\mathbb{Z}_p \bmod p^{t_j}$, n_{ij} runs over $\mathbb{Z}_p \bmod p^{t_i}$, and E runs over the set

$$\begin{pmatrix} 1 & k_{12} & \cdots & k_{1,b} \\ & \ddots & \ddots & \vdots \\ & & \ddots & k_{b-1,b} \\ & & & 1 \end{pmatrix} \begin{pmatrix} y_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & y_b \end{pmatrix} \begin{pmatrix} 1 & & & \\ \ell_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \ell_{b,1} & \cdots & \ell_{b,b-1} & 1 \end{pmatrix},$$

where y_i runs over $p^{-t_{i+a+b}}\mathbb{Z}_p^\times \bmod \mathbb{Z}_p$, k_{ij} runs over $\mathbb{Z}_p \bmod p^{t_{a+b+j}}$, and ℓ_{ij} runs over $\mathbb{Z}_p \bmod p^{t_{a+b+i}}$.

Proof. This is elementary and we omit it here. □

We also define, for $g \in \mathrm{GL}_{a+2b}(\mathbb{Q}_p)$,

$$g^\iota = \begin{pmatrix} & 1_{a\times a} & \\ & & 1_{b\times b} \\ 1_{b\times b} & & \end{pmatrix} g \begin{pmatrix} & & 1_{b\times b} \\ 1_{a\times a} & & \\ & 1_{b\times b} & \end{pmatrix}$$

and

$$g_\iota = \begin{pmatrix} & 1_{a\times a} & \\ & & 1_{b\times b} \\ 1_{b\times b} & & \end{pmatrix}^{-1} g \begin{pmatrix} & & 1_{b\times b} \\ 1_{a\times a} & & \\ & 1_{b\times b} & \end{pmatrix}^{-1}.$$

Corollary 4.30. *We have*

$$f^\dagger(z,g) = p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+i}} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1)$$
$$\times \sum_{A,B,C,D,E} \prod_{i=1}^{a} \bar{\xi}_i\left(\frac{\det A_i}{\det A_{i-1}} p^{t_i}\right) \prod_{i=1}^{b} \bar{\xi}_{a+i,a+i}\left(\frac{\det D_i}{\det D_{i-1}} p^{t_{a+i}}\right)$$
$$\times \prod_{i=1}^{b} \bar{\xi}_{a+b+1+i}\left(\frac{\det E_i}{\det E_{i-1}} p^{t_{a+b+i}}\right) \tilde{f}^\dagger\left(z, g\begin{pmatrix} 1_{a+2b+1} & \begin{matrix} A & B \\ C & D \\ & \\ & E \end{matrix} \\ & 1_{a+2b+1} \end{pmatrix}\right).$$

Here A_i is the i-th upper-left minor of A, D_i is the $(a+i)$-th upper-left minor of $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ (not of D), E_i is the i-th upper-left minor of E, and the sum runs over the set of representatives of Lemma 4.29.

Proof. We only need to check the Siegel Eisenstein sections on both sides coincide on $wN_{a+2b+1}(\mathbb{Q}_p)$, since the big cell $Q_{a+2b+1}(\mathbb{Q}_p)wN_{a+2b+1}(\mathbb{Q}_p)$ is dense in $\mathrm{GL}_{2a+4b+2}$. To see this we just need to know that they have the same β-th Fourier coefficients for all $\beta \in S_{a+2b+1}(\mathbb{Q}_p)$. But this is seen by (12) and Lemmas 4.28 and 4.29. ☐

Now we define several sets. Let $\mathfrak{B}'$ be the set of $(a+b)\times(a+b)$ upper-triangular matrices of the form

$$\begin{pmatrix} 1 & m_{12} & \cdots & m_{1,a+b} \\ & \ddots & \ddots & \vdots \\ & & \ddots & m_{a+b-1,a+b} \\ & & & 1 \end{pmatrix}\begin{pmatrix} x_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & x_{a+b} \end{pmatrix},$$

where x_i runs over $\mathbb{Z}_p^\times \bmod p^{t_i}$ and m_{ij} runs over $\mathbb{Z}_p \bmod p^{t_j}$.

Let $\mathfrak{D}'$ be the set of $b\times b$ lower-triangular matrices of the form

$$\begin{pmatrix} 1 & & & \\ n_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ n_{a+b,1} & \cdots & n_{a+b,a+b-1} & 1 \end{pmatrix},$$

where n_{ij} runs over $\mathbb{Z}_p \bmod p^{t_{i+a+b}}$.

Let $\mathfrak{E}'$ be the set of $b\times b$ upper-triangular matrices of the form

$$\begin{pmatrix} 1 & k_{12} & \cdots & k_{1,b} \\ & \ddots & \ddots & \vdots \\ & & \ddots & k_{b-1,b} \\ & & & 1 \end{pmatrix},$$

where k_{ij} runs over $\mathbb{Z}_p$ mod $p^{t_{a+b+j}}$.

Let $\mathfrak{C}'$ be the set of $(a+b)\times(a+b)$ lower-triangular matrices of the form

$$\begin{pmatrix} y_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & y_b \end{pmatrix}\begin{pmatrix} 1 & & & \\ \ell_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \ell_{b,1} & \cdots & \ell_{b,b-1} & 1 \end{pmatrix},$$

where y_i runs over $\mathbb{Z}_p^\times$ mod $p^{t_{i+a+b}}\mathbb{Z}_p$ and ℓ_{ij} runs over $\mathbb{Z}_p$ mod $p^{t_{a+b+i}}$.

Thus, if B', C', D', E' run over the set $\mathfrak{B}'$, $\mathfrak{C}'$, $\mathfrak{D}'$, $\mathfrak{E}'$, respectively, then

$$\begin{aligned} f^\dagger(z,g) &= p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+i}} \\ &\qquad \times \sum_{B',C',D',E'} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1) \\ &\qquad \times \sum_{B',C',D',E'} \prod_{i=1}^{a+b} \bar{\xi}_i(B'_{ii}) \prod_{i=1}^{b} \xi_{a+b+i}(C'_{ii}) \\ &\qquad \times \tilde{f}^\dagger\left(z, g\alpha\left(\mathrm{diag}(B',1,C'1), \begin{pmatrix} E' & \\ & D' \end{pmatrix}\right)^\iota\right)\begin{pmatrix} 2 & A' \\ & 1 \end{pmatrix} \\ &\qquad \times \alpha\left(\mathrm{diag}(B',1,C'1), \begin{pmatrix} E' & \\ & D' \end{pmatrix}\right)^{\iota\,-1}\Bigg) \\ &= p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+1+i}} \\ &\qquad \sum_{B',C',D',E'} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1) \\ &\qquad \times \sum_{B',C',D',E'} \prod_{i=1}^{a+b} \bar{\xi}_i(B'_{ii}) \prod_{i=1}^{b} \xi_{a+b+i}(C'_{ii}) \prod_{i=1}^{a+b} \bar{\tau}_1(B'_{ii}) \prod_{i=1}^{b} \bar{\tau}_2(C'_{ii}) \\ &\qquad \times \tilde{f}^\dagger\left(z, g\alpha\left(\mathrm{diag}(B',1,C'1), \begin{pmatrix} E' & \\ & D' \end{pmatrix}\right)^\iota\right)\begin{pmatrix} 1 & A' \\ & 1 \end{pmatrix}\Bigg), \end{aligned}$$

where

$$A' = \begin{pmatrix} p^{-t_1} & & & & & & & & \\ & \ddots & & & & & & & \\ & & p^{-t_a} & & & & & & \\ & & & & & & p^{-t_{a+1}} & & \\ & & & & & & & \ddots & \\ & & & & & & & & p^{-t_{a+b}} \\ & & & & & 0 & & & \\ & & & p^{-t_{a+b+1}} & & & & & \\ & & & & \ddots & & & & \\ & & & & & p^{-t_{a+2b}} & & & \end{pmatrix}. \tag{14}$$

We let

$$\gamma = \begin{pmatrix} \zeta^{-1} & & & & -\zeta^{-1} & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ \frac{1}{2} & & & & \frac{1}{2} & & & \\ & & & 1 & & 1 & & \\ & & & & & & 1 & \\ & 1 & & & & & & 1 \end{pmatrix} \quad \text{and} \quad w' = \begin{pmatrix} & & & & 1 & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ -1 & & & & & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{pmatrix}.$$

Definition 4.31 (pullback section). If f is a Siegel section and $\varphi \in \pi_p$, then

$$F_\varphi(z, f, g) := \int_{\mathrm{GL}_{a+2b}(\mathbb{Q}_p)} f(z, \gamma\alpha(g, g_1)\gamma^{-1})\bar{\tau}(\det g_1)\rho(g_1)\varphi \, dg_1.$$

Now we define a subset K of $\mathrm{GL}_{a+2b+2}(\mathbb{Z}_p)$ so that $k \in K$ if and only if p^{t_i} divides the below-diagonal entries of the i-th column for $1 \le i \le a+b$, p^{s_1} divides the below-diagonal entries of the $(a+b+1)$-th column, and $p^{t_{a+b+j}}$ divides the right-to-diagonal entries of the $(a+b+1+j)$-th row for $1 \le j \le b-1$. We also define ν, a character of K, by

$$\nu(k) = \tau_1(k_{a+b+1,a+b+1})\tau_2(k_{a+2b+2,a+2b+2}) \times \prod_{i=1}^{a+b} \chi_i(k_{ii}) \prod_{i=1}^{b} \chi_{a+b+i}(k_{a+b+i+1,a+b+i+1})$$

for any $k \in K$.

Definition 4.32. We define Υ to be the element in $\mathrm{U}(n, n)(F_v)$ (which equals $\mathrm{U}(n, n)(\mathbb{Q}_p)$) such that the projection to the first component of $\mathcal{K}_v = F_v \times F_v$ equals that of γ (note that $\gamma \notin \mathrm{U}(n, n)(F_v)$).

Lemma 4.33. *Let $K' \subset K$ be the compact subgroup defined by, for $k \in K$,*

$$k = \begin{pmatrix} a_1 & a_2 & a_3 & b_1 & b_2 \\ a_4 & a_5 & a_6 & b_3 & b_4 \\ a_7 & a_8 & a_9 & b_5 & b_6 \\ c_1 & c_2 & c_3 & d_1 & d_2 \\ c_4 & c_5 & c_6 & d_3 & d_4 \end{pmatrix} \in K'$$

(here the blocks are with respect to the partition $(a+b+1+b+1)$) if and only if $p^{t_{a+b+i}+t_j}$ divides the (i, j)-th entry of c_1 for $1 \le i \le b$, $1 \le j \le a$, and $p^{t_{a+b+i}+t_{a+j}}$ divides the (i, j)-th entry of c_2 for $1 \le i \le b$, $1 \le j \le b$. (It is not hard to check that this is a group.)

Then $F_\varphi(z, \rho(\Upsilon) f^\dagger, gk) = \nu(k) F_\varphi(z, \rho(\Upsilon) f^\dagger, g)$ for any $\varphi \in \pi$ and $k \in K'$.

Proof. This follows directly from the action of K' on the Godement section $f^\dagger$. □

We define K'' to be the subgroup of K that consists of matrices

$$\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ c_1 & c_2 & & 1 & \\ & & & & 1 \end{pmatrix}$$

such that p^{t_j} divides the (i, j)-th entry of c_1 for $1 \le i \le b$, $1 \le j \le a$, and $p^{t_{a+j}}$ divides the (i, j)-th entry of c_2 for $1 \le i \le b$, $1 \le j \le b$.

Definition 4.34. Let $\tilde{K} \subset \mathrm{GL}_{a+2b}(\mathbb{Z}_p)$ be the set of matrices

$$\begin{pmatrix} a_1 & a_3 & a_2 \\ a_7 & a_9 & a_8 \\ a_4 & a_6 & a_5 \end{pmatrix}$$

(the blocks are with respect to the partition $(b+a+b)$) such that the columns of a_3 and a_6 are divisible by $p^{t_1}, \ldots, p^{t_a}$, the columns of a_4 are divisible by $p^{t_{a+1}}, \ldots, p^{t_{a+b}}$, $p^{t_{a+i}}$ divides the below-diagonal entries of the i-th column of a_1 $(1 \le i \le b)$, p^{t_j} divides the below-diagonal entries of the j-th column of a_9 $(1 \le j \le a)$, and $p^{t_{a+b+k}}$ divides the above-diagonal entries of the k-th row of a_5.

Let $\tilde{K}' \subset \tilde{K}$ be those matrices such that $p^{t_{a+b+i}+t_{a+j}}$ divides the (i, j)-th entry of a_4 for $1 \le i \le b$, $1 \le j \le b$, and $p^{t_{a+b+i}+t_j}$ divides the (i, j)-th entry of a_6 for $1 \le i \le b$, $1 \le j \le a$. We also define $\tilde{K}''$ to be the subset of $\tilde{K}$ consisting of matrices

$$\begin{pmatrix} 1 & & \\ & 1 & \\ a_4 & a_6 & 1 \end{pmatrix}$$

such that $p^{t_{a+j}}$ divides the (i, j)-th entry of a_4 for $1 \le i \le b$, $1 \le j \le b$, and p^{t_j} divides the (i, j)-th entry of a_6 for $1 \le i \le b$, $1 \le j \le a$. We also define $\tilde{\nu}$, a character of $\tilde{K}$, by

$$\tilde{\nu}(k) = \prod_{i=1}^{b} \chi_{a+i}(k_{i,i}) \prod_{i=1}^{a} \chi_i(k_{b+i,b+i}) \prod_{i=1}^{b} \chi_{a+b+i}(k_{a+b+i,a+b+i}).$$

The following lemma will be useful in identifying our pullback section:

Lemma 4.35. *Suppose $F_\varphi(z, \rho(\Upsilon) f^\dagger, g)$ as a function of g is supported in PwK and*

$$F_\varphi(z, \rho(\Upsilon) f^\dagger, gk) = \nu(k) F_\varphi(z, \rho(\Upsilon) f^\dagger, g)$$

for $k \in K'$, and $F_\varphi(z, \rho(\Upsilon) f^\dagger, w)$ is invariant under the action of $(\tilde{K}'')^\iota$. Then $F_\varphi(a, \rho(\Upsilon) f^\dagger, g)$ is the unique section (up to scalar) whose action by $k \in K$ is given by multiplying by $\nu(k)$.

Proof. This is easy from the fact that $K = K'K'' = K''K'$. The uniqueness follows from Lemma 4.19. ☐

From now on in this subsection we use w to denote

$$\begin{pmatrix} 1_a & & \\ & & 1_{b+1} \\ & -1_{b+1} & \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1_a & & \\ & & 1_b \\ & -1_b & \end{pmatrix}.$$

Lemma 4.36. *If $\gamma\alpha(g, 1)\gamma^{-1} \in \mathrm{supp}(\rho(\Upsilon) f^\dagger)$ then $g \in PwK$. (Here ρ denotes the action of $\mathrm{GU}_{a+2b+1}(F_v)$ on the Siegel sections given by right-translation.)*

Proof. Since $f^\dagger$ is of the form $f^\dagger(g) = \sum_{A \in \mathfrak{X}} \tilde{f}^\dagger\big(g\big(\begin{smallmatrix} 1 & A \\ & 1 \end{smallmatrix}\big)\big)$, where $\mathfrak{X}$ is some set, we only have to check the lemma for each term in the summation.

Recall we defined A' in (14), where the blocks are with respect to the partition $(a+b+1+b)$. Let ζ_v and γ_v be the projection of ζ and γ_v to the first component of $\mathcal{K}_v \simeq F_v \times F_v$; then

$$\gamma_v = \begin{pmatrix} \zeta_v^{-1} & & & & -\zeta_v^{-1} & & & \\ & 1_b & & & & & & \\ & & 1 & & & & & \\ & & & 1_b & & & & \\ \frac{1}{2}1_a & & & & \frac{1}{2}1_a & & & \\ & & & 1_b & & 1_b & & \\ & & & & & & 1 & \\ & 1_b & & & & & & 1_b \end{pmatrix}$$

$$
= \begin{pmatrix}
2\zeta_v^{-1} & & & & -\zeta_v^{-1} & & & \\
& 1_b & & & & & & \\
& & 1 & & & & & \\
& & & 1_b & & & & \\
& & & & \frac{1}{2}1_a & & & \\
& & & & & 1_b & & \\
& & & & & & 1 & \\
& & & & & & & 1_b
\end{pmatrix}
\begin{pmatrix}
1_a & & & & & & & \\
& 1_b & & & & & & \\
& & 1 & & & & & \\
& & & 1_b & & & & \\
1_a & & & & 1_a & & & \\
& & & 1_b & & 1_b & & \\
& & & & & & 1 & \\
& 1_b & & & & & & 1_b
\end{pmatrix}.
$$

We denote the last term by $\tilde{\gamma}_v$ (different from the definition in the prime-to-p case).

Using the expression for $f^\dagger$ involving the various B', C', D' and E' as above and the fact that $\gamma(m(g,1),g) \in Q$ and that K is invariant under right-multiplication by any B or C, we only need to check that if $\tilde{\gamma}_v\alpha(g,1)\tilde{\gamma}_v^{-1} \in \mathrm{supp}\big(\rho(\Upsilon)\rho\big(\big(\begin{smallmatrix}1 & A' \\ & 1\end{smallmatrix}\big)\big)\big)\tilde{f}^\dagger$ then $g \in PwK$. Our calculations below are generalizations of the proof of [Skinner and Urban 2014, Proposition 11.16]. If

$$
gw = \begin{pmatrix}
a_1 & a_2 & a_3 & b_1 & b_2 \\
a_4 & a_5 & a_6 & b_3 & b_4 \\
a_7 & a_8 & a_9 & b_5 & b_6 \\
c_1 & c_2 & c_3 & d_1 & d_2 \\
c_4 & c_5 & c_6 & d_3 & d_4
\end{pmatrix}
$$

then this is equivalent to

$$
\begin{pmatrix}
1_a & & & & & & & \\
& 1_b & & & & & & \\
& & 1 & & & & & \\
& & & 1_b & & & & \\
1_a & & & & 1_a & & & \\
& & & 1_b & & 1_b & & \\
& & & & & & 1 & \\
& 1_b & & & & & & 1_b
\end{pmatrix}
\cdot
\begin{pmatrix}
a_1 & a_2 & a_3 & & & b_1 & b_2 & \\
a_4 & a_5 & a_6 & & & b_3 & b_4 & \\
a_7 & a_8 & a_9 & & & b_5 & b_6 & \\
& & & 1_b & & & & \\
& & & & 1_c & & & \\
c_1 & c_2 & c_3 & & & d_1 & d_2 & \\
c_4 & c_5 & c_6 & & & d_3 & d_4 & \\
& & & & & & & 1_b
\end{pmatrix}
\alpha(1,w^{-1})w'
$$

$$
\times \mathfrak{M}
\begin{pmatrix}
1_a & & & & & & & \\
& 1_b & & & & & & \\
& & 1 & & & & & \\
& & & 1_b & & & & \\
-1_a & & & & 1_a & & & \\
& & & -1_b & & 1_b & & \\
& & & & & & 1 & \\
& -1_b & & & & & & 1_b
\end{pmatrix}
\mathfrak{M}^{-1} w_{a+2b+1}^{-1}
$$

being in supp $\tilde{f}^\dagger$, where

$$\mathfrak{M} = \mathrm{diag}(p^{t_1}, \dots, p^{t_a}, 1_b, 1, p^{t_{a+1}}, \dots, p^{t_{a+b}}, 1_a, 1_b, 1, p^{-t_{a+b+1}}, \dots, p^{-t_{a+2b}})$$

temporarily, which is equivalent to

$$\tilde{\gamma}_v \alpha \left(gw, \begin{pmatrix} & & -1_b \\ & 1_a & \\ 1_b & & \end{pmatrix} \mathrm{diag}(p^{-t_{a+b+1}}, \dots, p^{t_1}, \dots, p^{t_{a+1}}, \dots) \right) w' \tilde{\gamma}_v^{-1}$$

belonging to

$$\mathrm{supp}\big(\rho(\mathfrak{M}^{-1} w_{a+2b+1}) \tilde{f}^\dagger\big).$$

The right-hand side is contained in

$$Q_t := Q \cdot \left\{ \begin{pmatrix} 1 & \\ S & 1 \end{pmatrix} \,\middle|\, S \in S_t = \begin{pmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{pmatrix} \right\},$$

where the blocks for S_t are with respect to the partition $a+b+1+b$ and consist of matrices $S_{ij} \in M(\mathbb{Z}_p)$ such that p^{t_i} divides the i-th column of the matrix S for $1 \le i \le a$, $p^{t_{a+i}}$ divides the $(a+b+1+i)$-th column for $1 \le i \le b$, $p^{t_{a+b+i}}$ divides the $(a+b+1+i)$-th row for $1 \le i \le b$, and the (i, j)-th entry of S_{41} and S_{44} is divisible by $p^{t_{a+b+i}+t_j}$ and $p^{t_{a+b+i}+t_{a+j}}$, respectively. Observe that we have only to show that if $\tilde{\gamma}\alpha(gw, 1)w'\tilde{\gamma}^{-1} \in Q_t$ then $g \in PwK$, i.e., $gw \in PK^w$ for $K^w := wKw$ (note that $\gamma(m(g_1, 1), g_1) \in Q$).

Let

$$\tilde{\gamma}_v \alpha(gw, 1) w' \tilde{\gamma}_v^{-1} = \begin{pmatrix} -a_1 & a_2 & a_3 & -b_1 & a_1 & b_1 & b_2 & \\ -a_4 & a_5 & a_6 & -b_3 & a_4 & b_3 & b_4 & \\ -a_7 & a_8 & a_9 & -b_5 & a_7 & b_5 & b_6 & \\ & & & 1 & & & & \\ -1-a_1 & a_2 & a_3 & -b_1 & a_1 & b_1 & b_2 & \\ -c_1 & c_2 & c_3 & 1-d_1 & c_1 & d_1 & d_2 & \\ -c_4 & c_5 & c_6 & -d_3 & c_4 & d_3 & d_4 & \\ -a_4 & a_5-1 & a_6 & -b_3 & a_4 & b_3 & b_4 & 1 \end{pmatrix} := H.$$

Thus, if $H \in Q_t$, then

$$\begin{pmatrix} a_1 & b_1 & b_2 & \\ c_1 & d_1 & d_2 & \\ c_4 & d_3 & d_4 & \\ a_4 & b_3 & b_4 & 1 \end{pmatrix}$$

is invertible and there exists $S \in S_t$ such that

$$\begin{pmatrix} -1-a_1 & a_2 & a_3 & -b_1 \\ -c_1 & c_2 & c_3 & 1-d_1 \\ -c_4 & c_5 & c_6 & -d_3 \\ -a_4 & a_5-1 & a_6 & -b_3 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 & b_2 & \\ c_1 & d_1 & d_2 & \\ c_4 & d_3 & d_4 & \\ a_4 & b_3 & b_4 & 1 \end{pmatrix} S.$$

By looking at the third row (blockwise), one finds $d_4 \neq 0$, so by left-multiplying g by a matrix

$$\begin{pmatrix} 1_a & & & & \times \\ & 1_b & & & \times \\ & & 1 & & \times \\ & & & 1_b & \times \\ & & & & d_4^{-1} \end{pmatrix}$$

(which does not change the assumption and conclusion) we may assume that $d_4 = 1$ and $d_2 = b_2 = b_4 = b_6 = 0$. So we assume that gw is of the form

$$\begin{pmatrix} a_1 & a_2 & a_3 & b_1 & \\ a_4 & a_5 & a_6 & b_3 & \\ a_7 & a_8 & a_9 & & \\ c_1 & c_2 & c_3 & d_1 & \\ c_4 & c_5 & c_6 & d_3 & 1 \end{pmatrix}.$$

Next, by looking at the second row (blockwise) and noting that $d_2 = 0$, we find that d_1 is of the form

$$\begin{pmatrix} \mathbb{Z}_p^\times & \mathbb{Z}_p & \cdots & \cdots & \mathbb{Z}_p \\ p^{t_{a+1}}\mathbb{Z}_p & \mathbb{Z}_p^\times & \ddots & \ddots & \mathbb{Z}_p \\ \vdots & p^{t_{a+2}}\mathbb{Z}_p & \mathbb{Z}_p^\times & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbb{Z}_p \\ p^{t_{a+1}}\mathbb{Z}_p & \cdots & \cdots & \cdots & \mathbb{Z}_p^\times \end{pmatrix}.$$

So, by multiplying by a matrix of the form

$$\begin{pmatrix} 1_a & & & & \\ & 1_b & & & \\ & & 1 & \times & \\ & & & 1_b & \\ & & & & 1 \end{pmatrix}$$

on the left we may assume that $b_5 = 0$. Also, by looking at the third row again we see $c_4 = (p^{t_1}\mathbb{Z}_p, \ldots, p^{t_a}\mathbb{Z}_p)$, $c_5, c_6 \in M(\mathbb{Z}_p)$ and $d_3 \in (p^{t_{a+1}}, \ldots, p^{t_{a+b}})$, while, from the second row, $c_1 \in (M_{b\times 1}(p^{t_1}\mathbb{Z}_p), M_{b\times 1}(p^{t_2}\mathbb{Z}_p), \ldots, M_{b\times 1}(p^{t_a}\mathbb{Z}_p))$, $c_2 \in M_{b\times b}(\mathbb{Z}_p)$ and $c_3 \in M_{b\times 1}(\mathbb{Z}_p)$.

By looking at the first row and noting that $b_2 = 0$, we know

$$a_1 \in \begin{pmatrix} \mathbb{Z}_p^\times & \mathbb{Z}_p & \cdots & \cdots & \mathbb{Z}_p \\ p^{t_1}\mathbb{Z}_p & \mathbb{Z}_p^\times & \ddots & \ddots & \mathbb{Z}_p \\ \vdots & p^{t_2}\mathbb{Z}_p & \mathbb{Z}_p^\times & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbb{Z}_p \\ p^{t_1}\mathbb{Z}_p & \cdots & \cdots & \cdots & \mathbb{Z}_p^\times \end{pmatrix},$$

a_2, $a_3 \in M(\mathbb{Z}_p)$ and $b_1 \in (M_{a\times 1}(p^{t_{a+1}}\mathbb{Z}_p), M_{a\times 1}(p^{t_{a+2}}\mathbb{Z}_p), \ldots, M_{a\times 1}(p^{t_{a+b}}\mathbb{Z}_p))$. Finally, looking at the fourth row (blockwise), we note that $b_4 = 0$. Similarly,

$$a_4 \in (M_{b\times 1}(p^{t_1}\mathbb{Z}_p), M_{b\times 1}(p^{t_2}\mathbb{Z}_p), \ldots, M_{b\times 1}(p^{t_a}\mathbb{Z}_p)),$$
$$b_3 \in (M_{b\times 1}(p^{t_{a+1}}\mathbb{Z}_p), M_{b\times 1}(p^{t_{a+2}}\mathbb{Z}_p), \ldots, M_{b\times 1}(p^{t_{a+b}}\mathbb{Z}_p)),$$
$$a_5 - 1 \in \begin{pmatrix} M_{1\times b}(p^{t_{a+b+1}}\mathbb{Z}_p) \\ M_{1\times b}(p^{t_{a+b+2}}\mathbb{Z}_p) \\ \vdots \\ M_{1\times b}(p^{t_{a+2b}}\mathbb{Z}_p) \end{pmatrix} \quad \text{and} \quad a_6 \in \begin{pmatrix} p^{t_{a+b+1}}\mathbb{Z}_p \\ p^{t_{a+b+2}}\mathbb{Z}_p \\ \vdots \\ p^{t_{a+2b}}\mathbb{Z}_p \end{pmatrix}.$$

Now we prove that $gw \in PK^w$ using the properties proven above. First we right-multiply gw by

$$\begin{pmatrix} 1_a & & & & \\ & 1_b & & & \\ & & 1 & & \\ -d_1^{-1}c_1 & -d_1^{-1}c_2 & -d_1^{-1}c_3 & d_1^{-1} & \\ -c_4 & -c_5 & -c_6 & -d_3 & 1 \end{pmatrix} \in K^w,$$

which does not change the above properties or what needs to be proven, so without loss of generality we assume that $c_4 = c_5 = c_6 = d_3 = c_1 = c_2 = c_3 = 0$ and $d_1 = 1$. Moreover, we set $\left(\begin{smallmatrix} a_1 & a_2 \\ a_4 & a_5 \end{smallmatrix}\right)^{-1}\left(\begin{smallmatrix} a_3 \\ a_6 \end{smallmatrix}\right) := T = \left(\begin{smallmatrix} T_1 \\ T_2 \end{smallmatrix}\right)$. Then

$$\begin{pmatrix} 1_a & & T_1 & & \\ & 1_b & T_2 & & \\ & & 1 & & \\ & & & 1_b & \\ & & & & 1 \end{pmatrix} \in K^w.$$

By multiplying

$$\begin{pmatrix} 1_a & & -T_1 & & \\ & 1_b & & -T_2 & \\ & & 1 & & \\ & & & 1_b & \\ & & & & 1 \end{pmatrix}$$

to the right we get an element in P. So it is clear that $gw \in PK^w$. $\square$

Now suppose that π is nearly ordinary with respect to $\underline{k}$. We define φ to be the unique (up to scalar) nearly ordinary vector in π with respect to the Borel subgroup $\tilde{B}$. Let $\varphi_w = \pi(w)\varphi$.

Now write

$$\varphi' = \pi\left(\mathrm{diag}(p^{-t_{a+b+1}}, \ldots, p^{t_1}, \ldots, p^{t_{a+1}}, \ldots)^\iota \begin{pmatrix} & & -1_b \\ & 1_a & \\ 1_b & & \end{pmatrix}^\iota\right)\varphi_w.$$

Compute the value $F_{\varphi'}(z, \rho(\Upsilon)f^\dagger, w)$. In fact, $F_{\varphi'}(z, \rho(\Upsilon)f^\dagger, w)$ is equal to

$$\sum_{B,C,D,E}\int_{\mathrm{GL}_{a+2b}(\mathbb{Q}_p)} \tilde{f}^\dagger\left(\tilde{\gamma}\alpha\left(w\begin{pmatrix} B & & & \\ & 1 & & \\ & & C & \\ & & & 1\end{pmatrix}w, \left(g_1\begin{pmatrix} E & \\ & D\end{pmatrix}\mathfrak{M}\begin{pmatrix} & & 1_b \\ & 1_a & \\ -1_b & & \end{pmatrix}\right)^\iota\right)\right.$$

$$\left. \times\, w'\tilde{\gamma}^{-1}\,\Xi\, w_{a+2b+1}^{-1}\right)\bar{\tau}(\det g_1)\rho(g_1^\iota)\varphi' dg_1$$

with (temporarily)

$$\mathfrak{M} = \mathrm{diag}(p^{-t_{a+b+1}}, \ldots, p^{t_1}, \ldots, p^{t_{a+1}}, \ldots),$$
$$\Xi = \mathrm{diag}(p^{-t_1}, \ldots, p^{-t_a}, 1_b, 1, p^{-t_{a+1}}, \ldots, p^{-t_{a+b}}, 1_a, 1_b, 1, p^{t_{a+b+1}}, p^{t_{a+2b}}),$$

where the sum is over $B \in \mathfrak{B}'$, $C \in \mathfrak{C}'$, $D \in \mathfrak{D}'$ and $E \in \mathfrak{E}'$. A direct computation gives

$$\tilde{\gamma}\alpha\left(1, \begin{pmatrix} a_1 & a_3 & a_2 \\ a_7 & a_9 & a_8 \\ a_4 & a_6 & a_5\end{pmatrix}^\iota\right)w'\tilde{\gamma}^{-1} = \begin{pmatrix} -1_a & & & 1_a & & \\ & 1_b & & & & \\ & & 1 & & & \\ -a_3 & -a_2 & a_1 & & & a_2 \\ -a_9-1_a & -a_8 & a_7 & 1_a & & a_8 \\ -a_3 & -a_2 & a_1-1_b & & 1_b & a_2 \\ & & & & 1 & \\ -a_6 & 1_b-a_5 & a_4 & & & a_5\end{pmatrix}.$$

Now we define $\mathfrak{Y}$ to be the subset of $\mathrm{GL}_{a+2b}(\mathbb{Z}_p)$ consisting of block matrices

$$\begin{pmatrix} a_1 & a_3 & a_2 \\ a_7 & a_9 & a_8 \\ a_4 & a_6 & a_5\end{pmatrix}$$

such that

$$\tilde{\gamma}\alpha\left(1, \begin{pmatrix} a_1 & a_3 & a_2 \\ a_7 & a_9 & a_8 \\ a_4 & a_6 & a_5 \end{pmatrix}^{\iota}\right) w'\tilde{\gamma}^{-1}$$

is in the Q_t defined in the proof of Lemma 4.36. It is not hard to prove that it can be described as follows: the i-th columns of $-a_9 - 1$ and a_3 (resp. a_7 and $a_1 - 1$) are divisible by p^{t_i} (resp. $p^{t_{a+i}}$) for $1 \le i \le a$, the (i, j)-th entry of a_6 (resp. a_4) is divisible by $p^{t_{a+b+i}+t_j}$ (resp. $p^{t_{a+b+i}+t_{a+j}}$), and the i-th row of $1 - a_5$ is divisible by $p^{t_{a+b+i}}$. The entries in a_2 and a_8 are in $\mathbb{Z}_p$. Then the pullback section is equal to

$$\sum_{B,C,D,E} \int \tilde{f}^{\dagger}\big(\tilde{\gamma}\alpha(1, g_1^{\iota})w'\tilde{\gamma}^{-1}\,\Xi w_{a+2b+1}^{-1}\big)\bar{\tau}(\det g_1)\pi(g_1^{\iota})\varphi\, dg_1,$$

where

$$\Xi = \operatorname{diag}(p^{-t_1}, \ldots, p^{-t_a}, 1, 1, p^{-t_{a+1}}, \ldots, p^{-t_{a+b}}, 1, 1, 1, p^{t_{a+b+1}}, \ldots, p^{t_{a+2b}})$$

and the integration is over elements (with superscript w meaning conjugation by w)

$$g_1 \in \begin{pmatrix} B & \\ & C \end{pmatrix}_{\iota}^{w} \mathfrak{Y} \begin{pmatrix} E & \\ & D \end{pmatrix}_{\mathrm{conj}} \begin{pmatrix} & & 1_b \\ & 1_a & \\ -1_b & & \end{pmatrix} \operatorname{diag}(p^{t_{a+b+1}}, \ldots, p^{-t_1}, \ldots, p^{-t_{a+b}}, \ldots)$$

for

$$\begin{pmatrix} E & \\ & D \end{pmatrix}_{\mathrm{conj}} := \begin{pmatrix} & & 1_b \\ & 1_a & \\ -1_b & & \end{pmatrix} \operatorname{diag}(p^{t_{a+b+1}}, \ldots, p^{-t_1}, \ldots, p^{-t_{a+1}}, \ldots) \begin{pmatrix} E & \\ & D \end{pmatrix}$$
$$\times \operatorname{diag}(p^{-t_{a+b+1}}, \ldots, p^{t_1}, \ldots, p^{t_{a+1}}, \ldots) \begin{pmatrix} & & -1_b \\ & 1_a & \\ 1_b & & \end{pmatrix}.$$

Lemma 4.37. *If φ_w is invariant under the action of $(\tilde{K}'')^{\iota}$, then*

$$F_{\varphi'}(z, \rho(\Upsilon) f^{\dagger}, w)$$

is such that the action of $\tilde{K}^{\iota}$ on it is given by $\tilde{\nu}$.

Proof. By the above two lemmas we only need to check that $F_{\varphi'}(z, \rho(\Upsilon) f^{\dagger}, w)$ is invariant under the action of $\tilde{K}''$. We first claim that $\sum_{D,E} \pi\big(\big(\begin{smallmatrix} E & \\ & D \end{smallmatrix}\big)_{\mathrm{conj}}^{\iota}\big)\varphi'$ is invariant under $(\tilde{K}'')^{\iota}$. The claim follows from direct checking. Also, for any $k_1 \in \tilde{K}''$, we can find a $k_2 \in \tilde{K}''$ such that $k_1\big(\begin{smallmatrix} B & \\ & C \end{smallmatrix}\big)_{\iota}^{w} k_2^{-1}$ runs over the same set of representatives as $\big(\begin{smallmatrix} B & \\ & C \end{smallmatrix}\big)_{\iota}^{w}$. For any $k_1 \in \tilde{K}''$, we can find a $k_2 \in \tilde{K}''$ such that $k_1 \mathfrak{Y} k_2^{-1} = \mathfrak{Y}$. The lemma follows from these observations. □

The value of $\tilde{f}^\dagger$ at

$$g_1 = \begin{pmatrix} & & 1_b \\ & 1_a & \\ -1_b & & \end{pmatrix} \mathrm{diag}(p^{t_{a+b+1}}, \dots, p^{-t_1}, \dots, p^{-t_{a+b}}, \dots)$$

is

$$\tau((p^{t_{a+b+1}+\cdots+t_{a+2b}}, p^{t_1+\cdots+t_{a+b}}))|p^{t_1+\cdots+t_{a+2b}}|^{-z-(a+2b+1)/2}.$$

So, a straightforward computation using the model for $\pi = \pi(\chi_1, \dots, \chi_{a+2b})$ tells us the following:

Lemma 4.38. *If φ and φ' are defined as after the proof of Lemma 4.36, then*:

$$\begin{aligned} &F_{\varphi'}(z, \rho(\Upsilon)f^\dagger, w) \\ &\quad = \tau((p^{t_1+\cdots+t_{a+b}}, p^{t_{a+b+1}+\cdots+t_{a+2b}}))|p^{t_1+\cdots+t_{a+2b}}|^{-z-(a+2b+1)/2}\,\mathrm{Vol}(\tilde{K}') \\ &\qquad \times p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+i}} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1)\varphi_w. \end{aligned}$$

Combining the three lemmas above, we get the following:

Proposition 4.39. *With assumptions as in the above lemma, $F_{\varphi'}(z, \rho(\Upsilon)f^\dagger, g)$ is the unique section supported in PwK such that the right action of K is given by multiplying the character ν, and its value at w is*

$$\begin{aligned} &F_{\varphi'}(z, \rho(\Upsilon)f^\dagger, w) \\ &\quad = \tau((p^{t_1+\cdots+t_{a+b}}, p^{t_{a+b+1}+\cdots+t_{a+2b}}))|p^{t_1+\cdots+t_{a+2b}}|^{-z-\frac{a+2b+1}{2}}\,\mathrm{Vol}(\tilde{K}') \\ &\qquad \times p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+1+i}} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1)\varphi_w. \end{aligned}$$

Proof. Clearly ϕ_w is invariant under $(\tilde{K}'')^\iota$. □

This $F_{\varphi'}(z, \rho(\Upsilon)f^\dagger, g)$ we constructed is not going to be the nearly ordinary vector unless we apply the intertwining operator to it. So now we start with some $\rho = (\pi, \tau)$. We define our Siegel section $f^0 \in I_{a+2b+1}(\tau)$ to be

$$f^0(z; g) := M(-z, f^\dagger)_z(g),$$

where $f^\dagger \in I_{a+2b+1}(\bar{\tau}^c)$. We recall the following generalization of a proposition from [Skinner and Urban 2014].

Proposition 4.40. *Suppose our data (π, τ) comes from the local component at v of a global data. Then there is a meromorphic function $\gamma^{(2)}(\rho, z)$ such that*

$$F_{\varphi^\vee}(-z, M(z, f), g) = \gamma^{(2)}(\rho, z)A(\rho, z, F_\varphi(f; z, -))_{-z}(g).$$

Moreover, if $\pi_v \simeq \pi(\chi_1, \ldots, \chi_{a+2b})$ then, if we write $\gamma^{(1)}(\rho, z) = \gamma^{(2)}\big(\rho, z+\tfrac{1}{2}\big)$, then

$$\gamma^{(1)}(\rho, z) = \psi(-1)c_n(\tau', z)\mathfrak{g}(\tau'_p)^n \epsilon\big(\pi, \tau^c, z+\tfrac{1}{2}\big)\frac{L\big(\tilde{\pi}, \bar{\tau}^c, \tfrac{1}{2}-z\big)}{L\big(\pi, \tau^c, z+\tfrac{1}{2}\big)},$$

where $c_n(\tau', z)$ is the constant appearing in Lemma 4.27.

Proof. The same as [Skinner and Urban 2014, Proposition 11.13]. □

Remark 4.41. Here we are using the L-factors for the base change from the unitary groups, while [Skinner and Urban 2014] uses the GL_2 L-factor for π, so our formula appears slightly different.

Now we are going to show that

$$F_v^0(z; g) := F_{\varphi'}(z, \rho(\Upsilon) f^0, g)$$

is a constant multiple of the nearly ordinary vector if our ρ comes from the local component of the global Eisenstein data (see Section 3A). Return to the situation of our Eisenstein data. Suppose that at the Archimedean places our representation is a holomorphic discrete series associated to the (scalar) weight $\underline{k} = (0, \ldots 0; \kappa, \ldots \kappa)$ with r zeroes and s kappas. Here $r = a+b$ and $s = b$. Suppose $\pi \simeq \mathrm{Ind}(\chi_1, \ldots, \chi_{a+2b})$ is nearly ordinary with respect to the weight $\underline{k}$. We may reorder the χ_i so that $\nu_p(\chi_1(p)) = s - \frac{1}{2}n + \frac{1}{2}, \ldots, \nu_p(\chi_r(p)) = r+s-1-\frac{1}{2}n+\frac{1}{2}$, $\nu_p(\chi_{r+s}(p)) = \kappa - \frac{1}{2}n + \frac{1}{2}, \ldots, \nu_p(\chi_{r+1}(p)) = \kappa+s-1-\frac{1}{2}n+\frac{1}{2}$, and $\tau = (\tau_1, \tau_2^{-1})$ is a character of $\mathbb{Q}_p^\times \times \mathbb{Q}_p^\times$ with $\nu_p(\tau_1(p)) = \nu_p(\tau_2(p)) = \frac{\kappa}{2}$, so

$$\begin{aligned}\nu_p(\chi_1(p)) < \cdots < \nu_p(\chi_{a+b}(p)) &< \nu_p(\tau_2(p)p^{-z_\kappa}) \\ &< \nu_p(\tau_1(p)p^{z_\kappa}) \\ &< \nu_p(\chi_{a+2b}(p)) < \cdots < \nu_p(\chi_{a+b+1}(p)),\end{aligned}$$

where $z_\kappa = \frac{1}{2}(\kappa - r - s - 1)$. It is easy to see that

$$I(\rho_v, z_\kappa) \simeq \mathrm{Ind}(\chi_1, \ldots \chi_{r+s}, \tau_2|\cdot|^{z_\kappa}, \tau_1|\cdot|^{-z_\kappa}).$$

By definition, $I(\rho_v, z_\kappa)$ is nearly ordinary with respect to the weight

$$(\underbrace{0, \ldots, 0}_{r+1}; \underbrace{\kappa, \ldots, \kappa}_{s+1}).$$

Definition 4.42. With assumptions and conventions as above, we say (π, τ) is generic if

$$\begin{aligned}\mathrm{cond}(\chi_1) > \cdots > \mathrm{cond}(\chi_{a+b}) > \mathrm{cond}(\tau_2)& \\ > \mathrm{cond}(\chi_{a+b+1}) > \cdots > \mathrm{cond}(\chi_{a+2b}) > \mathrm{cond}(\tau_1)&.\end{aligned}$$

We suppose also that the conductor of τ_i is p^{s_i}. Notice that we have $s_2 > s_1$ by our assumption, which is different from Definition 4.21 (since we have applied the intertwining operator here).

Let us record the following formula for the ϵ-factor in Proposition 4.40:

$$\epsilon\big(\pi,\tau^c,z+\tfrac{1}{2}\big)=\prod_{i=1}^{r}\mathfrak{g}(\chi_i^{-1}\tau_2)\chi_i\tau_2^{-1}(p^{t_i})\cdot\prod_{i=1}^{s}\mathfrak{g}(\chi_{r+i}^{-1}\tau_2)\chi_{r+i}\tau_2^{-1}(p^{s_2})$$
$$\times|p^{\sum_{i=1}^{r}t_i+s\cdot s_2}|^{z+\frac{1}{2}}\prod_{i=1}^{r+s}\mathfrak{g}(\chi_i\tau_1^{-1})\chi_i^{-1}\tau_1(p^{t_i})\cdot|p^{\sum_{i=1}^{r+s}t_i}|^{z+\frac{1}{2}}. \quad (15)$$

From the form of $F_{\varphi'}(z,\rho(\Upsilon)f^\dagger;g)$ and the above proposition we have a description in the "generic" case for $F_v^0(z,g)$ as in [Skinner and Urban 2014, Lemma 9.6]: it is supported in $P(\mathbb{Q}_p)K_v$, with

$$\begin{aligned}F_v^0(z,1)&=\gamma^{(2)}(\rho,-z)\bar{\tau}^c((p^{t_1+\cdots+t_{a+b}},p^{t_{a+b+1}+\cdots+t_{a+2b}}))\\&\quad\times|p^{t_1+\cdots+t_{a+2b}}|^{z-(a+2b+1)/2}\,\mathrm{Vol}(\tilde{K}')p^{-\sum_{i=1}^{a+b}(i+1)t_i-\sum_{i=1}^{b}(i+1)t_{a+b+1+i}}\\&\quad\times\prod_{i=1}^{a+b}\mathfrak{g}(\xi_i^\dagger)\xi_i(-1)\prod_{i=1}^{b}\mathfrak{g}(\xi_{a+b+1+i}^\dagger)\xi_{a+b+1+i}^\dagger(-1)\varphi\\&=c_n\big(\tau_p',-z-\tfrac{1}{2}\big)\mathfrak{g}(\tau_p')^n\bar{\tau}^c((p^{t_1+\cdots+t_{a+b}},p^{t_{a+b+1}+\cdots+t_{a+2b}}))\\&\quad\times|p^{t_1+\cdots+t_{a+2b}}|^{z-(a+2b+1)/2}\,\mathrm{Vol}(\tilde{K}')p^{-\sum_{i=1}^{a+b}it_i-\sum_{i=1}^{b}it_{a+b+1+i}}\\&\quad\times\prod_{i=r+1}^{r+s}\mathfrak{g}(\chi_i^{-1}\tau_2)\chi_i\tau_2^{-1}(p^{s_2})\prod_{j=1}^{r}\mathfrak{g}(\chi_j\tau_1^{-1})\chi_j^{-1}\tau_1(p^{t_j})\epsilon(\pi,\tau^c,z)\varphi\\&\quad\times|p^{\sum_{i=1}^{r}t_i+s\cdot s_2}|^{-z}\cdot|p^{\sum_{i=1}^{r+s}t_i}|^{-z},\end{aligned}$$

where the $\xi_i^\dagger$ are the ξ_i defined in Definition 4.21 but using $(\pi,\bar{\tau}^c)$ instead of (π,τ). Here we also used Proposition 4.40 and the formula for the epsilon factor there. Notice that we have absorbed a factor $p^{-\sum_{i=1}^{a+b}t_i-\sum_{i=1}^{b}t_{a+b+1+i}}$, which comes from computing the image under the intertwining operator of $F_{\varphi'}(z,\rho(\Upsilon)f^\dagger;g)$ to get the factor $p^{-\sum_{i=1}^{a+b}(i+1)t_i-\sum_{i=1}^{b}(i+1)t_{a+b+1+i}}$ in the above expression. The right action of K_v is given by the character

$$\chi_1(g_{11})\cdots\chi_{a+b}(g_{a+b\ a+b})\tau_2(g_{a+b+1\ a+b+1})\chi_{a+b+1}(g_{a+b+2\ a+b+2})\times\cdots$$
$$\times\chi_{a+2b}(g_{a+2b+1\ a+2b+1})\tau_1(g_{a+2b+2\ a+2b+2}).$$

(It is easy to compute $A(\rho,z,F_{\varphi'}(\rho(\Upsilon)f^\dagger;z,-))_{-z}(1)$ and we use the uniqueness of the vector with the required K_v action. Here, on the second row of the above formula for $F_v^0(z,1)$, the power for p is slightly different from that for the section

$F(z, f^\dagger, w)$. This comes from the computations for the intertwining operators for Klingen Eisenstein sections.)

Thus, Corollary 4.20 tells us that $F_v^0(z, g)$ is a nearly ordinary vector in $I(\rho)$.

Now we describe f^0:

Definition 4.43. Suppose $(p^t) = \mathrm{cond}(\tau')$ for $t \geq 1$, then define f_t to be the section supported in $Q(\mathbb{Q}_p)K_Q(p^t)$ with $f_t(k) = \tau(\det d_k)$ on $K_Q(p^t)$.

Lemma 4.44. $$\tilde{f}^0 := M(-z, \tilde{f}^\dagger)_z = f_{t,z}.$$

Proof. This is just [Skinner and Urban 2014, Lemma 11.10]. □

Corollary 4.45. *We have*

$$f^0(z, g) = p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+i}} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1)$$
$$\times \sum_{A,B,C,D,E} \prod_{i=1}^{a} \bar{\xi}_i\left(\frac{\det A_i}{\det A_{i-1}} p^{t_i}\right) \prod_{i=1}^{b} \bar{\xi}_{a+i,a+i}\left(\frac{\det D_i}{\det D_{i-1}} p^{t_{a+i}}\right)$$
$$\times \prod_{i=1}^{b} \bar{\xi}_{a+b+1+i}\left(\frac{\det E_i}{\det E_{i-1}} p^{t_{a+b+i}}\right) \tilde{f}_t\left(z, g\begin{pmatrix} 1_{a+2b+1} & \begin{matrix} A & B \\ C & D \end{matrix} \\ & \begin{matrix} E & \\ & \end{matrix} \\ & 1_{a+2b+1} \end{pmatrix}\right).$$

Here, A_i is the i-th upper-left minor of A, D_i is the $(a+i)$-th upper-left minor of $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ and E_i is the i-th upper-left minor of E.

We define the Siegel section $f^{0\prime} \in I_{a+2b}(\tau)$ by

$$f^{0\prime}(z, g) = p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+i}} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1)$$
$$\times \sum_{A,B,C,D,E} \prod_{i=1}^{a} \bar{\xi}_i\left(\frac{\det A_i}{\det A_{i-1}} p^{t_i}\right) \prod_{i=1}^{b} \bar{\xi}_{a+i,a+i}\left(\frac{\det D_i}{\det D_{i-1}} p^{t_{a+i}}\right)$$
$$\times \prod_{i=1}^{b} \bar{\xi}_{a+b+1+i}\left(\frac{\det E_i}{\det E_{i-1}} p^{t_{a+b+i}}\right)$$
$$\times \tilde{f}_t\left(z, g w'^{-1}_{\mathrm{Borel}} \begin{pmatrix} 1_{a+2b+1} & \begin{matrix} A & B \\ C & D \end{matrix} \\ & \begin{matrix} E & \\ & \end{matrix} \\ & 1_{a+2b+1} \end{pmatrix} w'_{\mathrm{Borel}}\right).$$

Then, similar to before, the corresponding pullback section $F'_{\varphi'}(z, \rho(\Upsilon')f^{0\prime}, 1)$ equals

$$c_n(\tau'_p, -z)\mathfrak{g}(\tau'_p)^n\bar{\tau}^c((p^{t_1+\cdots+t_{a+b}}, p^{t_{a+b+1}+\cdots+t_{a+2b}}))|p^{t_1+\cdots+t_{a+2b}}|^{z-(a+2b)/2}\operatorname{Vol}(\tilde{K}')$$
$$\times p^{-\sum(i-1)t_i-\sum(i-1)t_{a+b+i}}\prod_{i=r+1}^{r+s}\mathfrak{g}(\chi_i^{-1}\tau_2)\chi_i\tau_2^{-1}(p^{s_2})$$
$$\times\prod_{j=1}^{r}\mathfrak{g}(\chi_j\tau_1^{-1})\chi_j^{-1}\tau_1(p^{t_j})\epsilon\big(\pi, \tau^c, z+\tfrac{1}{2}\big)\varphi$$
$$\times|p^{\sum_{i=1}^{r}t_i+s\cdot s_2}|^{-z+\frac{1}{2}}\cdot|p^{\sum_{i=1}^{r+s}t_i}|^{-z+\frac{1}{2}}.$$

Fourier coefficients for f^0. We record a formula here for the Fourier coefficients for f^0 which will be used in p-adic interpolation.

Lemma 4.46. *Suppose* $|\det\beta|\neq 0$; *then*:

(i) *If $\beta\notin S_{a+2b+1}(\mathbb{Z}_p)$ then $f^0_\beta(z, 1)=0$.*

(ii) *Let $t:=\operatorname{ord}_p(\operatorname{cond}(\tau'))$. If $\beta\in S_{a+2b+1}(\mathbb{Z}_p)$ then*

$$f^0_\beta(z, 1)=\bar{\tau}'(\det\beta)|\det\beta|_p^{2z}\mathfrak{g}(\tau')^{a+2b+1}c_{a+2b+1}(\bar{\tau}', -z)\Phi_\xi({}^t\beta),$$

where $c_{a+2b+1}(-, -)$ is as defined in (13) and Φ_ξ is defined in (11).

Proof. This follows from [Skinner and Urban 2014, Lemma 11.12] and the argument of Corollary 4.30, where we deduce the form of $f^\dagger$ from the section $\tilde{f}^\dagger$. □

4D3. *Fourier–Jacobi coefficients.* Now let $m=b+1$. For $\beta\in S_m(F_v)\cap\mathrm{GL}_m(\mathcal{O}_v)$ we are going to compute the Fourier–Jacobi coefficient for f_t at β.

Lemma 4.47. *Let $x:=\left(\begin{smallmatrix}1&\\D&1\end{smallmatrix}\right)$ (this is a block matrix with respect to $(a+b)+(a+b)$). Then*:

(a) $\mathrm{FJ}_\beta(f_t; z, v, x\eta^{-1}, 1)=0$ *if $D\notin p^tM_{a+b}(\mathbb{Z}_p)$.*

(b) *If $D\in p^tM_n(\mathbb{Z}_p)$ then* $\mathrm{FJ}_\beta(f_t; z, v, x\eta^{-1}, 1)=c(\beta, \tau, z)\Phi_0(v)$, *where*

$$c(\beta, \tau, z):=\bar{\tau}(-\det\beta)|\det\beta|_v^{2z+n-m}\mathfrak{g}(\tau')^m c_m\big(\tau', -z-\tfrac{1}{2}(n-m)\big)$$

and c_m is as defined in Lemma 4.27.

Proof. Similar to the proof of [Skinner and Urban 2014, Lemma 11.20]. We only give the detailed proof for the case when $a=0$. The case when $a>0$ is even easier to treat.

Assuming $a=0$, we temporarily write n for b and save the letter b for other use. We have

$$w_{2n+1}\begin{pmatrix}1_{2n+1} & S & v\\ & {}^t\bar{v} & D\\ & 1_{2n+1} & \end{pmatrix}\alpha(1,\eta^{-1})=\begin{pmatrix} & & 1_{n+1} & \\ & & & -1_n \\ -1_{n+1} & v & -S & \\ & D & -{}^t\bar{v} & -1_n\end{pmatrix}.$$

This belongs to $Q_{2n+1}(\mathbb{Q}_p)K_{Q_{2n+1}}(p^t)$ (where $K_{Q_{2n+1}}(p^t)$ consists of matrices in $Q_{2n+1}(\mathbb{Z}_p)$ modulo p^t) if and only if S is invertible with $S^{-1}\in p^tM_{n+1}(\mathbb{O}_v)$, $S^{-1}v\in p^tM_{(n+1)\times n}(\mathbb{O}_v)$ and ${}^t\bar{v}S^{-1}v-D\in p^tM_n(\mathbb{Z}_p)$. Since $v=\gamma\,{}^t(b,0)$ for some $\gamma\in\mathrm{SL}_{n+1}(\mathbb{O}_v)$ and $b\in M_n(\mathcal{K}_v)$, we are reduced to the case $v={}^t(b,0)$. Writing $b=(b_1,b_2)$ with $b_i\in M_n(\mathbb{Q}_p)$, and $S=(T,{}^tT)$ with $T\in M_{n+1}(\mathbb{Q}_p)$ and $T^{-1}=\left(\begin{smallmatrix}a_1 & a_2\\ a_3 & a_4\end{smallmatrix}\right)$, where $a_1\in M_n(\mathbb{Q}_p)$, $a_2\in M_{n\times 1}(\mathbb{Q}_p)$, $a_3\in M_{1\times n}(\mathbb{Q}_p)$ and $a_4\in M_1(\mathbb{Q}_p)$, the conditions on S and v can be rewritten as

$$\begin{gathered}\det T\neq 0,\quad a_i\in p^tM_n(\mathbb{Z}_p),\quad a_1b_1\in p^tM_n(\mathbb{Z}_p),\quad a_3b_1\in p^tM_{1\times n}(\mathbb{Z}_p),\\ {}^ta_1b_2\in p^tM_n(\mathbb{Z}_p),\quad {}^ta_2b_2\in p^t\mathbb{Z}_p,\quad {}^tb_2a_1b_1-D\in p^tM_n(\mathbb{Z}_p).\end{gathered}\tag{*}$$

Now we prove that if the integral for $\mathrm{FJ}_\beta(f_t;z,v,x\eta^{-1},1)$ is nonzero then $b_1,b_2\in M_n(\mathbb{Z}_p)$. Suppose otherwise; then without loss of generality we assume b_1 has an entry which has the maximal p-adic absolute value among all entries of b_1 and b_2, Suppose it is p^w for $w>0$ (w means this only inside this lemma). Also, for any matrix A of given size, we say $A\in{}^tb_2^\vee$ if and only ${}^tb_2A$ has all entries in $\mathbb{Z}_p$ (of course we assume the sizes of the matrices are correct so that the product makes sense).

Now let

$$\Gamma:=\left\{\gamma=\begin{pmatrix}h & j\\ k & l\end{pmatrix}\in\mathrm{GL}_n(\mathbb{Z}_p)\;\middle|\; \begin{aligned}&h\in\mathrm{GL}_{n+1}(\mathbb{Z}_p),\ l\in\mathbb{Z}_p^\times,\ h-1\in{}^tb_2^\vee\cap p^tM_n(\mathbb{Z}_p),\\ &\qquad\qquad j\in\mathbb{Z}_p^n\cap{}^tb_2^\vee,\ k\in p^tM_{1\times n}(\mathbb{Z}_p)\end{aligned}\right\}.$$

Suppose that our b_1, b_2 and D are such that there exist a_i satisfying (*); then one can check that Γ is a subgroup and, if T satisfies (*), so does $T\gamma$ for any $\gamma\in\Gamma$. Let $\mathcal{T}$ denote the set of $T\in M_{n+1}(\mathbb{Q}_p)$ satisfying (*). Then

$$\begin{aligned}\mathrm{FJ}_\beta&\left(f_t;z,v,\begin{pmatrix}1 & \\ D & 1\end{pmatrix}\eta^{-1},1\right)\\ &=\sum_{T\in\mathcal{T}/\Gamma}|\det T|_p^{3n+2-2z}\int_\Gamma\tau'(-\det T\gamma)e_p(-\operatorname{Tr}\beta T\gamma)\,d\gamma.\end{aligned}$$

Let $T':=\beta T=\left(\begin{smallmatrix}c_1 & c_2\\ c_3 & c_4\end{smallmatrix}\right)$ (with blocks with respect to the partition $(n+1)$); then the above integral is zero unless we have $c_1\in p^{-t}M_n(\mathbb{Z}_p)\oplus[{}^tb_2]_{n\times n}$, $c_4\in p^{-t}\mathbb{Z}_p$,

$c_2 \in p^{-t}M_{n+1}(\mathbb{Z}_p)$ and $c_3 \in [{}^tb_2]_{1\times n} \oplus M_{1\times n}(\mathbb{Z}_p)$. Here $[{}^tb_2]_{i\times n}$ means the set of $i \times n$ matrices such that each row is a $\mathbb{Z}_p$-linear combination of the rows of ${}^tb_2$.

But then

$$\beta \begin{pmatrix} b_1 \\ 0 \end{pmatrix} = T'T^{-1} \begin{pmatrix} b_1 \\ 0 \end{pmatrix} = \begin{pmatrix} c_1a_1b_1 + c_2a_3b_1 \\ c_3a_1b_1 + c_4a_3b_1 \end{pmatrix}.$$

Since $\beta \in \mathrm{GL}_{n+1}(\mathbb{Z}_p)$, the left-hand side must contain some entry with p-adic absolute value p^w. But it is not hard to see that all entries on the right-hand side have p-adic values strictly less than p^w; a contradiction. Thus we conclude that $b_1 \in M_n(\mathbb{Z}_p)$ and $b_2 \in M_n(\mathbb{Z}_p)$. By (*), $b_2\,{}^ta_1b_1 - D \in p^tM_n(\mathbb{Z}_p)$ and $a_1 \in p^tM_n(\mathbb{Z}_p)$. So $D \in p^tM_n(\mathbb{Z}_p)$.

The value claimed in part (b) can be deduced similarly to in [Skinner and Urban 2014, Lemma 11.20] □

4D4. *Original basis.* Recall that we changed the basis at the beginning of this subsection. Now we go back. We define the corresponding sections (we use the same notations)

$$\begin{aligned} f^\dagger(z,g) = p^{-\sum_{i=1}^{a+b} it_i - \sum_{i=1}^{b} it_{a+b+i}} \prod_{i=1}^{a+b} \mathfrak{g}(\xi_i)\xi_i(-1) \prod_{i=1}^{b} \mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1) \\ \times \sum_{A,B,C,D,E} \prod_{i=1}^{a} \bar{\xi}_i\left(\frac{\det A_i}{\det A_{i-1}} p^{t_i}\right) \prod_{i=1}^{b} \bar{\xi}_{a+i,a+i}\left(\frac{\det D_i}{\det D_{i-1}} p^{t_{a+i}}\right) \\ \times \prod_{i=1}^{b} \bar{\xi}_{a+b+1+i}\left(\frac{\det E_i}{\det E_{i-1}} p^{t_{a+b+i}}\right) \\ \times \tilde{f}^\dagger\left(z, g w_{\mathrm{Borel}}^{-1} \begin{pmatrix} 1_b & & & & & & C & D \\ & 1 & & & & & & \\ & & 1_a & & & & A & B \\ & & & 1_b & E & & & \\ & & & & 1_b & & & \\ & & & & & 1 & & \\ & & & & & & 1_a & \\ & & & & & & & 1_b \end{pmatrix} w_{\mathrm{Borel}} \right), \end{aligned}$$

and $f^0(z,g)$ the same except using $\tilde{f}_t$ in place of $\tilde{f}^\dagger$. Here, A_i is the i-th upper-left minor of A, D_i is the $(a+i)$-th upper-left minor of $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ and E_i is the i-th upper-left minor of E. The w_{Borel} is the element in $G(F_p)$ such that, for any $v = w\bar{w}$ dividing p with $w \in \Sigma_p$, its projection to the first factor of $\mathcal{K}_v \simeq \mathcal{K}_w \times \mathcal{K}_{\bar{w}}$ is the

Weyl element defined at the beginning of Section 4D2. We also define

$$f^{\dagger\prime}(z,g)=p^{-\sum_{i=1}^{a+b} it_i-\sum_{i=1}^{b} it_{a+b+i}}\prod_{i=1}^{a+b}\mathfrak{g}(\xi_i)\xi_i(-1)\prod_{i=1}^{b}\mathfrak{g}(\xi_{a+b+1+i})\xi_{a+b+1+i}(-1)$$
$$\times\sum_{A,B,C,D,E}\prod_{i=1}^{a}\bar{\xi}_i\left(\frac{\det A_i}{\det A_{i-1}}\right)\prod_{i=1}^{b}\bar{\xi}_{a+i,a+i}\left(\frac{\det D_i}{\det D_{i-1}}\right)$$
$$\times\prod_{i=1}^{b}\bar{\xi}_{a+b+1+i}\left(\frac{\det E_i}{\det E_{i-1}}\right)$$
$$\times\tilde{f}^{\dagger}\left(z,\, g w_{\mathrm{Borel}}^{\prime -1}\begin{pmatrix}1_b &&&& C & D\\ & 1_a &&& A & B\\ && 1_b & E &&\\ &&& 1_b &&\\ &&&& 1_a &\\ &&&&& 1_b\end{pmatrix} w'_{\mathrm{Borel}}\right),$$

and $f^{0\prime}$ the same except with $\tilde{f}_t$ instead of $\tilde{f}^{\dagger}$. The corresponding pullback section $F_{\varphi'}(f^0,z,-)$ is the nearly ordinary section with respect to the Borel B_2 defined in Section 4D2 such that $F_{\varphi'}(f^0,z,w_{\mathrm{Borel}})$ is given by

$$c_{n+1}\big(\tau'_p,-z-\tfrac{1}{2}\big)\mathfrak{g}(\tau'_p)^{n+1}\bar{\tau}^c((p^{t_1+\cdots+t_{a+b}},\,p^{t_{a+b+1}+\cdots+t_{a+2b}}))$$
$$\times|p^{t_1+\cdots+t_{a+2b}}|^{-z-(a+2b+1)/2}\,\mathrm{Vol}(\tilde{K}')p^{-\sum it_i-\sum it_{a+b+i}}$$
$$\times\prod_{i=r+1}^{r+s}\mathfrak{g}(\chi_i^{-1}\tau_2)\chi_i\tau_2^{-1}(p^{t_i})\prod_{j=1}^{r}\mathfrak{g}(\chi_j\tau_1^{-1})\chi_j^{-1}\tau_1(p^{t_j})\epsilon(\pi,\tau^c,z)\varphi.$$

Also, we have that $F'_{\varphi'}(z,\rho(\Upsilon')f^{0'},w'_{\mathrm{Borel}})$ is given by

$$c_n(\tau'_p,-z)\mathfrak{g}(\tau'_p)^n\bar{\tau}^c((p^{t_1+\cdots+t_{a+b}},\,p^{t_{a+b+1}+\cdots+t_{a+2b}}))|p^{t_1+\cdots+t_{a+2b}}|^{-z-(a+2b)/2}\,\mathrm{Vol}(\tilde{K}')$$
$$\times p^{-\sum it_i-\sum it_{a+b+i}}\prod_{i=r+1}^{r+s}\mathfrak{g}(\chi_i^{-1}\tau_2)\chi_i\tau_2^{-1}(p^{t_i})$$
$$\times\prod_{j=1}^{r}\mathfrak{g}(\chi_j\tau_1^{-1})\chi_j^{-1}\tau_1(p^{t_j})\epsilon\big(\pi,\tau^c,z+\tfrac{1}{2}\big)\varphi.$$

5. Global computations

5A. *p-adic interpolation.*

5A1. *Weight space and Eisenstein datum.* Recall that we have the algebraic group $H=\prod_{v|p}\mathrm{GL}_r\times\mathrm{GL}_s$ such that $H(/\mathbb{Z}_p)$ is the Galois group of the Igusa tower over the ordinary locus of the toroidal compactified Shimura variety. Let $T_{/\mathbb{Z}_p}$ be the diagonal torus. Let $\boldsymbol{T} := T(1+\mathbb{Z}_p)$. We define the weight ring $\Lambda = \Lambda_{r,s}$ as $\mathbb{O}_L[\![\boldsymbol{T}]\!]$. Fix throughout a finite-order character χ_0 of $T(\mathbb{F}_p)$ (the torsion part of $T(\mathbb{Z}_p)$); a $\bar{\mathbb{Q}}_p$-point $\phi \in \operatorname{Spec}\Lambda$ is called arithmetic if there is a weight $\underline{k} = (c_{s+1}, \ldots, c_{s+r}; c_1, \ldots, c_s) = (0, \ldots, 0; \kappa, \ldots, \kappa)$ such that ϕ is given by a character $\chi_0\chi_\phi t_1^{c_{s+1}} \cdots t_r^{c_{r+s}} t_{r+1}^{-c_1} \cdots t_{r+s}^{-c_s}$ of $\boldsymbol{T}$ for χ_ϕ a character of order and conductor powers of p, with $\kappa \geq 2(a+b+1)$. We write this κ as κ_ϕ. Let $\Lambda_{\mathcal{K}} = \mathbb{O}_L[\![\Gamma_{\mathcal{K}}]\!]$.

Definition 5.1. For $\mathbb{I}$ a normal domain over Λ which is also a finite module over Λ, a $\bar{\mathbb{Q}}_p$-point $\phi \in \operatorname{Spec}\mathbb{I}$ is called arithmetic if its image in $\operatorname{Spec}\Lambda$ is arithmetic.

(i) If $s > 0$, let $V^N_{\infty,\infty}(K, \mathbb{I}, \chi_0)$ be the set of $\mathbb{I}$-adic formal Fourier–Jacobi expansions

$$\left\{ f_x = \sum_\beta a_\beta(x, f) q^\beta \right\}_x$$

such that, for a Zariski-dense set of generic arithmetic points $\phi \in \operatorname{Spec}\mathbb{I}$, the specialization $\boldsymbol{f}_\phi$ is the formal Fourier–Jacobi expansion of a form on $\mathrm{U}(r, s)$ whose p-part nebentype at $\operatorname{diag}(t_1, \ldots, t_{r+s})$ is given by

$$\chi_0\chi_\phi\omega(t_1^{c_{s+1}} \cdots t_r^{c_{s+r}} t_{r+1}^{-c_1} \cdots t_{r+s}^{-c_s})$$

for the weight $(c_{s+1}, \ldots, c_{s+r}; c_1, \ldots, c_s) = (0, \ldots, 0; \kappa_\phi, \ldots, \kappa_\phi)$. Here, by χ_ϕ we also mean the character of $T(\mathbb{Z}_p)$ restricting to χ_ϕ on $\boldsymbol{T}$ that is trivial on the torsion part of $T(\mathbb{Z}_p)$. We say $\boldsymbol{f} \in V^N_{\infty,\infty}(K, \mathbb{I})$ is a family of eigenforms if the specializations $\boldsymbol{f}_\phi$ above are eigenforms. We define $V^{N,\mathrm{ord}}_{\infty,\infty}(K, \mathbb{I}, \chi_0)$ for the subspace such that the specializations above are nearly ordinary.

(ii) If $s = 0$, then let $K = \prod_v K_v$ and let

$$K_0(p) = \prod_{v \nmid p} K_v \prod_{v \mid p} K_0(p)_v$$

(with $K_0(p)_v \subset G(\mathbb{O}_{F_v})$ be the set of matrices which are in $B(\mathbb{O}_{F,v})$ modulo p). Then $G(F)\backslash G(\mathbb{A}_F)/K_0(p)$ is a finite set with $\{g_i\}_i$ a set of representatives. We identify the set

$$S^N_G(K) := G(F)\backslash G(\mathbb{A}_F)/K^p N(\mathbb{O}_{F,p})$$

with the disjoint union of $g_i \cdot N^-(p\mathbb{O}_{F,p})T(\mathbb{O}_{F,p})$ and endow the latter with the p-adic topology on $N^-(p\mathbb{O}_{F,p})T(\mathbb{O}_{F,p})$. We define $V^N_{\infty,\infty}(K, \mathbb{I}, \chi_0)$ to be the set of continuous $\mathbb{I}$-valued functions on $S^N_G(K)$ such that, for a Zariski-dense set of arithmetic points $\phi \in \operatorname{Spec}\mathbb{I}$, the specialization $\boldsymbol{f}_\phi$ is a form on $\mathrm{U}(r, 0)$ whose p-part nebentype at $\operatorname{diag}(t_1, \ldots, t_r)$ is given by

$$\chi_0\chi_\phi\omega(t_1^{c_1} \cdots t_r^{c_r})$$

for the weight $(0,\dots,0)$. Note that, by the description of nebentypus at p, such a family is determined by its values on $g_i\cdot N^-(p\mathbb{O}_{F,p})$. Similarly we define $V^{N,\mathrm{ord}}_{\infty,\infty}(K,\mathbb{I},\chi_0)$ for the nearly ordinary part.

Remark 5.2. To see this is a good definition, we have to compare it with the notion of Hida families in the literature. We refer to [Hida 2004b, Chapter 8; Hsieh 2014, Sections 3–4] for the definition of Hida families. We have to check that a Hida family in Hsieh's terms does give a Hida family here. We need to show that, if $\kappa_\phi\gg 0$ (depending on the p-part of the conductor at ϕ) when $s>0$, then any nearly ordinary p-adic cusp form is classical. If $s>0$ this is proved by the argument of [Hsieh 2014, Theorem 4.19]. (It is assumed that $s=1$ in [loc. cit.]; however, the proof for this particular theorem works in the general case.) If $s=0$ the situation is even easier: the contraction property of the U_p operator [Hsieh 2014, Proposition 4.4] (which again works in our case as well) shows that the specialization at ϕ is right-invariant under an open subgroup of $\mathrm{U}(r)(\mathbb{Z}_p)$ depending only on the conductor of the nebentypus (note also that we have trivial weight if $s=0$), and is thus classical.

Definition 5.3. We define an Eisenstein datum as a quadruple $\boldsymbol{D}:=(\mathbb{I},\boldsymbol{f},\tau_0,\chi_0)$, where χ_0 is a finite-order character of $T(\mathbb{Z}_p)$, τ_0 is a finite-order character of $\mathcal{K}^\times\backslash\mathbb{A}^\times_{\mathcal{K}}$ whose conductors at primes above p divides (p), and $\boldsymbol{f}\in V^{N,\mathrm{ord}}_{\infty,\infty}(K,\mathbb{I})$ is a Hida family of eigenforms defined as above. We define $\Lambda_{\boldsymbol{D}}:=\Lambda\otimes_{\mathbb{O}_L}\Lambda_{\mathcal{K}}$. We call a $\overline{\mathbb{Q}}_p$-point $\phi\in\operatorname{Spec}\Lambda_{\boldsymbol{D}}$ is arithmetic if $\phi|_{\mathbb{I}}$ is arithmetic with some weight κ_ϕ and $\phi(\gamma^+)=(1+p)^{\kappa_\phi/2}\zeta_+$, $\phi(\gamma^-)=(1+p)^{\kappa_\phi/2}\zeta_-$ for p-power roots of unity $\zeta_\pm$. We define $\tau_\phi=\phi\circ\Psi_{\mathcal{K}}$.

Let $\mathcal{X}$ be the set of arithmetic points. If $\boldsymbol{f}_\phi$ is classical and generates an irreducible automorphic representation $\pi_{\boldsymbol{f}_\phi}$ of $\mathrm{U}(r,s)$, we say that ϕ is generic if $(\pi_{\boldsymbol{f}_\phi},\tau)$ is generic (see Definition 4.42). Let $\mathcal{X}^{\mathrm{gen}}$ be the set of generic arithmetic points.

5B. *Some assumptions.*

5B1. *Including types.* Consider the group $\mathrm{U}(s,r)$. Suppose $K^p=K_\Sigma K^\Sigma\subset G(\mathbb{A}^p_f)$ for a finite set of primes Σ and let W_Σ be a finite $\mathbb{O}_L$-module on which K_Σ acts through a finite quotient. Let $K'_\Sigma\subset K_\Sigma$ be a normal subgroup containing $\prod_{v\in\Sigma\backslash\{v|p\}}\mathfrak{Y}_v$, defined in Definition 4.11 and acting trivially on W_Σ, and let $K'=G(\mathbb{Z}_p)K'_\Sigma K^\Sigma$. The modules of modular forms of weight κ, type W_Σ and character ψ are

$$M_\kappa(K,W_\Sigma;\mathbb{O}_L)=(M_\kappa(K';\mathbb{O}_L)\otimes_{\mathbb{O}_L}W_\Sigma)^{K_\Sigma}.$$

Suppose for $v\in\Sigma\backslash\{v|p\}$ we have open compact subgroups $\tilde{K}'_v\subset\tilde{K}_v\subset G(F_v)$ such that $\tilde{K}'_v$ is a normal subgroup of $\tilde{K}_v$ and an irreducible finite-dimensional representation W_v of $\tilde{K}_v/\tilde{K}'_v$. Suppose $\varphi_v\in\pi_v$ is a vector in W_v. We fix a $\tilde{K}_v$-invariant measure and let $v_1, v_2, \ldots$ be a basis such that φ_v is v_1. We also assume that $\tilde{K}'_v$ includes the $\mathfrak{Y}^{\iota}_v$ defined in Section 4. We let $W_v^\vee$ be the dual representation and we write $v_1^\vee, v_2^\vee, \ldots$ for the dual basis. We first prove the following lemma:

Lemma 5.4. *Let G be a finite group and $\rho : G \to \mathrm{Aut}(V)$ an irreducible representation on an n-dimensional vector space V. We fix a G-invariant norm and a unitary basis $v_1, \ldots, v_n$. Let $\rho^\vee$ be the dual representation on $V^\vee$ with dual basis $v_1^\vee, \ldots, v_n^\vee$. Then, as elements in $V \otimes V^\vee$,*

$$\sum_{g\in G}(gv_i \otimes gv_j^\vee) = 0, \qquad\qquad i \neq j,$$

$$\sum_{g}(gv_i \otimes gv_i^\vee) = |G|\sum_{i=1}^{n} v_i \otimes v_i^\vee.$$

Proof. This is a straightforward application of the Schur orthogonal relation. □

Definition 5.5. We define $W_{\Sigma\setminus\{p\}} = \prod_{v\in\Sigma\setminus\{p\}} W_v$ and $v_1 = \prod_{v\in\Sigma\setminus\{p\}} v_{v,1} \in W_{\Sigma\setminus\{p\}}$.

We can also make a notion of $W_{\Sigma\setminus\{p\}}$-valued Hida families in a similar manner to Definition 5.1.

5B2. *Assumption* TEMPERED. Let $\boldsymbol{f}$ be a Hida family of eigenforms as defined in Definition 5.1. We say it satisfies the assumption "TEMPERED" if the specializations $\boldsymbol{f}_\phi$ in the definition are tempered eigenforms.

5B3. *Assumption* DUAL. We first define an $\mathcal{O}_L$-involution $\circ : \Lambda \to \Lambda$ sending any $\mathrm{diag}(t_1, \ldots, t_n) \in T(1+\mathbb{Z}_p)$ to $\mathrm{diag}(t_n^{-1}, \ldots, t_1^{-1})$. We define $\mathbb{I}^\circ$ to be the ring $\mathbb{I}$ but with the Λ-algebra structure given by composing the involution $\circ$ with the original Λ structure map of $\mathbb{I}$.

Let $\boldsymbol{f}$ be an $\mathbb{I}$-adic cuspidal eigenform on $\mathrm{U}(r, s)$ such that, for a Zariski-dense set of generic arithmetic points ϕ, the specialization $\boldsymbol{f}_\phi$ is classical and generates an irreducible automorphic representation $\pi_{\boldsymbol{f}_\phi}$ of $\mathrm{U}(r, s)$; we say it satisfies assumption DUAL if there is an $\mathbb{I}^\circ$-adic nearly ordinary cusp form $\boldsymbol{f}^\vee$ on $\mathrm{U}(s, r)$ such that $\boldsymbol{f}_\phi^\vee \in \pi_{\boldsymbol{f}_\phi}^\vee$ for all the arithmetic points $\phi \in \mathrm{Spec}\,\mathbb{I}$ that are in the image of some point in $\mathscr{X}^{\mathrm{gen}}$. (Here we identified $\mathrm{U}(r, s)$ and $\mathrm{U}(s, r)$ in the obvious way. At an arithmetic point both f_ϕ and $f_\phi^\vee$ have scalar weight κ. Note also that we only require the specialization $\boldsymbol{f}_\phi$ to be "generic" (not required for $\boldsymbol{f}_\phi^\vee$).)

5B4. *Assumptions* $\mathrm{Proj}_{f^\vee}$ *and* $\mathrm{Proj}_{\boldsymbol{f}^\vee}$. We say a nearly ordinary cuspidal eigenform $f^\vee$ on $\mathrm{U}(s, r)$ satisfies assumption $\mathrm{Proj}_{f^\vee}$ if $(\pi_{f^\vee} \otimes W_{\Sigma\setminus\{p\}})^K$ is 1-dimensional and there is a Hecke operator $1_{f^\vee}$ on $\mathrm{U}(s, r)$ that is an L-coefficient polynomial of Hecke operators outside Σ such that, for any $g \in M_\kappa(K, W_{\Sigma\setminus\{p\}})$, we have that $e^{\mathrm{ord}} \cdot g - 1_{f^\vee} e^{\mathrm{ord}} \cdot g$ is a sum of forms in irreducible automorphic representations which are orthogonal to $\pi_{f^\vee}$.

We say a nonzero nearly ordinary cuspidal $\mathbb{I}^\circ$-adic family of eigenforms $\boldsymbol{f}^\vee$ in $(V^N_{\infty,\infty}(K, \mathbb{I}^\circ, \chi_0^{-1}) \otimes W_{\Sigma\backslash\{p\}})^{K_\Sigma}$ satisfies assumption $\mathrm{Proj}_{\boldsymbol{f}^\vee}$ if there is an action e^{ord} acting on $(V^N_{\infty,\infty}(K, \mathbb{I}^\circ, \chi_0^{-1}) \otimes W_{\Sigma\backslash\{p\}})^{K_\Sigma}$ interpolating the e^{ord} of specializations and there is a Hecke operator $1_{\boldsymbol{f}^\vee}$ which is an $F_{\mathbb{I}}$ polynomial of Hecke operators outside Σ such that, for Zariski-dense set of arithmetic points $\phi \in \operatorname{Spec} \mathbb{I}^\circ$ in the image of $\mathscr{X}^{\mathrm{gen}}$, $(\pi_{\boldsymbol{f}^\vee_\phi} \otimes W_{\Sigma\backslash\{p\}})^K$ is 1-dimensional and, for any $\boldsymbol{g} \in (V^N_{\infty,\infty}(K, \mathbb{I}^\circ, \chi_0^{-1}) \otimes W_\Sigma)^{K_\Sigma}$, $(e^{\mathrm{ord}} \cdot \boldsymbol{g} - 1_{\boldsymbol{f}^\vee} e^{\mathrm{ord}} \boldsymbol{g})_\phi$ is a sum of forms in irreducible automorphic representations which are orthogonal to $\pi_{\boldsymbol{f}^\vee_\phi}$.

Remark 5.6. If $r+s=2$ then these assumptions often hold, since the unitary group is closely related to GL_2 or quaternion algebras. It is easy to see DUAL by simply taking $\boldsymbol{f}^\vee = \boldsymbol{f} \otimes (\boldsymbol{\chi})^{-1}$ for $\boldsymbol{\chi}$ the central character of $\boldsymbol{f}$. To see $\mathrm{Proj}_{\boldsymbol{f}}$ and $\mathrm{Proj}_{\boldsymbol{f}^\vee}$, we first suppose $r = s = 1$ and $\boldsymbol{f}$ is a Hida family of GL_2 newforms with tame level M such that $(M, p\delta_{\mathcal{K}}) = 1$ and trivial character. The existence of e^{ord} is as in [Skinner and Urban 2014, Lemma 12.2] Since we have an isomorphism of algebraic groups over F,

$$\mathrm{GU}(1, 1) \sim \mathrm{GL}_2 \times_{\mathbb{G}_m} \mathrm{Res}_{\mathcal{K}/F}\, \mathbb{G}_m,$$

we can obtain a family on $\mathrm{U}(1, 1)$ from $\boldsymbol{f}$ and the trivial character of $\mathbb{A}^\times_{\mathcal{K}}/\mathcal{K}^\times$, which we still denote by $\boldsymbol{f}$. Take an arithmetic point ϕ and a GL_2 Hecke operator t involving only Hecke operators T_v at primes v outside Σ which are split in $\mathcal{K}/F$ such that the t-eigenvalue $t(\boldsymbol{f}_\phi)$ is different from its eigenvalues on other forms on $S^{\mathrm{ord}}_{\kappa_\phi}(\Gamma_0(M) \cap \Gamma_1(p^{t_\phi}), \mathbb{C})$ (the space of ordinary cusp forms on $\mathrm{U}(1, 1)$ of weight $(0, \kappa_\phi)$ and level $\Gamma_0(M) \cap \Gamma_1(p^{t_\phi})$, with p^{t_ϕ} being the p-part level at ϕ. Also here we use the $\mathrm{U}(1, 1)$ Hecke operators at split primes $v = w\bar{w}$ which are associated to the elements $(\mathrm{diag}(\varpi_w, 1), \mathrm{diag}(1, \varpi_{\bar{w}}^{-1}))$. This is possible since any form in $S^{\mathrm{ord}}_{\kappa_\phi}(\Gamma_0(M) \cap \Gamma_1(p^{t_\phi}), \mathbb{C})$ is the restriction of a form on $\mathrm{GU}(1, 1)$ obtained from a GL_2 form of conductor dividing $\mathrm{Nm}_{\mathcal{K}/F}\delta_{\mathcal{K}/F} M p^{t_\phi}$ and a character of $\mathbb{A}^\times_{\mathcal{K}}/\mathcal{K}^\times$ unramified outside p. Note that any cuspidal automorphic representation on GL_2/F with the same Hecke eigenvalue with $\boldsymbol{f}_\phi$ on split primes is $\pi_{\boldsymbol{f}_\phi}$ or $\pi_{\boldsymbol{f}_\phi} \otimes \chi_{\mathcal{K}/F}$, and that any element $g \in \mathrm{GL}_2(F_v)$ such that $\det(g) \in \mathrm{Nm}_{\mathcal{K}/F}(\mathcal{K}^\times_v)$ can be written as ag' with $a \in \mathcal{K}^\times_v$ and $g' \in \mathrm{U}(1, 1)(F_v)$. A simple representation-theoretic argument shows that the only forms in $S^{\mathrm{ord}}_{\kappa_\phi}(\Gamma_0(M) \cap \Gamma_1(p^{t_\phi}), \mathbb{C})$ with the same Hecke eigenvalues with $\boldsymbol{f}_\phi$ at split primes are in the 1-dimensional space spanned by $\boldsymbol{f}_\phi$. Let Λ be the weight space for $\mathrm{U}(1, 1)$ and define

$$S^{\mathrm{ord}}(\Gamma_0(M), \mathbb{I}) := S^{\mathrm{ord}}(\Gamma_0(M), \Lambda) \otimes_\Lambda \mathbb{I}.$$

It follows from Hida's control theorem for unitary groups (see [Hsieh 2014, Theorem 4.21], for example) that this is a free module over $\mathbb{I}$ of finite rank, and the specialization of this free module to ϕ gives the space $S^{\mathrm{ord}}_{\kappa_\phi}(\Gamma_0(M) \cap \Gamma_1(p^{t_\phi}), \mathcal{O}_L)$ for some L finite over $\mathbb{Q}_p$ provided $\kappa_\phi \gg 0$ with respect to the p-part of the conductor of ϕ. We consider $\det(T-t)$, where T is a variable and we regard t as an operator on this free $\mathbb{I}$-module. We thus obtain an $\mathbb{I}$-coefficient polynomial of T. Moreover, we can write $\det(T-t)=(T-t(\boldsymbol{f}))\cdot g(T)$ for some polynomial $g(T)$. Then we define

$$1_{\boldsymbol{f}}=\frac{g(t)}{g(\boldsymbol{f})}$$

(note that $g(t(\boldsymbol{f}))$ is not identically zero.) This proves $\mathrm{Proj}_{\boldsymbol{f}}$, and $\mathrm{Proj}_{\boldsymbol{f}^\vee}$ is seen in a similar way. If $(r,s)=(2,0)$ we observe that if we set

$$D=\{g\in M_2(\mathcal{K})\mid g^t\zeta\bar{g}=\det(g)\zeta\}$$

then D is a definite quaternion algebra over $\mathbb{Q}$ with local invariants $\mathrm{inv}_v(D)=(-\mathfrak{s},-D_{\mathcal{K}/\mathbb{Q}})_v$ (the Hilbert symbol). The relation between $\mathrm{GU}(2)$ and D is explained by

$$\mathrm{GU}(2)=D^\times\times_{\mathbb{G}_m}\mathrm{Res}_{\mathcal{K}/\mathbb{Q}}\mathbb{G}_m.$$

We can similarly show that, if $\boldsymbol{f}$ is a Hida family of newforms on $D^\times$ with trivial character, tame level prime to p and all primes of $\delta_{\mathcal{K}}$ such that D is unramified, and is the trivial representation at primes where D is ramified, then we can produce a family $\boldsymbol{f}$ on $\mathrm{U}(2,0)$ from $\boldsymbol{f}$ and the trivial character of $\mathbb{A}_{\mathcal{K}}^\times/\mathcal{K}^\times$. A similar argument proves that $\mathrm{Proj}_{\boldsymbol{f}}$ and $\mathrm{Proj}_{\boldsymbol{f}^\vee}$ is true.

5C. *Klingen Eisenstein series and p-adic L-functions.*

5C1. *Construction.* Now we are going to construct the nearly ordinary Klingen Eisenstein series (and will p-adically interpolate them in families). First of all, let τ be a Hecke character which is of infinite type $\left(-\frac{1}{2}\kappa,\frac{1}{2}\kappa\right)$ at all infinite places (here the convention is that the first infinite place of $\mathcal{K}$ is inside our CM type). Recall that we write $\mathcal{D}:=\{\pi,\tau,\Sigma\}$ for the Eisenstein data (see Definition 3.2). We define the normalization factor

$$B_{\mathcal{D}}:=\frac{\Omega_p^{r\kappa\Sigma_\infty}}{\Omega_\infty^{r\kappa\Sigma_\infty}}\left(\frac{(-2)^{-d(a+2b+1)}(2\pi i)^{d(a+2b+1)\kappa}(2/\pi)^{d(a+2b+1)(a+2b)/2}}{\prod_{j=0}^{a+2b}(\kappa-j-1)^d}\right)^{-1}$$
$$\times\prod_{i=0}^{a+2b}L^\Sigma(2z_\kappa+a+2b+1-i,\bar{\tau}'\chi_{\mathcal{K}}^i)\prod_{v\mid p}(\mathfrak{g}(\bar{\tau}_v')^{a+2b+1}c_{a+2b+1}(\tau_v',-z_\kappa))^{-1},$$

$$B'_{\mathcal{D}}:=\frac{\Omega_p^{r\kappa\Sigma_\infty}}{\Omega_\infty^{r\kappa\Sigma_\infty}}\left(\frac{(-2)^{-d(a+2b)}(2\pi i)^{d(a+2b)\kappa}(2/\pi)^{d(a+2b)(a+2b-1)/2}}{\prod_{j=0}^{a+2b-1}(\kappa-j-1)^d}\right)^{-1}$$
$$\times\prod_{i=0}^{a+2b-1}L^\Sigma(2z_\kappa+a+2b-i,\bar{\tau}'\chi_{\mathcal{K}}^i)\prod_{v\mid p}(\mathfrak{g}(\bar{\tau}_v')^{a+2b}c_{a+2b}(\tau_v',-z_\kappa'))^{-1}.$$

Here, $z_\kappa=\frac{1}{2}(\kappa-a-2b-1)$ and $z_\kappa'=\frac{1}{2}(\kappa-a-2b)$, c_m is defined in (13), and Ω_∞ is the CM period in Section 2A.

We construct a Siegel Eisenstein series E_{sieg} associated to the Siegel section

$$f_{\mathfrak{D},\mathrm{sieg}} = B_{\mathfrak{D}} \prod_{v|\infty} f_\kappa \prod_{v|p} \rho(\Upsilon_v) f_v^0 \prod_{v\in\Sigma, v\nmid p} \tilde{f}_{v,\mathrm{sieg}} \prod_{v} f_v^{\mathrm{sph}} \in I_{a+2b+1}(\tau, z)$$

and E'_{sieg} associated to the section

$$f'_{\mathfrak{D},\mathrm{sieg}} = B'_{\mathfrak{D}} \prod_{v|\infty} f'_\kappa \prod_{v|p} \rho(\Upsilon'_v) f_v^{0\prime} \prod_{v\in\Sigma, v\nmid p} \tilde{f}'_{v,\mathrm{sieg}} \prod_{v} f_v^{\mathrm{sph},\prime} \in I_{a+2b}(\tau, z).$$

Here Υ_v and Υ'_v are as defined in Definition 4.32. First note that, since π is nearly ordinary with respect to the scalar weight κ, its contragradient is also nearly ordinary on $\mathrm{U}(s,r)$ with respect to the scalar weight κ. We denote this representation by $\tilde{\pi}$. We consider $E(\gamma(g,-))$ as an automorphic form on $\mathrm{U}(s,r)$. For each $v \nmid p$ we choose an open compact group $\tilde{K}_{v,s} \subset \mathrm{U}(s,r)_v$ such that

$$\prod_{v\in\Sigma, v\nmid p} \rho\big(\gamma(1, \eta\,\mathrm{diag}(\bar{x}_v^{-1}, 1, x_v).\tilde{S}_v^{-1})\big)\big(E(\gamma(g,-)) \otimes \bar{\tau}(\det -)\big)$$

is invariant under its action. We have the following lemma:

Lemma 5.7. *There is a bounded measure $\mathscr{E}_{\mathfrak{D},\mathrm{sieg}}$ on $\Gamma_{\mathcal{K}} \times T(1+\mathbb{Z}_p)$ with values in the space of p-adic automorphic forms on $\mathrm{U}(r+s+1, r+s+1)$ such that, for all arithmetic points $\phi \in \mathcal{X}^{\mathrm{gen}}$ with the associated character $\hat{\phi}$ on $\Gamma_{\mathcal{K}} \times T(1+\mathbb{Z}_p)$, we have*

$$\int_{\Gamma_{\mathcal{K}} \times T(1+\mathbb{Z}_p)} \hat{\phi}\, d\mathscr{E}_{\mathfrak{D},\mathrm{sieg}}$$

is the Siegel Eisenstein series $\rho\big(\prod_{v\in\Sigma, v\nmid p} \gamma(1, \eta\,\mathrm{diag}(\bar{x}_v^{-1}, 1, x_v)\tilde{S}_v^{-1})\big) E_{\mathrm{sieg},\mathfrak{D}_\phi}$, where $E_{\mathrm{sieg},\mathfrak{D}_\phi}$ is the Siegel Eisenstein series we construct using the characters $(\chi_{1,\phi}, \dots, \chi_{n,\phi}, \tau_\phi)$. Similarly, we can define a measure $\mathscr{E}'_{\boldsymbol{D},\mathrm{sieg}}$ interpolating the $E'_{\mathrm{sieg},\mathfrak{D}_\phi}$.

Proof. It follows from our computations for Fourier coefficients, Lemmas 4.2, 4.6, 4.12 and 4.46, and [Skinner and Urban 2014, Lemma 11.2], that all the Fourier coefficients of E_{sieg} and E'_{sieg} are interpolated by elements in $\Lambda_{r,s}[\![\Gamma_{\mathcal{K}}]\!]$. Then the lemma follows from the abstract Kummer congruence. We refer to [Hsieh 2011, Lemma 3.15, Theorem 3.16] for a detailed proof. □

Now we define our Klingen Eisenstein series using the pullback formula. Note that by (3) the pullback of the Siegel Eisenstein series are still holomorphic automorphic forms. Let β be the embedding given in Section 2B. Let $\tilde{K}_s$ be the open compact subgroup of $G(\mathcal{O}_{F,\Sigma})$, which is $\tilde{K}_{v,s}$ as above for $v \in \Sigma\backslash\{v|p\}$, $\tilde{K}_v$ for $v|p$ and spherical otherwise. We define $\boldsymbol{E}_{\boldsymbol{D},\mathrm{Kling}}$ by, for any points x and x_1 on the

Igusa schemes of $\mathrm{U}(r+1,s+1)$ and $\mathrm{U}(s,r)$,

$$e^{\mathrm{ord,low}}\cdot 1^{\mathrm{low}}_{\boldsymbol{f}^\vee}\,\mathrm{Tr}_{\tilde{K}/\tilde{K}_s}(e^{\mathrm{low}}(\beta^{-1}(\mathscr{E}_{\boldsymbol{D},\mathrm{sieg}})\cdot\bar{\tau}(\det(g_1)))\otimes v_1)(x,x_1)\\ =\boldsymbol{E}_{\boldsymbol{D},\mathrm{Kling}}(x)\boxtimes\boldsymbol{f}^\vee(x_1)$$

(as a $W_{\Sigma\backslash\{p\}}$-valued form — recall $v_1\in W_{\Sigma\backslash\{p\}}$; see Section 5B1). Here we let $\tilde{K}_{\Sigma\backslash\{p\}}$ act on both $\mathscr{E}_{\boldsymbol{D},\mathrm{sieg}}$ and $W_{\Sigma\backslash\{p\}}$. We get a $\Lambda_{\mathscr{D}}$-adic formal Fourier–Jacobi expansion from the measure $e^{\mathrm{low}}\beta^{-1}(\mathscr{E}_{\boldsymbol{D},\mathrm{sieg}})$ and then apply the Hecke operators to the expansion. We also define the Σ-primitive p-adic L-function $\mathscr{L}^{\Sigma}_{\boldsymbol{f},\mathscr{K},\tau_0}\in\mathbb{I}^{\mathrm{ur}}[\![\Gamma_{\mathscr{K}}]\!]$ by, for elements x and x_1 in the Igusa schemes of $\mathrm{U}(r,s)$ and $\mathrm{U}(s,r)$,

$$e^{\mathrm{ord,low}}\cdot 1^{\mathrm{low}}_{\boldsymbol{f}^\vee}\,\mathrm{Tr}_{\tilde{K}/\tilde{K}_s}(e^{\mathrm{low}}\beta^{-1}(\mathscr{E}'_{\boldsymbol{D},\mathrm{sieg}})\cdot\bar{\tau}(\det g_1)\otimes v_1)(x,x_1)\\ =\mathscr{L}^{\Sigma}_{\boldsymbol{f},\mathscr{K},\tau_0}\boldsymbol{f}_1(x)\boxtimes\boldsymbol{f}^\vee(x_1).$$

The $\boldsymbol{f}_1$ is the $v_1^\vee$-component of $\boldsymbol{f}$ (see Section 5B1). This is possible by Lemma 5.4. Here note that the necessity of enlarging the coefficient ring to include $\mathbb{O}^{\mathrm{ur}}_L$ is caused when specifying points on Igusa schemes (recall Section 2F).

Here we used the superscript "low" to mean that, under

$$\mathrm{U}(a+b+1,b+1)\times\mathrm{U}(b,a+b)\hookrightarrow\mathrm{U}(a+2b+1,a+2b+1),$$

the action is for the group $\mathrm{U}(b,a+b)$.

5C2. *Identify with Klingen Eisenstein series constructed before.* We define a Klingen Eisenstein section by

$$f^{\square}_{\mathscr{D}_\phi,\mathrm{Kling}}(z,g)=B_{\mathscr{D}}\prod_v F_{\varphi_v}(z;f_{v,\mathrm{sieg}},g),$$

where the $F_{\varphi_v}(z;f_{v,\mathrm{sieg}},g)$ are the pullback sections we computed in Section 5 and φ_v for $v\in\Sigma\backslash\{v|p\}$ is the $v_1^\vee$-component, as in Sections 5B1 and 5B4. We first look at places dividing p. The pairing $\langle\,,\rangle$ induces a natural pairing between π and $\tilde{\pi}$. Write

$$\varphi_w=\prod_{v|\infty}\varphi_v\prod_{v\notin\Sigma}\varphi^{\mathrm{sph}}\prod_{v\in\Sigma,v\nmid p}\varphi_v\prod_{v|p}\varphi_{w,v}.$$

Then

$$\Big\langle\prod_{v\nmid p}\mathrm{Tr}_{\tilde{K}_v/\tilde{K}_{v,s}}\,\rho\big(\gamma(1,\eta\,\mathrm{diag}(\bar{x}_v^{-1},1,x_v).\tilde{S}_v^{-1})\big)\big(E_{\mathrm{sieg}}(\gamma(g,-))\bar{\tau}(\det -)\big),\\ \rho\left(\prod_{v|p}\left(\mathrm{diag}(p^{-t_{a+b+1}},\dots,p^{t_1},\dots,p^{t_{a+1}},\dots)^{\iota}\begin{pmatrix}&&-1_b\\&1_a&\\1_b&&\end{pmatrix}^{\iota}\right)\right)\varphi_w\Big\rangle$$

equals

$$\Big\langle \rho^{\mathrm{low}}\Big(\prod_{v\mid p}\mathrm{diag}(p^{t_{a+b+1}},\ldots,p^{t_{a+2b}},1_a,1_b)^{\iota}\Big)$$
$$\times\prod_{v\nmid p}\mathrm{Tr}_{\tilde{K}_v/\tilde{K}_{v,s}}\,\rho(\gamma(1,\eta\,\mathrm{diag}(\bar{x}_v^{-1},1,x_v)\tilde{S}_v^{-1}))\big(E(\gamma(g,-))\bar{\tau}(\det -)\big),$$
$$\rho\left(\prod_{v\mid p}\left(\mathrm{diag}(1_b,p^{t_1},\ldots,p^{t_{a+1}},\ldots)^{\iota}\begin{pmatrix} & & -1_b\\ & 1_a & \\ 1_b & & \end{pmatrix}^{\iota}\right)\right)\varphi_w\Big\rangle.$$

Since $E_{\mathrm{sieg}}(\gamma(g,-))\bar{\tau}(\det -)$ satisfies the property that, if $\tilde{K}'''$ is the subgroup of $\mathrm{GL}_{a+2b}(\mathbb{Z}_p)$ (defined in the last section) consisting of matrices

$$\begin{pmatrix} a_1 & a_3 & a_2\\ a_7 & a_9 & a_8\\ a_4 & a_6 & a_5\end{pmatrix}$$

such that the (i,j)-th entries of a_7 and a_4 are divisible by $p^{t_i+t_{a+b+j}}$ and $p^{t_{a+i}+t_{a+b+j}}$, respectively, the i-th row of a_8 and the right-to-diagonal entries of a_9 are divisible by p^{t_i} for $i=1,\ldots,a$, the below-diagonal entries of the i-th column of a_1 are divisible by $p^{t_{a+b+i}}$, the up-to-diagonal entries of the i-th row of a_5 are divisible by $p^{t_{a+i}}$, and $a_2,a_3,a_6\in M(\mathbb{Z}_p)$, then the right action of h^{ι} for $h\in\tilde{K}'''$ on $E(\gamma(g,-))\bar{\tau}(\det -)$ is given by the character

$$\lambda(h^{\iota})=\bar{\chi}_{a+b+1}(h_{11})\cdots\bar{\chi}_{a+2b}(h_{bb})\bar{\chi}_1(h_{b+1,b+1})\cdots\bar{\chi}(h_{a+b,a+b})$$
$$\times\bar{\chi}_{a+1}(h_{a+b+1,a+b+1})\cdots\bar{\chi}_{a+b}(h_{a+2b,a+2b}).$$

(This is easily checked from the definition of the Godement section.) It is elementary to check that the above expression equals

$$\prod_{v\mid p}\frac{1}{\prod_{i=1}^{b}p^{t_{a+b+i}(a+b)}}$$
$$\times\Big\langle\prod_{v\mid p}\sum_{y}\rho^{\mathrm{low}}(y)\rho^{\mathrm{low}}(\mathrm{diag}(p^{t_{a+b+1}},\ldots,1_a,1_b)^{\iota})$$
$$\times\prod_{v\nmid p}\mathrm{Tr}_{\tilde{K}_v/\tilde{K}_{v,s}}\rho\big(\gamma(1,\eta\mathrm{diag}(\bar{x}_v^{-1},1,x_v)\tilde{S}_v^{-1})\big)\big(E_{\mathrm{sieg}}(\gamma(g,-))\bar{\tau}(\det -)\big),$$
$$\rho\left(\prod_{v\mid p}\mathrm{diag}(1_b,p^{t_1}\ldots,p^{t_{a+1}},\ldots)^{\iota}\begin{pmatrix} & & -1_b\\ & 1_a & \\ 1_b & & \end{pmatrix}^{\iota}\right)\varphi_w\Big\rangle,\tag{15}$$

where y runs over $N(\mathbb{Z}_p)/\beta N(\mathbb{Z}_p)\beta^{-1}$ for N consisting of matrices of the form $\left(\begin{smallmatrix} 1_s & 0 \\ * & 1_r \end{smallmatrix}\right)$ with $*$ having $\mathbb{Z}_p$-entries and $\beta = \mathrm{diag}(p^{t_{a+b+1}}, \dots, 1_a, 1_b)$. Write the expression

$$e^{\mathrm{ord,low}} \prod_{v|p} \sum_y \rho^{\mathrm{low}}(y)\rho^{\mathrm{low}}(\beta^\iota) \\ \times \prod_{v\nmid p} \mathrm{Tr}_{\tilde{K}_v/\tilde{K}_{v,s}}\, \rho\big(\gamma(1, \eta\, \mathrm{diag}(\bar{x}_v^{-1}, 1, x_v)\tilde{S}_v^{-1})\big)\big(E_{\mathrm{sieg}}(\gamma(g, -))\bar{\tau}(\det -)\big). \quad (16)$$

Now let $\tilde{K}^\flat$ consists of matrices in $\mathrm{GL}_b(\mathbb{Z}_p)$ whose below-diagonal entries of the i-th row are divisible by p^t_{a+b+i} for $1 \le i \le s$. Let $\tilde{K}^\sharp$ be the set of elements in $\mathrm{GL}_{a+2b}(\mathbb{Z}_p)$ whose right-to-diagonal entries of the i-th row are divisible by p^{t_i} for $1 \le i \le a+b$ and whose lower-right $b \times b$ block is in

$$\mathrm{diag}(p^{t_{a+b+1}}, \dots, p^{t_{a+2b}})\tilde{K}^\flat\, \mathrm{diag}(p^{t_{a+b+1}}, \dots, p^{t_{a+2b}})^{-1}.$$

Then a similar argument as in Section 4D1 shows that there is a unique (up to scalar) vector $\tilde{\varphi}_v^\sharp \in \pi(\chi_1^{-1}, \dots, \chi_{a+2b}^{-1})$ such that the action of $(k_{ij}) \in K^\sharp$ is given by the character $\mathrm{diag}(\chi_1^{-1}(k_{11}), \dots, \chi_{a+2b}^{-1}(k_{a+2ba+2b}))$. We use the model of the induced representation from $\chi_1^{-1} \otimes \cdots \otimes \chi_{a+2b}^{-1}$ on the space of smooth functions on $\mathrm{GL}_{a+2b}(\mathbb{Z}_p)$. We take $\tilde{\varphi}_v^\sharp$ such that, if $\tilde{\varphi}_v^{\mathrm{ord}}$ takes value 1 on identity in this model, then $\tilde{\varphi}_v^\sharp$ also takes value 1 on identity (and has support $K^\sharp \subset \mathrm{GL}_{a+2b}(\mathbb{Z}_p)$). From the action of the level group we know that the action of $\rho^{\mathrm{low}}(K^\sharp)$ on the left part of the inner product in (15) is given by the character $\mathrm{diag}(\chi_1^{-1}(k_{11}), \dots, \chi_{a+2b}^{-1}(k_{a+2ba+2b}))$. For $v|p$ define $T_{\beta,v}^{\mathrm{low}}$ to be the Hecke operator corresponding to β just in terms of double cosets acting on $\pi_\varphi^\vee$ (with no normalization factors involved). By checking the actions of the level groups at primes dividing p (certain open compact subgroups of $G(\mathbb{O}_{F,p})$) we can see that the $\tilde{\pi}$ component of the left part, when viewed as an automorphic form on $\mathrm{U}(a+b, b)$, is a multiple of $\tilde{\varphi}^{\mathrm{ord}}$. Suppose the eigenvalue for the Hecke operator $T_{\beta,v}^{\mathrm{low}}$ on $\tilde{\varphi}^{\mathrm{ord}}$ is $\tilde{\lambda}_{\beta,v}$. It is easy to compute that

$$\tilde{\lambda}_{\beta,v} = p^{\sum_{i=1}^b t_{a+b+i}((a+2b+1)/2-i)} \cdot \prod_{j=1}^b \chi_{a+2b+1-j}^{-1}(p^{t_{a+b+j}}) \quad (17)$$

with the convention on the χ_i after Remark 4.41.

Let

$$\varphi' = \prod_{v|\infty} \varphi_v \prod_{v\notin\Sigma} \varphi^{\mathrm{sph}} \prod_{v\in\Sigma, v\nmid p} \varphi_v \\ \times \prod_{v|p} \rho\left(\mathrm{diag}(p^{-t_{a+b+1}}, \dots, p^{t_1}, \dots, p^{t_{a+1}}, \dots)^\iota \begin{pmatrix} & & -1_b \\ & 1_a & \\ 1_b & & \end{pmatrix}^\iota\right)\varphi_{w,v}$$

and

$$\varphi'' = \prod_{v|\infty} \varphi_v \prod_{v\notin\Sigma} \varphi^{\mathrm{sph}} \prod_{v\in\Sigma, v\nmid p} \varphi_v$$
$$\times \prod_{v|p} \rho\left(\mathrm{diag}(1,\ldots,p^{t_1},\ldots,p^{t_{a+1}},\ldots)^{\iota}\begin{pmatrix} & & -1_b \\ & 1_a & \\ 1_b & & \end{pmatrix}^{\iota}\right)\varphi_{w,v}.$$

Here, for $v|\infty$, the φ_v is the unique vector mentioned before Definition 3.1. Define the Klingen Eisenstein section promised in the introduction as

$$f_{\mathfrak{D}_\phi,\mathrm{Kling}} = B_{\mathfrak{D}_\phi}\left(\prod_{v\in\Sigma, v\nmid p} |\tilde{K}_v/\tilde{K}_{v,s}|\right) f^{\square}_{\mathfrak{D}_\phi,\mathrm{Kling}}.$$

Then we have:

Proposition 5.8. *For a classical generic arithmetic point ϕ, we have*

$$\phi(\boldsymbol{E}_{\boldsymbol{D},\mathrm{Kling}})$$
$$= \prod_{v\in\Sigma, v\nmid p} |\tilde{K}_v/\tilde{K}_{v,s}| \frac{E_{\mathrm{Kling}}(f_{\mathfrak{D}_\phi,\mathrm{Kling}}, z_{\kappa_\phi}, g)}{\langle \tilde{\varphi}^{\mathrm{ord}}_\phi, \varphi_\phi\rangle}$$
$$\times \prod_{v|p}\left(\prod_{j=1}^{s} \chi_{r+j}(p^{t_{r+j}}) \prod_{j=1}^{r} \chi_j^{-1}(p^{t_j}) p^{\sum_{i=1}^{s} t_{a+b+i}((a-1)/2+i)} \cdot p^{-\sum_{j=1}^{r} t_j((a+1)/2-j)}\right).$$

Proof. Here, let Θ be the expression (15) and Ξ the expression (16). We have

$$\frac{\Theta}{\langle \tilde{\varphi}^{\sharp}, \varphi''\rangle} = \frac{\langle \Xi, \varphi''\rangle \tilde{\lambda}_{\beta,v}}{\prod_{v|p}\left(\prod_{1\le i\le j\le s} p^{t_{a+b+i}-t_{a+b+j}}\right)\langle \tilde{\varphi}^{\mathrm{ord}}, \varphi''\rangle}$$

and

$$\langle \tilde{\varphi}^{\sharp}, \varphi''\rangle = \langle \tilde{\varphi}^{\mathrm{ord}}, \varphi''\rangle \cdot \prod_{v|p}\left(\prod_{1\le i\le j\le s} p^{t_{a+b+i}-t_{a+b+j}}\right)$$

(e.g., using the model of the induced representation). So

$$\frac{\langle \Xi, \varphi''\rangle}{\langle \tilde{\varphi}^{\mathrm{ord}}, \varphi''\rangle} = \frac{\Theta \prod_{v|p}\left(\prod_{1\le i\le j\le s} p^{t_{a+b+i}-t_{a+b+j}}\right)}{\tilde{\lambda}_{\beta,v}\langle \tilde{\varphi}^{\sharp}, \varphi''\rangle}$$
$$= \frac{\Theta \prod_{v|p}\left(\prod_{1\le i\le j\le s} p^{t_{a+b+i}-t_{a+b+j}}\right)}{\prod_{v|p}\left(\prod_{1\le i\le j\le s} p^{t_{a+b+i}-t_{a+b+j}}\right)\tilde{\lambda}_{\beta,v}\langle \tilde{\varphi}^{\mathrm{ord}}, \varphi''\rangle}.$$

We also have

$$\langle \tilde{\varphi}^{\mathrm{ord}}, \varphi''\rangle = \prod_{j=1}^{r} \chi_j(p^{t_j}) \cdot p^{\sum_{j=1}^{r} t_j((a+1)/2-j)}.$$

The proposition follows. ∎

Then parts (i) and (ii) of Theorem 1.1 are just a corollary of the above proposition (except the statement in the $s=0$ case, which we are going to consider next).

Similarly, we obtain an interpolation formula for the p-adic L-function as in Theorem 1.1, using also the formula (15).

5C3. *Interpolating Petersson inner products for definite unitary groups.* To simplify the exposition we only discuss the case when $F=\mathbb{Q}$ in this subsubsection. In the case when $s=0$, we hope that the periods showing up are CM periods. Thus, by our assumption, the Archimedean components of π are trivial representations. For this purpose we prove that, under certain assumptions, the Petersson inner products of two families can be interpolated by elements in the Iwasawa algebra. Let $K=\prod_v K_v$ be an open compact subgroup of $\mathrm{U}(r,s)(\mathbb{A}_f)$ which is $G(\mathbb{Z}_p)$ at all primes dividing p and $K_0(p)$, obtained from K by replacing the v-component by K_0^1 at all primes v dividing p. Now we take a set $\{g_i\}_i$ of representatives for $\mathrm{U}(r,s)(F)\backslash\mathrm{U}(r,s)(\mathbb{A}_F)/K_0(p)$. We take K sufficiently small so that for all i we have $\mathrm{U}(r,s)(F)\cap g_iKg_i^{-1}=1$. For the nearly ordinary Hida family $\boldsymbol{f}^\vee$ of eigenforms (recall that this Hida family is nearly ordinary with respect to the lower-triangular Borel subgroup) we construct a bounded $\mathbb{I}$-valued measure μ_i on $N^-(p\mathbb{Z}_p)$ as follows. Let T^- be the set of elements $\mathrm{diag}(p^{a_1},\dots,p^{a_r})$ with $a_1\le\cdots\le a_r$. We only need to specify the measure for sets of the form $nt^-N^-(\mathbb{Z}_p)(t^-)^{-1}$, where $n\in N^-(\mathbb{Z}_p)$ and $t^-\in T^-$. We assign its measure $\mu_i(nt^-N^-(\mathbb{Z}_p)(t^-)^{-1})$ by $\boldsymbol{f}^\vee(g_in\cdot t^-)\lambda(t^-)^{-1}$, where $\lambda(t^-)$ is the Hecke eigenvalue of $\boldsymbol{f}^\vee$ for U_{t^-}. This does define a measure. We briefly explain the point when $r=2$ (the general case is only notationally more complicated). Write $\pi_{\boldsymbol{f}_\phi^\vee,p}=\pi(\chi_{1,p},\chi_{2,p})$ such that $v_p(\chi_{1,p}(p))=\frac{1}{2}$, $v_p(\chi_{2,p}(p))=-\frac{1}{2}$. Then $\lambda(\mathrm{diag}(1,p^n))=(\chi_{2,p}(p)\cdot p^{1/2})^n$. One checks that

$$\sum_{m\in p^{n-1}\mathbb{Z}_p/p^n\mathbb{Z}_p}\pi\left(\begin{pmatrix}1&\\m&1\end{pmatrix}_p\right)\pi(\mathrm{diag}(1,p^n)_p)\boldsymbol{f}^\vee_{\phi,p}=(\chi_{2,p}(p)\cdot p^{1/2})\pi(\mathrm{diag}(1,p^{n-1})_p)\boldsymbol{f}^\vee_{\phi,p}.$$

This implies that, for any $m_1\in p\mathbb{Z}_p/p^{n-1}\mathbb{Z}_p$,

$$\sum_{m_2\in p^{n-1}\mathbb{Z}_p/p^n\mathbb{Z}_p}\mu_i(m_1m_2\,\mathrm{diag}(1,p^n)N^-(\mathbb{Z}_p)\,\mathrm{diag}(1,p^{-n}))=\mu_i(m_1\,\mathrm{diag}(1,p^{n-1})N^-(\mathbb{Z}_p)\,\mathrm{diag}(1,p^{1-n})),$$

i.e., this μ_i does define a measure.

Proposition 5.9. *If we define*

$$\langle\boldsymbol{f},\boldsymbol{f}^\vee\rangle:=\sum_i\int_{n\in N^-(p\mathbb{Z}_p)}\boldsymbol{f}(g_in)\,d\mu_i\in\mathbb{I}$$

then, for all $\phi\in\mathcal{X}^{\mathrm{gen}}$, the specialization of $\langle\boldsymbol{f},\boldsymbol{f}^\vee\rangle$ to ϕ is $\langle\boldsymbol{f}_\phi,\boldsymbol{f}_\phi^\vee\rangle\cdot\mathrm{Vol}(\tilde{K}_\phi)^{-1}$.

Proof. For each $\phi \in \mathcal{X}^{\mathrm{gen}}$, we choose t^- such that $t^- N^-(p\mathbb{Z}_p)(t^-)^{-1} \subseteq \tilde{K}_\phi$. We consider

$$\langle \boldsymbol{f}_\phi, \pi^\vee_{\boldsymbol{f}_\phi}(t^-)\boldsymbol{f}^\vee_\phi\rangle.$$

Unfolding the definitions, note $\chi_\phi^{-1}(t^-)\delta_B(t^-)$ gives the Hecke eigenvalue $\lambda(t^-)$; this gives $\delta_B(t^-)\chi_\phi^{-1}(t^-)\sum_i\int_{n\in N^-(p\mathbb{Z}_p)} \boldsymbol{f}(g_i n)\,d\mu_i\cdot \mathrm{Vol}(\tilde{K}_\phi)$. On the other hand, using the model of $\pi_{\boldsymbol{f}_\phi,p}$ and $\pi_{\boldsymbol{f}^\vee_\phi,p}$ as the induced representation $\pi(\chi_{1,\phi},\ldots,\chi_{r,\phi})$ and $\pi(\chi^{-1}_{1,\phi},\ldots,\chi^{-1}_{r,\phi})$ of $\mathrm{GL}_r(\mathbb{Q}_p)$, we get that

$$\langle \boldsymbol{f}_\phi, \pi^\vee_{\boldsymbol{f}_\phi}(t^-)\boldsymbol{f}^\vee_\phi\rangle = \delta_B(t^-)\chi_\phi^{-1}(t^-)\langle \boldsymbol{f}_\phi, \boldsymbol{f}^\vee_\phi\rangle.$$

This proves that the specialization of $\langle \boldsymbol{f}, \boldsymbol{f}^\vee\rangle$ to ϕ is $\langle \boldsymbol{f}_\phi, \boldsymbol{f}^\vee_\phi\rangle\cdot\mathrm{Vol}(\tilde{K}_\phi)^{-1}$. □

So, to see the main theorem in the case when $s=0$, instead of applying the Hecke operator $e^{\mathrm{ord}}\cdot 1_{\boldsymbol{f}^\vee}$ we pair the pullback of Siegel Eisenstein series ($\mathbb{I}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]$-valued) with the measure determined by the Hida family $\boldsymbol{f}$ using the above lemma. That is, considering

$$E_{\mathrm{Kling}}(g, z_\kappa) = \sum_i \int_{n\in N^-(p\mathcal{O}_{F,p})} E_{\mathrm{sieg}}(S^{-1}\alpha(g, g_i n)S, z_\kappa)\,d\mu_i,$$

where the $\{d\mu_i\}_i$ are the measures constructed from $\boldsymbol{f}$ as above. In our situation, when restricting to $\mathrm{U}(s,r)$, the level group at p for Eisenstein series is lower-triangular modulo a certain power of p while that for $\boldsymbol{f}$ is upper-triangular modulo a certain power of p. The above construction works in the same way. The powers of CM and p-adic periods enter when applying the comparison between the standard basis and the Néron basis for differentials of CM abelian varieties while doing pullback (see [Hsieh 2014, (3.14)]).

5D. ***Constant terms.*** We explain part (iii) of the main theorem.

5D1. *p-adic L-functions for Dirichlet characters.* There is an element $\mathcal{L}_{\bar{\tau}'}$ in $\Lambda_{\mathcal{K},\mathcal{O}_L}$ such that $\phi(\mathcal{L}_{\bar{\tau}'}) = L(\bar{\tau}'_\phi, \kappa_\phi - r).\tau'_\phi(p^{-1})p^{\kappa_\phi - r}\mathfrak{g}(\bar{\tau}'_\phi)^{-1}$ at each arithmetic point ϕ in $\mathcal{X}^{\mathrm{pb}}$. For more details see [Skinner and Urban 2014, §3.4.3].

5D2. *Archimedean computation.* As in [Skinner and Urban 2014], we calculate the Archimedean part of the intertwining operator for Klingen Eisenstein sections and prove the "intertwining operator" part (see Lemma 3.4) of the constant term vanishes. Suppose π is associated to the weight $(0,\ldots,0;\kappa,\ldots,\kappa)$; then it is well known that there is a unique (up to scalar) vector $v\in\pi$ such that $k\cdot v = \det\mu(k,i)^{-\kappa}$ for any $k\in K^{+,\prime}_\infty v$ (with notations as in Section 3A1). Recall we defined $c(\rho, z)$ in Section 3A1.

Lemma 5.10. *With assumptions as above,*

$$c(\rho,z)=\pi^{a+2b+1}\prod_{i=0}^{b-1}\left(\frac{1}{z+\frac{1}{2}\kappa-\frac{1}{2}-i-a}\,\frac{1}{z-\frac{1}{2}\kappa+\frac{1}{2}-i}\right)\prod_{i=0}^{a-1}\frac{1}{-1+i-2z+2b}$$
$$\times\frac{\Gamma(2z+a)2^{-1-2z+2b}}{\Gamma\left(\frac{1}{2}(a+1)+z+\frac{1}{2}\kappa\right)\Gamma\left(\frac{1}{2}(a+1)+z-\frac{1}{2}\kappa\right)}\det\left(\tfrac{1}{2}i\zeta\right)^{-2}.$$

Proof. This follows the same way as [Skinner and Urban 2014, Lemma 9.3]. □

Corollary 5.11. *When $\kappa>\frac{3}{2}a+2b$, or $\kappa\geq 2b$ and $a=0$, we have $c(\rho,z)=0$ at the point $z=\frac{1}{2}(\kappa-a-2b-1)$.*

In the case when κ is sufficiently large, the intertwining operator

$$A(\rho,z_\kappa,F)=A(\rho_\infty,z_\kappa,F_\kappa)\otimes A(\rho_f,z_\kappa,F_f)$$

and all terms are absolutely convergent. Thus, as a consequence of the above corollary we have $A(\rho,z_\kappa,F)=0$. Therefore the constant term of E_{Kling} is essentially

$$\frac{L^\Sigma(\tilde{\pi},\bar{\tau}^c,z_\kappa+1)}{\Omega_\infty^{2\kappa\Sigma}\langle\tilde{\varphi}^{\mathrm{ord}},\varphi''\rangle}.L^\Sigma(2z_\kappa+1,\bar{\tau}'\chi_{\mathcal{K}}^{a+2b})\varphi,$$

up to a product of normalization factors at local places. Interpolating the calculations in p-adic families, part (iii) of Theorem 1.1 follows from the above discussion, Lemma 3.4 and our local descriptions for the $F_{\varphi_v}(z;f_{v,\mathrm{sieg}},g)$ in Section 4. (See also the proof of [Skinner and Urban 2014, Theorem 12.11].)

Index of symbols

Appendix: Boundary strata of connected components in positive characteristics

by Kai-Wen Lan

Under the assumption that the PEL datum involves no factor of type D and that the integral model has good reduction, we show that all boundary strata of the toroidal or minimal compactifications of the integral model (constructed in earlier works of the author) have nonempty pullbacks to connected components of geometric fibers, even in positive characteristics.

A.1. ***Introduction.*** Toroidal and minimal compactifications of Shimura varieties and their integral models have played important roles in the study of arithmetic properties of cohomological automorphic representations. While all known models of them are equipped with natural stratifications, they often suffer from some imprecisions or redundancies due to their constructions. The situation is especially subtle in positive or mixed characteristics, or when we need purely algebraic constructions even in characteristic zero (for example, when we study the degeneration of abelian varieties), where the constructions are much less direct than algebraizing complex manifolds created by unions of explicit double coset spaces.

For example, integral models of Shimura varieties defined by moduli problems of PEL structures suffer from the so-called *failure of Hasse's principle*, because there is no known way to tell the difference between two moduli problems associated with algebraic groups which are everywhere locally isomorphic to each other. Similarly, when their toroidal and minimal compactifications are constructed using the theory of degeneration, the data for describing them are also local in nature. Unlike in the complex analytic construction, one cannot just express all the boundary points as the disjoint unions of some double coset spaces labeled by certain standard maximal (rational) parabolic subgroups. (Even the nonemptiness of the *whole boundaries* in positive characteristics was not straightforward — see the introduction to [Lan 2011].) As we shall see, in Example A.7.2, when factors of type D are allowed, it is unrealistic to expect that the boundary stratifications in the algebraic and complex analytic constructions match with each other.

Our goal here is a simple-minded one — to show that the strata of good reduction integral models of toroidal and minimal compactifications constructed as in [Lan 2013a] have nonempty pullbacks to each connected component of each geometric fiber, under the assumption that the data defining them involve no factors of type D (in a sense we will make precise). We will also answer the analogous question for the integral models constructed by normalization in [Lan 2014], allowing arbitrarily deep levels and ramifications (that is, *bad reductions* in general).

This goal is motivated by the study of p-adic families of Eisenstein series, for which it is crucial to know that the strata on connected components of the characteristic-p fibers are all nonempty. For example, this is useful for the consideration of algebraic Fourier–Jacobi expansions. We expect it to play foundational roles in other applications of a similar nature.

A.2. ***Main result.*** We shall formulate our results in the notation system of [Lan 2013a] — henceforth abbreviated [KWL] — which we shall briefly review. (We shall follow [KWL, Notation and conventions, pp. xxi–xxiii] unless otherwise specified. While for practical reasons we cannot explain everything we need from there, we recommend the reader to make use of the reasonably detailed index and table of contents there when looking for the numerous definitions.)

Let $(\mathcal{O}, \star, L, \langle\,\cdot\,,\cdot\,\rangle, h_0)$ be an *integral PEL datum*, where $\mathcal{O}$, $\star$, and $(L, \langle\,\cdot\,,\cdot\,\rangle, h_0)$ are as in [KWL, Definition 1.2.1.3], satisfying [KWL, Condition 1.4.3.10], which defines a group functor G over $\mathbb{Z}$ as in [KWL, Definition 1.2.1.6], and the *reflex field* F_0 (as a subfield of $\mathbb{C}$), as in [KWL, Definition 1.2.5.4], with ring of integers $\mathcal{O}_{F_0}$. Let p be any *good prime*, as in [KWL, Definition 1.4.1.1]. Let $\mathcal{H}^p$ be any open compact subgroup of $\mathrm{G}(\hat{\mathbb{Z}}^p)$ that is *neat*, as in [KWL, Definition 1.4.1.8]. Then we have a moduli problem $\mathsf{M}_{\mathcal{H}^p}$ over $\mathsf{S}_0 = \mathrm{Spec}(\mathcal{O}_{F_0,(p)})$, as in [KWL, Definition 1.4.1.4], which is representable by a scheme that is quasiprojective and smooth over S_0, by [KWL, Theorem 1.4.1.11 and Corollary 7.2.3.10]. By [KWL, Theorem 7.2.4.1 and Proposition 7.2.4.3], we have the *minimal compactification* $\mathsf{M}^{\min}_{\mathcal{H}^p}$ of $\mathsf{M}_{\mathcal{H}^p}$, which is a scheme that is projective and flat over S_0, with geometrically normal fibers. Moreover, for each compatible collection Σ^p of cone decompositions for $\mathsf{M}_{\mathcal{H}^p}$, as in [KWL, Definition 6.3.3.4], we also have the *toroidal compactification* $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p}$ of $\mathsf{M}_{\mathcal{H}^p}$, which is an algebraic space that is proper and smooth over S_0, by [KWL, Theorem 6.4.1.1], and which is representable by a scheme projective over M_0 when Σ^p is *projective*, as in [KWL, Definition 7.3.1.3], by [KWL, Theorem 7.3.3.4]. Any such $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p}$ admits a canonical surjection $\oint_{\mathcal{H}^p} : \mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p} \to \mathsf{M}^{\min}_{\mathcal{H}^p}$, which is constructed by Stein factorization as in [KWL, Section 7.2.3], whose fibers are all geometrically connected. (The superscript "p" indicates that the objects are defined using level structures "away from p". We will also encounter their variants without the superscript "p", which also involve level structures "at p".)

By [KWL, Theorem 7.2.4.1(4)], there is a stratification of $\mathsf{M}^{\min}_{\mathcal{H}^p}$ by locally closed subschemes $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$, where $[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]$ runs through the (finite) set of *cusp labels* for $\mathsf{M}_{\mathcal{H}^p}$ (see [KWL, Definition 5.4.2.4]). The open dense subscheme $\mathsf{M}_{\mathcal{H}^p}$ is the stratum labeled by $[(0,0)]$; we call all the other strata the *cusps* of $\mathsf{M}_{\mathcal{H}^p}$. Similarly, by [KWL, Theorem 6.4.1.1(2)], there is a stratification of $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p}$ by locally closed subschemes $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p},\sigma^p)]}$, where $[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p},\sigma^p)]$ runs through equivalence classes, as in [KWL, Definition 6.2.6.1], with $\sigma^p \subset \boldsymbol{P}^+_{\Phi_{\mathcal{H}^p}}$ and $\sigma^p \in \Sigma_{\Phi_{\mathcal{H}^p}} \in \Sigma^p$. By [KWL, Theorem 7.2.4.1(5)], the surjection $\oint_{\mathcal{H}^p}$ induces a surjection from the $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]$-stratum $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]}$ of $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p, \Sigma^p}$ to the $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$-stratum $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]}$ of $\mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p}$.

Let $s \to \mathsf{S}_0$ be any geometric point with residue field $k(s)$, and let U be any connected component of the fiber $\mathsf{M}_{\mathcal{H}^p} \times_{\mathsf{S}_0} s$. Since $\mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p} \to \mathsf{S}_0$ is proper and has geometrically normal fibers, the closure U^{min} of U in $\mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p} \times_{\mathsf{S}_0} s$ is a connected component of $\mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p} \times_{\mathsf{S}_0} s$. Similarly, since $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p, \Sigma^p} \to \mathsf{S}_0$ is proper and smooth, the closure U^{tor} of U in $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p, \Sigma^p} \times_{\mathsf{S}_0} s$ is a connected component of $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p, \Sigma^p} \times_{\mathsf{S}_0} s$. (In these cases the connected components are also the irreducible components of the ambient spaces.)

The stratifications of $\mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p}$ and $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p, \Sigma^p}$ induce stratifications of U^{min} and U^{tor}, respectively, by pullback. We shall denote the pullback of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]}$ to U^{min} by $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]}$ and call it the $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$-stratum of U^{min}. Similarly, we shall denote the pullback of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]}$ to U^{tor} by $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]}$, and call it the $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]$-stratum of U^{tor}. By construction, the surjection $\oint_{\mathcal{H}^p}$ induces a surjection $U^{\mathrm{tor}} \to U^{\mathrm{min}}$, which maps the $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]$-stratum $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]}$ of U^{tor} surjectively onto the $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$-stratum $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]}$ of U^{min}. It is natural to ask whether a particular stratum of U^{min} or U^{tor} is nonempty.

From now on, we shall assume the following:

Assumption A.2.1. The semisimple algebra $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ over $\mathbb{Q}$ involves no factor of type D (in the sense of [KWL, Definition 1.2.1.15]).

Our main result is the following:

Theorem A.2.2. *With the setting as above, all strata of U^{min} are nonempty.*

An immediate consequence is the following:

Corollary A.2.3. *With the setting as above, all strata of U^{tor} are nonempty.*

Proof. Since the canonical morphism $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]} \to U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]}$ is surjective for each equivalence class $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]$ with underlying cusp label $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$ as above, the nonemptiness of $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]}$ implies that of $U_{[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]}$. $\square$

Remark A.2.4. Each stratum $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p}, \mathsf{Z}_{\mathcal{H}^p})]}$ (resp. $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p}, \mathsf{Z}_{\mathcal{H}^p}, \sigma^p)]}$) is nonempty by [KWL, Theorem 7.2.4.1(4)–(5), Corollary 6.4.1.2, and the explanation of the existence of complex points as in Remark 1.4.3.14]. The question is whether its pullback to U^{min} (resp. U^{tor}) is still nonempty for every U as above.

Remark A.2.5. It easily follows from Theorem A.2.2 and Corollary A.2.3 that their analogues are also true when the geometric point $s \to \mathsf{S}_0$ is replaced with morphisms from general schemes, although we shall omit their statements. In particular, we can talk about connected components of fibers rather than geometric fibers.

The proof of Theorem A.2.2 will be carried out in Sections A.3, A.4, and A.5. In Sections A.5 and A.6, we will also state and prove analogues of Theorem A.2.2 in zero and arbitrarily ramified characteristics, respectively (see Theorems A.5.1 and A.6.1). We will give some examples in Section A.7, including one (see Example A.7.2) showing that we cannot expect Theorem A.2.2 to be true without the requirement (in Assumption A.2.1) that $\mathbb{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ involves no factor of type D.

A.3. ***Reduction to the case of characteristic zero.*** The goal of this section is to prove the following:

Proposition A.3.1. *Suppose Theorem A.2.2 is true when* $\mathrm{char}(k(s)) = 0$*. Then it is also true when* $\mathrm{char}(k(s)) = p > 0$.

Remark A.3.2. Proposition A.3.1 holds regardless of Assumption A.2.1.

Remark A.3.3. It might seem that everything in characteristic zero is well known and straightforward. But Proposition A.3.1, which is insensitive to the crucial Assumption A.2.1, shows that the key difficulty is in fact in characteristic zero.

By [KWL, Theorem 7.2.4.1(4)], each $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ is isomorphic to a boundary moduli problem $\mathsf{M}_{\mathcal{H}^p}^{\mathsf{Z}_{\mathcal{H}^p}}$ defined in the same way as $\mathsf{M}_{\mathcal{H}^p}$ (but with certain integral PEL datum associated with $\mathsf{Z}_{\mathcal{H}^p}$). Then it makes sense to consider the minimal compactification $\mathsf{Z}^{\min}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$, which is proper flat and has geometrically normal fibers over $\mathsf{M}_{\mathcal{H}}$, as in [KWL, Theorem 7.2.4.1 and Proposition 7.2.4.3]. (So the connected components of the geometric fibers of $\mathsf{Z}^{\min}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \to \mathsf{S}_0$ are closures of those of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \to \mathsf{S}_0$.) By considering the Stein factorizations of the structural morphisms $\mathsf{Z}^{\min}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \to \mathsf{S}_0$ (see [EGA III$_1$ 1961, Corollaire (4.3.3) and Remarque (4.3.4), pp. 131–132]), we obtain the following:

Lemma A.3.4 (cf. [KWL, Corollary 6.4.1.2] and [Deligne and Mumford 1969, Theorem 4.17]). *Suppose* $\mathrm{char}(k(s)) = p > 0$*. Then there exists some discrete valuation ring R that is flat over $\mathbb{O}_{F_0,(p)}$, with fraction field K and residue field $k(s)$, the latter lifting the structural homomorphism $\mathbb{O}_{F_0,(p)} \to k(s)$ such that, for each cusp label $[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]$ and each connected component V of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \otimes_{\mathbb{O}_{F_0,(p)}} R$, the induced flat morphism $V \to$* $\mathrm{Spec}(R)$ *has connected special fiber over* $\mathrm{Spec}(k(s))$.

Proof of Proposition A.3.1. Let R be as in Lemma A.3.4. Let $\tilde{U}$ denote the connected component of $\mathsf{M}_{\mathcal{H}^p} \otimes_{\mathbb{O}_{F_0,(p)}} R = \mathsf{Z}_{[(0,0)]} \otimes_{\mathbb{O}_{F_0,(p)}} R$ such that $\tilde{U} \otimes_R k(s) = U$ as subsets of $\mathsf{M}_{\mathcal{H}^p} \otimes_{\mathbb{O}_{F_0,(p)}} k(s) = \mathsf{M}_{\mathcal{H}^p} \times_{\mathsf{S}_0} s$, and let $\tilde{U}^{\min}$ denote its closure in $\mathsf{M}^{\min}_{\mathcal{H}^p} \otimes_{\mathbb{O}_{F_0,(p)}} R$, which is a connected component of $\mathsf{M}^{\min}_{\mathcal{H}^p} \otimes_{\mathbb{O}_{F_0,(p)}} R$ because $\mathsf{M}^{\min}_{\mathcal{H}^p} \otimes_{\mathbb{O}_{F_0,(p)}} R$ is normal, by [KWL, Proposition 7.2.4.3(4)]. For each cusp label $[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]$, let $\tilde{U}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ denote the pullback of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ to $\tilde{U}^{\min}$. Then $\tilde{U}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ is an open and closed subscheme of $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \otimes_{\mathbb{O}_{F_0,(p)}} R$ such that $\tilde{U}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \otimes_R k(s) = U_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ as subsets of $\mathsf{M}^{\min}_{\mathcal{H}^p} \otimes_{\mathbb{O}_{F_0,(p)}} k(s)$. By Lemma A.3.4, it suffices to show that $\tilde{U}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \otimes_R \overline{K} \neq \varnothing$ for some algebraic closure $\overline{K}$ of K. Also by Lemma A.3.4, $\tilde{U} \otimes_R \overline{K} \neq \varnothing$, and so $\tilde{U}^{\min} \otimes_R \overline{K}$ contains at least one connected component of $\mathsf{M}^{\min}_{\mathcal{H}^p} \otimes_{\mathcal{O}_{F_0,(p)}} \bar{K}$. Thus, $\tilde{U}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \otimes_R \bar{K} \neq \varnothing$ under the assumption of the proposition, as desired. $\square$

A.4. ***Comparison of cusp labels.*** Let $\mathcal{H}_p := \mathrm{G}(\mathbb{Z}_p)$ and $\mathcal{H} := \mathcal{H}^p\mathcal{H}_p$, the latter being a neat open compact subgroup of $\mathrm{G}(\hat{\mathbb{Z}})$. By the same references to [KWL] as in Section A.2, we have the moduli problem $\mathsf{M}_{\mathcal{H}}$ and its minimal compactification $\mathsf{M}^{\min}_{\mathcal{H}}$ over $\mathsf{S}_{0,\mathbb{Q}} := \mathsf{S}_0 \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathrm{Spec}(F_0)$. For each compatible collection Σ' of cone decompositions for $\mathsf{M}_{\mathcal{H}}$, we also have a toroidal compactification $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma'}$, together with a canonical morphism $\oint_{\mathcal{H}} : \mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma'} \to \mathsf{M}^{\min}_{\mathcal{H}}$, over $\mathsf{S}_{0,\mathbb{Q}}$. (Here Σ' does not have to be related to the Σ^p above.)

Each cusp label $[(\mathrm{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$ (where $\mathrm{Z}_{\mathcal{H}}$ has been suppressed in the notation for simplicity) can be described as an equivalence class of the $\mathcal{H}$-orbit $(\mathrm{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$ of some triple $(\mathrm{Z}, \Phi, \delta)$, where:

(1) $\mathrm{Z} = \{\mathrm{Z}_{-i}\}_{i\in\mathbb{Z}}$ is an admissible filtration on $L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ that is fully symplectic, as in [KWL, Definition 5.2.7.1]. In particular, $\mathrm{Z}_{-i} = (\mathrm{Z}_{-i} \otimes_{\mathbb{Z}} \mathbb{Q}) \cap (L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})$, the symplectic filtration $\mathrm{Z} \otimes_{\mathbb{Z}} \mathbb{Q}$ on $L \otimes_{\mathbb{Z}} \mathbb{A}^\infty$ extends to a symplectic filtration $\mathrm{Z}_{\mathbb{A}}$ on $\mathrm{Z} \otimes_{\mathbb{Z}} \mathbb{A}$, and each graded piece of Z or $\mathrm{Z} \otimes_{\mathbb{Z}} \mathbb{Q}$ is integrable, as in [KWL, Definition 1.2.1.23], that is, it is the base extension of some $\mathcal{O}$-lattice.

(2) $\Phi = (X, Y, \phi, \varphi_{-2}, \varphi_0)$ is a torus argument, as in [KWL, Definition 5.4.1.3], where $\phi : Y \hookrightarrow X$ is an embedding of $\mathcal{O}$-lattices with finite cokernel, and where $\varphi_{-2} : \mathrm{Gr}^{\mathrm{Z}}_{-2} \xrightarrow{\sim} \mathrm{Hom}_{\hat{\mathbb{Z}}}(X \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \hat{\mathbb{Z}}(1))$ and $\varphi_0 : \mathrm{Gr}^{\mathrm{Z}}_0 \xrightarrow{\sim} Y \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ are isomorphisms matching the pairing $\langle \cdot , \cdot \rangle_{20} : \mathrm{Gr}^{\mathrm{Z}}_{-2} \times \mathrm{Gr}^{\mathrm{Z}}_0 \to \hat{\mathbb{Z}}(1)$ induced by $\langle \cdot , \cdot \rangle$ with the pairing $\langle \cdot , \cdot \rangle_\phi : \mathrm{Hom}_{\hat{\mathbb{Z}}}(X \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \hat{\mathbb{Z}}(1)) \times (Y \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}) \to \hat{\mathbb{Z}}(1)$ induced by ϕ.

(3) $\delta : \mathrm{Gr}^{\mathrm{Z}} \xrightarrow{\sim} L$ is an $\mathcal{O}$-equivariant splitting of the filtration Z.

(4) Two triples $(\mathrm{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$ and $(\mathrm{Z}'_{\mathcal{H}}, \Phi'_{\mathcal{H}}, \delta'_{\mathcal{H}})$ are equivalent (as in [KWL, Definition 5.4.2.2]) if $\mathrm{Z}_{\mathcal{H}} = \mathrm{Z}'_{\mathcal{H}}$ and there exists a pair of isomorphisms, $\gamma_X : X' \xrightarrow{\sim} X$ and $\gamma_Y : Y \xrightarrow{\sim} Y'$, matching $\Phi_{\mathcal{H}}$ with $\Phi'_{\mathcal{H}}$.

Since $\mathcal{H} = \mathcal{H}^p\mathcal{H}_p$, it makes sense to consider the p-part of $(\mathrm{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$, which is the $\mathcal{H}_p$-orbit of some triple $(\mathrm{Z}_{\mathbb{Z}_p}, (\varphi_{-2,\mathbb{Z}_p}, \varphi_{0,\mathbb{Z}_p}), \delta_{\mathbb{Z}_p})$, where:

(1) $\mathrm{Z}_{\mathbb{Z}_p} = \{\mathrm{Z}_{\mathbb{Z}_p,-i}\}_{i\in\mathbb{Z}}$ is a symplectic admissible filtration on $L \otimes_{\mathbb{Z}} \mathbb{Z}_p$, which determines and is determined by a symplectic admissible filtration $\mathrm{Z}_{\mathbb{Q}_p} = \{\mathrm{Z}_{\mathbb{Q}_p,-i}\}_{i\in\mathbb{Z}}$ of $L \otimes_{\mathbb{Z}} \mathbb{Q}_p$ by $\mathrm{Z}_{\mathbb{Q}_p,-i} = \mathrm{Z}_{\mathbb{Z}_p,-i} \otimes_{\mathbb{Z}} \mathbb{Q}$ and $\mathrm{Z}_{\mathbb{Z}_p,-i} = \mathrm{Z}_{\mathbb{Q}_p,-i} \cap (L \otimes_{\mathbb{Z}} \mathbb{Z}_p)$ for all $i \in \mathbb{Z}$.

(2) $\varphi_{-2,\mathbb{Z}_p} : \mathrm{Gr}^{\mathrm{Z}_{\mathbb{Z}_p}}_{-2} \xrightarrow{\sim} \mathrm{Hom}_{\mathbb{Z}_p}(X \otimes_{\mathbb{Z}} \mathbb{Z}_p, \mathbb{Z}_p(1))$ and $\varphi_0 : \mathrm{Gr}^{\mathrm{Z}_{\mathbb{Z}_p}}_0 \xrightarrow{\sim} Y \otimes_{\mathbb{Z}} \mathbb{Z}_p$ are isomorphisms matching the pairing $\langle \cdot , \cdot \rangle_{20,\mathbb{Z}_p} : \mathrm{Gr}^{\mathrm{Z}_{\mathbb{Z}_p}}_{-2} \times \mathrm{Gr}^{\mathrm{Z}_{\mathbb{Z}_p}}_0 \to \mathbb{Z}_p(1)$ induced by $\langle \cdot , \cdot \rangle$ with the pairing $\langle \cdot , \cdot \rangle_{\phi,\mathbb{Z}_p} : \mathrm{Hom}_{\mathbb{Z}_p}(X \otimes_{\mathbb{Z}} \mathbb{Z}_p, \mathbb{Z}_p(1)) \times (Y \otimes_{\mathbb{Z}} \mathbb{Z}_p) \to \mathbb{Z}_p(1)$ induced by ϕ.

(3) $\delta_{\mathbb{Z}_p} : \mathrm{Gr}^{\mathrm{Z}_{\mathbb{Z}_p}} \xrightarrow{\sim} L \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is a splitting of the filtration $\mathrm{Z}_{\mathbb{Z}_p}$.

By forgetting its p-part, each representative $(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$ for $\mathsf{M}_{\mathcal{H}}$ induces a representative $(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})$ for $\mathsf{M}_{\mathcal{H}^p}$, and this assignment is compatible with the formation of equivalence classes. Therefore, we have well-defined assignments

$$(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}}) \mapsto (\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}) \tag{A.4.1}$$

and

$$[(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})] \mapsto [(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]. \tag{A.4.2}$$

By construction, these assignments are compatible with surjections on both their sides (see [KWL, Definition 5.4.2.12]). We would like to show that they are both bijective.

Lemma A.4.3. *Let k be any field over $\mathbb{Z}_{(p)}$. Consider the assignment to each flag W of totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} k$-submodules of $L \otimes_{\mathbb{Z}} k$ (with respect to $\langle \cdot, \cdot \rangle \otimes_{\mathbb{Z}} k$) its stabilizer subgroup P_W in $\mathrm{G} \otimes_{\mathbb{Z}} k$. Then each such P_W is a parabolic subgroup of $\mathrm{G} \otimes_{\mathbb{Z}} k$ and the assignment is bijective. Moreover, given any minimal parabolic subgroup P_{W_0} of $\mathrm{G} \otimes_{\mathbb{Z}} k$, which is the stabilizer of some maximal flag W_0 of totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} k$-submodules of $L \otimes_{\mathbb{Z}} k$, every parabolic subgroup of $\mathrm{G} \otimes_{\mathbb{Z}} k$ is conjugate under the action of $\mathrm{G}(k)$ to some parabolic subgroup of $\mathrm{G} \otimes_{\mathbb{Z}} k$ containing P_{W_0}, which is the stabilizer of some subflag of W_0.*

Although the assertions in this lemma are well known, we provide a proof because we cannot find a convenient reference in the literature in the generality we need.

Proof. Let k^{sep} be a separable closure of k. Since the characteristic of k is either 0 or p, the latter being a good prime by assumption, it follows from [KWL, Proposition 1.2.3.11] that each of the simple factors of the adjoint quotient of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ is isomorphic to one of the groups of standard type listed in the proof of [KWL, Proposition 1.2.3.11]. Then we can make an explicit choice of a Borel subgroup B of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ stabilizing a flag of totally isotropic submodules, with a maximal torus T of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ contained in B which is isomorphic to the group of automorphisms of the graded pieces of this flag. By [Springer 1998, Theorem 6.2.7 and Theorem 8.4.3(iv)], since all parabolic subgroups of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ are conjugate to one containing B, the parabolic subgroups of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ are exactly the stabilizers of flags of totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$-submodules of $L \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$. Then the analogous assertion over k follows, because the assignment of maximal parabolic subgroups of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ is compatible with the actions of $\mathrm{Gal}(k^{\mathrm{sep}}/k)$ on the set of flags of totally isotropic submodules of $L \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$ and on the set of parabolic subgroups of $\mathrm{G} \otimes_{\mathbb{Z}} k^{\mathrm{sep}}$. The last assertion of the lemma follows from [Springer 1998, Theorem 15.1.2(ii) and Theorem 15.4.6(i)]. □

Lemma A.4.4. *The assignment*

$$\mathsf{Z}_{\mathcal{H}} \mapsto \mathsf{Z}_{\mathcal{H}_p} \tag{A.4.5}$$

is bijective.

Proof. Let $\mathrm{Z}_{\mathbb{Z}_p} = \{\mathrm{Z}_{\mathbb{Z}_p,-i}\}_{i\in\mathbb{Z}}$ be a symplectic admissible filtration on $L \otimes_{\mathbb{Z}} \mathbb{Z}_p$ as above, which determines and is determined by a symplectic filtration $\mathrm{Z}_{\mathbb{Q}_p} = \{\mathrm{Z}_{\mathbb{Q}_p,-i}\}_{i\in\mathbb{Z}}$ on $L \otimes_{\mathbb{Z}} \mathbb{Q}_p$. By Lemma A.4.3, the action of $\mathrm{G}(\mathbb{Q}_p)$ on the set of such filtrations $\mathrm{Z}_{\mathbb{Q}_p}$ is transitive, because the $\mathcal{O}$-multirank (see [KWL, Definition 1.2.1.25]) of the bottom piece $\mathrm{Z}_{\mathbb{Q}_p,-2}$ of any such $\mathrm{Z}_{\mathbb{Q}_p}$ is determined by the existence of some isomorphism

$$\varphi_{-2,\mathbb{Z}_p} : \mathrm{Gr}_{-2}^{\mathrm{Z}_{\mathbb{Z}_p}} \xrightarrow{\sim} \mathrm{Hom}_{\mathbb{Z}_p}(X \otimes_{\mathbb{Z}} \mathbb{Z}_p, \mathbb{Z}_p(1)).$$

Let P denote the parabolic subgroup of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}_p$ stabilizing any such $\mathrm{Z}_{\mathbb{Q}_p}$ (see Lemma A.4.3). Since p is a good prime by assumption, the pairing $\langle\cdot,\cdot\rangle \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is self-dual, and hence $\mathrm{G}(\mathbb{Z}_p)$ is a maximal open compact subgroup of $\mathrm{G}(\mathbb{Q}_p)$, by [Bruhat and Tits 1972, Corollary 3.3.2]. Since $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}_p$ is connected under Assumption A.2.1 (because the kernel of the similitude character of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}_p$ factorizes over an algebraic closure of $\mathbb{Q}_p$ as a product of connected groups, by the proof of [KWL, Proposition 1.2.3.11]), we have the Iwasawa decomposition $\mathrm{G}(\mathbb{Q}_p) = \mathrm{G}(\mathbb{Z}_p)\mathrm{P}(\mathbb{Q}_p)$, by [Bruhat and Tits 1972, Proposition 4.4.3] (see also [Casselman 1980, (18) on p. 392] for a more explicit statement). Consequently, $\mathcal{H}_p = \mathrm{G}(\mathbb{Z}_p)$ acts transitively on the set of possible filtrations $\mathrm{Z}_{\mathbb{Z}_p}$ as above, and hence the assignment (A.4.5) is injective.

As for the surjectivity of (A.4.5), it suffices to show that, for some symplectic admissible filtration $\mathrm{Z}_{\mathbb{Z}_p}$, an isomorphism $\varphi_{-2,\mathbb{Z}_p} : \mathrm{Gr}_{-2}^{\mathrm{Z}_{\mathbb{Z}_p}} \xrightarrow{\sim} \mathrm{Hom}_{\mathbb{Z}_p}(X \otimes_{\mathbb{Z}} \mathbb{Z}_p, \mathbb{Z}_p(1))$ exists. By [Reiner 1975, Theorem 18.10] and [KWL, Corollary 1.1.2.6], it suffices to show that there exists some symplectic filtration $\mathrm{Z}_{\mathbb{Q}_p}$ such that $\mathrm{Z}_{\mathbb{Q}_p,-2}$ and $\mathrm{Hom}_{\mathbb{Q}_p}(X \otimes_{\mathbb{Z}} \mathbb{Q}_p, \mathbb{Q}_p(1))$ have the same $\mathcal{O}$-multirank. Or, rather, we just need to notice that the $\mathcal{O}$-multirank of a totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}_p$-submodule can be any $\mathcal{O}$-multirank below a maximal one (with respect to the natural partial order), by Assumption A.2.1 and by the classification in [KWL, Proposition 1.2.3.7 and Corollary 1.2.3.10]. □

Lemma A.4.6. *The assignment (A.4.1) is bijective.*

Proof. It is already explained in the proof of Lemma A.4.4 that an isomorphism $\varphi_{-2,\mathbb{Z}_p} : \mathrm{Gr}_{-2}^{\mathrm{Z}_{\mathbb{Z}_p}} \xrightarrow{\sim} \mathrm{Hom}_{\mathbb{Z}_p}(X \otimes_{\mathbb{Z}} \mathbb{Z}_p, \mathbb{Z}_p(1))$ exists for any $\mathrm{Z}_{\mathbb{Z}_p}$ considered there. Since p is a good prime, which forces both $[L^{\#} : L]$ and $[X : \phi(Y)]$ to be prime to p, any choice of $\varphi_{-2,\mathbb{Z}_p}$ above uniquely determines an isomorphism $\varphi_0 : \mathrm{Gr}_0^{\mathrm{Z}_{\mathbb{Z}_p}} \xrightarrow{\sim} Y \otimes_{\mathbb{Z}} \mathbb{Z}_p$. Also, by the explicit classification in [KWL, Proposition 1.2.3.7 and Corollary 1.2.3.10] as in the proof of Lemma A.4.4, there exists a splitting $\delta_{\mathbb{Z}_p} : \mathrm{Gr}^{\mathrm{Z}_{\mathbb{Z}_p}} \xrightarrow{\sim} L \otimes_{\mathbb{Z}} \mathbb{Z}_p$ and the action of $\mathrm{G}(\mathbb{Z}_p) \cap \mathrm{P}(\mathbb{Q}_p)$ acts transitively on the set of possible triples $(\varphi_{-2,\mathbb{Z}_p}, \varphi_{0,\mathbb{Z}_p}, \delta_{\mathbb{Z}_p})$. Hence the assignment (A.4.1) is bijective, as desired. □

Lemma A.4.7. *The assignment (A.4.2) is bijective.*

Proof. By Lemma A.4.6, it suffices to show that (A.4.2) is injective. Suppose two representatives $(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$ and $(\mathsf{Z}'_{\mathcal{H}}, \Phi'_{\mathcal{H}}, \delta'_{\mathcal{H}})$ with $\Phi_{\mathcal{H}} = (X, Y, \phi, \varphi_{-2,\mathcal{H}}, \varphi_{0,\mathcal{H}})$ and $\Phi'_{\mathcal{H}} = (X', Y', \phi', \varphi'_{-2,\mathcal{H}}, \varphi'_{0,\mathcal{H}})$ are such that the induced $(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})$ and $(\mathsf{Z}'_{\mathcal{H}^p}, \Phi'_{\mathcal{H}^p}, \delta'_{\mathcal{H}^p})$ are equivalent to each other. By definition, $\mathsf{Z}_{\mathcal{H}^p} = \mathsf{Z}'_{\mathcal{H}^p}$, so that $\mathsf{Z}_{\mathcal{H}} = \mathsf{Z}'_{\mathcal{H}}$ by Lemma A.4.4, and there exists a pair $(\gamma_X : X' \xrightarrow{\sim} X, \gamma_Y : Y \xrightarrow{\sim} Y')$ matching $\Phi_{\mathcal{H}^p}$ with $\Phi'_{\mathcal{H}^p}$. Hence we may assume that $(X, Y, \phi) = (X', Y', \phi')$, take any Z in $\mathsf{Z}_{\mathcal{H}^p} = \mathsf{Z}'_{\mathcal{H}^p}$, and take any pairs

$$(\varphi_{-2} : \mathrm{Gr}^{\mathsf{Z}}_{-2} \xrightarrow{\sim} \mathrm{Hom}_{\hat{\mathbb{Z}}}(X \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \hat{\mathbb{Z}}(1)), \varphi_0 : \mathrm{Gr}^{\mathsf{Z}}_0 \xrightarrow{\sim} Y \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})$$

and

$$(\varphi'_{-2} : \mathrm{Gr}^{\mathsf{Z}}_{-2} \xrightarrow{\sim} \mathrm{Hom}_{\hat{\mathbb{Z}}}(X \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \hat{\mathbb{Z}}(1)), \varphi'_0 : \mathrm{Gr}^{\mathsf{Z}}_0 \xrightarrow{\sim} Y \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})$$

inducing $(\varphi_{-2,\mathcal{H}}, \varphi_{0,\mathcal{H}})$ and $(\varphi'_{-2,\mathcal{H}}, \varphi'_{0,\mathcal{H}})$, respectively, and inducing the same $(\varphi_{-2,\mathcal{H}^p}, \varphi_{0,\mathcal{H}^p})$ and $(\varphi'_{-2,\mathcal{H}^p}, \varphi'_{0,\mathcal{H}^p})$. Then the injectivity of (A.4.2) follows from that of (A.4.1). □

Lemma A.4.8. *If* $(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})$ *is assigned to* $(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$ *under* (A.4.1), *then we have a canonical isomorphism*

$$\Gamma_{\Phi_{\mathcal{H}}} \xrightarrow{\sim} \Gamma_{\Phi_{\mathcal{H}^p}} \tag{A.4.9}$$

(see [KWL, Definition 6.2.4.1]). *Moreover, we have a canonical isomorphism*

$$\boldsymbol{S}_{\Phi_{\mathcal{H}^p}} \xrightarrow{\sim} \boldsymbol{S}_{\Phi_{\mathcal{H}}}, \tag{A.4.10}$$

which induces a canonical isomorphism

$$(\boldsymbol{S}_{\Phi_{\mathcal{H}}})^{\vee}_{\mathbb{R}} \xrightarrow{\sim} (\boldsymbol{S}_{\Phi_{\mathcal{H}^p}})^{\vee}_{\mathbb{R}} \tag{A.4.11}$$

matching $\boldsymbol{P}_{\Phi_{\mathcal{H}}}$ *with* $\boldsymbol{P}_{\Phi_{\mathcal{H}^p}}$ *and* $\boldsymbol{P}^+_{\Phi_{\mathcal{H}}}$ *with* $\boldsymbol{P}^+_{\Phi_{\mathcal{H}^p}}$, *both isomorphisms being equivariant with the actions of the two sides of* (A.4.9) *above.*

Proof. Since p is a good prime, with $\mathcal{H}_p = \mathrm{G}(\mathbb{Z}_p)$, the levels at p are not needed in the constructions of $\Gamma_{\Phi_{\mathcal{H}}}$ and $\boldsymbol{S}_{\Phi_{\mathcal{H}}}$ in [KWL, Sections 6.2.3–6.2.4], and hence we have the desired isomorphisms (A.4.9) and (A.4.10). The induced morphism (A.4.11) matches $\boldsymbol{P}_{\Phi_{\mathcal{H}}}$ with $\boldsymbol{P}_{\Phi_{\mathcal{H}^p}}$ and $\boldsymbol{P}^+_{\Phi_{\mathcal{H}}}$ with $\boldsymbol{P}^+_{\Phi_{\mathcal{H}^p}}$ because both sides of (A.4.11) can be canonically identified with the space of Hermitian forms over $Y \otimes_{\mathbb{Z}} \mathbb{R}$, as explained in the beginning of [KWL, Section 6.2.5], regardless of the levels $\mathcal{H}$ and $\mathcal{H}^p$. □

Therefore, we also have assignments

$$(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma) \mapsto (\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p) \tag{A.4.12}$$

and

$$[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)] \mapsto [(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)] \tag{A.4.13}$$

(see [KWL, Definition 6.2.6.2]), which are compatible with (A.4.1) and (A.4.2). Here we have suppressed $\mathsf{Z}_{\mathcal{H}}$ and $\mathsf{Z}_{\mathcal{H}^p}$ from the notation; also, $\sigma \subset (\boldsymbol{S}_{\Phi_{\mathcal{H}}})^\vee_{\mathbb{R}}$ and $\sigma^p \subset (\boldsymbol{S}_{\Phi_{\mathcal{H}^p}})^\vee_{\mathbb{R}}$ is the image of σ under the isomorphism (A.4.11).

Lemma A.4.14. *The assignment* (A.4.12) *is bijective.*

Proof. This follows from Lemma A.4.6 and the definition of (A.4.12), based on Lemma A.4.8. □

Lemma A.4.15. *The assignment* (A.4.13) *is bijective.*

Proof. By [KWL, Definition 6.2.6.2], given any representative $(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$ of a cusp label, the collection of the cones $\sigma \subset (\boldsymbol{S}_{\Phi_{\mathcal{H}}})^\vee_{\mathbb{R}}$ defining the same equivalence class $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]$ form a $\Gamma_{\Phi_{\mathcal{H}}}$-orbit. Similarly, the collection of the cones $\sigma^p \subset (\boldsymbol{S}_{\Phi_{\mathcal{H}^p}})^\vee_{\mathbb{R}}$ defining the same equivalence class $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]$ form a $\Gamma_{\Phi_{\mathcal{H}^p}}$-orbit. Hence, given (A.4.9), the lemma follows from Lemma A.4.7. □

Definition A.4.16. Σ is *induced* by Σ^p if, for each cusp label $[(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ of $\mathsf{M}_{\mathcal{H}}$ represented by some $(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})$, with assigned $(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})$ as in (A.4.1), the cone decomposition $\Sigma_{\Phi_{\mathcal{H}}}$ of $\boldsymbol{P}_{\Phi_{\mathcal{H}}}$ is the pullback of the cone decomposition $\Sigma_{\Phi_{\mathcal{H}^p}}$ of $\boldsymbol{P}_{\Phi_{\mathcal{H}^p}}$ under (A.4.11).

By forgetting the p-parts of level structures, we obtain a canonical isomorphism

$$\mathsf{M}_{\mathcal{H}} \xrightarrow{\sim} \mathsf{M}_{\mathcal{H}^p} \otimes_{\mathbb{Z}} \mathbb{Q} \tag{A.4.17}$$

over $\mathsf{S}_{0,\mathbb{Q}}$ (as in [KWL, 1.4.4.1]), by [KWL, Proposition 1.4.4.3 and Remark 1.4.4.4] and by Assumption A.2.1. Given any Σ^p for $\mathsf{M}_{\mathcal{H}^p}$, with induced Σ for $\mathsf{M}_{\mathcal{H}}$ as in Definition A.4.16, by comparing the universal properties of $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma}$ and $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p}$ as in [KWL, Theorem 6.4.1.1(5)–(6)], the isomorphism (A.4.17) above extends to a canonical isomorphism

$$\mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma} \xrightarrow{\sim} \mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p} \otimes_{\mathbb{Z}} \mathbb{Q} \tag{A.4.18}$$

over $\mathsf{S}_{0,\mathbb{Q}}$, mapping $\mathsf{Z}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}},\sigma)]}$ isomorphically to $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p},\sigma^p)]} \otimes_{\mathbb{Z}} \mathbb{Q}$ when $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p}, \sigma^p)]$ is assigned to $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]$ under (A.4.13), such that the pullback of the tautological semiabelian scheme over $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p} \otimes_{\mathbb{Z}} \mathbb{Q}$ is canonically isomorphic to the pullback of the tautological semiabelian scheme over $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma}$. Consequently, by [KWL, Theorem 7.2.4.1(3)–(4)] and the fact that the pullback of the Hodge invertible sheaf over $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p} \otimes_{\mathbb{Z}} \mathbb{Q}$ is canonically isomorphic to the pullback of the Hodge invertible sheaf over $\mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma}$ (because their definitions only use the tautological semiabelian schemes), the canonical isomorphism (A.4.18) induces a canonical isomorphism

$$\mathsf{M}^{\mathrm{min}}_{\mathcal{H}} \xrightarrow{\sim} \mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p} \otimes_{\mathbb{Z}} \mathbb{Q} \tag{A.4.19}$$

over $\mathsf{S}_{0,\mathbb{Q}}$, extending (A.4.17), compatible with (A.4.18) (under the canonical morphisms $\oint_{\mathcal{H}} : \mathsf{M}^{\mathrm{tor}}_{\mathcal{H},\Sigma} \to \mathsf{M}^{\mathrm{min}}_{\mathcal{H}}$ and $\oint_{\mathcal{H}^p} \otimes_{\mathbb{Z}} \mathbb{Q} : \mathsf{M}^{\mathrm{tor}}_{\mathcal{H}^p,\Sigma^p} \otimes_{\mathbb{Z}} \mathbb{Q} \to \mathsf{M}^{\mathrm{min}}_{\mathcal{H}^p} \otimes_{\mathbb{Z}} \mathbb{Q}$), and mapping $\mathsf{Z}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ isomorphically to $\mathsf{Z}_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]} \otimes_{\mathbb{Z}} \mathbb{Q}$ when $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$ is assigned to $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ under (A.4.2) (where we have suppressed $\mathsf{Z}_{\mathcal{H}}$ and $\mathsf{Z}_{\mathcal{H}^p}$ from the notation).

A.5. ***Complex analytic construction.*** By Proposition A.3.1, in order to prove Theorem A.2.2 we may and we shall assume that $\mathrm{char}(k(s)) = 0$. Thanks to the isomorphisms (A.4.17) and (A.4.19), we shall identify U with a connected component of $\mathsf{M}_{\mathcal{H}} \otimes_{F_0} k(s)$, $U^{\min}$ with the connected component of $\mathsf{M}_{\mathcal{H}}^{\min} \otimes_{F_0} k(s)$ that is the closure of U, and $U_{[(\Phi_{\mathcal{H}^p},\delta_{\mathcal{H}^p})]}$ with $U_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$, the pullback of the stratum $\mathsf{Z}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ of $\mathsf{M}_{\mathcal{H}}^{\min}$ under the canonical morphism $U^{\min} \to \mathsf{M}_{\mathcal{H}}^{\min}$, when $[(\Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$ is assigned to $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ under (A.4.2).

Now, in characteristic zero we no longer need $\mathcal{H}$ to be of the form $\mathcal{H} = \mathcal{H}^p\mathcal{H}_p$ as in Section A.4. We shall allow $\mathcal{H}$ to be any neat open compact subgroup of $\mathrm{G}(\hat{\mathbb{Z}})$. Then $\mathsf{M}_{\mathcal{H}}$ and $\mathsf{M}_{\mathcal{H}}^{\min}$ are still defined over $\mathsf{M}_{0,\mathbb{Q}} = \mathrm{Spec}(F_0)$, with the stratification on the latter by locally closed subschemes $\mathsf{Z}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ labeled by cusp labels $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$ (see the same references as in Section A.2). For any geometric point $s \to \mathsf{S}_{0,\mathbb{Q}}$ with residue field $k(s)$ and for any connected component U of the fiber $\mathsf{M}_{\mathcal{H}^p} \times_{\mathsf{S}_0} s$, we define $U^{\min}$ to be the closure of U in $\mathsf{M}_{\mathcal{H}}^{\min} \times_{\mathsf{S}_0} s$ and $U_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ to be the pullback of $\mathsf{Z}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ to of $U^{\min}$ for each cusp label $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$. (These are consistent with what we have done before, when the settings overlap.)

Then we have the following analogue of Theorem A.2.2:

Theorem A.5.1. *With the setting as above, every stratum* $U_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ *is nonempty.*

Since $\mathsf{M}_{\mathcal{H}}^{\min}$ is projective over $\mathsf{S}_{0,\mathbb{Q}}$, we may and we shall assume that $k(s) \cong \mathbb{C}$. We shall denote base changes to $\mathbb{C}$ with a subscript, such as $\mathsf{M}_{\mathcal{H},\mathbb{C}} = \mathsf{M}_{\mathcal{H}} \otimes_{F_0} \mathbb{C}$.

Let X denote the $\mathrm{G}(\mathbb{R})$-orbit of h_0, which is a finite disjoint union of Hermitian symmetric domains, and let X_0 denote the connected component of X containing h_0. Let $\mathrm{G}(\mathbb{Q})_0$ denote the finite index subgroup of $\mathrm{G}(\mathbb{Q})$ stabilizing X_0. Let $\mathsf{Sh}_{\mathcal{H}} := \mathrm{G}(\mathbb{Q})\backslash\mathsf{X} \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H}$. By [Lan 2012, Lemma 2.5.1], we have a canonical bijection $\mathrm{G}(\mathbb{Q})_0\backslash\mathsf{X}_0 \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H} \to \mathrm{G}(\mathbb{Q})\backslash\mathsf{X} \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H}$. Let $\{g_i\}_{i\in I}$ be any finite set of elements of $\mathrm{G}(\mathbb{A}^\infty)$ such that $\mathrm{G}(\mathbb{A}^\infty) = \coprod_{i\in I} \mathrm{G}(\mathbb{Q})_0 h_i \mathcal{H}$, which exists because of [Borel 1963, Theorem 5.1] and because $\mathrm{G}(\mathbb{Q})_0$ is of finite index in $\mathrm{G}(\mathbb{Q})$. Then we have

$$\mathsf{Sh}_{\mathcal{H}} = \mathrm{G}(\mathbb{Q})_0\backslash\mathsf{X}_0 \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H} = \coprod_{i\in I} \Gamma^{(g_i)}\backslash\mathsf{X}_0, \tag{A.5.2}$$

where $\Gamma^{(g_i)} := (g_i\mathcal{H}g_i^{-1}) \cap \mathrm{G}(\mathbb{Q})_0$ for each $i \in I$. By applying [Baily and Borel 1966, Theorem 10.11] to each $\Gamma^{(g_i)}\backslash\mathsf{X}_0$, we obtain the minimal compactification $\mathsf{Sh}_{\mathcal{H}}^{\min}$ of $\mathsf{Sh}_{\mathcal{H}}$, which is the complex analytification of a normal projective variety $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{\min}$ over $\mathbb{C}$. Thus, $\mathsf{Sh}_{\mathcal{H}}$ is the analytification of a quasiprojective variety $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}$ (embedded in $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{\min}$).

By [Lan 2012, Lemma 3.1.1], the rational boundary components $\mathsf{X}_{\mathtt{V}}$ of X_0 (see [Baily and Borel 1966, Section 3.5]) correspond to parabolic subgroups of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ stabilizing symplectic filtrations $\mathtt{V}$ on $L \otimes_{\mathbb{Z}} \mathbb{Q}$ with $\mathtt{V}_{-3} = 0 \subset \mathtt{V}_{-2} \subset \mathtt{V}_{-1} = \mathtt{V}_{-2}^{\perp} \subset \mathtt{V}_0 = L \otimes_{\mathbb{Z}} \mathbb{Q}$. Consider the rational boundary components of $\mathsf{X} \times \mathrm{G}(\mathbb{A}^\infty)$ as in [Lan 2012, Definition 3.1.2], which are $\mathrm{G}(\mathbb{Q})$-orbits of pairs $(\mathtt{V}, g)$, where the $\mathtt{V}$ are as above and $g \in \mathrm{G}(\mathbb{A}^\infty)$. Consider the boundary components $\mathrm{G}(\mathbb{Q}) \backslash (\mathrm{G}(\mathbb{Q})\mathsf{X}_{\mathtt{V}}) \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H} = \mathrm{G}(\mathbb{Q})_0 \backslash (\mathrm{G}(\mathbb{Q})_0 \mathsf{X}_{\mathtt{V}}) \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H}$ of $\mathsf{Sh}_{\mathcal{H}} = \mathrm{G}(\mathbb{Q})_0 \backslash \mathsf{X}_0 \times \mathrm{G}(\mathbb{A}^\infty)/\mathcal{H}$. By the construction in [Baily and Borel 1966], each such component defines a nonempty, locally closed subset and meets all connected components of $\mathsf{Sh}_{\mathcal{H}}^{\min}$, corresponding to a nonempty, locally closed subscheme of $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{\min}$, called its $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)\mathcal{H}$-stratum. Thus, we obtain the following:

Proposition A.5.3 (Satake, Baily–Borel). *Each* $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)\mathcal{H}$*-stratum as above meets every connected component of* $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{\min}$.

For each $g \in \mathrm{G}(\mathbb{A}^\infty)$, let $L^{(g)}$ denote the $\mathcal{O}$-lattice in $L \otimes_{\mathbb{Z}} \mathbb{Q}$ such that $L^{(g)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} = g(L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})$ in $L \otimes_{\mathbb{Z}} \mathbb{A}^\infty$. Let $r \in \mathbb{Q}_{>0}^\times$ be the unique element such that $\nu(g) = ru$ for some $u \in \hat{\mathbb{Z}}$, and let $\langle \cdot , \cdot \rangle^{(g)} : L^{(g)} \times L^{(g)} \to \mathbb{Z}(1)$ denote the pairing induced by $r\langle \cdot , \cdot \rangle \otimes_{\mathbb{Z}} \mathbb{Q}$ (see [Lan 2012, Section 2.4], the key point being that $\langle \cdot , \cdot \rangle^{(g)}$ is valued in $\mathbb{Z}(1)$).

Construction A.5.4. As explained in [Lan 2012, Section 3.1], we have an assignment of a fully symplectic admissible filtration $\mathtt{Z}^{(g)}$ on $\mathtt{Z} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ and a torus argument $\Phi^{(g)} = (X^{(g)}, Y^{(g)}, \phi^{(g)}, \varphi_{-2}^{(g)}, \varphi_0^{(g)})$ to $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)$, by setting:

(1) $\mathtt{F}^{(g)} := \{\mathtt{F}_{-i}^{(g)} := \mathtt{V}_{-i} \cap L^{(g)}\}_{i \in \mathbb{Z}}$.

(2) $\mathtt{Z}^{(g)} := \{\mathtt{Z}_{-i}^{(g)} := g^{-1}(\mathtt{F}_{-i}^{(g)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})\}_{i \in \mathbb{Z}} = \{g^{-1}(\mathtt{V}_{-i} \otimes_{\mathbb{Q}} \mathbb{A}^\infty) \cap (L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})\}_{i \in \mathbb{Z}}$.

(3) $X^{(g)} := \mathrm{Hom}_{\mathbb{Z}}(\mathtt{F}_{-2}^{(g)}, \mathbb{Z}(1)) = \mathrm{Hom}_{\mathbb{Z}}(\mathrm{Gr}_{-2}^{\mathtt{F}^{(g)}}, \mathbb{Z}(1))$.

(4) $Y^{(g)} := \mathrm{Gr}_0^{\mathtt{F}^{(g)}} = \mathtt{F}_0^{(g)}/\mathtt{F}_{-1}^{(g)}$.

(5) $\phi^{(g)} : Y^{(g)} \hookrightarrow X^{(g)}$, equivalent to the nondegenerate pairing

$$\langle \cdot , \cdot \rangle_{20}^{(g)} : \mathrm{Gr}_{-2}^{\mathtt{F}^{(g)}} \times \mathrm{Gr}_0^{\mathtt{F}^{(g)}} \to \mathbb{Z}(1)$$

induced by $\langle \cdot , \cdot \rangle^{(g)} : L^{(g)} \times L^{(g)} \to \mathbb{Z}(1)$.

(6) $\varphi_{-2}^{(g)} : \mathrm{Gr}_{-2}^{\mathtt{Z}^{(g)}} \xrightarrow{\sim} \mathrm{Hom}_{\hat{\mathbb{Z}}}(X^{(g)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \hat{\mathbb{Z}}(1))$, the composition

$$\mathrm{Gr}_{-2}^{\mathtt{Z}^{(g)}} \xrightarrow[\sim]{\mathrm{Gr}_{-2}(g)} \mathrm{Gr}_{-2}^{\mathtt{F}^{(g)}} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \xrightarrow{\sim} \mathrm{Hom}_{\hat{\mathbb{Z}}}(X^{(g)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \hat{\mathbb{Z}}(1)).$$

(7) $\varphi_0^{(g)} : \mathrm{Gr}_0^{\mathtt{Z}^{(g)}} \xrightarrow{\sim} Y^{(g)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$, the composition

$$\mathrm{Gr}_0^{\mathtt{Z}^{(g)}} \xrightarrow[\sim]{\mathrm{Gr}_0(g)} \mathrm{Gr}_0^{\mathtt{F}^{(g)}} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \xrightarrow{\sim} Y^{(g)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}.$$

By the assumption that our integral PEL datum satisfies [Lan 2013a, Condition 1.4.3.10] and by the fact that maximal orders over Dedekind domains are *hereditary* (see [Reiner 1975, Theorem 21.4 and Corollary 21.5]), there exists a splitting $\varepsilon^{(g)} : \mathrm{Gr}^{\mathtt{F}^{(g)}} \xrightarrow{\sim} L^{(g)}$, whose base extension from $\mathbb{Z}$ to $\hat{\mathbb{Z}}$ defines, by pre- and post-compositions with $\mathrm{Gr}(g)$ and g^{-1}, a splitting $\delta^{(g)} : \mathrm{Gr}^{\mathtt{Z}^{(g)}} \xrightarrow{\sim} L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$. These define an assignment

$$\mathrm{G}(\mathbb{Q})(\mathtt{V}, g) \mapsto [(\mathtt{Z}^{(g)}, \Phi^{(g)}, \delta^{(g)})], \tag{A.5.5}$$

which is compatible with the formation of $\mathcal{H}$-orbits and induces an assignment

$$\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)\mathcal{H} \mapsto [(\mathtt{Z}^{(g)}_{\mathcal{H}}, \Phi^{(g)}_{\mathcal{H}}, \delta^{(g)}_{\mathcal{H}})]. \tag{A.5.6}$$

Definition A.5.7. For each cusp label $[(\mathtt{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$, the $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$-stratum of $\mathsf{Sh}^{\min}_{\mathcal{H},\mathrm{alg}}$ is the union of all the $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)\mathcal{H}$-strata such that $[(\mathtt{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ is assigned to $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)\mathcal{H}$ under (A.5.6).

Proposition A.5.8. *Given the $\mathcal{H}$-orbit $\mathtt{Z}_{\mathcal{H}}$ of any $\mathtt{Z} = \{\mathtt{Z}_{-i}\}_{i\in\mathbb{Z}}$ as above, there exists some totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$-submodule $\mathtt{V}_{-2}$ of $L \otimes_{\mathbb{Z}} \mathbb{Q}$ such that $\mathtt{V}_{-2} \otimes_{\mathbb{Q}} \mathbb{A}^\infty$ lies in the $\mathcal{H}$-orbit of $\mathtt{Z}_{-2} \otimes_{\mathbb{Z}} \mathbb{Q}$.*

Proof. Up to replacing $\mathcal{H}$ with an open compact subgroup, which is harmless for proving this proposition, we may and we shall assume that $\mathcal{H} = \mathcal{H}^S\mathcal{H}_S$, where S is a finite set of primes containing all bad ones for the integral PEL datum (see [KWL, Definition 1.4.1.1]), such that $\mathcal{H}^S = \mathrm{G}(\hat{\mathbb{Z}}^S) = \prod_{\ell\notin S} \mathrm{G}(\mathbb{Z}_\ell)$ and $\mathcal{H}_S \subset \mathrm{G}(\hat{\mathbb{Z}}_S) = \prod_{\ell\in S} \mathrm{G}(\mathbb{Z}_\ell)$, where $\ell \notin S$ means that ℓ runs through all prime numbers not in S.

By Assumption A.2.1, by reduction to the case where $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ is a product of division algebras, by Morita equivalence (see [KWL, Proposition 1.2.1.14]) and, by the local-global principle for isotropy in [Scharlau 1985, table on p. 347 and its references], it follows that, if $\mathtt{Z}_{-2} \otimes_{\mathbb{Z}} \mathbb{Q}$ is nonzero and extends to some isotropic $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{A}$-submodule of $L \otimes_{\mathbb{Z}} \mathbb{A}$ isomorphic to the base extension of some $\mathcal{O}$-lattice, then there exists some nonzero isotropic element in $L \otimes_{\mathbb{Z}} \mathbb{Q}$. By induction on the $\mathcal{O}$-multirank of $\mathtt{Z}_{-2} \otimes_{\mathbb{Z}} \mathbb{Q}$ — by replacing $L \otimes_{\mathbb{Z}} \mathbb{Q}$ (resp. $L \otimes_{\mathbb{Z}} \mathbb{A}^\infty$) with the orthogonal complement modulo the span of a nonzero isotropic element in $L \otimes_{\mathbb{Z}} \mathbb{Q}$ (resp. $L \otimes_{\mathbb{Z}} \mathbb{A}^\infty$) — there exists some totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$-submodule $\mathtt{V}^0_{-2}$ of $L \otimes_{\mathbb{Z}} \mathbb{Q}$ such that $\mathtt{V}^0_{-2} \otimes_{\mathbb{Q}} \mathbb{A}^\infty$ and $\mathtt{Z}_{-2} \otimes_{\mathbb{Z}} \mathbb{Q}$ have the same $\mathcal{O}$-multirank.

Let G' denote the derived subgroup of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ (see [SGA 3_{I} 1970, Définition 7.2(vii), p. 364 and Corollaire 7.10, p. 373]). Then the pullback to G' induces a bijection between the parabolic subgroups of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ and those of G' (see [SGA 3_{III} 1970, Propositions 6.2.4 and 6.2.8, pp. 264–266; Springer 1998, Theorem 15.1.2(ii) and Theorem 15.4.6(i)]), and they both are in bijection with the stabilizers of flags of totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$-submodules, as in Lemma A.4.3. Therefore, there exists some element $h = (h_\ell) \in \mathrm{G}'(\mathbb{A}^\infty)$, where the index ℓ runs through all prime numbers, such that $\mathtt{V}^0_{-2} \otimes_{\mathbb{Q}} \mathbb{A}^\infty = h(\mathtt{Z}_{-2} \otimes_{\mathbb{Z}} \mathbb{Q})$.

Since G' is simply connected, by Assumption A.2.1 (because the kernel of the similitude character of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ factorizes over an algebraic closure of $\mathbb{Q}$ as a product of groups with simply connected derived groups, by the proof of [KWL, Proposition 1.2.3.11]), by weak approximation (see [Platonov and Rapinchuk 1994, Theorem 7.8]) there exists $\gamma \in \mathrm{G}'(\mathbb{Q})$ such that $\gamma(h_\ell)_{\ell \in S} \in \mathcal{H}_S$. On the other hand, by using the Iwasawa decomposition at the places $\ell \in S$ as in the proof of Lemma A.4.4, up to replacing h_ℓ with a right-multiple of h_ℓ by an element of $\mathrm{G}'(\mathbb{Q}_\ell)$ stabilizing $\mathtt{Z}_{-2} \otimes_{\hat{\mathbb{Z}}} \mathbb{Q}_\ell$, we may assume that $\gamma h_\ell \in \mathrm{G}(\mathbb{Z}_\ell)$ for all $\ell \notin S$. Thus, we can conclude by taking $\mathtt{V}_{-2} := \gamma(\mathtt{V}^0_{-2})$. □

Proposition A.5.9. *For each cusp label* $[(\mathtt{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$, *there exists some rational boundary component* $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)$ *of* $\mathsf{X} \times \mathrm{G}(\mathbb{A}^\infty)$ *such that* $[(\mathtt{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ *is assigned to* $\mathrm{G}(\mathbb{Q})(\mathtt{V}, g)\mathcal{H}$ *under* (A.5.6).

Proof. Let $(\mathtt{Z}, \Phi = (X, Y, \phi, \varphi_{-2}, \varphi_0), \delta)$ be any triple whose $\mathcal{H}$-orbit induces $[(\mathtt{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ and let $\mathtt{V}_{-2}$ be as in Proposition A.5.8. Up to replacing $(\mathtt{Z}, \Phi, \delta)$ with another such triple, we may and we shall assume that

$$\mathtt{Z}_{-2} = (\mathtt{V}_{-2} \otimes_{\mathbb{Q}} \mathbb{A}^\infty) \cap (L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}) = \mathtt{Z}^{(1)}_{-2}, \tag{A.5.10}$$

where $\mathtt{F}^{(1)} = \{\mathtt{F}^{(1)}_{-i}\}_{i \in \mathbb{Z}}$, $\mathtt{Z}^{(1)} = \{\mathtt{Z}^{(1)}_{-i}\}_{i \in \mathbb{Z}}$ and $\Phi^{(1)} = (X^{(1)}, Y^{(1)}, \phi^{(1)}, \varphi^{(1)}_{-2}, \varphi^{(1)}_0)$ are assigned to $(\mathtt{V}, 1)$ as in Construction A.5.4, together with some noncanonical choices of $\varepsilon^{(1)}$ and $\delta^{(1)}$.

Let P denote the parabolic subgroup of $\mathrm{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ stabilizing $\mathtt{V}_{-2}$ (see Lemma A.4.3). By (A.5.10), the elements of $\mathrm{P}(\mathbb{A}^\infty)$ also stabilize $\mathtt{Z}_{-2} \otimes_{\mathbb{Z}} \mathbb{Q}$. Therefore, for each $g \in \mathrm{P}(\mathbb{A}^\infty)$, the filtration $\mathtt{Z}^{(g)}$ defined as in Construction A.5.4 coincides with $\mathtt{Z}$.

Using (A.5.10) and the compatibility among the objects, both $\phi \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ and $\phi^{(1)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ can be identified (under $(\varphi_{-2}, \varphi_0)$ and $(\varphi^{(1)}_{-2}, \varphi^{(1)}_0)$) with the canonical morphism

$$\langle \cdot, \cdot \rangle^*_{20} : \mathrm{Gr}^{\mathtt{Z}}_0 \to \mathrm{Hom}_{\hat{\mathbb{Z}}}(\mathrm{Gr}^{\mathtt{Z}}_{-2}, \hat{\mathbb{Z}}(1)) \tag{A.5.11}$$

induced by the pairing $\langle \cdot, \cdot \rangle$, which induce compatible isomorphisms

$${}^t(\varphi^{(1)}_{-2} \circ \varphi^{-1}_{-2}) : X^{(1)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \xrightarrow{\sim} X \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \tag{A.5.12}$$

and

$$\varphi^{(1)}_0 \circ \varphi^{-1}_0 : Y \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \xrightarrow{\sim} Y^{(1)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}. \tag{A.5.13}$$

By [KWL, Condition 1.4.3.10], there exists some maximal order $\mathcal{O}'$ in $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$, containing $\mathcal{O}$, such that the $\mathcal{O}$-action on L extends to an $\mathcal{O}'$-action; hence the $\mathcal{O}$-actions on Y and $Y^{(1)}$ also extend to $\mathcal{O}'$-actions. Using the local isomorphisms given by (A.5.13), by [Reiner 1975, Theorem 18.10] (which is applicable because we are now considering modules of the maximal order $\mathcal{O}'$) and [KWL, Corollary 1.1.2.6] there exists an element $g_0 \in \mathrm{GL}_{\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{A}^\infty}(\mathrm{Gr}^{\mathtt{Z}}_0 \otimes_{\mathbb{Z}} \mathbb{Q})$ and an $\mathcal{O}$-equivariant embedding $h_0 : Y^{(1)} \hookrightarrow Y \otimes_{\mathbb{Z}} \mathbb{Q}$ such that $(h_0(Y^{(1)})) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} = (\varphi_0 \otimes_{\mathbb{Z}} \mathbb{Q})(g_0(\mathrm{Gr}_0^{\mathrm{Z}}))$ in $Y \otimes_{\mathbb{Z}} \mathbb{A}^\infty$. Let $g_{-2} := {}^t g_0^{-1} \in \mathrm{GL}_{\mathcal{O}\otimes_{\mathbb{Z}}\mathbb{A}^\infty}(\mathrm{Gr}_{-2}^{\mathrm{Z}} \otimes_{\mathbb{Z}} \mathbb{Q})$, where the transposition is induced by (A.5.11). Then there is a corresponding $\mathcal{O}$-equivariant embedding $h_{-2} : \mathrm{Hom}_{\mathbb{Z}}(X^{(1)}, \mathbb{Z}(1)) \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(X, \mathbb{Z}(1)) \otimes_{\mathbb{Z}} \mathbb{Q}$ such that

$$\big(h_{-2}\big(\mathrm{Hom}_{\mathbb{Z}}(X^{(1)}, \mathbb{Z}(1))\big)\big) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} = (\varphi_{-2} \otimes_{\mathbb{Z}} \mathbb{Q})(g_{-2}(\mathrm{Gr}_{-2}^{\mathrm{Z}}))$$

in $\mathrm{Hom}_{\mathbb{Z}}(X, \mathbb{Z}(1)) \otimes_{\mathbb{Z}} \mathbb{A}^\infty$.

Take $g \in \mathrm{P}(\mathbb{A}^\infty)$ such that $\mathrm{Gr}_{-2}(g) = g_{-2}$, $\mathrm{Gr}_0(g) = g_0$, and $\nu(g) = 1$, which exists thanks to the splitting δ. Then $X^{(g)}$ and $Y^{(g)}$ are realized as the preimages of X and Y under ${}^t h_{-2} \otimes_{\mathbb{Z}} \mathbb{Q}$ and $h_0^{-1} \otimes_{\mathbb{Z}} \mathbb{Q}$, respectively, and the induced pair $(\gamma_X : X^{(g)} \xrightarrow{\sim} X, \gamma_Y : Y \xrightarrow{\sim} Y^{(g)})$ matches $\Phi^{(g)}$ with Φ. Such a $(\mathtt{V}, g)$ is what we want. □

As explained in [Lan 2012, Section 2.5], there is a canonical open and closed immersion

$$\mathsf{Sh}_{\mathcal{H},\mathrm{alg}} \hookrightarrow \mathsf{M}_{\mathcal{H},\mathbb{C}}. \tag{A.5.14}$$

As explained in [Kottwitz 1992, §8, p. 399] (see also [KWL, Remark 1.4.3.12]), $\mathsf{M}_{\mathcal{H},\mathbb{C}}$ is the disjoint union of the images of morphisms like (A.5.14), from certain $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{(j)}$ defined by some $(\mathcal{O}, \star, L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}, h_0)$ such that $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \cong (L, \langle\cdot,\cdot\rangle) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ and $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}) \otimes_{\mathbb{Z}} \mathbb{R} \cong (L, \langle\cdot,\cdot\rangle) \otimes_{\mathbb{Z}} \mathbb{R}$, but not necessarily satisfying $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}) \otimes_{\mathbb{Z}} \mathbb{Q} \cong (L, \langle\cdot,\cdot\rangle) \otimes_{\mathbb{Z}} \mathbb{Q}$, for all j in some index set J (whose precise description is not important for our purpose). (Each $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)})$ is determined by its rational version $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}) \otimes_{\mathbb{Z}} \mathbb{Q}$ by taking the intersection of the latter with $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \cong (L, \langle\cdot,\cdot\rangle) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ in $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)}) \otimes_{\mathbb{Z}} \mathbb{A}^\infty \cong (L, \langle\cdot,\cdot\rangle) \otimes_{\mathbb{Z}} \mathbb{A}^\infty$. Due to the failure of Hasse's principle, J might have more than one element.)

By [Lan 2012, Theorem 5.1.1], (A.5.14) extends to a canonical open and closed immersion

$$\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{\min} \hookrightarrow \mathsf{M}_{\mathcal{H},\mathbb{C}}^{\min} \tag{A.5.15}$$

respecting the stratifications on both sides labeled by cusp labels (see Definition A.5.7). Again, $\mathsf{M}_{\mathcal{H},\mathbb{C}}^{\min}$ is the disjoint union of the images of morphisms like (A.5.15), from the minimal compactifications $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{(j),\min}$ of $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{(j)}$ for all $j \in J$.

Everything we have proved remains true after replacing the objects defined by $(L, \langle\cdot,\cdot\rangle)$ with those defined by $(L^{(j)}, \langle\cdot,\cdot\rangle^{(j)})$ for each $j \in J$. Thus, in order to show that $U_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ is nonempty, it suffices to note that, by Propositions A.5.3 and A.5.9, the $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$-stratum of $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{(j),\min}$ meets every connected component of $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}^{(j),\min}$ for all $j \in J$. The proof of Theorem A.5.1 is now complete.

By Proposition A.3.1, and by the explanations in Section A.4 and in the beginning of this section, the proof of Theorem A.2.2 is also complete.

A.6. *Extension to cases of ramified characteristics.* In this section, we shall no longer assume that p is a good prime for the integral PEL datum $(\mathcal{O}, \star, L, \langle \cdot, \cdot \rangle, h_0)$, but we shall assume that the image $\mathcal{H}^p$ of $\mathcal{H}$ under the canonical homomorphism $\mathrm{G}(\hat{\mathbb{Z}}) \to \mathrm{G}(\hat{\mathbb{Z}}^p)$ is *neat*.

Even for such general $\mathcal{H}$ and p, for any collections of lattices stabilized by $\mathcal{H}$ as in [Lan 2014, Section 2] we still have an integral model $\vec{\mathsf{M}}_{\mathcal{H}}$ of $\mathsf{M}_{\mathcal{H}}$ that is flat over S_0, constructed by "taking normalization" (see [Lan 2014, Proposition 6.1 and also the introduction]). Moreover, we have an integral model $\vec{\mathsf{M}}_{\mathcal{H}}^{\min}$ of $\mathsf{M}_{\mathcal{H}}^{\min}$ that is projective and flat over S_0 (see [Lan 2014, Proposition 6.4]), with a stratification by locally closed subschemes $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ labeled by cusp labels $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$, which extends the stratification of $\mathsf{M}_{\mathcal{H}}$ by the locally closed subschemes $\mathsf{Z}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ (see [Lan 2014, Theorem 12.1]). For certain (possibly nonsmooth) compatible collections Σ (not the same ones for which we can construct $\mathsf{M}_{\mathcal{H},\Sigma}^{\mathrm{tor}}$ over $\mathsf{M}_{0,\mathbb{Q}}$), we also have the toroidal compactifications $\vec{\mathsf{M}}_{\mathcal{H},\Sigma}^{\mathrm{tor}}$ of $\vec{\mathsf{M}}_{\mathcal{H}}$ that are projective and flat over S_0 (see [Lan 2014, Section 7]), with a stratification by locally closed subschemes $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]}$ (see [Lan 2014, Theorem 9.13]) and a canonical surjection $\vec{\oint}_{\mathcal{H}} : \vec{\mathsf{M}}_{\mathcal{H},\Sigma}^{\mathrm{tor}} \to \vec{\mathsf{M}}_{\mathcal{H}}^{\min}$ with geometrically connected fibers (see [Lan 2014, Lemma 12.9 and its proof]), inducing surjections $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]} \to \vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ (see [Lan 2014, Theorem 12.16]).

As in Section A.2, consider a geometric point $s \to \mathsf{S}_0 = \mathrm{Spec}(\mathcal{O}_{F_0,(p)})$ with algebraically closed residue field $k(s)$ and consider a connected component $U^{\min}$ of the fiber $\vec{\mathsf{M}}_{\mathcal{H}}^{\min} \times_{\mathsf{S}_0} s$. For each cusp label $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$, we define $U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ to be the pullback of $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ to $U^{\min}$. Since the fibers of $\vec{\oint}_{\mathcal{H}}$ are geometrically connected, the preimage of $U^{\min}$ under $\vec{\oint}_{\mathcal{H}} \times_{\mathsf{S}_0} s$ is a connected component U^{tor} of $\vec{\mathsf{M}}_{\mathcal{H},\Sigma}^{\mathrm{tor}} \times_{\mathsf{S}_0} s$. (In general, neither $\vec{\mathsf{M}}_{\mathcal{H}}^{\min} \times_{\mathsf{S}_0} s$ nor $\vec{\mathsf{M}}_{\mathcal{H},\Sigma}^{\mathrm{tor}} \times_{\mathsf{S}_0} s$ is normal.) For each equivalence class $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]$ defining a stratum $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]}$ of $\vec{\mathsf{M}}_{\mathcal{H},\Sigma}^{\mathrm{tor}}$, we define $U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]}$ to be the pullback of $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]}$. Then we also have a canonical surjection $U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]} \to U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ induced by $\vec{\oint}_{\mathcal{H}}$.

Theorem A.6.1. *With the setting as above, all strata of $U^{\min}$ are nonempty.*

By using the canonical surjection $U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}}, \sigma)]} \to U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]}$ (as in the proof of Corollary A.2.3), Theorem A.6.1 implies the following:

Corollary A.6.2. *With the setting as above, all strata of U^{tor} are nonempty.*

As in Section A.3, it suffices to prove the following:

Proposition A.6.3. *Suppose Theorem A.6.1 is true when* $\mathrm{char}(k(s)) = 0$. *Then it is also true when* $\mathrm{char}(k(s)) = p > 0$.

Remark A.6.4. Since $\vec{\mathsf{M}}_{\mathcal{H}} \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathsf{M}_{\mathcal{H}}$ and $\vec{\mathsf{M}}_{\mathcal{H}}^{\min} \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathsf{M}_{\mathcal{H}}^{\min}$ by construction, by Theorem A.5.1 the assumption in Proposition A.6.3 always holds. Nevertheless, the proof of Proposition A.6.3 will clarify that the deduction of Theorem A.6.1 from Theorem A.5.1 does not require Assumption A.2.1 (cf. Remark A.3.2).

The remainder of this section will be devoted to the proof of Proposition A.6.3. We shall assume that $\mathrm{char}(k(s)) = p > 0$.

While each $\mathsf{Z}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ is isomorphic to some boundary moduli problem $\mathsf{M}_{\mathcal{H}}^{\mathsf{Z}_{\mathcal{H}}}$, each stratum $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ of $\vec{\mathsf{M}}_{\mathcal{H}}^{\min}$ is similarly isomorphic to some integral model $\vec{\mathsf{M}}_{\mathcal{H}}^{\mathsf{Z}_{\mathcal{H}}}$ defined by taking normalization (see [Lan 2014, Proposition 7.4 and Theorems 12.1 and 12.16]). Hence it also makes sense to consider the minimal compactification $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}^{\min}$ of $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$, which is proper flat (with possibly nonnormal geometric fibers) over S_0, and we obtain the following:

Lemma A.6.5 (cf. Lemma A.3.4 and [Deligne and Mumford 1969, Theorem 4.17(ii)]). *There exists some discrete valuation ring R that is flat over $\mathcal{O}_{F_0,(p)}$, with fraction field K and residue field $k(s)$, the latter lifting the structural homomorphism $\mathcal{O}_{F_0,(p)} \to k(s)$, such that, for each cusp label $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ and each connected component V of $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}^{\min} \otimes_{\mathcal{O}_{F_0,(p)}} R$, the induced flat morphism $V \to \mathrm{Spec}(R)$ has connected special fiber over $\mathrm{Spec}(k(s))$.*

Proof of Proposition A.6.3. By [Lan 2014, Corollary 12.4], it suffices to show that $U_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]} \neq \varnothing$ when $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ is *maximal* with respect to the surjection relations, as in [KWL, Definition 5.4.2.13]. In this case, by [Lan 2014, Theorem 12.1], $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}$ is a closed stratum of $\vec{\mathsf{M}}_{\mathcal{H}}^{\min}$ and so $\vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]} = \vec{\mathsf{Z}}_{[(\Phi_{\mathcal{H}},\delta_{\mathcal{H}})]}^{\min}$. Hence the lemma follows from Theorem A.5.1 and the same argument as in the proof of Proposition A.3.1, with the reference to Lemma A.3.4 replaced with an analogous reference to Lemma A.6.5. □

As explained in Remark A.6.4, the proof of Theorem A.6.1 is now complete.

A.7. *Examples.*

Example A.7.1. Suppose $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ is a CM field F with maximal totally subfield F^+, with positive involution given by the complex conjugation of F over F^+. Suppose $L = \mathcal{O}_F^{\oplus a+b}$, where $a \geq b \geq 0$ are integers. Suppose $(2\pi\sqrt{-1})^{-1}\langle\,\cdot\,,\cdot\,\rangle$ is the skew-Hermitian pairing defined in block matrix form

$$\begin{pmatrix} & & 1_b \\ & S & \\ -1_b & & \end{pmatrix},$$

where S is some $(a-b)\times(a-b)$ matrix over F such that $\sqrt{-1}S$ is Hermitian and either positive or negative definite. Then, for each $0 \leq r \leq b$, the $\mathcal{O}$-submodule $\mathsf{Z}_{-2}^{(r)}$ of $L = \mathcal{O}_F^{\oplus(a+b)}$ with the last $a+b-r$ entries zero is totally isotropic, and $\mathsf{V}_{-2}^{(r)} := \mathsf{F}_{-2}^{(r)} \otimes_{\mathbb{Z}} \mathbb{Q}$ is a totally isotropic F-submodule of $L \otimes_{\mathbb{Z}} \mathbb{Q} = F^{\oplus(a+b)}$, which is maximal when $r = b$. The stabilizer of $\mathsf{V}_{-2}^{(r)}$ either is the whole group (when $r = 0$) or defines a maximal (proper) parabolic subgroup $\mathsf{P}^{(r)}$ of $\mathsf{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ (when $r > 0$), and all maximal parabolic subgroups of $\mathsf{G} \otimes_{\mathbb{Z}} \mathbb{Q}$ are conjugate to one of these *standard* ones, by Lemma A.4.3. Similarly, $\mathsf{Z}_{-2}^{(r)} := \mathsf{F}_{-2}^{(r)} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ is a totally isotropic $\mathcal{O} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$-submodule of $L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$, and the left $\mathsf{G}(\mathbb{Q})$- and right $\mathcal{H}$-double orbits of $\mathsf{Z}_{-2}^{(r)}$, for $0 \leq r \leq b$, exhaust all the possible $\mathsf{Z}_{\mathcal{H}}$ appearing in cusp labels $[(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$, by Proposition A.5.8. By Lemma A.4.7, by forgetting their p-parts, their left $\mathsf{G}(\mathbb{Q})$- and right $\mathcal{H}^p$-double orbits also exhaust all the possible $\mathsf{Z}_{\mathcal{H}^p}$ appearing in cusp labels $[(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$ for $\mathsf{M}_{\mathcal{H}^p}$. Let us say that a cusp label $[(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$ is *of rank* r if $\mathsf{Z}_{\mathcal{H}}$ is in the double orbit of $\mathsf{Z}_{-2}^{(r)}$, and that a cusp $[(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$ for $\mathsf{M}_{\mathcal{H}^p}$ is *of rank* r if it is assigned to one of rank r under (A.4.1). (This is consistent with [KWL, Definitions 5.4.1.12 and 5.4.2.7].) On the other hand, as a byproduct of the proof of Proposition A.5.9, any $\mathsf{Z}_{\mathcal{H}}$ in the double orbit of $\mathsf{Z}_{-2}^{(r)}$ does extend to some cusp label $[(\mathsf{Z}_{\mathcal{H}}, \Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$, inducing some cusp label $[(\mathsf{Z}_{\mathcal{H}^p}, \Phi_{\mathcal{H}^p}, \delta_{\mathcal{H}^p})]$ for $\mathsf{M}_{\mathcal{H}^p}$ under (A.4.1). Then Theorem A.2.2 shows that, in the boundary stratification of every connected component of every geometric fiber of $\mathsf{M}_{\mathcal{H}^p}^{\min} \to \mathsf{S}_0 = \mathrm{Spec}(\mathcal{O}_{F_{0,(p)}})$, there exist nonempty strata labeled by cusp labels for $\mathsf{M}_{\mathcal{H}^p}$ of all possible ranks $0 \leq r \leq b$. (The theorem shows the more refined nonemptiness for strata labeled by cusp labels, not just by ranks.)

The next example shows that we cannot expect Theorem A.2.2 to be true without the requirement (in Assumption A.2.1) that $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ involves no factor of type D.

Example A.7.2. Suppose $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ is a central division algebra D over a totally real field F, as in [KWL, Proposition 1.2.1.13] such that $D \otimes_{F,\tau} \mathbb{R} \cong \mathbb{H}$, the real Hamiltonian quaternion algebra, for every embedding $\tau : F \to \mathbb{R}$, with $\star = \diamond$ given by $x \mapsto x^\diamond := \mathrm{Tr}_{D/F}(x) - x$. Suppose that D is nonsplit at strictly more than two places. Suppose L is chosen such that $L \otimes_{\mathbb{Z}} \mathbb{Q} \cong D^{\oplus 2}$. By the Gram–Schmidt process, as in [KWL, Section 1.2.4] and by [KWL, Corollary 1.1.2.6], there is up to isomorphism only one isotropic skew-Hermitian pairing over $L \otimes_{\mathbb{Z}} \mathbb{Q}$. But we do know the failure of Hasse's principle (see [Kottwitz 1992, §7, p. 393]) in this case (see [Scharlau 1985, Remark 10.4.6]), which means there exists a choice of $(L, \langle\,\cdot\,,\,\cdot\,\rangle)$ as above that is globally anisotropic but locally isotropic everywhere. Thus, even when $k(s) \cong \mathbb{C}$, there exists some connected component U of $\mathsf{Sh}_{\mathcal{H},\mathrm{alg}}$ and some nonzero cusp label $[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]$ for $\mathsf{M}_{\mathcal{H}}$ such that $U_{[(\Phi_{\mathcal{H}}, \delta_{\mathcal{H}})]} = \varnothing$.

Acknowledgements

This paper is part of Wan's thesis under the supervision of Professor Christopher Skinner. The author would like to thank him for suggesting the problem and many helpful conversations. Wan thanks Kai-Wen Lan for answering many questions about the compactification of Shimura varieties, and would also like to thank the anonymous referees for carefully reading the paper, giving numerous suggestions on correcting some mistakes and improving the writing.

Lan was partially supported by the National Science Foundation under agreement No. DMS-1352216 and by an Alfred P. Sloan Research Fellowship.

References

[Baily and Borel 1966] W. L. Baily, Jr. and A. Borel, "Compactification of arithmetic quotients of bounded symmetric domains", *Ann. of Math.* (2) **84** (1966), 442–528. MR 35 #6870 Zbl 0154.08602

[Borel 1963] A. Borel, "Some finiteness properties of adele groups over number fields", *Inst. Hautes Études Sci. Publ. Math.* **16** (1963), 5–30. MR 34 #2578 Zbl 0135.08902

[Bruhat and Tits 1972] F. Bruhat and J. Tits, "Groupes réductifs sur un corps local", *Inst. Hautes Études Sci. Publ. Math.* **41** (1972), 5–251. MR 48 #6265 Zbl 0254.14017

[Casselman 1980] W. Casselman, "The unramified principal series of $\mathfrak{p}$-adic groups, I: The spherical function", *Compositio Math.* **40**:3 (1980), 387–406. MR 83a:22018

[Deligne and Mumford 1969] P. Deligne and D. Mumford, "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* **36** (1969), 75–109. MR 41 #6850 Zbl 0181.48803

[EGA III_1 1961] A. Grothendieck, "Eléments de géométrie algébrique, III: Étude cohomologique des faisceaux cohérents, I", *Inst. Hautes Études Sci. Publ. Math.* **11** (1961), 5–167. MR 0163910 Zbl 0118.36206

[Eischen and Wan 2014] E. Eischen and X. Wan, "p-adic L-functions of finite slope forms on unitary groups and Eisenstein series", *J. Math. Inst. Jussieu* (online publication November 2014).

[Eischen et al. $\geq$ 2015] E. Eischen, M. Harris, J. Li, and C. Skinner, "p-adic L-functions for unitary Shimura varieties, II", in preparation.

[Garrett 1984] P. B. Garrett, "Pullbacks of Eisenstein series; applications", pp. 114–137 in *Automorphic forms of several variables* (Katata, 1983), edited by I. Satake and Y. Morita, Progr. Math. **46**, Birkhäuser, Boston, 1984. MR 86f:11039 Zbl 0544.10023

[Garrett 1989] P. B. Garrett, "Integral representations of Eisenstein series and L-functions", pp. 241–264 in *Number theory, trace formulas and discrete groups* (Oslo, 1987), edited by K. E. Aubert et al., Academic Press, Boston, 1989. MR 90e:11073 Zbl 0671.10024

[Gelbart et al. 1987] S. Gelbart, I. Piatetski-Shapiro, and S. Rallis, *Explicit constructions of automorphic L-functions*, Lecture Notes in Mathematics **1254**, Springer, Berlin, 1987. MR 89k:11038 Zbl 0612.10022

[Harris 1984] M. Harris, "Eisenstein series on Shimura varieties", *Ann. of Math.* (2) **119**:1 (1984), 59–94. MR 85j:11052 Zbl 0589.10030

[Hida 2004a] H. Hida, "Non-vanishing modulo p of Hecke L-values", pp. 735–784 in *Geometric aspects of Dwork theory*, vol. I–II, edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR 2006b:11055 Zbl 1162.11342

[Hida 2004b] H. Hida, *p-adic automorphic forms on Shimura varieties*, Springer, New York, 2004. MR 2005e:11054 Zbl 1055.11032

[Hida 2011] H. Hida, "Irreducibility of the Igusa tower over unitary Shimura varieties", pp. 187–203 in *On certain L-functions*, edited by J. Arthur et al., Clay Math. Proc. **13**, Amer. Math. Soc., 2011. MR 2012c:11131 Zbl 1272.11081

[Hsieh 2011] M.-L. Hsieh, "Ordinary p-adic Eisenstein series and p-adic L-functions for unitary groups", *Ann. Inst. Fourier* (*Grenoble*) **61**:3 (2011), 987–1059. MR 2918724 Zbl 1271.11051

[Hsieh 2014] M.-L. Hsieh, "Eisenstein congruence on unitary groups and Iwasawa main conjectures for CM fields", *J. Amer. Math. Soc.* **27**:3 (2014), 753–862. MR 3194494

[Ikeda 1994] T. Ikeda, "On the theory of Jacobi forms and Fourier–Jacobi coefficients of Eisenstein series", *J. Math. Kyoto Univ.* **34**:3 (1994), 615–636. MR 95h:11044 Zbl 0822.11041

[Kottwitz 1992] R. E. Kottwitz, "Points on some Shimura varieties over finite fields", *J. Amer. Math. Soc.* **5**:2 (1992), 373–444. MR 93a:11053 Zbl 0796.14014

[Kudla 1994] S. S. Kudla, "Splitting metaplectic covers of dual reductive pairs", *Israel J. Math.* **87**:1–3 (1994), 361–401. MR 95h:22019 Zbl 0840.22029

[Lan 2008] K.-W. Lan, *Arithmetic compactifications of PEL-type Shimura varieties*, Ph.D. thesis, Harvard University, 2008, http://search.proquest.com/docview/304599957. MR 2711676

[Lan 2011] K.-W. Lan, "Elevators for degenerations of PEL structures", *Math. Res. Lett.* **18**:5 (2011), 889–907. MR 2875862 Zbl 1260.14026

[Lan 2012] K.-W. Lan, "Comparison between analytic and algebraic constructions of toroidal compactifications of PEL-type Shimura varieties", *J. Reine Angew. Math.* **664** (2012), 163–228. MR 2980135 Zbl 1242.14022

[Lan 2013a] K.-W. Lan, *Arithmetic compactifications of PEL-type Shimura varieties*, London Mathematical Society Monographs Series **36**, Princeton University Press, NJ, 2013. MR 3186092 Zbl 1284.14004

[Lan 2013b] K.-W. Lan, "Compactifications of PEL-Type Shimura Varieties and Kuga Families with Ordinary Loci", preprint, 2013, http://www.math.umn.edu/~kwlan/articles/cpt-ram-ord.pdf.

[Lan 2014] K.-W. Lan, "Compactifications of PEL-type Shimura varieties in ramified characteristics", preprint, 2014, http://www.math.umn.edu/~kwlan/articles/cpt-ram-flat.pdf.

[Lapid and Rallis 2005] E. M. Lapid and S. Rallis, "On the local factors of representations of classical groups", pp. 309–359 in *Automorphic representations, L-functions and applications: progress and prospects*, edited by J. W. Cogdell et al., Ohio State Univ. Math. Res. Inst. Publ. **11**, de Gruyter, Berlin, 2005. MR 2006j:11071 Zbl 1188.11023

[Platonov and Rapinchuk 1994] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics **139**, Academic Press, Boston, 1994. MR 95b:11039 Zbl 0841.20046

[Reiner 1975] I. Reiner, *Maximal orders*, London Mathematical Society Monographs **5**, Academic Press, London–New York, 1975. MR 52 #13910 Zbl 0305.16001

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Math. Wissenschaften **270**, Springer, Berlin, 1985. MR 86k:11022 Zbl 0584.10010

[SGA 3_{I} 1970] M. Demazure and A. Grothendieck, *Schémas en groupes, Tome I: Propriétés générales des schémas en groupes, Exposés I–VII* (Séminaire de Géométrie Algébrique du Bois Marie 1962–1964), Lecture Notes in Math. **151**, Springer, Berlin, 1970. MR 43 #223a Zbl 0207.51401

[SGA 3_{III} 1970] M. Demazure and A. Grothendieck, *Schémas en groupes, Tome III: Structure des schémas en groupes réductifs, Exposés XIX–XXVI* (Séminaire de Géométrie Algébrique du Bois Marie 1962–1964), Lecture Notes in Math. **153**, Springer, Berlin, 1970. MR 43 #223c Zbl 0212.52810

[Shimura 1997] G. Shimura, *Euler products and Eisenstein series*, CBMS Regional Conference Series in Mathematics **93**, Amer. Math. Soc., 1997. MR 98h:11057 Zbl 0906.11020

[Skinner 2014] C. Skinner, "A converse of Gross, Zagier and Kolyvagin", preprint, 2014. arXiv 1405.7294

[Skinner and Urban 2014] C. Skinner and E. Urban, "The Iwasawa main conjectures for GL_2", *Invent. Math.* **195**:1 (2014), 1–277. MR 3148103 Zbl 1301.11074

[Springer 1998] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progress in Mathematics **9**, Birkhäuser, Boston, 1998. MR 99h:20075 Zbl 0927.20024

[Wan 2013] X. Wan, "Iwasawa main conjecture for Hilbert modular forms", preprint, 2013, http://www.math.columbia.edu/~xw2295/Hilbert%20Modular%20Forms.pdf.

[Wan 2015] X. Wan, "The Iwasawa main conjecture for Hilbert modular forms", *Forum Math. Sigma* **3** (2015), e18.

Communicated by Shou-Wu Zhang
Received 2014-02-12 Revised 2015-06-27 Accepted 2015-08-18

xw2295@math.columbia.edu *Department of Mathematics, Columbia University, Room 509, MC 4406, 2990 Broadway, New York, NY 10027, United States*

kwlan@math.umn.edu *School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA*

ALGEBRA AND NUMBER THEORY 9:9 (2015)
dx.doi.org/10.2140/ant.2015.9.2055

Classifying orders in the Sklyanin algebra

Daniel Rogalski, Susan J. Sierra and J. Toby Stafford

Let S denote the 3-dimensional Sklyanin algebra over an algebraically closed field $\Bbbk$ and assume that S is not a finite module over its centre. (This algebra corresponds to a generic noncommutative $\mathbb{P}^2$.) Let $A = \bigoplus_{i\geq 0} A_i$ be any connected graded $\Bbbk$-algebra that is contained in and has the same quotient ring as a Veronese ring $S^{(3n)}$. Then we give a reasonably complete description of the structure of A. This is most satisfactory when A is a maximal order, in which case we prove, subject to a minor technical condition, that A is a noncommutative blowup of $S^{(3n)}$ at a (possibly noneffective) divisor on the associated elliptic curve E. It follows that A has surprisingly pleasant properties; for example, it is automatically noetherian, indeed strongly noetherian, and has a dualising complex.

Rogalski was partially supported by NSF grants DMS-0900981 and DMS-1201572. Sierra was partially supported by an NSF Postdoctoral Research Fellowship, grant DMS-0802935. Stafford was a Royal Society Wolfson Research Merit Award Holder.

MSC2010: primary 14A22; secondary 16P40, 16W50, 16S38, 14H52, 16E65, 18E15.

Keywords: noncommutative projective geometry, noncommutative surfaces, Sklyanin algebras, noetherian graded rings, noncommutative blowing-up.

1. Introduction

Noncommutative (projective) algebraic geometry has been very successful in using techniques and intuition from algebraic geometry to study noncommutative graded algebras, and many classes of algebras have been classified using these ideas. In particular, noncommutative irreducible curves (or connected graded domains of Gelfand–Kirillov dimension 2) have been classified [Artin and Stafford 1995] as have large classes of noncommutative irreducible surfaces (or connected graded noetherian domains of Gelfand–Kirillov dimension 3).

Indeed, the starting point of this subject was really the classification by Artin, Tate, and Van den Bergh [Artin et al. 1990; 1991] of noncommutative projective planes (noncommutative analogues of a polynomial ring $\Bbbk[x, y, z]$). The generic example here is the Sklyanin algebra

$$S = \mathrm{Skl}(a, b, c) = \Bbbk\{x_1, x_2, x_3\}/(ax_ix_{i+1} + bx_{i+1}x_i + cx_{i+2}^2 : i \in \mathbb{Z}_3),$$

where $(a, b, c) \in \mathbb{P}^2 \smallsetminus \mathscr{S}$ for a (known) finite set $\mathscr{S}$. The geometric methods of [Artin et al. 1990] were necessary to understand this algebra. See [Stafford and Van den Bergh 2001] for a survey of many of these results.

In the other direction, one would like to classify *all* noncommutative surfaces, and a programme for this has been suggested by Artin [1997]. This paper completes a significant case of this programme by classifying the graded noetherian orders contained in the Sklyanin algebra. In this introduction we will first describe our main results and then discuss the historical background and give an idea of the proofs.

The main results. Fix a Sklyanin algebra $S = \mathrm{Skl}(a, b, c)$ defined over an algebraically closed base field $\Bbbk$. For technical reasons we mostly work inside the 3-Veronese ring $T = S^{(3)}$; thus $T = \bigoplus T_n$ with $T_n = S_{3n}$ for each n, under the natural graded structure of S. The difference between these algebras is not particularly significant; for example, the quotient category qgr-T of graded noetherian right T-modules modulo those of finite length, is equivalent to qgr-S. Then T contains a canonical central element $g \in T_1 = S_3$ such that the factor $B = T/gT$ is a *TCR or twisted homogeneous coordinate ring* $B = B(E, \mathcal{M}, \tau)$ of an elliptic curve E. Here $\mathcal{M}$ is a line bundle of degree 9 and $\tau \in \mathrm{Aut}_{\Bbbk}(E)$ (see Section 2 for the definition). We assume throughout the paper that $|\tau| = \infty$; equivalently, that T is not a finite module over its centre.

Our main results are phrased in terms of certain *blowups* $T(\boldsymbol{d}) \subset T$, where $\boldsymbol{d}$ is a divisor on E. These are discussed in more detail later in this introduction. Here we will just note that, when $\boldsymbol{p}$ is a closed point of E, the blowup $T(\boldsymbol{p})$ is the subring of T generated by those elements $x \in T_1$ whose images in T/gT vanish at $\boldsymbol{p}$. For an effective divisor $\boldsymbol{d}$ (always of degree at most 8), $T(\boldsymbol{d})$ has properties similar to those of a (commutative) anticanonical homogeneous coordinate ring of the blowup of $\mathbb{P}^2$ along the divisor $\boldsymbol{d}$. However, we also need algebras that should be considered as blowups $T(\boldsymbol{d}')$ of T at *noneffective divisors* of the form $\boldsymbol{d}' = \boldsymbol{x} - \boldsymbol{y} + \tau^{-1}(\boldsymbol{y})$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are effective divisors on E, $0 \leq \deg \boldsymbol{d}' \leq 8$ and certain combinatorial conditions hold (see Definition 7.1 for the details). Such a divisor will be called *virtually effective*.

Given domains U, U' with the same Goldie quotient ring $Q(U) = Q(U') = Q$, we say that U and U' are *equivalent orders* if $aUb \subseteq U'$ and $a'U'b' \subseteq U$ for some $a, b, a', b' \in Q \smallsetminus \{0\}$. If $Q(U) = Q(V)$ for some ring $V \supseteq U$, then U is called a *maximal V-order* if there exists no ring U' equivalent to U such that $U \subsetneq U' \subseteq V$. When $V = Q(U)$, U is simply termed a *maximal order*. These can be regarded as the appropriate noncommutative analogues of integrally closed domains. The algebra T is a maximal order. When $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$ the concepts of maximal orders and maximal T-orders are essentially the only cases that will concern us and, as the next result shows, they are closely connected.

In this result, an $\mathbb{N}$-graded $\Bbbk$-algebra $A = \bigoplus_{n \geq 0} A_n$ is called *connected graded* (*cg*) if $A_0 = \Bbbk$ and $\dim_{\Bbbk} A_n < \infty$ for all n. Also, for a cg algebra $U \subseteq T$, we write $\overline{U} = (U + gT)/gT$.

Proposition 1.1 (combine Theorem 8.11 with Proposition 6.4). *Let U be a cg maximal T-order, such that $\overline{U} \neq \Bbbk$. Then there exists a unique maximal order $F = F(U) \supseteq U$ equivalent to U. Moreover, F is a finitely generated U-module with* $\mathrm{GKdim}_U(F/U) \leq 1$.

We remark that there do exist graded maximal T-orders U with $U \neq F(U)$ (see Proposition 10.3).

Our results are most satisfactory for maximal T-orders, and our main result is the following complete classification of such algebras.

Theorem 1.2 (Theorem 8.11). *Let U be a cg maximal T-order with $\overline{U} \neq \Bbbk$. Then there exists a virtually effective divisor $\boldsymbol{d}' = \boldsymbol{d} - \boldsymbol{y} + \tau^{-1}(\boldsymbol{y})$ with* $\deg(\boldsymbol{d}') \leq 8$ *such that the associated maximal order $F(U)$ is a blowup $F(U) = T(\boldsymbol{d}')$ of T at $\boldsymbol{d}'$.*

Remarks 1.3. (1) Although in this introduction we are restricting our attention to the Sklyanin algebra $S = \mathrm{Skl}$, this theorem and indeed all the results of this paper are proved simultaneously for certain related algebras; see Assumption 2.1 and Examples 2.2 for the details.

(2) Theorem 1.2 is actually proved in the context of graded maximal $T^{(n)}$-orders, but as that result is a little more complicated to state, the reader is referred to Theorem 8.11 for the details.

(3) The assumption that $\overline{U} \neq \Bbbk$ in the theorem is annoying but necessary (see Example 10.8). It can be bypassed at the expense of passing to a Veronese ring and

then regrading the algebra. However, the resulting theorems are not as strong as Theorem 1.2 (see Section 9 for the details).

One consequence of Theorem 1.2 is that maximal T-orders have very pleasant properties. The undefined terms in the next result are standard concepts and are defined in the body of the paper.

Corollary 1.4. *Let U and $F = F(U) = T(\boldsymbol{d}')$ be as in Theorem 1.2.*

(1) (Proposition 2.9 and Theorem 8.11(1)) *Both U and F are finitely generated $\Bbbk$-algebras and are strongly noetherian: in other words, $U \otimes_{\Bbbk} C$ and $F \otimes_{\Bbbk} C$ are noetherian for any commutative, noetherian $\Bbbk$-algebra C.*

(2) (Corollary 8.12) *Both U and F satisfy the Artin–Zhang χ conditions, have finite cohomological dimension and possess balanced dualising complexes.*

(3) (Proposition 4.10 and Example 10.4) *If F is the blowup at an effective divisor then $U = F$. In this case F also satisfies the Auslander–Gorenstein and Cohen–Macaulay conditions. These conditions do not necessarily hold when $\boldsymbol{d}'$ is virtually effective.*

In the other direction, we prove:

Theorem 1.5 (Theorem 7.4(3)). *For any virtually effective divisor $\boldsymbol{d}'$ there exists a blowup of T at $\boldsymbol{d}'$ in the sense described above.*

The fact that U is automatically noetherian in Theorem 1.2 is one of the result's most striking features. In general, nonnoetherian graded subalgebras of T can be rather unpleasant and so, in order to classify reasonable classes of nonmaximal orders in T, we make a noetherian hypothesis. Given a connected graded noetherian algebra U, one can easily obtain further noetherian rings by taking Veronese rings, idealiser subrings $\mathbb{I}(J) = \{\theta \in U : \theta J \subseteq J\}$ for a right ideal J of U, or equivalent orders $U' \subseteq U$ containing an ideal K of U. We show that this suffices:

Corollary 1.6 (Corollary 9.5). *Let U be a cg noetherian subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T^{(n)})$ for some n. Assume that $\overline{U} \neq \Bbbk$ (as in Remarks 1.3, this can be assumed at the expense of taking a Veronese ring and regrading).*

Then U can be obtained from some virtual blowup $R = T(\boldsymbol{d}')$ by a combination of Veronese rings, idealisers and equivalent orders $K \subseteq U \subseteq V$, where K is an ideal of V with $\operatorname{GKdim}(V/K) \leq 1$.

History. We briefly explain the history behind these results and their wider relevance. As we mentioned earlier, noncommutative curves and noncommutative analogues of the polynomial ring $\Bbbk[x, y, z]$ have been classified. Motivated by these results, Artin suggested a program for classifying all noncommutative surfaces, but in order to outline this program we need some notation.

Given a cg domain A of finite Gelfand–Kirillov dimension, one can invert the nonzero homogeneous elements to obtain the *graded quotient ring* $Q_{\mathrm{gr}}(A) \cong D[t, t^{-1}; \alpha]$, for some automorphism α of the division ring $D = Q(A)_0 = D_{\mathrm{gr}}(A)$. This division ring will be called the *function skewfield of* A.

Let A be a noetherian, cg $\Bbbk$-algebra. A useful intuition is to regard qgr-A as the coherent sheaves over the (nonexistent) noncommutative projective scheme $\mathrm{Proj}(A)$, although we will slightly abuse notation by regarding qgr-A itself as that scheme. Under this intuition, a noncommutative surface is qgr-A for a noetherian cg domain A with $\mathrm{GKdim}\, A = 3$. (In fact, one should probably weaken this last condition to the assumption that $D_{\mathrm{gr}}(A)$ has lower transcendence degree two in the sense of [Zhang 1998], but that is not really relevant here.) There are strong arguments for saying that noncommutative projective planes are the categories qgr-A, as A ranges over the Artin–Schelter regular rings of dimension 3 with the Hilbert series $(1-t)^{-3}$ of a polynomial ring in three variables (see [Stafford and Van den Bergh 2001, §11.2] for more details). These are the algebras classified in [Artin et al. 1990] and for which the Sklyanin algebra $S = S(a, b, c)$ is the generic example. Van den Bergh [2011; 2012] has similarly classified noncommutative analogues of quadrics and related surfaces.

Artin's Conjectures 1.7. Artin conjectured that the only function skewfields of noncommutative surfaces are the following:

(i) division rings D finite-dimensional over their centres $F = Z(D)$, which are then fields of transcendence degree two;

(ii) division rings of fractions D of Ore extensions $\Bbbk(X)[z; \sigma, \delta]$ for some curve X, where D is not a finite module over its centre; and

(iii) the function skewfield $D = D_{\mathrm{gr}}(S)$ of a Sklyanin algebra $S = S(a, b, c)$, where S is not a finite module over its centre.

Artin then asked for a classification of the noncommutative surfaces qgr-A within each birational class; that is, the cg noetherian algebras A with $D_{\mathrm{gr}}(A)$ being a fixed division ring from this list.

The case of Artin's programme when $D = \Bbbk(Y)$ is the function field of a surface and $\mathrm{GKdim}\, A = 3$ has been completed in [Rogalski and Stafford 2009; Sierra 2011] (if one strays from algebras of Gelfand–Kirillov dimension 3, then things become more complicated, as [Rogalski and Sierra 2012] shows). As explained earlier, in this paper we are interested in the other extreme, that of case (iii) from Artin's list.

The first main results in this direction come from [Rogalski 2011], of which this paper is a continuation. In particular, [ibid., Theorem 1.2] shows that the maximal orders $U \subseteq T = S^{(3)}$ that have $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$ and are generated in degree one are just the blowups $T(\boldsymbol{d})$ for an effective divisor $\boldsymbol{d}$ on E with $\deg(\boldsymbol{d}) \leq 7$. We remark that in this case $T(\boldsymbol{d})$ is simply the subalgebra of T generated by those elements of T_1 whose images in T/gT vanish on $\boldsymbol{d}$. As such, $T(\boldsymbol{d})$ is quite similar to a commutative blowup and qgr-$T(\boldsymbol{d})$ also coincides with the more categorical version of a blowup in [Van den Bergh 2001]. In this paper we will also need $T(\boldsymbol{d})$ when $\deg(\boldsymbol{d}) = 8$, and this is harder to describe as it is not generated in degree one. Its construction and basic properties are described in the companion paper [RSS 2015].

The proofs. For simplicity we assume here that U is a cg subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$.

A key strategy in the description of the Sklyanin algebra S, and in the classification of noncommutative projective planes in [Artin et al. 1990], was to understand the factor ring S/gS, where $g \in S_3 = T_1$ is the central element mentioned earlier. Indeed, one of the main steps in that paper was to show that $S/gS \cong B(E, \mathscr{L}, \sigma)$ for the appropriate $\mathscr{L}$ and σ. We apply a similar strategy. The nicest case is when $U \subseteq T$ is *g-divisible* in the sense that $g \in U$ and $U \cap gT = gU$. In particular, $\overline{U} = U/gU$ is then a subalgebra of $\overline{T}$ with $\mathrm{GKdim}(\overline{U}) = 2$. As such $\overline{U}$ *and hence* U *are automatically noetherian* (see Proposition 2.9). Much of this paper concerns the classification of g-divisible algebras U, and the starting point is the following result.

Theorem 1.8 (Theorem 5.24). *Let U be a g-divisible subalgebra of T such that $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Then U is an equivalent order to some blowup $T(\boldsymbol{d})$ at an effective divisor $\boldsymbol{d}$ on E with* $\deg \boldsymbol{d} \leq 8$.

It follows easily from this result that *a g-divisible maximal T-order U equals* $\mathrm{End}_{T(\boldsymbol{d})}(M)$ *for some finitely generated right T-module M* (see Corollary 6.6). When U is g-divisible, the rest of the proof of Theorem 1.2 amounts to showing that, *up to a finite-dimensional vector space, $\overline{U} = B(E, \mathcal{M}(-\boldsymbol{d}'), \tau)$, for some virtually effective divisor* $\boldsymbol{d}' = \boldsymbol{d} - \boldsymbol{y} + \tau^{-1}(\boldsymbol{y})$ (see Theorem 6.7). This is also the key property in the definition of a blowup at such a divisor (see Definitions 6.9 and 7.1 for more details).

Now suppose that U is not necessarily g-divisible and set $C = U\langle g\rangle$ with *g-divisible hull*

$$\widehat{C} = \{\theta \in T : g^m\theta \in C \text{ for some } m \geq 0\}.$$

The remaining step in the proof of Theorem 1.2 is to show that U, C and $\widehat{C}$ are equivalent orders. This in turn follows from the following fact. *Let V be a graded subalgebra of T with $g \in V$ and $Q_{\mathrm{gr}}(V) = Q_{\mathrm{gr}}(T)$. Then V has a minimal sporadic ideal in the sense that V has a unique ideal I minimal with respect to* $\mathrm{GKdim}(V/I) \leq 1$ *and V/I being g-torsionfree* (see Corollary 8.8).

Further results. The g-divisible subalgebras of T are closely related to subalgebras of the (ungraded) localised ring $T^\circ = T[g^{-1}]_0$. The algebra T° is a hereditary noetherian domain of GK-dimension 2 and can be thought of as a noncommutative coordinate ring of the affine space $\mathbb{P}^2 \smallsetminus E$. By [RSS 2014], any subalgebra of T° is noetherian and so the algebras $U^\circ = U[g^{-1}]_0 \subseteq T^\circ$ give a plentiful supply of noetherian domains of GK-dimension 2. All the above results have parallel versions for orders in T°. For example:

Corollary 1.9 (Corollary 7.10 and Corollary 8.5). *Let A be a subalgebra of T° with $Q(A) = Q(T^\circ)$.*

(1) *The algebra A has finitely many prime ideals and DCC on ideals.*

(2) *If A is a maximal T°-order then $A = T(\boldsymbol{d}')^\circ$ for some virtually effective divisor $\boldsymbol{d}'$.*

Organisation of the paper. In Section 2 we prove basic technical results, including the important, though easy, fact that any g-divisible subalgebra of T is strongly noetherian (see Proposition 2.9). Section 3 is devoted to studying finitely generated graded orders in $\Bbbk(E)[t;\tau]$. The main result (Theorem 3.1) shows that any such order is (up to finite dimension) an idealiser in a twisted homogeneous coordinate ring. This improves on one of the main results from [Artin and Stafford 1995] and has useful applications to the study of point modules over such an algebra. Section 4 incorporates needed results from [RSS 2015] about right ideals of T and the blowups $T(\boldsymbol{d})$ at effective divisors.

Sections 5–7 are devoted to g-divisible algebras in T. The main result of Section 5 is Theorem 1.8 from above. Section 6 is concerned with the structure of $V = \operatorname{End}_{T(\boldsymbol{d})}(M)$, where $M \subset T$ is a reflexive $T(\boldsymbol{d})$-module and $\boldsymbol{d}$ is effective. Most importantly, Theorem 6.7 describes the factor V/gV. Section 7 pulls these results together, proves Theorem 1.5 for g-divisible algebras and draws various conclusions.

In Section 8 we show that various algebras have minimal sporadic ideals. This is then used to complete the proof of Theorem 1.2. Section 9 studies subalgebras of the Veronese rings $T^{(m)}$ and algebras U with $\overline{U} = \Bbbk$. We apply this to prove Corollary 1.6. Finally, Section 10 is devoted to examples. At the end of the paper we also provide an index of notation.

2. Basic results

In this section we collect the basic definitions and results that will be used throughout the paper.

Throughout the paper $\Bbbk$ is an algebraically closed field and all rings will be $\Bbbk$-algebras. If X is a projective $\Bbbk$-scheme, $\mathcal{L}$ is an invertible sheaf on X, and $\sigma : X \to X$ is an automorphism, then there is a *TCR* or *twisted homogeneous coordinate ring* $B = B(X, \mathcal{L}, \sigma)$ associated to this data and defined as follows. Write $\mathcal{F}^\sigma = \sigma^*(\mathcal{F})$

for a pullback of a sheaf $\mathcal{F}$ on X and set $\mathcal{F}_n = \mathcal{F} \otimes \mathcal{F}^{\sigma} \otimes \cdots \otimes \mathcal{F}^{\sigma^{n-1}}$ for $n \geq 1$. Then

$$B = \bigoplus_{n \geq 0} H^0(X, \mathcal{L}_n), \quad \text{with product } x * y = x \otimes (\sigma^m)^*(y) \text{ for } x \in B_m,\, y \in B_n.$$

In this paper $X = E$ will usually be a smooth elliptic curve, and a review of some of the important properties of $B(E, \mathcal{L}, \sigma)$ in this case can be found in [Rogalski 2011]. It is well known, going back to [Artin et al. 1990], that much of the structure of the Sklyanin algebra S is controlled by the factor ring $S/gS \cong B(E, \mathcal{L}, \sigma)$, and this in turn can be analysed geometrically.

In fact, there are several different families of Sklyanin algebras, and we first set up a framework which will allow our results to apply to subalgebras of any of these (and, indeed, more generally). Recall that for an $\mathbb{N}$-graded ring $R = \bigoplus_{n \geq 0} R_n$ the *d-th Veronese ring*, for $d \geq 1$, is $R^{(d)} = \bigoplus_{n \geq 0} R_{nd}$. Usually this is graded by setting $R^{(d)}_n = R_{nd}$. However, we will sometimes want to regard $R^{(d)}$ as a graded subring of R, in which case each R_{nd} maintains its degree nd; we will call this the *unregraded* Veronese ring. In this paper it will be easier to work with the 3-Veronese ring of the Sklyanin $T = S^{(3)} = \bigoplus_{n \in \mathbb{Z}} T_n$, largely because this ensures that the canonical central element g lies in T_1. Similar comments will apply to the other families, and so in the body of the paper we will work with algebras satisfying the following hypotheses.

Assumption 2.1. Let T be a cg $\Bbbk$-algebra which is a domain with a central element $0 \neq g \in T_1$, such that there is a graded isomorphism $T/gT \cong B = B(E, \mathcal{M}, \tau)$ for a smooth elliptic curve E, invertible sheaf $\mathcal{M}$ with $\mu = \deg \mathcal{M} \geq 2$, and infinite-order automorphism τ. Such a T is called an *elliptic algebra of degree* μ.

This assumption holds throughout the paper. In the language of [Van den Bergh 2001], the assumption can be interpreted geometrically to say that the surface qgr-T contains the commutative elliptic curve qgr-$B \simeq \operatorname{coh} E$ as a divisor. We will need stronger conditions on T in the main results of Section 8 (see Assumption 8.2).

Examples 2.2. The hypotheses of Assumption 2.1 are satisfied in a number of examples, in particular for Veronese rings of the following types of Sklyanin algebras.

(1) Let S be the quadratic Sklyanin algebra

$$S(a, b, c) = \Bbbk\{x_0, x_1, x_2\}/(ax_i x_{i+1} + bx_{i+1}x_i + cx_{i+2}^2 : i \in \mathbb{Z}_3),$$

for appropriate $[a, b, c] \in \mathbb{P}^2_k$, and let $T = S^{(d)}$ for $d = 3$.

(2) Let S be the cubic Sklyanin algebra

$$S(a, b, c) = \Bbbk\{x_0, x_1\}/(ax_{i+1}^2 x_i + bx_{i+1}x_i x_{i+1} + ax_i x_{i+1}^2 + cx_i^3 : i \in \mathbb{Z}_2),$$

for appropriate $[a, b, c] \in \mathbb{P}^2_k$ and let $T = S^{(d)}$, for $d = 4$.

(3) Let x have degree 1 and y degree 2, and set

$$S = S(a,b,c) = \Bbbk\{x,y\}/(ay^2x + cyxy + axy^2 + bx^5,\ ax^2y + cxyx + ayx^2 + by^2),$$

for appropriate $[a,b,c] \in \mathbb{P}^2_k$, and let $T = S^{(d)}$, for $d = 6$.

(4) There are other examples satisfying these hypotheses; for example, take $T = B(E, \mathcal{M}, \tau)[g]$, where $\mathcal{M}$ is an invertible sheaf on the elliptic curve E with $\deg \mathcal{M} \geq 2$ and $|\tau| = \infty$.

The detailed properties of the examples above can be found in [Artin et al. 1990; 1991; Stephenson 1997]. In particular, the restrictions on the parameters $\{a, b, c\}$ in (1–3) are determined as follows. In each case, there exists a central element $g \in S_d$ such that $S/gS \cong B = B(E, \mathcal{L}, \sigma)$, for some $\mathcal{L}$ and σ. This factor ring also determines the Sklyanin algebra, since g is the unique relation for B of degree d [Artin et al. 1990, Theorem 6.8(1); Stephenson 1997, Theorem 4.1]. The requirements on $\{a, b, c\}$ are that E is an elliptic curve and that $|\sigma| = \infty$. Explicit criteria on the parameters are known for E to be an elliptic curve but not for $|\sigma| = \infty$; nevertheless, this will be the case when the parameters are generic. In these examples, $\deg \mathcal{L} = 3, 2, 1$, respectively, and hence $T/gT \cong B(E, \mathcal{M}, \sigma^d)$, where $\mathcal{M} = \mathcal{L}_d$ has degree $\mu = d \cdot (\deg \mathcal{L}) = 9, 8, 6$, respectively.

Notation 2.3. All algebras A considered in this paper are domains of finite Gelfand–Kirillov dimension, written $\mathrm{GKdim}(A)$. If A is graded, then the set $\mathscr{C}$ of nonzero homogeneous elements therefore forms an Ore set (see [McConnell and Robson 2001, Corollary 8.1.21] and [Năstăsescu and van Oystaeyen 1982, C.I.1.6]). By [ibid., A.14.3], the localisation $Q_{\mathrm{gr}}(A) = A\mathscr{C}^{-1}$ is a *graded division ring* in the sense that $Q_{\mathrm{gr}}(A)$ is an Ore extension $Q_{\mathrm{gr}}(A) = D[z, z^{-1}; \alpha]$ of a division ring D by an automorphism α; thus $zd = d^{\alpha} z$ for all $d \in D$. The algebra D will be denoted $D = D_{\mathrm{gr}}(A)$ and called the *function skewfield of* A, while $Q_{\mathrm{gr}}(A)$ will be called the *graded quotient ring of* A.

Notation 2.4. For the most part, the algebras A considered in this paper will be connected graded, in which case we usually work in the category Gr-A of $\mathbb{Z}$-graded right A-modules, with homomorphisms $\mathrm{Hom}_{\text{Gr-}A}(M, N)$ being graded of degree zero. In particular, an isomorphism of graded modules or rings will be assumed to be graded of degree zero, unless otherwise stated. The category of noetherian graded right A-modules will be written gr-A, while the category of ungraded modules will be written Mod-A, and we reserve the term $\mathrm{Hom}(M, N) = \mathrm{Hom}_A(M, N)$ for homomorphisms in the ungraded category. For $M, N \in$ Gr-A, the *shift* $M[n]$ is defined by $M[n] = \bigoplus M[n]_i$ for $M[n]_i = M_{n+i}$. Similar comments apply to $\mathrm{Ext}_{\text{Gr-}A}$ and Ext_A as well as to $\mathrm{End}_A(M) = \mathrm{Hom}_A(M, M)$. If fd-$A$ denotes the category of finite-dimensional (right) A-modules, then we write qgr-A for the quotient category gr-A/ fd-A. Similarly, A-qgr $= A$-gr $/ A$-fd is the quotient category of noetherian graded left modules modulo finite-dimensional modules. The basic properties of this construction can be found in [Artin and Zhang 1994].

Notation 2.5. Write $T_{(g)}$ for the homogeneous localisation of T at the completely prime ideal gT; thus $T_{(g)} = T\mathscr{C}^{-1}$ for $\mathscr{C}$ the set of homogeneous elements in $T \smallsetminus gT$. Note that $T_{(g)}/gT_{(g)} \cong Q_{\mathrm{gr}}(B) = \Bbbk(E)[t, t^{-1}; \tau]$, a ring of twisted Laurent polynomials over the function field of E. In particular, $T_{(g)}/gT_{(g)}$ is a graded division ring and by [Goodearl and Warfield 1989, Exercise 1Q] it is also simple as an ungraded ring. Also, as will be used frequently in the body of the paper,

$$\textit{the only graded right or left ideals of } T_{(g)} \textit{ are the } g^n T_{(g)}. \tag{2.6}$$

For any graded vector subspace $X \subseteq T_{(g)}$, set

$$\widehat{X} = \{t \in T_{(g)} : tg^n \in X \text{ for some } n \in \mathbb{N}\}.$$

We say that X is *g-divisible* if $X \cap gT_{(g)} = gX$. Note that if X is g-divisible and $1 \in X$ (as happens when X is a subring of $T_{(g)}$), then $g \in X$. For any $\Bbbk$-subspace Y of $T_{(g)}$, write $\overline{Y} = (Y + gT_{(g)})/gT_{(g)}$ for the image of Y in $T_{(g)}/gT_{(g)}$.

If $R \subseteq T_{(g)}$ is a subalgebra with $g \in R$, then the *g-torsion* submodule of a right R-module M is $\mathrm{tors}_g(M) = \{m \in M : g^n m = 0 \text{ for some } n \geq 1\}$. We say that M is *g-torsionfree* if $\mathrm{tors}_g(M) = 0$ and *g-torsion* if $\mathrm{tors}_g(M) = M$.

We notice that the rings T automatically satisfy some useful additional properties. An algebra C is called *just infinite* if every nonzero ideal I of C satisfies $\dim_{\Bbbk} C/I < \infty$.

Lemma 2.7. *Let T satisfy Assumption 2.1. Then*:

(1) *T is generated as an algebra in degree* 1.

(2) *Any finitely generated, cg subalgebra of $Q_{\mathrm{gr}}(T/gT) = \Bbbk(E)[z, z^{-1}; \tau]$, in particular T/gT itself, is just infinite.*

Proof. (1) Since $\mu \geq 2$, the ring $B = T/gT \cong B(E, \mathcal{M}, \tau)$ is generated in degree 1 [Rogalski 2011, Lemma 3.1]. Thus $T_2 = (T_1)^2 + gT_1 = (T_1)^2$ and, by induction, $(T_1)^n = T_n$ for all $n \geq 1$.

(2) This follows from [RSS 2014, Corollary 2.10 and §3]. □

As the next few results show, g-divisible algebras and modules have pleasant properties. The first gives a useful, albeit easy, alternative characterisation of $\widehat{X}$ that will be used without particular reference.

Lemma 2.8. *Let $R \subseteq T_{(g)}$ be a cg subalgebra with $g \in R$, and let $X \subseteq T_{(g)}$ be a graded right R-module. Then $X \subseteq \widehat{X}$, and $\widehat{X}$ is also a right R-module. Moreover*:

$$X \text{ is } g\text{-divisible} \iff X = \widehat{X} \iff T_{(g)}/X \text{ is a } g\text{-torsionfree } R\text{-module.} \quad \square$$

Proposition 2.9. (1) *If R is any g-divisible cg subalgebra of T, then R is finitely generated as a $\Bbbk$-algebra.*

(2) *Let R be a finitely generated g-divisible cg subalgebra of $T_{(g)}$. Then R is strongly noetherian.*

Proof. (1) We have $\bar{R} \cong (R+gT)/gT \subseteq \bar{T} \cong B(E, \mathcal{M}, \tau)$ and so [RSS 2014, Theorem 2.9] implies that $\bar{R}$ is noetherian. By [Artin et al. 1990, Lemma 8.2], R is noetherian. Since the generators of $R_{\geq 1}$ as an R-module also generate R as a $\Bbbk$-algebra, R is finitely generated as a $\Bbbk$-algebra.

(2) In this case, $\bar{R} = R/gR \cong (R+gT_{(g)})/gT_{(g)} \subseteq Q_{\text{gr}}(B) = \Bbbk(E)[t, t^{-1}; \tau]$. By [RSS 2014, Corollary 2.10] $\bar{R}$ is noetherian. Also GKdim $\bar{R} \leq 2$, for instance by [Artin and Stafford 1995, Theorem 0.1], and so $\bar{R}$ is strongly noetherian by [Artin et al. 1999, Theorem 4.24]. Thus R is strongly noetherian by [Artin et al. 1990, Lemma 8.2]. □

Lemma 2.10. *Let R be a g-divisible cg subalgebra of $T_{(g)}$ with $D_{\text{gr}}(R) = D_{\text{gr}}(T)$. Then*

(1) *$Q_{\text{gr}}(R) = Q_{\text{gr}}(T)$, and*

(2) *$Q_{\text{gr}}(\bar{R}) = Q_{\text{gr}}(\bar{T})$.*

Proof. (1) As $g \in R_1$ we have

$$Q_{\text{gr}}(T) = D_{\text{gr}}(T)[g, g^{-1}] = D_{\text{gr}}(R)[g, g^{-1}] = Q_{\text{gr}}(R).$$

(2) Since $Q_{\text{gr}}(R) = Q_{\text{gr}}(T)$, there exists $0 \neq x \in R_d$ such that $xT_1 \subseteq R_{d+1}$. Then $\bar{x}\bar{T}_1 \subseteq \bar{R}$. As long as $\bar{x} \neq 0$, this shows that the graded quotient ring of $\bar{R}$ contains a generating set for $\bar{T}$ and we are done. On the other hand, if $\bar{x} = 0$, then write $x = g^i y$ with $y \in T_{(g)} \smallsetminus gT_{(g)}$; equivalently $y \in R \smallsetminus gR$ by g-divisibility. Then $g^i yT_1 \subseteq R \cap g^i T_{(g)} = g^i R$, and so $yT_1 \subseteq R$. Thus we are again done. □

If A is a cg domain with graded quotient ring $Q = Q_{\text{gr}}(A)$ and $M \subseteq Q$ is a finitely generated graded right A-submodule, we can and always will identify

$$\begin{aligned} \operatorname{End}_A(M) = \{q \in Q : qM \subseteq M\} \quad \text{and} \\ M^* = \operatorname{Hom}_A(M, A) = \{q \in Q : qM \subseteq A\}. \end{aligned} \tag{2.11}$$

Clearly both $\operatorname{End}_A(M)$ and M^* are graded subspaces of Q.

Lemma 2.12. *Let R be any g-divisible subring of $T_{(g)}$ with $Q_{\text{gr}}(R) = Q_{\text{gr}}(T_{(g)})$, and let $M, M' \subseteq T_{(g)}$ be finitely generated nonzero right R-modules.*

(1) *If $M \not\subseteq gT_{(g)}$, then we can identify*

$$\operatorname{Hom}_R(M, M') = \{x \in T_{(g)} : xM \subseteq M'\} \subseteq T_{(g)}.$$

(2) *If M' is g-divisible, and $M \not\subseteq gT_{(g)}$ (in particular if M is g-divisible) then* $\mathrm{Hom}_R(M, M') \subseteq T_{(g)}$ *is also g-divisible.*

(3) *If M is g-divisible, then $U = \mathrm{End}_R(M) \subseteq T_{(g)}$ is g-divisible, and M is a finitely generated left U-module. Moreover, $\overline{U} \subseteq \mathrm{End}_{\overline{R}}(\overline{M})$.*

Proof. (1) Since $M \not\subseteq gT_{(g)}$, it follows from (2.6) that $MT_{(g)} = T_{(g)}$. In particular,

$$N = \mathrm{Hom}_R(M, M') \subseteq \mathrm{Hom}_{T_{(g)}}(MT_{(g)}, M'T_{(g)}) \subseteq T_{(g)}.$$

(2) Part (1) applies, and so $N = \mathrm{Hom}_R(M, M') \subseteq T_{(g)}$. Next, let $\theta \in N \cap gT_{(g)}$; say $\theta = gs$ for some $s \in T_{(g)}$. Then $sgM = \theta M \in M' \cap gT_{(g)} = M'g$ since M' is g-divisible. Hence $sM \subseteq M'$ and $s \in N$. Thus $N \cap gT_{(g)} = gN$.

(3) By part (2), U is g-divisible, and hence is noetherian by Proposition 2.9. As $Q_{\mathrm{gr}}(R) = Q_{\mathrm{gr}}(T_{(g)})$, there exists $x \in T_{(g)} \smallsetminus \{0\}$ so that $xM \subseteq R$. Then $MxM \subseteq MR = M$. Hence (up to a shift) $M \cong Mx \subseteq U$ is finitely generated as a left U-module.

Now

$$\overline{U} = (U + gT_{(g)})/gT_{(g)} \subseteq \overline{T_{(g)}} = \Bbbk(E)[t, t^{-1}; \tau].$$

Since $Q_{\mathrm{gr}}(\overline{U}) = Q_{\mathrm{gr}}(\overline{T_{(g)}})$ by Lemma 2.10, as in (2.11) we identify $\mathrm{End}_{\overline{R}}(\overline{M})$ with $\{x \in \overline{T_{(g)}} : x\overline{M} \subseteq \overline{M}\}$. But since $UM \subseteq M$, clearly $(\overline{U})(\overline{M}) \subseteq \overline{M}$. $\square$

Lemma 2.13. *Let R be a graded subalgebra of $T_{(g)}$ with $Q_{\mathrm{gr}}(R) = Q_{\mathrm{gr}}(T_{(g)})$ and let $M \subseteq T_{(g)}$ be a graded right R-submodule of $T_{(g)}$ such that $M \not\subseteq gT_{(g)}$. Then:*

(1) *For any $x \in T_{(g)} \smallsetminus gT_{(g)}$, we have $\widehat{xM} = x\widehat{M}$.*

(2) *If R is g-divisible and M is a finitely generated R-module, then so is $\widehat{M}$.*

(3) *If R is g-divisible, then $T_{(g)} \supseteq M^* = \widehat{M^*}$ and $M^* \not\subseteq gT_{(g)}$. Hence $T_{(g)} \supseteq M^{**} = \widehat{M^{**}}$. Moreover, we have $(\widehat{M})^* = M^*$ and $(\widehat{M})^{**} = M^{**}$.*

Proof. (1) Let $r \in \widehat{M}$. For some n we have $rg^n \in M$, so $xrg^n \in xM$. Since $xr \in T_{(g)}$ it follows that $xr \in \widehat{xM}$. Conversely, if $r \in T_{(g)}$ with $rg^n \in xM$, then $rg^n = g^n r \in g^n T_{(g)} \cap xT_{(g)}$. As $gT_{(g)}$ is a completely prime ideal and $x \notin gT_{(g)}$, clearly $g^n T_{(g)} \cap xT_{(g)} = g^n xT_{(g)}$. Thus $r = xs$ for some $s \in T_{(g)}$ and $xM \ni rg^n = xsg^n$. Therefore $sg^n \in M$, whence $s \in \widehat{M}$ and $r \in x\widehat{M}$. Thus $\widehat{xM} = x\widehat{M}$, as claimed.

(2) As in the proof of Lemma 2.12, there exists $x \in T_{(g)} \smallsetminus \{0\}$ so that $xM \subseteq R$. If $x = gy$ for some $y \in T_{(g)}$, then $g(yM) \subseteq R$ and so $yM \subseteq R$ since R is g-divisible. Thus we can assume that $x \in T_{(g)} \smallsetminus gT_{(g)}$. Again by g-divisibility, $\widehat{xM} \subseteq \widehat{R} = R$. By Proposition 2.9 $\widehat{xM}$ is a finitely generated right ideal of R. Up to a shift, $\widehat{M} \cong x\widehat{M} = \widehat{xM}$ by (1). This is finitely generated as an R-module.

(3) By Lemma 2.12(2), M^* is equal to $\mathrm{Hom}_R(M, R) \subseteq T_{(g)}$ and is g-divisible, i.e., $M^* = \widehat{M^*}$. Clearly then $M^* \not\subseteq gT_{(g)}$, and so by the left-handed analogue of the same argument, $M^{**} = \widehat{M^{**}} \subseteq T_{(g)}$ also.

Now as $M \subseteq \widehat{M}$, certainly $(\widehat{M})^* \subseteq M^*$. On the other hand, if $\theta \in M^*$ and $x \in \widehat{M}$, say with $xg^n \in M$, then $(\theta x)g^n = \theta(xg^n) \in R = \widehat{R}$. Hence $\theta x \in R$. Thus $\theta \in (\widehat{M})^*$ and $(\widehat{M})^* = M^*$. Taking a second dual gives $(\widehat{M})^{**} = M^{**}$. $\square$

We note next some special properties of modules of GK-dimension 1.

Lemma 2.14. *Let R be a cg g-divisible subalgebra of $T_{(g)}$ and suppose that M is a finitely generated, g-torsionfree R-module with* $\operatorname{GKdim}(M) \leq 1$. *Then the Hilbert series of M is eventually constant; that is,* $\dim_{\Bbbk} M_n = \dim_{\Bbbk} M_{n+1}$ *for all $n \gg 0$. Moreover, M is a finitely generated $\Bbbk[g]$-module.*

Proof. By [Krause and Lenagan 1985, Proposition 5.1(e)], $\operatorname{GKdim}(M/Mg) \leq 0$ and so $\dim_{\Bbbk} M/Mg < \infty$. Thus $M_r g = M_{r+1}$ for all $r \gg 0$; say for $r \geq n_0$. In particular, $M = M_{\leq n_0}\Bbbk[g]$. Moreover, since multiplication by g is an injective map from M_r to M_{r+1}, it follows that $\dim_{\Bbbk} M_r = \dim_{\Bbbk} M_{r+1}$ for all $r \geq n_0$. $\square$

A graded ideal I in a cg algebra R is called a *sporadic ideal* if $\operatorname{GKdim}(R/I) = 1$ (these are called special ideals in [Rogalski 2011]). The name is justified since, as will be shown in Section 8, orders in T have very few such ideals. The next lemma will be useful in understanding them.

Lemma 2.15. *Let R be a g-divisible finitely generated cg subring of $T_{(g)}$ with $Q_{\mathrm{gr}}(R) = Q_{\mathrm{gr}}(T)$. Then*:

(1) *If J is a nonzero g-divisible graded ideal of R, then* $\operatorname{GKdim}(R/J) \leq 1$.

(2) *Conversely, if J is a graded ideal of R such that* $\operatorname{GKdim}(R/J) \leq 1$, *then $\widehat{J}/J$ is finite-dimensional.*

(3) *If K is any ideal of R, then $K = g^n I$ for some $n \geq 1$ and ideal I satisfying* $\operatorname{GKdim}(R/I) \leq 1$.

(4) *Suppose that L, M are graded subspaces of $T_{(g)}$ with $L \not\subseteq gT_{(g)}$ and $M \not\subseteq gT_{(g)}$ and assume that $I = LM$ is an ideal of R. Then* $\operatorname{GKdim}(R/I) \leq 1$.

Proof. (1) By Lemma 2.10, $\bar{R} \subseteq k(E)[t, t^{-1}; \tau] = Q_{\mathrm{gr}}(\bar{R})$ and, by Lemma 2.7(2), $\bar{R}$ is just infinite. Since J is g-divisible, $J \not\subseteq gR$ and so $\bar{J} \neq 0$; thus $\dim_{\Bbbk} \bar{R}/\bar{J} < \infty$. Equivalently, if $R' = R/J$ then $\dim_{\Bbbk} R'/gR' < \infty$. It follows that $R'_m = gR'_{m-1}$ for all $m \gg 0$, and hence that $\operatorname{GKdim} R' \leq 1$.

(2) Once again, $\bar{R}$ is just infinite. Thus, since $J \subseteq gR$ would lead to the contradiction $\operatorname{GKdim}(R/J) \geq 2$, we must have $\dim_{\Bbbk} R/(gR + J) = \dim_{\Bbbk} \bar{R}/\bar{J} < \infty$. Since $\widehat{J}$ is noetherian, $g^n \widehat{J} \subseteq J$ for some n. If J' is the largest right ideal inside $\widehat{J}$ such that J'/J is finite-dimensional, then J' is an ideal and we can replace J by J' without loss. If we still have $J \neq \widehat{J}$, then there exists $x \in \widehat{J} \smallsetminus J$ such that $xg \in J$. Thus $x(gR+J) \subseteq J$, and left multiplication by x defines a surjection $R/(gR + J) \twoheadrightarrow (xR + J)/J$. We have $\dim_{\Bbbk}(xR + J)/J = \infty$ and $\dim_{\Bbbk} R/(gR + J) < \infty$, a contradiction. Thus $\widehat{J} = J$.

(3) Write $K = g^n J$ with n as large as possible and J an ideal of R. Then $J \not\subseteq gR$, and so Lemma 2.7(2) again implies that $\dim_{\Bbbk} \bar{R}/\bar{J} < \infty$ and hence GKdim $R/J \leq 1$.

(4) Since $gT_{(g)}$ is completely prime, $I = LM \not\subseteq gT_{(g)}$ and hence $I \not\subseteq gR$. Now apply part (3). $\square$

Next, we want to prove some general results about equivalent orders that will be useful elsewhere. We recall that two cg domains A and B with a common (graded) quotient ring $Q = Q_{\text{gr}}(A) = Q_{\text{gr}}(B)$ are *equivalent orders* if $aAb \subseteq B$ and $cBd \subseteq A$ for some $a, b, c, d \in Q \smallsetminus \{0\}$. Clearing denominators on the appropriate sides, one can always assume that $a, b, c, d \in B$. One can also assume that a, b, c, d are homogeneous; indeed, if a and b have leading terms a_n and b_m, then $a_n A b_m \subseteq B$.

Proposition 2.16. *Suppose that $U \subseteq R$ are g-divisible cg finitely generated subalgebras of $T_{(g)}$ such that $Q_{\text{gr}}(U) = Q_{\text{gr}}(R) = Q_{\text{gr}}(T_{(g)})$. Then the following are equivalent*:

(1) *U and R are equivalent orders in $Q_{\text{gr}}(U) = Q_{\text{gr}}(T)$.*

(2) *U/gU and R/gR are equivalent orders in $Q_{\text{gr}}(U/gU)$.*

Proof. (1) $\Rightarrow$ (2) Choose nonzero homogeneous elements $a, b \in U$ such that $aRb \subseteq U$. Write $a = g^n a'$ where $a' \in U \smallsetminus gU$. Then $g^n a' Rb \subseteq U$ and so $a' Rb \subseteq U$ since U is g-divisible. Replacing a by a', we can assume that $a \notin gU$ and, similarly, that $b \notin gU$. Then $\bar{a}\bar{R}\bar{b} \subseteq \bar{U}$, with $\bar{a}, \bar{b} \neq 0$, as required.

(2) $\Rightarrow$ (1) Set $\bar{U} = U/gU \subseteq \bar{R} = R/gR$. We first note that there is a subalgebra $\bar{U} \subseteq S \subseteq \bar{R}$ so that S is a noetherian right $\bar{U}$-module and $\bar{R}$ is a noetherian left S-module. Indeed, write $a\bar{R}b \subseteq \bar{U}$ for some nonzero $a, b \in \bar{U}$ and set $S = \bar{U} + \bar{R}b\bar{U}$. Clearly $aS \subseteq \bar{U}$ and $\bar{R}b \subseteq S$. As in the proof of Proposition 2.9, all subalgebras of $\bar{R}$ are noetherian. In particular, S and $\bar{U}$ are noetherian and so these inclusions ensure that $S_{\bar{U}}$ and ${}_S\bar{R}$ are finitely generated, as claimed.

Let $F \subseteq R$ be a finite-dimensional vector space, containing 1, such that $\overline{F}\bar{U} = S$. Set $M = \widehat{FU}$ and $V = \text{End}_U(M)$. Clearly $Q_{\text{gr}}(V) = Q_{\text{gr}}(U) = Q_{\text{gr}}(T)$. Since $1 \in M$ and hence $M \not\subseteq gT_{(g)}$, we can and will use Lemma 2.12(1) to identify $V = \{q \in Q_{\text{gr}}(U) : qM \subseteq M\} \subseteq T_{(g)}$. By Lemma 2.12(3), $V = \widehat{V}$ and ${}_V M$ is finitely generated, while, by Lemma 2.13, M_U is finitely generated. As R is g-divisible and $FU \subseteq R$, we have $M \subseteq R$. Since $1 \in M$ this implies that $MR = R$. Hence $VR = VMR = MR = R$ and $V \subseteq R$.

Let $G, H \subseteq R$ be finite-dimensional vector spaces with $VG = M$ and $S\overline{H} = \bar{R}$. Then

$$\bar{R} \supseteq \overline{VGH} \supseteq \overline{FUH} = S\overline{H} = \bar{R}.$$

Thus $\bar{R} = \overline{MH} = \overline{VGH}$ is finitely generated as a left $\bar{V}$-module. Since $g \in V_+ = \bigoplus_{i>0} V_i \subseteq R_+$, this implies that $R/(V_+)R$ is a finitely generated left module over

V/V_+. By the graded analogue of Nakayama's lemma, this implies that R is finitely generated as a left V-module. Thus R and V are equivalent orders. As V and U are equivalent orders (via the bimodule M), it follows that U and R are equivalent. $\square$

3. Curves

The main result of [Artin and Stafford 1995] shows that any cg domain A of Gelfand–Kirillov dimension two has a Veronese ring that is an idealiser inside a TCR. In this section we strengthen this result for elliptic curves by proving that, for subalgebras of a TCR over such a curve corresponding to an automorphism of infinite order, the result holds without taking a Veronese ring, although at the cost of replacing the idealiser by an algebra which is isomorphic to an idealiser in large degree.

Given graded modules $M, N \subseteq P$ over a cg algebra A, we write $M \doteq N$ if M and N agree up to a finite-dimensional vector space. If $M, N \in \text{gr-}A$, this is equivalent to $M_{\geq n} = N_{\geq n}$ for some $n \geq 0$.

Theorem 3.1. *Let A be a cg ring such that $Q_{\mathrm{gr}}(A) = \Bbbk(E)[z, z^{-1}; \tau]$ for some infinite-order automorphism τ of a smooth elliptic curve E and $z \in Q_{\mathrm{gr}}(A)_1$. Then there are an ideal sheaf $\mathcal{A}$ and an ample invertible sheaf $\mathcal{H}$ on E so that*

$$A \doteq \bigoplus_{n \geq 0} H^0(E, \mathcal{A}\mathcal{H}_n).$$

Remarks 3.2. (1) The *idealiser* $\mathbb{I}(J) = \mathbb{I}_U(J)$ of a right ideal J in a ring U is the subring

$$\mathbb{I}(J) = \{u \in U : uJ \subseteq J\}.$$

In the notation of the theorem, $J = \bigoplus_{n \geq 0} H^0(E, \mathcal{A}\mathcal{H}_n)$ is a right ideal of the TCR $B(E, \mathcal{H}, \tau)$; further, $\mathbb{I}_U(J) \doteq \Bbbk + J$. So, an equivalent way of phrasing the theorem is to assert that (up to a finite-dimensional vector space) A is equal to the idealiser $\mathbb{I}(J)$ inside $B(E, \mathcal{H}, \tau)$.

(2) The assertion that $z \in Q_{\mathrm{gr}}(A)_1$ can be avoided at the expense of regrading A, although in the process one must replace τ by some τ^m in the definition of the $\mathcal{H}_n$.

(3) The sheaf $\mathcal{H}$ is ample if and only if it has positive degree [Hartshorne 1977, Corollary 3.3], if and only if $\mathcal{H}$ is τ-*ample*: that is, for any coherent $\mathcal{F}$ and for $n \gg 0$, the sheaf $\mathcal{F} \otimes \mathcal{H}_n$ is globally generated with $H^1(E, \mathcal{F} \otimes \mathcal{H}_n) = 0$ [Artin and Van den Bergh 1990, Corollary 1.6].

Proof. The hypothesis on z ensures that $A_p \neq 0 \neq A_{p+1}$ for all $p \gg 0$. Fix some such p.

The conclusion of the theorem is, essentially, the same as that of [Artin and Stafford 1995, Theorem 5.11], although that result has two hypotheses we need to remove. The first, [ibid., Hypothesis 2.1] requires that the ring in question has a nonzero element in degree one, so does at least hold for the Veronese rings $A^{(p)}$ and $A^{(q)}$, for $q = p+1$. The remaining hypothesis, [ibid., Hypothesis 2.15], concerns τ-fixed points of E. In our situation, this automatically holds as E has no such fixed points (see the discussion before [ibid., (2.9)]).

By the discussion above, [ibid., Theorem 5.11 and Remark 5.12(2)] can be applied to the Veronese rings $A^{(p)}$ and $A^{(q)}$. This provides invertible sheaves $\mathscr{A}, \mathscr{B}, \mathscr{F}, \mathscr{G}$ with $\mathscr{F}, \mathscr{G}$ ample such that

$$A^{(p)} \doteq \bigoplus_{n\geq 0} H^0(E, \mathscr{A}\otimes\mathscr{F}_{p,n}) \quad \text{and} \quad A^{(q)} \doteq \bigoplus_{n\geq 0} H^0(E, \mathscr{B}\otimes\mathscr{G}_{q,n}),$$

where in order to take account of the Veronese rings we have written $\mathscr{M}_{r,n} = \mathscr{M}\otimes\mathscr{M}^{\tau^r}\otimes\cdots\otimes\mathscr{M}^{\tau^{(n-1)r}}$ for an invertible sheaf $\mathscr{M}$. For $n \gg 0$ the sheaves $\mathscr{A}\otimes\mathscr{F}_{p,nq}$ and $\mathscr{B}\otimes\mathscr{G}_{q,np}$ are generated by their sections A_{npq} and so $\mathscr{A}\otimes\mathscr{F}_{p,nq} = \mathscr{B}\otimes\mathscr{G}_{q,np}$ for such n. Replacing n by $n+m$, we obtain

$$\begin{aligned}\mathscr{A}\otimes\mathscr{F}_{p,nq}\otimes\mathscr{F}_{p,mq}^{\tau^{npq}} &= \mathscr{A}\otimes\mathscr{F}_{p,(n+m)q} = \mathscr{B}\otimes\mathscr{G}_{q,(n+m)p} \\ &= \mathscr{B}\otimes\mathscr{G}_{q,np}\otimes\mathscr{G}_{q,mp}^{\tau^{npq}} = \mathscr{A}\otimes\mathscr{F}_{p,nq}\otimes\mathscr{G}_{q,mp}^{\tau^{npq}}\end{aligned}$$

for all $n+m > n \gg 0$. Cancelling the first two terms and applying τ^{-npq} gives $\mathscr{F}_{p,mq} = \mathscr{G}_{q,mp}$ for all $m \geq 1$. In particular, it holds for $m = n$, and hence $\mathscr{A} = \mathscr{B}$.

Next, set $\mathscr{H} = \mathscr{G}\otimes(\mathscr{F}^{\tau})^{-1}$; thus the equation $\mathscr{F}_{p,q} = \mathscr{G}_{q,p}$ gives

$$\mathscr{H}_{q,p} = \mathscr{F}_{p,q}\otimes(\mathscr{F}^{\tau})^{-1}_{q,p}. \tag{3.3}$$

We claim that $\mathscr{F}$ is the unique invertible sheaf $\widetilde{\mathscr{F}}$ satisfying $\mathscr{H}_{q,p} = \widetilde{\mathscr{F}}_{p,q}\otimes(\widetilde{\mathscr{F}}^{\tau})^{-1}_{q,p}$. To see this, suppose that $\widetilde{\mathscr{F}}$ is a second sheaf satisfying this property and consider associated divisors. Pick a closed point $x \in E$ and write $\mathbb{O}_x = \{x(i) = \tau^{-i}(x) : i \in \mathbb{Z}\}$ for the orbit of x under τ. Writing $\mathscr{F} = \mathscr{O}_E(F)$ and $\mathscr{H} = \mathscr{O}_E(H)$ for some divisors F and H and restricting to $\mathbb{O} = \mathbb{O}_x$ gives $F|_{\mathbb{O}} = \sum m(i)x(i)$ and $H|_{\mathbb{O}} = \sum r(i)x(i)$, for some integers $m(i), r(i)$ (the notation is chosen to avoid excessive subscripts). Now, in terms of divisors, (3.3) gives

$$\begin{aligned}&\sum r(i)x(i) + r(i)x(i+q) + \cdots + r(i)x(i+(p-1)q) \\ &\quad = \sum m(i)x(i) + m(i)x(i+p) + \cdots + m(i)x(i+(q-1)p) \\ &\qquad - \sum m(i)x(i+1) + m(i)x(i+1+q) + \cdots + m(i)x(i+1+(p-1)q).\end{aligned}$$

Equating coefficients of $x(t)$ in the last displayed equation gives

$$\begin{aligned}&m(t) + m(t-p) + \cdots + m(t-(q-1)p) \\ &\qquad - m(t-1) - m(t-1-q) - \cdots - m(t-1-(p-1)q) \\ &\qquad\qquad = r(t) + r(t-q) + \cdots + r(t-(p-1)q).\end{aligned}$$

Recall that $m(i) = r(i) = 0$ for $|i| \gg 0$. Therefore, solving this system from $t \ll 0$ through to $t \gg 0$ gives a unique solution for the $m(i)$ in terms of the $r(j)$. Finally, doing this for every orbit involved in the divisors F and H shows that F is uniquely determined by H, and so $\mathcal{F}$ is uniquely determined as claimed.

A direct calculation shows that if $\widetilde{\mathcal{F}} = \mathcal{H}_{1,p} = \mathcal{H}_p$ then $\mathcal{H}_{q,p} = \widetilde{\mathcal{F}}_{p,q} \otimes (\widetilde{\mathcal{F}}^{\tau})^{-1}_{q,p}$. Thus $\widetilde{\mathcal{F}} = \mathcal{F}$ and consequently $\mathcal{H}_{1,mp} = \mathcal{F}_{p,m}$ for all $m \geq 1$. It follows from the equation $\mathcal{H} = \mathcal{G} \otimes (\mathcal{F}^{\tau})^{-1}$ that $\mathcal{G} = \mathcal{H}_{1,q}$, and thus $\mathcal{H}_{1,mq} = \mathcal{G}_{q,m}$ for all $m \geq 1$ as well. To summarise, we have found sheaves $\mathcal{A}$ and $\mathcal{H}$ such that

$$A^{(s)} \doteq \bigoplus_n H^0(E, \mathcal{A}\mathcal{H}_{1,ns}) = \bigoplus_n H^0(E, \mathcal{A}\mathcal{H}_{ns}) \quad \text{for } s = p, p+1. \tag{3.4}$$

It follows that (3.4) holds for all $s \gg 0$, but this is not quite enough to prove the theorem since, as s increases, one has no control over the finitely many values of $n = n(s)$ for which $A_n^{(s)} \neq H^0(E, \mathcal{A}\mathcal{H}_{1,ns})$. So we take a slightly different tack.

For $0 \leq r \leq p-1$, write $M(r) = \bigoplus_{n \geq 0} A_{np+r}$; thus $A = \bigoplus_{r=0}^{p-1} M(r)$. Fix some such r. We can find $0 \neq x \in A_{2p-r}$, since $2p - r > p$. Thus $xM(r) \subseteq A^{(p)}$ and so, by [ibid., Proposition 5.4] and (3.4), there exists an ideal sheaf $\mathcal{I} \subseteq \mathcal{O}_E$ such that $xM(r) \doteq \bigoplus_{n \geq 0} H^0(E, \mathcal{I} \otimes \mathcal{H}_{1,np}^{\tau^{2p}})$ (in this formula, the twist by τ^{2p} is for convenience only but it will simplify the computations). Since A is a domain, $M(r)$ is isomorphic to the shift $xM(r)[2p-r]$. Hence, for some integer n_0 independent of r, [Keeler et al. 2005, Lemma 5.5] implies that

$$M(r)_{\geq n_0} = \bigoplus_{n \geq n_0} H^0(E, \mathcal{I}' \otimes \mathcal{H}_{1,np}^{\tau^r}) = \bigoplus_{n \geq n_0} H^0(E, \mathcal{J}(r) \otimes \mathcal{H}_r \otimes \mathcal{H}_{1,np}^{\tau^r}) \tag{3.5}$$

for some invertible sheaves $\mathcal{I}'$ and $\mathcal{J}(r) = \mathcal{I}' \otimes (\mathcal{H}_r)^{-1}$. Possibly after increasing n_0, we may also assume that the sheaves in (3.4) and (3.5) are generated by their sections for $n \geq n_0$. Now pick $n \geq n_0$ such that $r + np = (p+1)m$ for some m. Then comparing (3.4) and (3.5) shows that $M(r)_{r+np}$ generates the sheaves

$$\mathcal{J}(r) \otimes \mathcal{H}_{r+np} = \mathcal{J}(r) \otimes \mathcal{H}_r \otimes \mathcal{H}_{1,np}^{\tau^r} = \mathcal{A} \otimes \mathcal{H}_{1,(p+1)m} = \mathcal{A} \otimes \mathcal{H}_{r+np}.$$

Hence $\mathcal{J}(r) = \mathcal{A}$. Since this holds for all $0 \leq r \leq p-1$, it follows that $A_n = H^0(E, \mathcal{A}\mathcal{H}^{\tau^n})$, for all $n \geq n_0 p$. □

Remark 3.6. We note that [Rogalski 2011, Lemma 3.2(2)] states a result similar to Theorem 3.1, but the proof erroneously quotes the relevant theorems from [Artin and Stafford 1995] without removing the hypothesis that rings should have a nonzero element in degree one. Thus the above proof also corrects this oversight. In any case, [Rogalski 2011, Lemma 3.2(2)] was only used in that paper for rings generated in degree one.

If A is a cg algebra generated in degree one, then we define a *point module* to be a cyclic module $M = \bigoplus_{i \geq 0} M_i$, with $\dim M_i = 1$ for all $i \geq 0$. When A is not generated in degree one, a point module has this asymptotic structure, but the precise definition can vary depending on circumstances, and so we will be careful to explain which definition we mean should the distinction be important.

To end this section we give some applications of the previous theorem to the structure of point modules, for which we need a definition. If $M = \bigoplus_n M_n$ is a graded module over a cg algebra A, we write $s^n(M) = (M_n A)[n]$. The largest artinian submodule of a noetherian module M is written $S(M)$.

Corollary 3.7. *Let A satisfy the hypotheses of Theorem 3.1. Let M and M' be 1-critical graded right A-modules generated in degree zero. Then*:

1. *the isomorphism classes of such modules are in one-to-one correspondence with the closed points of E*;
2. $\dim M_n \le 1$ *for all* $n \ge 0$, *with* $\dim M_n = 1$ *for* $n \gg 0$;
3. *for* $n \ge 0$, *either* $M_n = 0$ *or* $s^n M$ *is cyclic and* 1-*critical*;
4. *if* $s^n M \cong s^n M' \neq 0$ *for some* $n \in \mathbb{N}$, *then* $M \cong M'$.

Proof. It is well known that there is an equivalence of categories qgr-$A \sim \mathrm{coh}(E)$, and much of the corollary follows from this; thus we first review the details of the equivalence. By [Artin and Stafford 1995, Theorem 5.11] and the left-right analogue of Theorem 3.1, we can write

$$A \doteq \bigoplus_{n\ge 0} H^0(E, \mathcal{H}_n \mathcal{A}^{\tau^{n-1}}) \subseteq B = B(E, \mathcal{H}, \tau)$$

for some ideal sheaf $\mathcal{A}$ and invertible sheaf $\mathcal{H}$. For $n_0 \gg 0$, the ideal

$$J = A_{\ge n_0} = \bigoplus_{n\ge n_0} H^0(E, \mathcal{H}_n \mathcal{A}^{\tau^{n-1}})$$

is a left ideal of B. By [Stafford and Zhang 1994, Proposition 2.7] and its proof, qgr-$A \sim$ qgr-B under the maps $\alpha : N \mapsto N \otimes_A B$ and $\beta : N' \mapsto N' \otimes_B J$. Moreover, by [Artin and Van den Bergh 1990, Theorem 1.3], qgr-$B \sim \mathrm{coh}(E)$. Under that equivalence, for a closed point p of E the skyscraper sheaf $\Bbbk(p) \in \mathrm{coh}(E)$ maps to the module

$$M'_p = \bigoplus_{n\ge 0} H^0(E, k(p) \otimes \mathcal{H}_n) \in \text{qgr-}B;$$

thus if $M_p = M'_p / S(M'_p)$ then M_p is a 1-critical B-module with $\dim(M_p)_n = 1$ for $n \gg 0$. By [Stafford and Zhang 1994, Lemma 2.6] the same is true of the 1-critical A-module $N_p = \beta(M_p)/S(\beta(M_p))$. Furthermore, the image in qgr-A of any 1-critical graded A-module is a simple object, and so every 1-critical A-module is equal in qgr-A to some N_p.

(2) We will reduce to the case of a TCR generated in degree one, where the result is standard. If the result fails, there exists a 1-critical A-module M such that (possibly after shifting) $\dim M_n \le 1$ for all $n \ge 0$ but $\dim M_0 > 1$. By replacing M by any submodule generated by a 2-dimensional subspace of M_0 we may assume that $\dim(M_0) = 2$. Write $M = (A \oplus A)/F$.

Now consider $W = \alpha(M)/S(\alpha(M))$. Since W is equal in qgr-B to some M_p, certainly $\dim(W_n) \le 1$ for all $n \gg 0$. Moreover, the natural A-module map $M \to W$ must be injective since M is 1-critical, and so $\dim W_0 \ge 2$. As $\alpha(M)$ is a factor of $B \oplus B/FB$ it follows that $\dim W_0 = 2$. Unfortunately, B need not be generated in degree 1. However, for $\ell \gg 0$ (indeed $\ell \ge 2$) the Veronese ring $C = B^{(\ell)} = B(E, \mathcal{H}_\ell, \tau^\ell)$ will be generated in degree one (see [Rogalski 2011, Lemma 3.1(2)]). We claim that $X = W^{(\ell)}$ will still be a critical C-module. If not, then, picking an element $0 \ne x \in X_m$ in the socle of X, we will have $xC_{\ge 1} = 0$, and so $x \in W_{m\ell}$ satisfies $xB_{i\ell} = 0$ for all $i \ge 1$. Since $B_i B_j = B_{i+j}$ for all $i, j \gg 0$ [Rogalski 2011, Lemma 3.1(1)], it follows that $xB_m = 0$ for all $m \gg 0$, contradicting the 1-criticality of W. Thus X is indeed a critical C-module, with $\dim X_0 = 2$; say $X_0 = a\Bbbk \oplus b\Bbbk$.

Finally, given X, or any 1-critical C-module, then [Artin and Van den Bergh 1990] again implies that $\dim X_n = 1$ for all $n \ge n_0 \gg 0$. By [Keeler et al. 2005, Proposition 9.2] the map $N \mapsto N_{\ge 1}[1]$ is an automorphism on the set of isomorphism classes of C-point modules. Applying the inverse of this map to the shift of $X_{\ge n_0}$ shows that the two point modules aC and bC must be equal to this image and hence be isomorphic; say $bC = \phi(aC)$. Set $n = n_0 + 1$. As $\dim_\Bbbk X_n = 1$, we can write $C_{n-1} = c\Bbbk + \operatorname{ann}_C(a)_{n-1}$ for some $c \in C_{n-1}$. Since $\operatorname{ann}_C(a) = \operatorname{ann}_C(b)$, it follows that $ac = \lambda bc$ for some $\lambda \in \Bbbk$. Hence $(a - \lambda b)c = 0$, which implies that $(a - \lambda b)C_{n-1} = 0$. As C is generated in degree one, this forces $a - \lambda b \in S(X)$. This contradicts the criticality of X and proves the result.

(3) This is immediate from part (2).

(4) If not, pick 1-critical modules $M \not\cong M'$ such that there is an isomorphism $\gamma : s^n M \cong s^n M' \ne 0$ for some $n > 0$. Let n be the smallest integer with this property and then let $W \subset M$ be as large as possible a submodule of M for which γ extends to an isomorphism $\gamma : W \to W' \subset M'$. Set

$$N = \frac{M \oplus M'}{Z} \quad \text{and} \quad Z = \{(a, \gamma(a)) : a \in W\}.$$

We claim that N is 1-critical. If not, pick a homogeneous element $(u, u') \in M \oplus M'$ such that $[(u, u') + Z]$ is a nonzero element of the socle of N. If $p \in \text{r-ann}(u)$, then $(u, u')p \in (u, u')A_{\ge 1} \subseteq Z$. As $up = 0$, this forces $(0, u') \in Z$, and hence $u'p = 0$. Similarly, $u'p = 0$ forces $up = 0$, and hence $\text{r-ann}(u) = \text{r-ann}(u')$. Thus, there is an isomorphism $\gamma' : uA \cong u'A$, which restricts to an isomorphism $\gamma'' : uA \cap W \to u'A \cap W'$.

We claim that any other isomorphism $\psi : uA \cap W \to u'A \cap W'$ must be a scalar multiple of γ''. Put $P = uA \cap W$, which is 1-critical. As we have already proved, $\dim_{\Bbbk} P_n \leq 1$ for all n with equality for $n \gg 0$. Choose n such that $P_n \neq 0$ and fix $0 \neq x \in P_n$. Then xA is also 1-critical and so $xA \doteq P$. Now $\psi(x) = \lambda\gamma''(x)$ for some $\lambda \in \Bbbk^{\times}$, and this forces ψ to equal $\lambda\gamma''$ on all of xA. Given homogeneous $y \in P$ with $y \notin xA$, then $0 \neq yz \subseteq xA$ for some $z \in A_m$, $m \gg 0$, and so it is easy to see that this forces $\psi(y) = \lambda\gamma(y)$ also. Thus $\psi = \lambda\gamma''$ and the claim follows.

Therefore, possibly after multiplying by a scalar, we can assume that $\gamma'' = \gamma'|_{uA\cap W} = \gamma|_{uA\cap W}$. Thus, we can extend γ to $W + uA$, contradicting the maximality of W. Hence N is indeed critical. Finally, as $M \ncong M'$ with $\dim M_0 = 1 = \dim M'_0$, certainly $W \subseteq M_{\geq 1}$, and so $\dim N_0 = 2$, contradicting part (2).

(1) Since the tails $M_{\geq n_0}$ of 1-critical A-modules are in one-to-one correspondence with the points of E, this follows from part (4). □

We end the section with a technical consequence of these results for subalgebras of T.

Lemma 3.8. *Let U be a noetherian cg algebra and M a finitely generated, graded 1-critical right U-module. Then* $\text{r-ann}_U(M)$ *is prime and* $\text{r-ann}_U(M) = \text{r-ann}_U(N)$ *for every nonzero submodule $N \subseteq M$.*

Proof. This is a standard application of ideal invariance; use, for example, [McConnell and Robson 2001, Corollary 8.3.16 and the proof of (iii) ⇒ (iv) of Theorem 6.8.26]. □

Corollary 3.9. *Assume that T satisfies Assumption 2.1 and let U be a g-divisible subalgebra of $T_{(g)}$ with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Suppose that M and N are 1-critical right U-modules which are cyclic, generated in degree 0, with $g \in \text{r-ann}_U(M)$. For some $n \geq 0$ with $M_n \neq 0$, suppose that there exists $m \geq 0$ such that $(\text{r-ann}_U M_n)_{\geq m} = (\text{r-ann}_U N_n)_{\geq m}$. Then $M \cong N$.*

Proof. By hypothesis, $g^m \in (\text{r-ann}_U M_n)_{\geq m} = (\text{r-ann}_U N_n)_{\geq m}$. Then $N_0 g^{n+m} \subseteq N_n g^m = 0$. As g is central and N is generated by N_0, it follows that $g^{n+m} \in \text{r-ann}_U N$ and hence $g \in \text{r-ann}_U N$ by Lemma 3.8. Thus, both M and N are modules over $A = \overline{U}$, and to prove the lemma it suffices to consider modules over that ring. By Lemma 2.10, $Q_{\mathrm{gr}}(A) = Q_{\mathrm{gr}}(\overline{T}) = k(E)[t, t^{-1}; \tau]$, and so A satisfies the hypotheses of Theorem 3.1.

Clearly, $N_n \neq 0$. By Corollary 3.7(2), $\dim M_n = 1 = \dim N_n$, and so

$$M_n A[n] \cong A/I \quad \text{and} \quad N_n A[n] \cong A/J$$

for some graded right ideals I, J. By hypothesis, $I_{\geq m} = J_{\geq m}$. However, as $I/I_{\geq m}$ is finite-dimensional and A/I is 1-critical, $I/I_{\geq m}$ is the unique largest finite-dimensional submodule of $A/I_{\geq m} = A/J_{\geq m}$. Hence $I = J$ and $M_n A \cong N_n A$. By Corollary 3.7(4), $M \cong N$. □

4. Right ideals of T and the rings $T(\boldsymbol{d})$

Throughout this section, let T satisfy Assumption 2.1. Our first aim in this section is to describe certain graded right ideals J of T such that T/J is filtered by shifted point modules. In fact, the main method we use in the next section to understand a general subalgebra U of T is to compare its graded pieces with the graded pieces of these right ideals J and their left-sided analogues. The easiest way to construct the required right ideals J is to use some machinery from [Van den Bergh 2001]. The details will appear in a companion paper to this one [RSS 2015].

Definitions 4.1. Given a right ideal I of a cg algebra R, the *saturation* I^{sat} of I is the sum of the right ideals $L \supseteq I$ with $\dim_{\Bbbk} L/I < \infty$. If $I = I^{\text{sat}}$, we say that I is *saturated*.

Recall that $T/gT \cong B = B(E, \mathcal{M}, \tau)$, where $\deg \mathcal{M} = \mu$. For divisors $\boldsymbol{b}, \boldsymbol{c}$ on E, we write $\boldsymbol{b} \geq \boldsymbol{c}$ if $\boldsymbol{b} - \boldsymbol{c}$ is effective. A list of divisors $(\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1})$ on E is an *allowable divisor layering* if $\tau^{-1}(\boldsymbol{d}^{i-1}) \geq \boldsymbol{d}^i$ for all $1 \leq i \leq k-1$. By convention, we define $\boldsymbol{d}^i = 0$ for all $i \geq k$. Given an allowable divisor layering $\boldsymbol{d}^\bullet = (\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1})$ on E, let $J(\boldsymbol{d}^\bullet)$ be the saturated right ideal of T defined in [RSS 2015, Definition 3.4].

We omit the precise definition of $J(\boldsymbol{d}^\bullet)$ because it is technical, and not essential in this paper. Instead, what matters are the following properties of this right ideal, which help explain the name "divisor layering". For any graded right T-module M, we think of the B-module Mg^j/Mg^{j+1} as the j-th layer of M. Recall that we write $\pi(N)$ for the image of a finitely generated graded B-module N in the quotient category qgr-B. Recall also from the proof of Corollary 3.7(1) that there is an equivalence of categories $\operatorname{coh} E \simeq$ qgr-B given by $\mathcal{F} \mapsto \pi\left(\bigoplus_{n\geq 0} H^0(E, \mathcal{F} \otimes \mathcal{M}_n)\right)$.

Lemma 4.2 [RSS 2015, Lemma 3.5]. *Let $\boldsymbol{d}^\bullet$ be an allowable divisor layering and let $J = J(\boldsymbol{d}^\bullet)$ and $M = T/J$.*

(1) *If $M^j = Mg^j/Mg^{j+1}$, then as objects in* qgr-B *we have*

$$\pi(M^j) \cong \pi\left(\bigoplus_{n\geq 0} H^0(E, (\mathcal{O}_E/\mathcal{O}_E(-\boldsymbol{d}^j)) \otimes \mathcal{M}_n)\right).$$

In particular, the divisor $\boldsymbol{d}^j$ determines the point modules that occur in a filtration of M^j by (tails of) point modules.

(2) $(\bar{J})^{\text{sat}} = \bigoplus_{n\geq 0} H^0(E, \mathcal{M}_n(-\boldsymbol{d}^0))$.

(3) *If $\boldsymbol{d}^\bullet = (\boldsymbol{d})$ has length 1, then $J(\boldsymbol{d}) = \bigoplus_{n\geq 0}\{x \in T_n : \bar{x} \in H^0(E, \mathcal{M}_n(-\boldsymbol{d}))\}$.*

Note that, as a special case of part (3) of the lemma, if $p \in E$ and $\boldsymbol{d} = p$ then $J(p)$ is simply the right ideal of T such that $P(p) = T/J(p)$ is *the point module* corresponding to the point p. (We note that this will coincide with the earlier definition of a point module, should T be generated in degree one.)

We will require primarily the following two special cases of the construction above. Starting now, it will be sometimes convenient to employ the notation

$$p_i = \tau^{-i}(p) \quad \text{for any } p \in E. \tag{4.3}$$

Definition 4.4. Given any $p \in E$, $i \geq 1$, and $0 \leq r \leq d \leq \mu$, we define $Q(i, d, r, p) = J(\boldsymbol{d}^\bullet)$, where

$$\begin{aligned}
\boldsymbol{d}^0 &= dp + dp_1 + \cdots + dp_{i-1},\\
\boldsymbol{d}^1 &= dp_1 + \cdots + dp_{i-1},\\
&\vdots\\
\boldsymbol{d}^{i-2} &= dp_{i-2} + dp_{i-1},\\
\boldsymbol{d}^{i-1} &= rp_{i-1}.
\end{aligned}$$

Intuitively, the divisor layers for Q are in the form of a triangle, but the vanishing in the last layer is allowed to be of lower multiplicity than in the others. The other special case we need is a similar triangle shape which allows for the involvement of points from different orbits.

Definition 4.5. For any divisor $\boldsymbol{d}$ and $k \geq 1$, we define $M(k, \boldsymbol{d}) = J(\boldsymbol{c}^\bullet)$, where

$$\begin{aligned}
\boldsymbol{c}^0 &= \boldsymbol{d} + \tau^{-1}(\boldsymbol{d}) + \cdots + \tau^{-k+1}(\boldsymbol{d}),\\
\boldsymbol{c}^1 &= \tau^{-1}(\boldsymbol{d}) + \cdots + \tau^{-k+1}(\boldsymbol{d}),\\
&\vdots\\
\boldsymbol{c}^{k-1} &= \tau^{-k+1}(\boldsymbol{d}).
\end{aligned}$$

It is useful to also define $M(k, \boldsymbol{d}) = T$ by convention, for any $k \leq 0$.

Note that $M(k, dp) = Q(k, d, d, p)$ for any $k, d \geq 0$. The right ideals $M(k, \boldsymbol{d})$ are also useful for defining important subalgebras of T.

Definition 4.6. For any divisor $\boldsymbol{d}$ with $\deg \boldsymbol{d} < \mu$ we set

$$T(\boldsymbol{d}) := \bigoplus_{n \geq 0} M(n, \boldsymbol{d})_n,$$

which by [RSS 2015, Theorem 5.3(2)] is a g-divisible subalgebra of T. More generally, for any $\ell \geq 0$ we define

$$T_{\leq \ell} * T(\boldsymbol{d}) := \bigoplus_{n \geq 0} M(n - \ell, \boldsymbol{d}^{\tau^\ell})_n,$$

which by [RSS 2015, Proposition 5.2(2)] is a right g-divisible $T(\boldsymbol{d})$-module.

When $\deg \boldsymbol{d} \le \mu - 2$, but not in general, the module $T_{\le \ell} * T(\boldsymbol{d})$ is equal to the right $T(\boldsymbol{d})$-module $T_{\le \ell} T(\boldsymbol{d}) \subseteq T$ [RSS 2015, Theorem 5.3(6)], so the notation is chosen to suggest multiplication. As is discussed in [Rogalski 2011] and [RSS 2015, §5], the ring $T(\boldsymbol{d})$ should be thought of as corresponding geometrically to a blowup of T at the divisor $\boldsymbol{d}$.

There are left-sided versions of all of the above definitions and results, because Assumption 2.1 is left-right symmetric. We quickly state these analogues, because there are some nonobvious differences in the statements, which result from the fact that the equivalence of categories $\operatorname{coh} E \simeq B\text{-qgr}$ has the slightly different form $\mathcal{F} \to \pi\big(\bigoplus_{n \ge 0} H^0(E, \mathcal{M}_n \otimes \mathcal{F}^{\tau^{n-1}})\big)$. Generally, τ^{-1} appears in the left-sided results wherever τ appears in the right-sided version. A list of divisors $\boldsymbol{d}^\bullet = (\boldsymbol{d}^0, \boldsymbol{d}^1, \dots, \boldsymbol{d}^{k-1})$ on E is a *left allowable divisor layering* if $\tau(\boldsymbol{d}^{i-1}) \ge \boldsymbol{d}^i$ for all $1 \le i \le k-1$. We indicate left-sided versions by a prime in the notation. In particular, given a left allowable divisor layering, there is a corresponding saturated left ideal $J'(\boldsymbol{d}^\bullet)$ of T, defined in [RSS 2015, §6], which satisfies the following analogue of Lemma 4.2.

Lemma 4.7 [RSS 2015, Lemma 6.1]. *Let $\boldsymbol{d}^\bullet$ be a left allowable divisor layering and let $J' = J'(\boldsymbol{d}^\bullet)$ and $M = T/J'$.*

(1) *If $M^j = Mg^j/Mg^{j+1}$ is the j-th layer of M, then in B-qgr we have*
$$\pi(M^j) \cong \pi\left(\bigoplus_{n \ge 0} H^0(E, \mathcal{M}_n \otimes (\mathcal{O}_E/\mathcal{O}_E(-\tau^{-n+1}(\boldsymbol{d}^j))))\right).$$

(2) $(\overline{J'})^{\text{sat}} = \bigoplus_{n \ge 0} H^0(E, \mathcal{M}_n(-\tau^{-n+1}(\boldsymbol{d}^0)))$.

(3) *If $\boldsymbol{d}^\bullet = (\boldsymbol{d})$ has length 1, then*
$$J'(\boldsymbol{d}) = \bigoplus_{n \ge 0} \{x \in T_n : \bar{x} \in H^0(E, \mathcal{M}_n(-\tau^{-n+1}(\boldsymbol{d})))\}. \qquad \square$$

Similarly as on the right, as a special case of part (3) we have that $P'(p) = T/J'(p)$ is the left point module of T corresponding to p.

Of course, we also have left-sided analogues of Definitions 4.4 and 4.5, but we only need the former. Namely, given any $p \in E$, $i \ge 1$, and $0 \le r \le d \le \mu$, we define $Q'(i, d, r, p) = J'(\boldsymbol{d}^\bullet)$, where
$$\begin{aligned}
\boldsymbol{d}^0 &= dp + dp_{-1} + \cdots + dp_{-i+1}, \\
\boldsymbol{d}^1 &= dp_{-1} + \cdots + dp_{-i+1}, \\
&\;\;\vdots \\
\boldsymbol{d}^{i-2} &= dp_{-i+2} + dp_{-i+1}, \\
\boldsymbol{d}^{i-1} &= rp_{-i+1}.
\end{aligned}$$

The right ideals Q and their left-sided analogues Q' will be used below to define filtrations in which every factor is a shifted point module; we will then study how an arbitrary subalgebra U of T intersects such filtrations. The relevant result for this is as follows.

Lemma 4.8 [RSS 2015, Lemma 6.5]. *Let $i, r, d, n \in \mathbb{N}$, with $i < n$ and $1 \le r \le d \le \mu$, and $p \in E$. Then*:

(1) $Q(i, r, d, p) \subseteq Q(i, r-1, d, p)$, *with factor*

$$[Q(i, r-1, d, p)/Q(i, r, d, p)]_{\ge n} \cong P(p_{i-n-1})[-n].$$

(2) $Q'(i, r, d, p) \subseteq Q'(i, r-1, d, p)$, *with factor*

$$[Q'(i, r-1, d, p)/Q'(i, r, d, p)]_{\ge n} \cong P'(p_{-i+n+1})[-n]. \qquad \square$$

The left and right ideals defined above are actually closely related. In fact, by [RSS 2015, Proposition 6.8] one always has $Q(i, r, d, p)_n = Q'(i, r, d, p_{i-n})_n$, as we will exploit in the next section.

We conclude this section with a review of some important homological concepts.

Definition 4.9. A ring A is called *Auslander–Gorenstein* if it has finite injective dimension and satisfies the *Gorenstein condition*: if $p < q$ are nonnegative integers and M is a finitely generated A-module, then $\operatorname{Ext}^p_A(N, A) = 0$ for every submodule N of $\operatorname{Ext}^q_A(M, A)$. Set $j(M) = \min\{r : \operatorname{Ext}^r_A(M, A) \neq 0\}$ for the *homological grade* of M. Then an Auslander–Gorenstein ring A of finite Gelfand–Kirillov dimension is called *Cohen–Macaulay* (or CM) provided that $j(M) + \operatorname{GKdim}(M) = \operatorname{GKdim}(A)$ holds for every finitely generated A-module M. A cg $\Bbbk$-algebra A is called *Artin–Schelter (AS) Gorenstein* if A has injective dimension d and $\dim_{\Bbbk} \operatorname{Ext}^j_A(\Bbbk, A) = \delta_{j,d}$ for all $j \ge 0$. An AS Gorenstein algebra is called *AS regular* if it is also has finite global dimension d.

As the next two results show, many of the algebras appearing in this paper do satisfy these conditions, and this automatically leads to some nice consequences.

Proposition 4.10. *Let $R = T(\boldsymbol{d}) \subseteq T$ for some effective divisor $\boldsymbol{d}$ with $\deg \boldsymbol{d} \le \mu - 1$, in the notation of Assumption 2.1. Then the following hold*:

(1) $R/gR = B(E, \mathcal{M}(-\boldsymbol{d}), \tau)$.

(2) *If $\deg \boldsymbol{d} < \mu - 1$ then R is generated as an algebra in degree 1, while if $\deg \boldsymbol{d} = \mu - 1$ then R is generated as an algebra in degrees 1 and 2.*

(3) *Both R and R/gR are Auslander–Gorenstein and CM.*

(4) *R is a maximal order in $Q_{\mathrm{gr}}(R) = Q_{\mathrm{gr}}(T)$.*

Proof. Combine [RSS 2015, Theorem 5.3] and [Levasseur 1992, Theorem 6.6]. $\square$

Lemma 4.11. *Fix a cg noetherian domain A that is Auslander–Gorenstein and CM. Set* $\mathrm{GKdim}(A) = \alpha$.

(1) *If N is a finitely generated graded right (or left) A-submodule of $Q = Q_{\mathrm{gr}}(A)$ then N^{**} is the unique largest submodule $M \subseteq Q$ with* $\mathrm{GKdim}(M/N) \leq \alpha - 2$.

(2) *In particular, there is no graded A-module $A \subsetneq N \subset Q$ with* $\mathrm{GKdim}(N/A) \leq \mathrm{GKdim}(A) - 2$.

(3) *If $J = J^{**} \neq A$ is a proper reflexive right ideal of A then A/J is* $(\alpha - 1)$-pure *in the sense that* $\mathrm{GKdim}(I/J) = \mathrm{GKdim}(A/J) = \alpha - 1$ *for every nonzero A-module $I/J \subseteq A/J$.*

(4) *If N is a finitely generated A-module, then $\mathrm{Ext}_A^{j(N)}(N, A)$ is a pure module with Gelfand–Kirillov dimension equal to* $\mathrm{GKdim}(N)$.

Proof. Part (1) follows from [Björk and Ekström 1990, Theorem 3.6 and Example 3.2]. Parts (2) and (3) are special cases of (1), while part (4) follows by [ibid., Lemma 2.8]. □

5. An equivalent $T(\boldsymbol{d})$

Throughout this section, T will be an algebra satisfying Assumption 2.1, and we maintain all of the notation introduced in Section 4. In this section we prove that if U is a g-divisible graded subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$ then U is an equivalent order to some $T(\boldsymbol{d})$. This should be compared with [Rogalski 2011, Theorem 1.2]: the rings $T(\boldsymbol{d})$ with $\boldsymbol{d}$ effective of degree $< \mu - 1$ are precisely the maximal orders $U \subseteq T$ with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$ that are generated in degree 1.

We begin by studying $\bar{U}$ and related subalgebras of $\Bbbk(E)[t, t^{-1}; \tau]$. We say that two divisors $\boldsymbol{x}$ and $\boldsymbol{y}$ are *τ-equivalent* if, for every orbit $\mathbb{O}$ of τ on E, one has $\deg(\boldsymbol{x}|_{\mathbb{O}}) = \deg(\boldsymbol{y}|_{\mathbb{O}})$. Two invertible sheaves $\mathcal{O}_E(\boldsymbol{x})$ and $\mathcal{O}_E(\boldsymbol{y})$ are then *τ-equivalent* if the divisors $\boldsymbol{x}$ and $\boldsymbol{y}$ are τ-equivalent.

Lemma 5.1. *Let $\mathcal{N}, \mathcal{N}'$ be ample invertible sheaves on E of the same degree. Let $R := B(E, \mathcal{N}, \tau)$ and $R' := B(E, \mathcal{N}', \tau)$, and let F be an (R', R)-subbimodule of $\Bbbk(E)[t, t^{-1}; \tau]$. Then F_R is finitely generated if and only if ${}_{R'}F$ is finitely generated. In this case, $\mathcal{N}$ and $\mathcal{N}'$ are τ-equivalent.*

Proof. Suppose that F is a finitely generated right R-module. By [Artin and Van den Bergh 1990, Theorem 1.3] there is an invertible sheaf $\mathcal{F}$ on E so that $F \doteq \bigoplus_n H^0(E, \mathcal{F}\mathcal{N}_n)$. For $m, n \gg 0$, ampleness ensures that the sheaves $\mathcal{N}'_m$ and $\mathcal{F}\mathcal{N}_n$ are generated by their sections and, by construction, those sections are $R'_m = H^0(E, \mathcal{N}'_m)$ and $F_n = H^0(E, \mathcal{F}\mathcal{N}_n)$, respectively, again for $m, n \gg 0$. Since F is a left R'-submodule, $R'_m F_n \subseteq F_{n+m}$ for all m, n, and so these observations imply that

$$\mathcal{N}'_m \mathcal{F}^{\tau^m} \mathcal{N}_n^{\tau^m} \subseteq \mathcal{F}\mathcal{N}_{n+m}$$

for all $n, m \gg 0$. By hypothesis, $\mathcal{N}'_m\mathcal{F}^{\tau^m}\mathcal{N}_n^{\tau^m}$ and $\mathcal{F}\mathcal{N}_{n+m}$ have the same degree, and therefore they are equal. In addition, for $n, m \gg 0$ the sheaves $\mathcal{N}'_m$ and $\mathcal{F}^{\tau^m}\mathcal{N}_n^{\tau^m}$ have degree ≥ 3. Thus, by [Rogalski 2011, Lemma 3.1], the map

$$H^0(E, \mathcal{N}'_m) \otimes H^0(E, \mathcal{F}^{\tau^m}\mathcal{N}_n^{\tau^m}) \to H^0(E, \mathcal{N}'_m\mathcal{F}^{\tau^m}\mathcal{N}_n^{\tau^m})$$

is surjective. Thus, $R'_m F_n = F_{n+m}$ for all $n, m \gg 0$ and ${}_{R'}F$ is finitely generated. By symmetry, if ${}_{R'}F$ is finitely generated then so is F_R.

In either case, it follows that $\mathcal{N}'_m = \mathcal{F}\mathcal{N}_m(\mathcal{F}^{-1})^{\tau^m}$ for all $m \gg 0$. The identity $\mathcal{N}'(\mathcal{N}'_m)^\tau = \mathcal{N}'_{m+1}$ gives

$$\mathcal{N}'\mathcal{F}^\tau\mathcal{N}_m^\tau(\mathcal{F}^{-1})^{\tau^{m+1}} = \mathcal{N}'(\mathcal{N}'_m)^\tau = \mathcal{N}'_{m+1} = \mathcal{F}\mathcal{N}_{m+1}(\mathcal{F}^{-1})^{\tau^{m+1}} = \mathcal{F}\mathcal{N}\mathcal{N}_m^\tau(\mathcal{F}^{-1})^{\tau^{m+1}}.$$

Rearranging gives $\mathcal{N}' = \mathcal{F}\mathcal{N}(\mathcal{F}^{-1})^\tau$, which is certainly τ-equivalent to $\mathcal{N}$. $\square$

We next need two technical results on subalgebras of $B(E, \mathcal{M}, \tau)$ that modify the data given by Theorem 3.1.

Notation 5.2. Recall from (4.3) that given a closed point $p \in E$ we write $p_0 = p$ and $p_n = \tau^{-n}(p)$ for all $n \in \mathbb{Z}$. We will also write $\boldsymbol{x}^\tau = \tau^{-1}(\boldsymbol{x})$ when $\boldsymbol{x}$ is a divisor (or closed point) on E, to distinguish left and right actions, and set $\boldsymbol{x}_n = \boldsymbol{x} + \boldsymbol{x}^\tau + \cdots + \boldsymbol{x}^{\tau^{n-1}}$.

We start with a routine consequence of Theorem 3.1.

Corollary 5.3. *Let $A \subseteq B = B(E, \mathcal{M}, \tau)$ be a cg algebra with $Q_{\mathrm{gr}}(A) = Q_{\mathrm{gr}}(B)$. Then there exist $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{Div}(E)$ and $k \in \mathbb{Z}_{\geq 1}$ so that*

$$A_n = H^0(E, \mathcal{M}_n(-\boldsymbol{y} - \boldsymbol{x}_n)) \quad \textit{for all } n \geq k. \tag{5.4}$$

Furthermore, $\mu > \deg \boldsymbol{x} \geq 0$, and

$$\textit{for any } n \geq k \textit{ and divisor } \boldsymbol{c} > \boldsymbol{y} + \boldsymbol{x}_n \textit{ we have } A_n \not\subseteq H^0(E, \mathcal{M}_n(-\boldsymbol{c})). \tag{5.5}$$

Proof. By Theorem 3.1, there exist an integer $k \geq 1$, an ideal sheaf $\mathcal{Y}$ and an ample invertible sheaf $\mathcal{N}$ on E so that

$$A_n = H^0(E, \mathcal{Y}\mathcal{N}_n) \quad \text{for all } n \geq k. \tag{5.6}$$

Let $\mathcal{Y} = \mathcal{O}_E(-\boldsymbol{y})$ for some divisor $\boldsymbol{y}$ and write $\mathcal{N} = \mathcal{M}(-\boldsymbol{x})$ for the appropriate divisor $\boldsymbol{x}$ on E; thus (5.4) is just a restatement of (5.6). Further, $\deg \boldsymbol{x} = \mu - \deg \mathcal{N}$, which, as $\mathcal{N}$ is ample, implies that $\deg \boldsymbol{x} < \mu$. On the other hand, Riemann–Roch implies that

$$\mu n = \deg \mathcal{M}_n = \dim B_n \geq \dim A_n = n \deg(\mathcal{N}) - \deg \boldsymbol{y} = n(\mu - \deg \boldsymbol{x}) - \deg \boldsymbol{y}$$

for $n \gg 0$. Therefore, $\deg \boldsymbol{x} \geq 0$.

Finally, since $\mathcal{N}$ is τ-ample, after possibly increasing k we can assume that $\mathcal{M}_n(-\boldsymbol{y}-\boldsymbol{x}_n)$ is generated by its sections A_n for all $n \geq k$ (see for example [Artin and Stafford 1995, Lemma 4.2(1)]). Thus for any larger divisor $\boldsymbol{c} > \boldsymbol{y}+\boldsymbol{x}_n$ we will have $H^0(E, \mathcal{M}_n(-\boldsymbol{c})) \subsetneq H^0(E, \mathcal{M}_n(-\boldsymbol{y}-\boldsymbol{x}_n))$. Thus (5.5) also holds. □

We next want to modify Corollary 5.3 so that $\boldsymbol{x}$ is replaced by an effective divisor, although this will result in a weaker version of (5.4).

Proposition 5.7. *Let $A \subseteq B = B(E, \mathcal{M}, \tau)$ be a cg algebra with $Q_{\mathrm{gr}}(A) = Q_{\mathrm{gr}}(B)$. Then there is an effective divisor $\boldsymbol{d}$ on E, supported at points with distinct orbits and with $\deg \boldsymbol{d} < \mu$, so that A and $C = B(E, \mathcal{M}(-\boldsymbol{d}), \tau)$ are equivalent orders. Moreover, $\boldsymbol{d}$ and $k \in \mathbb{Z}_{\geq 1}$ can be chosen so that*

$$\begin{aligned} A_n \subseteq H^0(E, \mathcal{M}_n(-\boldsymbol{d}^{\tau^k} - \cdots - \boldsymbol{d}^{\tau^{n-1}})) \\ = H^0(E, \mathcal{O}_E(\boldsymbol{d}_k) \otimes \mathcal{M}(-\boldsymbol{d})_n) \quad \textit{for all } n \geq k. \end{aligned} \tag{5.8}$$

Proof. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the divisors constructed in the proof of Corollary 5.3, and let k be the integer from that result. Fix an orbit $\mathbb{O}$ of τ on E. By possibly enlarging k we can pick $p \in \mathbb{O}$ so that, using the notation of 5.2,

$$\begin{aligned} &\boldsymbol{x}|_{\mathbb{O}} \text{ is supported on } \{p = p_0, \ldots, p_k\}, \quad \text{and} \\ &\boldsymbol{y}|_{\mathbb{O}} \text{ is supported on } \{p_0, \ldots, p_{k-1}\}. \end{aligned} \tag{5.9}$$

Thus

$$\boldsymbol{y}|_{\mathbb{O}} = \sum_{i=0}^{k-1} y_i p_i \quad \text{and} \quad \boldsymbol{x}|_{\mathbb{O}} = \sum_{i=0}^{k} x_i p_i$$

for some integers y_i and x_j. For $n \in \mathbb{N}$ we have

$$(\boldsymbol{x}_n)|_{\mathbb{O}} = \sum_{i=0}^{k} x_i (p_i + p_{i+1} + \cdots + p_{i+n-1}).$$

Thus, for $n \geq k$ we calculate that

$$\begin{aligned} (\boldsymbol{y}+\boldsymbol{x}_n)|_{\mathbb{O}} = (y_0+x_0)p_0 + \cdots + \Big(y_j + \sum_{i \leq j} x_i\Big) p_j + \cdots + \Big(y_{k-1} + \sum_{i \leq k-1} x_i\Big) p_{k-1} \\ + \Big(\sum_{i \geq 0} x_i\Big)(p_k + \cdots + p_{n-1}) + \Big(\sum_{i \geq 1} x_i\Big) p_n + \cdots + \Big(\sum_{i \geq j} x_i\Big) p_{n+j-1} \\ + \cdots + x_k p_{n+k-1}. \end{aligned} \tag{5.10}$$

Let $e_p = \sum x_i$. Since $A \subseteq B(E, \mathcal{M}, \tau)$, the divisor $\boldsymbol{y}+\boldsymbol{x}_n$ is effective for $n \gg 0$, and so

$$y_j + \sum_{i \leq j} x_i \geq 0 \quad \text{and} \quad \sum_{i \geq j} x_i \geq 0 \tag{5.11}$$

for all $0 \le j \le k$. In particular, $e_p \ge 0$. Let $\boldsymbol{d} = \sum_p e_p p$, where the sum is taken over one closed point p in each orbit $\mathbb{O}$ of τ. Take the maximum of the values of k occurring for the different orbits in the support of $\boldsymbol{x}$ and $\boldsymbol{y}$, and call this also k. From (5.10) and (5.11) we see that, on each orbit $\mathbb{O}$ and hence in general,

$$\boldsymbol{y} + \boldsymbol{x}_n \ge \boldsymbol{d}^{\tau^k} + \cdots + \boldsymbol{d}^{\tau^{n-1}}$$

for all $n \ge k$. In other words, (5.8) holds for this $\boldsymbol{d}$ and k. By construction, $\deg \boldsymbol{d} = \deg \boldsymbol{x} < \mu$.

Finally, let $\mathcal{N} = \mathcal{M}(-\boldsymbol{x})$ and let $\mathcal{Y} = \mathcal{O}_E(-\boldsymbol{y})$. Let $C = B(E, \mathcal{M}(-\boldsymbol{d}), \tau)$ and $C' = B(E, \mathcal{N}, \tau)$. Equation (5.8) can be rephrased as saying that

$$\mathcal{Y}\mathcal{N}_n \subseteq \mathcal{M}_n(-\boldsymbol{d}^{\tau^k} - \cdots - \boldsymbol{d}^{\tau^{n-1}}) = \mathcal{O}(\boldsymbol{d}_k) \otimes \mathcal{M}(-\boldsymbol{d})_n \quad \text{for all } n \ge k.$$

Thus, for $n_0 \gg 0$,

$$C'_{\ge n_0} \subseteq N = \bigoplus_{n \ge n_0} H^0(E, (\mathcal{Y}^{-1} \otimes \mathcal{O}(\boldsymbol{d}_k)) \otimes \mathcal{M}(-\boldsymbol{d})_n).$$

Since $\mathcal{M}(-\boldsymbol{d})$ is τ-ample (because it has positive degree) and $\mathcal{Y}^{-1} \otimes \mathcal{O}(\boldsymbol{d}_k)$ is coherent, [Artin and Stafford 1995, Lemma 4.2(ii)] implies that N is a finitely generated right C-module. Hence, so is $C'C$. Since $\mathcal{N} = \mathcal{M}(-\boldsymbol{x})$ with $\deg \boldsymbol{d} = \deg \boldsymbol{x}$, we can apply Lemma 5.1 to conclude that $C'C$ is a finitely generated left C'-module. Thus C and C' are equivalent orders. By the proof of [ibid., Theorem 5.9(2)], C' is a finitely generated right A-module. Thus C' and A are equivalent orders, and so C and A are also equivalent. □

Definition 5.12. We say that $(\boldsymbol{y}, \boldsymbol{x}, k)$ as given by Corollary 5.3 is *geometric data for* A. If (5.9) holds for $p \in \mathbb{O}$, we say p is a *normalised orbit representative* for this data, and we say that $\boldsymbol{d} = \sum e_p p$ is a *normalised divisor* for $(\boldsymbol{y}, \boldsymbol{x}, k)$. To avoid trivialities, the only orbits considered here are the (finite number of) orbits containing the support of $\boldsymbol{x}$ and $\boldsymbol{y}$. By construction, $\deg \boldsymbol{d} = \deg \boldsymbol{x} < \mu$.

We now use these results to study subalgebras of T, and begin with a general idea of the strategy. Let U be a g-divisible graded subalgebra of T so that $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. By Proposition 2.9, U is automatically a finitely generated, noetherian $\Bbbk$-algebra, so the earlier results of the paper are available to us. Let $(\boldsymbol{y}, \boldsymbol{x}, k)$ be geometric data for $\overline{U}$ and let $\boldsymbol{d}$ be a normalised divisor for $(\boldsymbol{y}, \boldsymbol{x}, k)$. We will show that U and $T(\boldsymbol{d})$ are equivalent orders.

Recall that the right $T(\boldsymbol{d})$-module $T_{\le k} * T(\boldsymbol{d}) = \bigoplus_{n \ge 0} M(n-k, \boldsymbol{d}^{\tau^k})_n$ from Definition 4.6 is g-divisible, with

$$\overline{T_{\le k} * T(\boldsymbol{d})} = \bigoplus_{n \ge 0} H^0(E, \mathcal{M}_n(-\boldsymbol{d}^{\tau^k} - \cdots - \boldsymbol{d}^{\tau^{n-1}})),$$

by Lemma 4.2. In other words, by (5.8),

$$\bar{U} \subseteq \overline{T_{\leq k} * T(\boldsymbol{d})}.$$

Our next goal is to show that this holds without working modulo g: that is, that

$$U \subseteq T_{\leq k} * T(\boldsymbol{d}). \tag{5.13}$$

This will force $\widehat{UT(\boldsymbol{d})}$ to be finitely generated as a right $T(\boldsymbol{d})$-module, which is a key step towards proving that U and $T(\boldsymbol{d})$ are equivalent orders.

Suppose therefore that (5.13) fails, and so $U_{n_0} \not\subseteq (T_{\leq k} * T(\boldsymbol{d}))_{n_0}$ for some n_0. Necessarily, $n_0 > k$. We will find a right T-ideal $Q(i, r, e, p)$ and a left T-ideal $Q'(i, r, e, q)$ such that if we set $I = U \cap Q(i, r, e, p)$ and $J = U \cap Q'(i, r, e, q)$ then U/I and U/J are isomorphic to point modules in large degree. Further, we can choose p and q so that $I_{n_0} = J_{n_0}$. However, Corollary 3.9 can be used to derive precise formulæ for I_{n_0} and J_{n_0}, and we will see that these formulæ are inconsistent, leading to a final contradiction.

In the next few results, we carry out this argument, using induction and a filtration of T by the right ideals Q defined in Section 4. Recall the definition of I^{sat} from Definitions 4.1, and the definitions of $J(\boldsymbol{d}^\bullet)$, $Q(i, r, d, p)$, and their left-sided analogues from Section 4.

Lemma 5.14. *Let U be a g-divisible graded subalgebra of T with $Q_{\text{gr}}(U) = Q_{\text{gr}}(T)$. Suppose that $n > i \geq 1$, $1 \leq r \leq e \leq \mu$, and $j \in \mathbb{Z}$. Suppose further that*

(A) *$U_{\geq n} \subseteq Q(i, r-1, e, p_j)$, but $U_n \not\subseteq Q(i, r, e, p_j)$; and*

(B) *$\bar{U}_m \not\subseteq \overline{J(p_{i+j-n-1})_m} = H^0(E, \mathcal{M}_m(-p_{i+j-n-1}))$ for all $m \geq n$.*

Let $I = U \cap Q(i, r, e, p_j)$, and let $M = U/I^{\text{sat}}$. Then:

(1) *$M_n \neq 0$.*

(2) *M is 1-critical and $Mg = 0$.*

(3) *$(\text{r-ann}_U(M_n))_m = (U \cap J(p_{i+j-n-1}))_m$ for all $m \gg 0$.*

Proof. (1) Let $L = U/I$, so that $M = L/L'$, where L' is the largest finite-dimensional submodule of L. Since $n > i$, it follows from hypothesis (A) and Lemma 4.8(1) that $\dim L_m \leq 1$ for all $m \geq n$ and that

$$U_n J(p_{i+j-n-1}) \subseteq Q(i, r, e, p_j). \tag{5.15}$$

If $M_n = 0$, then $L_n U_m = 0$ for all $m \gg 0$. Then $U_n U_m \subseteq Q(i, r, e, p_j)$ for all $m \gg 0$. By hypothesis (B), $U_m + J(p_{i+j-n-1})_m = T_m$ for $m \geq n$, so $U_n T_m \subseteq Q(i, r, e, p_j)$ for $m \gg 0$ also. Since $Q(i, r, e, p_j)$ is a saturated right T-ideal, $U_n \subseteq Q(i, r, e, p_j)$, contradicting the hypotheses.

(2) By Lemma 4.2(3), $g \in J(p_{i+j-n-1})$, whence $M_{\geq n} \cdot g = 0$ and $\text{r-ann}_U(M) \supseteq U_{\geq n} g = gU_{\geq n}$. By construction, M has no finite-dimensional submodules, and so $Mg = 0$. Thus M is a $\overline{U}$-module. Also, $\dim_{\Bbbk} M_m \leq \dim_{\Bbbk} L_m \leq 1$ for all $m \geq n$ and $M \neq 0$ by part (1), so $\operatorname{GKdim}(M) = 1$. Since M is noetherian, it has a $\overline{U}$-submodule M' maximal with respect to the property $\operatorname{GKdim}(M/M') = 1$. Then M/M' is 1-critical. However, by Corollary 3.7(2) any 1-critical $\overline{U}$-module N has $\dim N_m = 1$ for all $m \gg 0$. Thus M' is finite-dimensional; hence $M' = 0$ and M is 1-critical.

(3) Since M is 1-critical, its cyclic submodule $N = M_n U$ must also be 1-critical. Thus $\dim_{\Bbbk} M_n = 1 = \dim_{\Bbbk} N_n$ for $n \gg 0$, forcing $M \doteq N$. In particular, we must have $\text{r-ann}_U(M_n)_m \subsetneq U_m$ for all $m \gg 0$. By (5.15), $\text{r-ann}_U(M_n) \supseteq U \cap J(p_{i+j-n-1})$. Now Lemma 4.2(3) implies that $J(p_{i+j-n-1})_m$ has codimension 1 in T_m for all $m \in \mathbb{N}$. Thus $\text{r-ann}_U(M_n)_m = (U \cap J(p_{i+j-n-1}))_m$ for all $m \gg 0$. □

Corollary 5.16. *Assume that we have the hypotheses of Lemma 5.14. Assume in addition to* (A), (B) *that we have $e < \mu$ and*

(C) $\overline{U}_{\geq n} \subseteq \overline{J(ep_{j+i-1})} = H^0(E, \mathcal{M}_n(-ep_{j+i-1}))$, *but*
$\overline{U}_n \not\subseteq \overline{J((e+1)p_{j+i-1})} = H^0(E, \mathcal{M}_n(-(e+1)p_{j+i-1}))$.

Then $U_n \cap Q(i, r, e, p_j) = U_n \cap J((e+1)p_{i+j-1})$.

Proof. Let

$$I = (U \cap Q(i, r, e, p_j))^{\text{sat}} \quad \text{and} \quad M = U/I.$$

Similarly, let

$$H = (U \cap J((e+1)p_{i+j-1}))^{\text{sat}} \quad \text{with} \quad N = U/H.$$

Note that $Q(1, d, d, p_{j+i-1}) = J(dp_{j+i-1})$ for any d. Also, since $g \in J(dp_{j+i-1})$, hypothesis (C) is equivalent to $U_{\geq n} \subseteq J(ep_{j+i-1})$ but $U_n \not\subseteq J((e+1)p_{j+i-1})$. Thus, hypothesis (C) implies that the hypothesis (A) of Lemma 5.14 also holds for $(i', r', e') = (1, e+1, e+1)$ and $j' = i + j - 1$. Also, hypothesis (B) for these values is the same as hypothesis (B) for the old values. Since $e < \mu$, the hypotheses of Lemma 5.14 hold for (i', r', e').

We may now apply Lemma 5.14 to M and N. Thus, $M_n, N_n \neq 0$, both M, N are 1-critical and killed by g, and $\text{r-ann}_U(M_n)$ and $\text{r-ann}_U(N_n)$ are both equal to $U \cap J(p_{i+j-n-1})$ in large degree. By Corollary 3.9, we have $M \cong N$ and so $I = H$. Thus, since $U_n \cap Q(i, r, e, p_j)$ and $U_n \cap J((e+1)p_{i+j-1})$ are already saturated in degree n by Lemma 5.14(1), we have

$$U_n \cap Q(i, r, e, p_j) = I_n = H_n = U_n \cap J((e+1)p_{i+j-1}). \qquad \square$$

We also need the left-sided versions of the two preceding results. Since the statements and proofs of these are largely symmetric, we give a combined statement of the left-sided versions, with an abbreviated proof. We note that a consequence of [RSS 2015, Lemmas 3.5 and 6.1] is that

$$\begin{aligned} &J'(dp_j)_n = J(dp_{j+n-1})_n \quad \text{and} \\ &\overline{J'(dp_j)_n} = H^0(E, \mathcal{M}_n(-dp_{j+n-1})) = \overline{J(dp_{j+n-1})_n}. \end{aligned} \tag{5.17}$$

Lemma 5.18. *Let U be a g-divisible graded subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Suppose that $n > i \geq 1$, $1 \leq r \leq e < \mu$, and $h \in \mathbb{Z}$. Suppose further that*

(A′) $U_{\geq n} \subseteq Q'(i, r-1, e, p_h)$, *but* $U_n \not\subseteq Q'(i, r, e, p_h)$;

(B′) $\overline{U}_m \not\subseteq \overline{J'(p_{h-i+n+1})_m} = H^0(E, \mathcal{M}_n(-p_{h-i+n+m}))$ *for* $m \geq n$; *and*

(C′) $\overline{U}_{\geq n} \subseteq \overline{J'(ep_{h-i+1})}$, *but*

$$\overline{U}_n \not\subseteq \overline{J'((e+1)p_{h-i+1})_n} = H^0(E, \mathcal{M}_n(-(e+1)p_{h-i+n})).$$

Then $U_n \cap Q'(i, r, e, p_h) = U_n \cap J'((e+1)p_{h-i+1})$.

Proof. The equalities in (B′), (C′) follow from (5.17). The rest of the proof is symmetric to the proofs of Lemma 5.14 and Corollary 5.16. In particular, one uses part (2) of Lemma 4.8 in place of part (1). □

The next result is the heart of the proof that U and $T(\boldsymbol{d})$ are equivalent orders.

Proposition 5.19. *Let U be a g-divisible cg subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Let $(\boldsymbol{y}, \boldsymbol{x}, k)$ be geometric data for $\overline{U}$ and let $\boldsymbol{d} = \sum e_p p$ be a normalised divisor for this data. Then*

$$U \subseteq T_{\leq k} * T(\boldsymbol{d}).$$

Proof. If $\boldsymbol{d} = 0$ the result is trivial, so we may assume that $\boldsymbol{d} > 0$. Suppose that $U \not\subseteq T_{\leq k} * T(\boldsymbol{d})$.

By [RSS 2015, Lemma 6.6], $T_{\leq k} * T(\boldsymbol{d}) = \bigcap_p T_{\leq k} * T(e_p p)$, where the intersection is over the normalised orbit representatives p. Thus there is some such p so that $U \not\subseteq T_{\leq k} * T(e_p p)$. Let $e = e_p < \mu$. By [RSS 2015, Lemma 6.6], again, for $n \in \mathbb{N}$ we have

$$(T_{\leq k} * T(ep))_n = \bigcap \{Q(i, r, e, p_j)_n : i \geq 1,\ k \leq j \leq n-i,\ 1 \leq r \leq e\}.$$

Thus, there are $i \geq 1$, $1 \leq r \leq e$, and $n, j \in \mathbb{N}$ with $1 \leq k \leq j \leq n-i$ such that

$$U_n \not\subseteq Q(i, r, e, p_j)_n. \tag{5.20}$$

Without loss of generality we can assume that i is minimal such that we can achieve this for some such n, j, r. Note that $i \geq 2$, since $\overline{Q(1, r, e, p_j)} = H^0(E, \mathcal{M}(-rp_j))$ by Lemma 4.2(2), and the sections in $\overline{U_n}$ vanish to multiplicity e at p_j by (5.10). Then choose r minimal (for this i) so that (5.20) holds for some such n, j. Intuitively, we are finding a “divisor triangle” of minimal size i such that the corresponding right ideal does not contain U_n, with deepest layer vanishing condition in this triangle to be of multiplicity r as small as possible.

Claim 1. $U_n \cap Q(i, r, e, p_j) = U_n \cap J((e+1)p_{j+i-1})$.

Proof. We check the hypotheses of Corollary 5.16. Hypothesis (A) follows by minimality of r when $r > 1$. When $r = 1$, then we need $U_{\geq n} \subseteq Q(i, 0, e, p_j)$. Now, by [RSS 2015, (6.7)],

$$Q(i, 0, e, p_j) = Q(i-1, e, e, p_j) \cap Q(i-1, e, e, p_{j+1}). \tag{5.21}$$

Since $U_{\geq n}$ is contained in both $Q(i-1, e, e, p_j)$ and $Q(i-1, e, e, p_{j+1})$ by the minimality of i, hypothesis (A) holds in this case as well.

Note that, by (5.5), the equation (5.10) gives exactly the vanishing (with multiplicities) at points on the τ-orbit of p for the sections in $\bar{U}_n \subseteq H^0(E, \mathcal{M}_n)$. In particular, (B) holds because $i + j - n - 1 < 0$. Similarly, (C) holds by (5.10) since $k \leq j + i - 1 \leq n - 1$. Thus Corollary 5.16 gives the result. □

Claim 2. $U_n \cap Q'(i, r, e, p_h) = U_n \cap J'((e+1)p_{h-i+1})$ for $h = j + i - n$.

Proof. This similarly follows from Lemma 5.18 once we verify the hypotheses of that result. For (B′), note that $h - i + n + m = j + m \geq k + m$ and use (5.10). Hypothesis (C′) follows again from (5.10) since $h - i + n = j$ satisfies $k \leq j \leq n - 1$.

It remains to verify (A′). We will use the equality

$$Q(k, r, m, p)_n = Q'(k, r, m, p_{k-n})_n, \tag{5.22}$$

proven in [RSS 2015, Proposition 6.8(3)]. Thus, $U_n \not\subseteq Q'(i, r, e, p_h)_n$. Now let $n' \geq n$. Suppose that $r > 1$. The minimality hypothesis on r means that, for any j' with $k \leq j' \leq n' - i$, we have $U_{n'} \subseteq Q(i, r-1, e, p_{j'})_{n'}$. In particular, since $k \leq j + n' - n \leq n' - i$, we have

$$U_{n'} \subseteq Q(i, r-1, e, p_{j+n'-n})_{n'} = Q'(i, r-1, e, p_h)_{n'}$$

by (5.22). Thus $U_{\geq n} \subseteq Q'(i, r-1, e, p_h)$. If instead $r = 1$, then

$$\begin{aligned} Q'(i, 0, e, p_h)_{n'} &= Q(i, 0, e, p_{j-n+n'}) \\ &= Q(i-1, e, e, p_{j-n+n'}) \cap Q(i-1, e, e, p_{j-n+n'+1}) \end{aligned}$$

by (5.21) and (5.22). But $U_{n'}$ is contained in both $Q(i-1, e, e, p_{j-n+n'+1})$ and $Q(i-1, e, e, p_{j-n+n'})$ by minimality of i. Thus $U_{\geq n} \subseteq Q'(i, r-1, e, p_h)$ in this case as well, and (A′) holds as needed. □

Claim 3. $U_n \cap Q(i, r, e, p_j) = U_n \cap J((e+1)p_j)$.

Proof. As in the proof of Claim 2, we have $Q(i,r,e,p_j)_n = Q'(i,r,e,p_{j+i-n})_n$, and so that claim gives

$$\begin{aligned} U_n \cap Q(i,r,e,p_j) &= U_n \cap Q'(i,r,e,p_{j+i-n}) \\ &= U_n \cap J'((e+1)p_{j-n+1}) \\ &= U_n \cap J((e+1)p_j), \end{aligned}$$

where we use (5.17) in the last step. □

We can now complete the proof of Proposition 5.19. Combining Claims 1 and 3, we have

$$U_n \cap J((e+1)p_j) = U_n \cap Q(i,r,e,p_j) = U_n \cap J((e+1)p_{i+j-1}). \tag{5.23}$$

Recall that $\overline{U}_n = H^0(E, \mathcal{M}_n(-\boldsymbol{y}-\boldsymbol{x}_n))$ and $i \geq 2$. Thus, by (5.10) and (5.5) we see that, after taking the image of (5.23) in B, the right-hand side vanishes to order e at p_j, while the left-hand side vanishes to order $e+1$ at p_j. This contradiction completes the proof of the proposition. □

We can now quickly prove our first main theorem.

Theorem 5.24. *Let U be a g-divisible graded subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Then there is an effective divisor $\boldsymbol{d}$ on E, supported on points with distinct orbits and with $\deg \boldsymbol{d} < \mu$, so that U is an equivalent order to $T(\boldsymbol{d})$.*

In more detail, for some $\boldsymbol{d}$ the $(U, T(\boldsymbol{d}))$-bimodule $M = \widehat{UT(\boldsymbol{d})}$ is a finitely generated g-divisible right $T(\boldsymbol{d})$-module with $MT = T$. Set $W = \mathrm{End}_{T(\boldsymbol{d})}(M)$. Then $U \subseteq W \subseteq T$, the bimodule M is finitely generated as a left W-module, while W, U, and $T(\boldsymbol{d})$ are equivalent orders.

Remark 5.25. Recall from Lemma 2.10 that, if U be a g-divisible graded subalgebra of T with $D_{\mathrm{gr}}(U) = D_{\mathrm{gr}}(T)$, then $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$ also holds. However, some condition on quotient rings is required for the theorem, since clearly $U = \Bbbk[g]$ is not equivalent to any $T(\boldsymbol{d})$.

Proof. By Lemma 2.10, $Q_{\mathrm{gr}}(\overline{U}) = Q_{\mathrm{gr}}(\overline{T})$ and so we can apply Proposition 5.7 to $A = \overline{U}$. Let $\boldsymbol{d}, k$ be as defined there; thus if $R = T(\boldsymbol{d})$ then $\overline{U}$ and $\overline{R}$ are equivalent orders. By Proposition 5.19, $U \subseteq T_{\leq k} * R$.

Let $M = \widehat{UR}$ and $W = \mathrm{End}_R(M)$. By [RSS 2015, Theorem 5.3(5)], $T_{\leq k} * R$ is a noetherian right R-module and so $M \subseteq T_{\leq k} * R$ is a finitely generated right R-module. Clearly $MT = T$ since $1 \in M \subseteq T$ and so $W \subseteq T$. Thus, by Lemma 2.12(3), ${}_W M$ is finitely generated, and so W and R are equivalent orders. A routine calculation shows that M is a left U-module and so $U \subseteq W$.

Consider the $(\overline{W}, \overline{R})$-bimodule $\overline{M}$. This is finitely generated on both sides, since the same is true of ${}_W M_R$. Thus $\overline{W}$ and $\overline{R}$ are equivalent orders, which, as $\overline{R}$ and $\overline{U}$ are equivalent orders, implies that $\overline{W}$ and $\overline{U}$ are likewise. Finally, as $U \subseteq W \subseteq T$, the hypotheses of the theorem ensure that $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(W) = Q_{\mathrm{gr}}(T)$. Thus, by Proposition 2.16, U and W are equivalent orders, and hence so are U and R. $\square$

Corollary 5.26. *Suppose that $\boldsymbol{u}$ and $\boldsymbol{v}$ are two effective, τ-equivalent divisors with degree* $\deg \boldsymbol{u} \leq \mu - 1$. *Then $T(\boldsymbol{u})$ and $T(\boldsymbol{v})$ are equivalent orders.*

Proof. Consider the construction of the divisor $\boldsymbol{d}$ in Theorem 5.24 starting from the algebra $U = T(\boldsymbol{u})$. Thus $\boldsymbol{d} = \sum e_p p$ is the divisor constructed in Proposition 5.7 and there is considerable flexibility in its choice. To begin, in the proof of Equation (5.4), one sees that $\boldsymbol{y} = 0$ and $\boldsymbol{x} = \boldsymbol{u}$. For each orbit $\mathbb{O}$ of τ, a point p is then chosen such that $\boldsymbol{u}|_{\mathbb{O}}$ is supported on $X_{\mathbb{O}} = \{p_0 = p(\mathbb{O}), p_1, \ldots, p_k\}$. For each such orbit, we can replace p_0 by some p_{-r} and increase k so that both $\boldsymbol{u}|_{\mathbb{O}}$ and $\boldsymbol{v}|_{\mathbb{O}}$ are supported on $X_{\mathbb{O}}$. Then $\boldsymbol{d} = \sum_{\mathbb{O}} e_p p(\mathbb{O})$, for these choices of points $p(\mathbb{O})$, and $e_p = \deg(\boldsymbol{u}|_{\mathbb{O}})$. As $\boldsymbol{u}$ and $\boldsymbol{v}$ are τ-equivalent, $\deg(\boldsymbol{u}|_{\mathbb{O}}) = \deg(\boldsymbol{v}|_{\mathbb{O}})$ for each orbit $\mathbb{O}$, and hence the divisor $\boldsymbol{d}$ is the same whether we started with $T(\boldsymbol{u})$ or $T(\boldsymbol{v})$. Hence, by Theorem 5.24, $T(\boldsymbol{u})$ and $T(\boldsymbol{v})$ are both equivalent to $T(\boldsymbol{d})$ and hence to each other. $\square$

Remark 5.27. One disadvantage of Theorem 5.24 is that the $(U, T(\boldsymbol{d}))$-bimodule M constructed there need not be finitely generated as a left U-module. Using [McConnell and Robson 2001, Proposition 3.1.14] and the fact that our rings are noetherian, one can easily produce such a bimodule. However, this typically lacks the extra structure inherent in M (notably that $MT = T$) and so is less useful for our purposes. As will be seen in the next section, this problem disappears when one works with maximal orders (see Corollary 6.6, for example) and this will in turn give extra information about the structure of such an algebra.

6. On endomorphism rings of $T(\boldsymbol{d})$-modules

Given a g-divisible algebra $U \subseteq T$, Theorem 5.24 provides a module M over some blowup $T(\boldsymbol{d})$ with $U \subseteq \operatorname{End}_{T(\boldsymbol{d})}(M)$. In this section, we reverse this procedure by obtaining detailed properties of such endomorphism rings (see Proposition 6.4 and Theorem 6.7). These results provide important information about the structure of maximal T-orders that will in turn be refined over the next two sections to prove the main result Theorem 1.2 from the introduction.

We begin with an expanded version of a definition from the introduction.

Definition 6.1. Let $U \subseteq V$ be Ore domains with the same quotient ring $Q(U)$. We say that U is a *maximal V-order* if there exists no order $U \subsetneqq U' \subseteq V$ that is equivalent to U. We note that if U and V are graded (in which case requiring that $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(V)$ is sufficient) then this is the same as being maximal among *graded* orders equivalent to U and contained in V. Indeed, suppose that U has the latter property, but that $U \subsetneq A \subseteq V$ for some equivalent order A. If A is given the filtration induced from the graded structure of V, then the associated graded ring

$\operatorname{gr} A$ will still satisfy $U \subsetneq \operatorname{gr} A \subseteq V$ and be equivalent to U, giving the required contradiction.

When $V = Q(U)$, or $V = Q_{\mathrm{gr}}(U)$ if U is graded, a maximal V-order is simply called *a maximal order*.

We are mostly interested in maximal T-orders. We introduce this concept because maximal T-orders need not be maximal orders (see Proposition 10.3), although the difference is not large (see Corollary 6.6). We first want to study the endomorphism ring $\operatorname{End}_{T(\boldsymbol{d})}(M)$ arising from Theorem 5.24, and we begin with two useful lemmas.

Lemma 6.2. *Let A be a noetherian domain with quotient division ring D. If N is a finitely generated right A-submodule of D then $\operatorname{End}_A(N^{**})$ is the unique maximal order among orders containing and equivalent to $\operatorname{End}_A(N)$.*

Proof. This is what is proved in [Cozzens 1976, Theorem 2.7], since $\operatorname{End}_A(N^*) = \operatorname{End}_A(N^{**})$. □

Lemma 6.3. *Let A and B be rings such that A is left noetherian and suppose that M is an (A, B)-bimodule that is finitely generated on both sides, and that N is a finitely generated right B-module. Then $\operatorname{Hom}_B(N, M)$ is a finitely generated left A-module. In particular, $\operatorname{End}_B(M)$ is a finitely generated left A-module, and if B is left noetherian then $N^* = \operatorname{Hom}_B(N, B)$ is a finitely generated left B-module.*

Proof. A surjective B-module homomorphism $B^{\oplus n} \to N$ induces an injective left A-module homomorphism $\operatorname{Hom}_B(N, M) \hookrightarrow \operatorname{Hom}_B(B^{\oplus n}, M) \cong M^{\oplus n}$. Since M is a noetherian left A-module, $\operatorname{Hom}_B(N, M)$ is a finitely generated left A-module. □

We are now ready to prove the first significant result of the section. Until further notice, all duals N^* will be taken as R-modules, for $R = T(\boldsymbol{d})$.

Proposition 6.4. *Let $\boldsymbol{d}$ be an effective divisor on E with $\deg \boldsymbol{d} < \mu$ and let $R = T(\boldsymbol{d})$. Let $M \subseteq T_{(g)}$ be a g-divisible finitely generated graded right R-module with $MT = T$ and set $W = \operatorname{End}_R(M)$ and $F = \operatorname{End}_R(M^{**})$. Then:*

1. *F, $V = F \cap T$ and W are g-divisible algebras with $Q_{\mathrm{gr}}(W) = Q_{\mathrm{gr}}(V) = Q_{\mathrm{gr}}(F) = Q_{\mathrm{gr}}(T)$.*
2. *F is the unique maximal order containing and equivalent to W, while V is the unique maximal T-order containing and equivalent to W.*
3. *There is an ideal K of F with $K \subseteq W$ and $\operatorname{GKdim} F/K \leq 1$.*
4. *$R = \operatorname{End}_W(M) = \operatorname{End}_F(M^{**})$.*

Proof. Since $Q_{\mathrm{gr}}(R) = Q_{\mathrm{gr}}(T)$ by Proposition 4.10, clearly the same is true for W, V and F. As in (2.11), given a right R-module $N \subset Q_{\mathrm{gr}}(R)$ we identify

$$N^* = \operatorname{Hom}_R(N, R) = \{\theta \in Q_{\mathrm{gr}}(R) : \theta N \subseteq R\},$$

and similarly for left modules. By Lemma 2.12(3), W is g-divisible and ${}_W M$ is finitely generated. Thus the left-sided version of Lemma 6.3 shows that $\mathrm{End}_W(M)$ is a finitely generated right R-module. Moreover, by Proposition 4.10, R is a maximal order and so $R = \mathrm{End}_W(M)$.

By Lemma 2.13(3), M^{**} is g-divisible with $M^{**} \subset T_{(g)}$. Since M^{**} is clearly a finitely generated right R-module, the same logic ensures that F is g-divisible, ${}_F M^{**}$ is finitely generated and $\mathrm{End}_F(M^{**}) = R$. By Lemma 6.2, $F \supseteq W$ and F is the unique maximal order containing and equivalent to W. This automatically ensures that $V = F \cap T$ is maximal among T-orders containing and equivalent to W. Clearly V is also g-divisible.

It remains to find the ideal K. By Proposition 2.9, both W and F are noetherian. By Proposition 4.10 and Lemma 4.11(1), $\mathrm{GKdim}_R(M^{**}/M) \le \mathrm{GKdim}(R) - 2 = 1$. Since M is g-divisible, $X = M^{**}/M$ is g-torsionfree and so, by Lemma 2.14, X is a finitely generated right $\Bbbk[g]$-module. Since $M \subseteq M^{**} \subset T_{(g)}$ the action of g is central on X and so X is also a finitely generated left $\Bbbk[g]$-module. Now, it is routine to check that M^{**} and hence X are left W-modules, while $\Bbbk[g] \subseteq W$ since W is g-divisible. Thus, X and hence M^{**} are finitely generated left W-modules. Moreover, $\mathrm{GKdim}_W(X) \le \mathrm{GKdim}_{\Bbbk[g]}(X) \le 1$ and so, by [Krause and Lenagan 1985, Lemma 5.3], $I = \ell\text{-ann}_W(X)$ satisfies $\mathrm{GKdim}(W/I) \le 1$.

Now consider F. First,

$$(IF)M \subseteq IFM^{**} \subseteq IM^{**} \subseteq M$$

and hence $IF = I \subseteq W$. Thus F is a finitely generated right W-module and (on the left) $\mathrm{GKdim}_W(F/W) \le \mathrm{GKdim}(W/I) \le 1$. On the other hand, as ${}_W M^{**}$ is finitely generated, Lemma 6.3 implies that $F = \mathrm{End}_R(M^{**})$ is a finitely generated left W-module. Thus, by [ibid., Lemma 5.3], again, the right annihilator $I' = \text{r-ann}_W(F/W)$ satisfies $\mathrm{GKdim}\, W/I' \le 1$. Thus $K = I'I$ is an ideal of both F and W. By the symmetry of the GK-dimension of bimodules finitely generated on both sides [ibid., Corollary 5.4] and the exactness of the GK-dimension [ibid., Theorem 6.14], $\mathrm{GKdim}(F/K) \le 1$. □

Pairs of algebras (V, F) satisfying the conclusions of the proposition will appear multiple times in this paper and so we turn those properties into a definition. For a case when $F \neq V$, see Proposition 10.3.

Definition 6.5. A pair (V, F) is called a *maximal order pair* if

(1) F and V are g-divisible, cg algebras with $V \subseteq F \subseteq T_{(g)}$ and $V \subseteq T$;

(2) F is a maximal order in $Q_{\mathrm{gr}}(F) = Q_{\mathrm{gr}}(T)$ and $V = F \cap T$ is a maximal T-order;

(3) there is an ideal K of F with $K \subseteq V$ and $\mathrm{GKdim}\, F/K \le 1$.

The next result illustrates the significance of Proposition 6.4 to the structure of maximal T-orders.

Corollary 6.6. *Let $U \subseteq T$ be a g-divisible cg maximal T-order.*

(1) *There exists an effective divisor $\boldsymbol{d}$ on E, with $\deg \boldsymbol{d} < \mu$, and a g-divisible $(U, T(\boldsymbol{d}))$-module $M \subseteq T$ with $MT = T$ that is finitely generated as both a left U-module and a right $T(\boldsymbol{d})$-module. Moreover, $U = \operatorname{End}_{T(\boldsymbol{d})}(M)$ and $T(\boldsymbol{d}) = \operatorname{End}_U(M)$.*

(2) *$(U, F = \operatorname{End}_R(M^{**}))$ is a maximal order pair; in particular, if U is a maximal order then $U = F$.*

(3) *Suppose that every ideal I of $T(\boldsymbol{d})$ satisfying $\operatorname{GKdim}(T(\boldsymbol{d})/I) = 1$ satisfies $\operatorname{GKdim} T/IT \leq 1$ (in particular, this holds if $T(\boldsymbol{d})$ has no such ideals I). Then $U = F$ is a maximal order.*

Proof. (1) By Theorem 5.24, there is an effective divisor $\boldsymbol{d}$ with $\deg \boldsymbol{d} < \mu$ so that

$$U \subseteq V = \operatorname{End}_{T(\boldsymbol{d})}(M) \subseteq T,$$

where $M = \widehat{UT(\boldsymbol{d})}$ is a finitely generated g-divisible graded right $T(\boldsymbol{d})$-module with $MT = T$. By Theorem 5.24 again, V and U are equivalent orders. Since U is a maximal T-order, this forces $U = V$. Finally, $T(\boldsymbol{d}) = \operatorname{End}_U(M)$ by Proposition 6.4.

(2) As $U = V$, this is a restatement of Proposition 6.4(2).

(3) Just as in the proof of Proposition 6.4, $J = \operatorname{r-ann}_R M^{**}/M$ is an ideal of R with $\operatorname{GKdim}(R/J) \leq 1$. Note that since M is g-divisible, either $M = M^{**}$ and $J = R$, or else $\operatorname{GKdim}(R/J) = 1$.

In either case, the hypotheses imply that $\operatorname{GKdim} T/JT \leq 1$. Now $M^{**}JT \subseteq MT = T$. Thus

$$\operatorname{GKdim}(\alpha T + T)/T \leq \operatorname{GKdim} T/JT \leq 1$$

for any $\alpha \in M^{**}$. By Proposition 4.10 and Lemma 4.11(1), this implies that $M^{**} \subseteq T$. This in turn implies that $M^{**}T = T$ and hence that $F \subseteq T$. Since U is a maximal T-order, $U = F$ is a maximal order. □

We now turn to the second main aim of this section, which is to describe the structure of $\overline{U}$ for suitable endomorphism rings $U = \operatorname{End}_{T(\boldsymbol{d})}(M)$. The importance of this result is that the pleasant properties of $\overline{U}$ can be pulled back to U.

Theorem 6.7. *Let $\boldsymbol{d}$ be an effective divisor on E with $\deg \boldsymbol{d} < \mu$, and let $R = T(\boldsymbol{d})$. Let M be a finitely generated g-divisible graded right R-module with $R \subseteq M \subseteq T$. Let $U = \operatorname{End}_R(M)$ and $F = \operatorname{End}_R(M^{**})$. Then there is an effective divisor $\boldsymbol{y}$ on E so that*

$$\overline{F} \doteq \overline{U} \doteq \operatorname{End}_{\overline{R}}(\overline{M}) \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau) \quad \text{for } \boldsymbol{x} = \boldsymbol{d} - \boldsymbol{y} + \tau^{-1}(\boldsymbol{y}). \tag{6.8}$$

Moreover, if $V = F \cap T$ then $U \subseteq V \subseteq F$ and (V, F) is a maximal order pair.

The proof of Theorem 6.7 depends on a series of lemmas that will take the rest of this section. Before getting to those results we make some comments and a definition. We first want to regard the ring F from the theorem as a blowup of T at the divisor $\boldsymbol{x}$ on E, even if $\boldsymbol{x}$ is not effective. We formalise this as follows.

Definition 6.9. Let $\boldsymbol{x}$ be a (possibly noneffective) divisor on E with $0 \le \deg \boldsymbol{x} < \mu = \deg \mathcal{M}$. We say that a cg algebra $F \subseteq T_{(g)}$ is *a blowup of* T *at* $\boldsymbol{x}$ if

(i) F is part of a maximal order pair (V, F) with $Q_{\mathrm{gr}}(F) = Q_{\mathrm{gr}}(T)$; and

(ii) $\overline{F} \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$.

Remarks 6.10. (i) The reader should regard this definition of a blowup as temporary in the sense that it will be refined in Definition 7.1 and justified in Remark 7.5. One caveat about the concept is that there may not be a unique blowup of T at the divisor $\boldsymbol{x}$; in the context of Theorem 6.7 there may be different R-modules M leading to distinct blowups F, which nonetheless have factors $\overline{F}$ which are equal in large degree. See Example 10.4 and Remark 10.7(2).

(ii) It follows easily from Theorem 6.7 that a maximal order pair (V, F) does give a blowup of T at an appropriate (possibly noneffective) divisor $\boldsymbol{x}$. The details are given in Theorem 7.4 which also gives a converse to Theorem 6.7.

(iii) We conjecture that, generically, the blowup $T(\boldsymbol{d})$ will have no sporadic ideals in Theorem 6.7 and so, by Corollary 6.6(3), $U = F$ will then be a maximal order. For an example where this happens see Example 10.4, and, conversely, for an example when $U \neq F$ and $F \not\subseteq T$ see Proposition 10.3.

Notation 6.11. For the rest of the section, we write $N^* = \mathrm{Hom}_U(N, U)$ provided that the ring U is clear from the context. In particular, given a g-divisible left ideal I of R, we have $\overline{I}^* = \mathrm{Hom}_{\overline{R}}(I/gI, \overline{R})$ while $\overline{I^*} = \overline{\mathrm{Hom}_R(I, R)}$. Recall from Lemma 4.11 that a R-module M is *α-pure* provided $\mathrm{GKdim}(M) = \mathrm{GKdim}(N) = \alpha$ for all nonzero submodules $N \subseteq M$.

The main technical result we will need is the following, showing that "bar and star commute" (up to a finite-dimensional vector space).

Proposition 6.12. *Let $R = T(\boldsymbol{d})$ for an effective divisor $\boldsymbol{d}$ with* $\deg \boldsymbol{d} < \mu$.

(1) *Let I be a proper, g-divisible left ideal of R for which R/I is 2-pure. Then I^*/R is a g-torsionfree, 2-pure right module; further, $I^* \subseteq T_{(g)}$ and $\overline{I^*} \doteq \overline{I}^*$.*

(2) *If M is a finitely generated g-divisible graded right R-module with $R \subseteq M \subseteq T$, then $\overline{M^*} \doteq \overline{M}^*$.*

Proof. (1) By Lemma 4.11(2), I^*/R is 2-pure. By Lemma 2.12, $I^* \subseteq T_{(g)}$ and since R is g-divisible, $T_{(g)}/R$ and hence I^*/R are g-torsionfree.

From the exact sequence $0 \to Rg \to R \to \bar{R} \to 0$ we obtain the long exact sequence of right R-modules

$$0 \to \mathrm{Hom}_R(R/I, \bar{R}) \to \mathrm{Ext}^1_R(R/I, Rg) \to \mathrm{Ext}^1_R(R/I, R) \to \mathrm{Ext}^1_R(R/I, \bar{R}) \\ \xrightarrow{\phi} \mathrm{Ext}^2_R(R/I, Rg) \xrightarrow{\psi} \mathrm{Ext}^2_R(R/I, R) \to \mathrm{Ext}^2_R(R/I, \bar{R}) \to \cdots . \quad (6.13)$$

By Proposition 4.10, $\bar{R}$ is Auslander–Gorenstein and CM. Thus $N = \mathrm{Ext}^2_{\bar{R}}(\bar{R}/\bar{I}, \bar{R})$ has grade $j(N) \geq 2$ and hence $\mathrm{GKdim}(N) \leq 2-2 = 0$. Therefore, by [RSS 2015, Lemma 7.9], $\mathrm{Ext}^2_R(R/I, \bar{R}) = N$ is finite-dimensional and the map ψ in (6.13) is surjective in large degree. If $E = \mathrm{Ext}^2_R(R/I, R)$, this says that $\psi : E[-1] \to E$ is surjective in large degree. Since $\dim_{\Bbbk} E_n < \infty$ for each n, this forces $\dim_{\Bbbk} E_n \geq \dim_{\Bbbk} E_{n+1}$ for all $n \gg 0$ and so $\dim_{\Bbbk} E_n$ is eventually constant. In turn, this forces ϕ to be zero in large degree.

Next, observe that $\mathrm{Hom}(R/I, \bar{R}) = 0$ since R/I is g-torsionfree. Since ϕ is zero in high degree, the complex

$$0 \longrightarrow \mathrm{Ext}^1_R(R/I, Rg) \longrightarrow \mathrm{Ext}^1_R(R/I, R) \longrightarrow \mathrm{Ext}^1_R(R/I, \bar{R}) \longrightarrow 0$$

is exact in high degree. Using [RSS 2015, Lemma 7.9] this can be identified with the complex

$$0 \longrightarrow (I^*/R)[-1] \xrightarrow{\alpha} I^*/R \longrightarrow \mathrm{Ext}^1_{\bar{R}}(\bar{R}/\bar{I}, \bar{R}) \longrightarrow 0,$$

where α is multiplication by g. As I^* is g-divisible by Lemma 2.12(2), it follows that

$$\overline{I^*}/\bar{R} \cong I^*/(R + I^*g) = \mathrm{coker}(\alpha) \doteq \mathrm{Ext}^1_{\bar{R}}(\bar{R}/\bar{I}, \bar{R}) = \bar{I}^*/\bar{R}.$$

In particular, $\dim_{\Bbbk} \overline{I^*} = \dim_{\Bbbk} \bar{I}^*$ for all $n \gg 0$, and as there is an obvious inclusion $\overline{I^*} \subseteq \bar{I}^*$ we conclude that $\overline{I^*} \doteq \bar{I}^*$.

(2) Note that M^{**}/M is a g-torsionfree module of GK-dimension 1, as in the proof of Proposition 6.4. By Lemma 2.14, $\dim_{\Bbbk}((M^{**}/M) \otimes_R \bar{R}) < \infty$. Thus $\overline{M^{**}} \doteq \overline{M}$.

Let $J = M^*$. Since J is a reflexive left ideal of R, the module R/J is 2-pure by Lemmas 4.10(3) and 4.11(3). Thus part (1) applies and shows that $\bar{J}^* \doteq \overline{J^*}$. Next, $\bar{J} \doteq \bar{J}^{**}$ by another use of Lemmas 4.10(3) and 4.11(3). Finally, it is easy to see that for any finitely generated graded $\bar{R}$-modules N and Q contained in $Q_{\mathrm{gr}}(\bar{R})$, if $N \doteq Q$ then $N^* \doteq Q^*$. Putting the pieces above together, we conclude that

$$\overline{M^*} = \bar{J} \doteq \bar{J}^{**} \doteq (\overline{J^*})^* \doteq \overline{M}^*. \qquad \square$$

The last ingredient we need for the proof of Theorem 6.7 is the following description of the endomorphism ring of a torsion-free rank-one module over a twisted homogeneous coordinate ring.

Lemma 6.14. *Let $B = B(E, \mathscr{L}, \tau)$, where E is a smooth elliptic curve*, $\deg \mathscr{L} \geq 1$, *and τ is of infinite order. Let N be a finitely generated, graded right B-submodule of $\Bbbk(E)[t, t^{-1}; \tau]$; by* [Artin and Van den Bergh 1990, Theorem 1.3],

$$N \doteq \bigoplus_{r \geq 0} H^0(E, \mathcal{O}(\boldsymbol{q}) \otimes \mathscr{L}_r)$$

for some divisor $\boldsymbol{q}$. Let $N^* = \operatorname{Hom}_B(N, B) \subseteq \Bbbk(E)[t, t^{-1}; \tau]$. Then:

(1) $\operatorname{End}_B(N) \doteq B(E, \mathscr{L}(\boldsymbol{q} - \tau^{-1}(\boldsymbol{q})), \tau)$.

(2) $NN^* \doteq \operatorname{End}_B(N)$.

(3) $N^* \doteq \bigoplus_{n \geq 0} H^0(E, \mathscr{L}_n \otimes \mathcal{O}(-\tau^{-n}(\boldsymbol{q})))$.

Proof. (1) Write $G = \operatorname{End}_B(N) \subseteq \Bbbk(E)[t, t^{-1}; \tau]$ and, for each n, let $\mathscr{G}_n$ be the subsheaf of the constant sheaf $\Bbbk(E)$ generated by $G_n \subseteq \Bbbk(E)$. Let $\mathcal{N}_n = \mathcal{O}(\boldsymbol{q}) \otimes \mathscr{L}_n$; thus $N_n = H^0(E, \mathcal{N}_n)$, and N_n generates the sheaf $\mathcal{N}_n$, for $n \gg 0$, say $n \geq n_0$.

For $n \geq n_0$ and $r \geq 0$, the equation $G_r N_n \subseteq N_{n+r}$ forces $\mathscr{G}_r \mathcal{N}_n^{\tau^r} \subseteq \mathcal{N}_{n+r}$ and thus

$$\mathscr{G}_r \otimes (\mathcal{O}(\boldsymbol{q}) \otimes \mathscr{L}_n)^{\tau^r} \subseteq \mathcal{O}(\boldsymbol{q}) \otimes \mathscr{L}_{n+r}.$$

Equivalently,

$$\mathscr{G}_r \subseteq \mathscr{L}_r(\boldsymbol{q} - \tau^{-r}(\boldsymbol{q})) = (\mathscr{L}(\boldsymbol{q} - \tau^{-1}(\boldsymbol{q})))_r.$$

This shows that

$$G \subseteq B(E, \mathscr{L}(\boldsymbol{q} - \tau^{-1}(\boldsymbol{q})), \tau).$$

Reversing this calculation shows that

$$(\mathscr{L}(\boldsymbol{q} - \tau^{-1}(\boldsymbol{q})))_r \mathcal{N}_n^{\tau^r} \subseteq \mathcal{N}_{n+r}$$

for $r, n \geq 0$ and taking sections for $n \geq n_0$ shows that

$$B(E, \mathscr{L}(\boldsymbol{q} - \tau^{-1}(\boldsymbol{q})), \tau) \subseteq \operatorname{End}_B(N_{\geq n_0}).$$

To complete the proof we need to prove that $G \doteq \operatorname{End}_B(N_{\geq n_0})$. This follows by [Rogalski 2011, Lemma 2.2(2)] and [Artin and Zhang 1994, Proposition 3.5] or by a routine computation.

(2) Clearly NN^* is an ideal of $\operatorname{End}_B(N)$. However, by Lemma 2.7(2), $\operatorname{End}_B(N)$ is just infinite, and so $NN^* \doteq \operatorname{End}_B(N)$.

(3) The proof is similar to that of (1) and, as it will not be used in the paper, is left to the reader. □

Proof of Theorem 6.7. We first check that $\overline{F} \doteq \overline{U}$. By Proposition 6.4 there exists an ideal K of F contained in U and satisfying $\operatorname{GKdim}(F/K) \leq 1$. In particular, $\operatorname{GKdim}(F/U) \leq 1$. By Lemma 2.12(3), U is g-divisible, and so $N = F/U$ is g-torsionfree. It follows from Lemma 2.14 that $\operatorname{GKdim}(\overline{F}/\overline{U}) = 0$, and so $\overline{U} \doteq \overline{F}$.

Now it is obvious that $U \supseteq MM^*$. Thus, using Proposition 6.12(2),

$$\bar{U} \supseteq (\overline{M})(\overline{M^*}) \doteq (\overline{M})(\overline{M}^*).$$

Conversely, by Lemma 2.12(3), $\bar{U} = \overline{\mathrm{End}_R(M)} \subseteq \mathrm{End}_{\bar{R}}(\overline{M})$. We also have $\bar{R} = B(E, \mathcal{M}(-\boldsymbol{d}), \tau)$. Applying Lemma 6.14 to $\mathcal{L} = \mathcal{M}(-\boldsymbol{d})$ and $N = \overline{M}$ gives

$$(\overline{M})(\overline{M}^*) \doteq \mathrm{End}_{\bar{R}}(\overline{M}) \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau),$$

where, in the notation of that lemma, $\boldsymbol{y} = \boldsymbol{q}$ and $\boldsymbol{x} = \boldsymbol{d} - \boldsymbol{y} + \tau^{-1}(\boldsymbol{y})$. That $\boldsymbol{y}$ is effective follows from $\bar{R} \subseteq \overline{M}$. Combining the last two displayed equations gives (6.8).

Since $R \subseteq M$, necessarily $MT = T$. Thus the second paragraph of the theorem is just a restatement of Proposition 6.4. □

7. The structure of g-divisible orders

In this section we first refine the results from the last two sections to give strong structural results for a g-divisible maximal T-order U (see Theorem 7.4). Then we use these results to analyse both arbitrary g-divisible orders and ungraded subalgebras of $D = D_{\mathrm{gr}}(T)$ (see Corollaries 7.6 and 7.10, respectively). In particular, we show that U is part of a maximal order pair (U, F) for which F is a blowup of T at a (possibly noneffective) divisor $\boldsymbol{x} = \boldsymbol{d} - \boldsymbol{y} + \tau^{-1}(\boldsymbol{y})$ in the sense of Definition 6.9. Here, the divisor $\boldsymbol{y}$ can have arbitrarily high degree but is not arbitrary, as we first explain.

Definition 7.1. Let $\boldsymbol{x}$ be a divisor on E. For each τ-orbit $\mathbb{O}$ in E pick $p = p_0 \in \mathbb{O}$ such that $\boldsymbol{x}|_{\mathbb{O}} = \sum_{i=0}^{k} x_i p_i$, where $p_i = \tau^{-i}(p)$. Then $\boldsymbol{x}$ is called a *virtually effective* divisor if for each orbit $\mathbb{O}$ and all $j \in \mathbb{Z}$ the divisor $\boldsymbol{x}$ satisfies

$$\sum_{i \le j} x_i \ge 0 \quad \text{and} \quad \sum_{i \ge j} x_i \ge 0. \tag{7.2}$$

If F is a blowup of T at a virtually effective divisor $\boldsymbol{x}$ then F is called a *virtual blowup* of T.

The relevance of this condition is shown by the next result, in which the notation $\boldsymbol{u}_k$ for a divisor $\boldsymbol{u}$ comes from Notation 5.2.

Proposition 7.3. (1) *The divisor $\boldsymbol{x}$ in Theorem 6.7 is virtually effective.*

(2) *A divisor $\boldsymbol{x}$ is virtually effective if and only if $\boldsymbol{x}$ can be written as*

$$\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v}),$$

where $\boldsymbol{u}$ is an effective divisor supported on distinct τ-orbits and $\boldsymbol{v}$ is an effective divisor such that $0 \le \boldsymbol{v} \le \boldsymbol{u}_k$ for some k.

Proof. (1) By Theorem 6.7, $\overline{F} \doteq B(E, \mathcal{N}, \tau)$, where $\mathcal{N} = \mathcal{M}(-\boldsymbol{x})$. Since $\overline{F} \doteq \overline{U} \subseteq \overline{T} = B(E, \mathcal{M}, \tau)$, we must have $\mathcal{N}_n \subseteq \mathcal{M}_n$ for $n \gg 0$. Now compare this with the computations in the proof of Corollary 5.3. In the notation of that proof, $\mathcal{Y} = \mathcal{O}_E$ and hence $\boldsymbol{y} = 0$. Therefore, as is explained in the proof of (5.11), this forces (7.2) to hold.

(2) It is enough to prove this in the case that $\boldsymbol{x}$ is supported on a single τ-orbit $\mathbb{O}$ in E.

($\Rightarrow$) As in Definition 7.1, write $\boldsymbol{x} = \sum_{i=0}^{k} x_i p_i$ for a suitable point $p_0 \in \mathbb{O}$. Set $e = \sum_{i \in \mathbb{Z}} x_i$ and $\boldsymbol{u} = ep$. For $j \in \mathbb{N}$, let $v_j = \sum_{i \geq j+1} x_i$ and put $\boldsymbol{v} = \sum_{j \geq 0} v_j p_j$.
By (7.2), $\boldsymbol{v}$ is effective. Also, since $\sum_{i \leq j} x_i \geq 0$ for all j, we have

$$v_j = e - \sum_{i \leq j} x_i \leq e \quad \text{for } 0 \leq j \leq k-1,$$

while $v_j = 0$ for $j \geq k$. Therefore, $0 \leq \boldsymbol{v} \leq \boldsymbol{u}_k = \sum_{i=0}^{k-1} ep_i$. Finally,

$$\begin{aligned}
\boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v}) &= ep_0 - \sum_{j \geq 0} \Bigl(\sum_{i \geq j+1} x_i \Bigr) p_j + \sum_{j \geq 0} \Bigl(\sum_{i \geq j+1} x_i \Bigr) p_{j+1} \\
&= ep_0 - \biggl(\Bigl(\sum_{i \geq 1} x_i \Bigr) p_0 + \sum_{j \geq 1} \Bigl(\sum_{i \geq j+1} x_i \Bigr) p_j \biggr) + \sum_{j \geq 1} \Bigl(\sum_{i \geq j} x_i \Bigr) p_j \\
&= \sum x_i p_i = \boldsymbol{x}.
\end{aligned}$$

($\Leftarrow$) Although this is similar to part (1), it seems easiest to give a direct proof.
Write $\boldsymbol{u} = ep = ep_0$ and $\boldsymbol{v} = \sum v_i p_i$ for some point p and some $v_j \geq 0$. By definition, $\boldsymbol{u}_k = \sum_{i=0}^{k-1} ep_i$, and so, by our assumptions, $0 \leq v_i \leq e$ for $0 \leq i \leq k-1$, and $v_i = 0$ for all other i. Therefore,

$$\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v}) = (e - v_0) p_0 + \sum_{i \geq 1} (v_{i-1} - v_i) p_i.$$

If $j \leq -1$ then $x_j = 0$ and $\sum_{i \leq j} x_i = 0$. If $j \geq 0$, then $\sum_{i \leq j} x_i = e - v_j \geq 0$. Similarly, if $j \leq 0$ then $\sum_{i \geq j} x_i = e \geq 0$, while if $j \geq 1$ then

$$\sum_{i \geq j} x_i = \sum_{i=j}^{k} (v_{i-1} - v_i) = v_{j-1} - v_k = v_{j-1} \geq 0.$$

Thus (7.2) is satisfied. $\square$

We are now ready to state our main result on the structure of g-divisible maximal T-orders.

Theorem 7.4. (1) *Let $V \subseteq T$ be a g-divisible cg maximal T-order. Then the following hold*:

(a) *There is a maximal order $F \supseteq V$ such that (V, F) is a maximal order pair.*
(b) *F is a virtual blowup of T at a virtually effective divisor $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v})$ satisfying $0 \le \deg \boldsymbol{x} < \mu$.*
(c) $\overline{V} \doteq \overline{F} \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$.

(2) *If $U \subseteq T$ is any g-divisible cg subalgebra with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$, there exists a maximal order pair (V, F) as in* (1) *such that U is contained in and equivalent to V.*

(3) *Conversely, let $\boldsymbol{x}$ be a virtually effective divisor with* $\deg \boldsymbol{x} < \mu$. *Then there exists a blowup F of T at $\boldsymbol{x}$.*

Proof. (1) By definition, $Q_{\mathrm{gr}}(V) = Q_{\mathrm{gr}}(T)$. Now combine Corollary 6.6(1–2), Theorem 6.7 and Proposition 7.3.

(2) By Theorem 5.24, U is contained in and equivalent to some $\mathrm{End}_{T(\boldsymbol{d})}(M)$ which, in turn, is contained in and equivalent to a maximal T-order by Proposition 6.4.

(3) Write $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v})$, where $\boldsymbol{u}$, $\boldsymbol{v}$, k are defined by applying Proposition 7.3 to $\boldsymbol{x}$. By [RSS 2015, Lemma 5.10], there is a g-divisible finitely generated right $T(\boldsymbol{u})$-module M with $T(\boldsymbol{u}) \subseteq M \subseteq MT = T$ so that

$$\overline{M} \doteq \bigoplus_n H^0(E, \mathcal{M}_n(-\boldsymbol{u}_n + \boldsymbol{v})).$$

Let $F = \mathrm{End}_R(M^{**}) \supseteq U = \mathrm{End}_R(M)$. By Theorem 6.7 and Lemma 6.14(1–2), we have

$$\overline{F} \doteq \overline{U} \doteq \overline{M}(\overline{M})^* \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau),$$

and $(F \cap T, F)$ is a maximal order pair. □

Remark 7.5. We should explain why F is called a virtual blowup of T at $\boldsymbol{x}$ both in this theorem and in Definition 7.1. When $\boldsymbol{x}$ is effective this is amply justified in [Rogalski 2011] and, in that case, $T(\boldsymbol{x})$ satisfies many of the basic properties of a commutative blowup; in particular, it agrees with Van den Bergh's more categorical blowup [2001]. For noneffective $\boldsymbol{x}$ there are several reasons why the notation is reasonable.

(1) As we have shown repeatedly in this paper, the factor $\overline{U}$ of a g-divisible algebra U controls much of U's behaviour and so Theorem 7.4(1c) shows that F will have many of the basic properties of a blowup at an effective divisor.

(2) This is also supported by the fact that, by Theorem 5.24, F and $T(\boldsymbol{u})$ are equivalent maximal orders and, again, many properties pass through such a Morita context.

(3) Finally, in the commutative case virtual blowups are blowups, both because virtually effective divisors are then effective and because equivalent maximal orders are then equal.

Theorem 7.4 can be easily used to describe arbitrary g-divisible subalgebras of T. We recall that the *idealiser* of a left ideal L in a ring A is the subring $\mathbb{I}(L) = \{\theta \in A : L\theta \subseteq L\}$.

Corollary 7.6. *Let $U \subseteq T$ be a g-divisible subalgebra with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Then U is an iterated subidealiser inside a virtual blowup of T. More precisely, we have the following chain of rings:*

(1) *There is a virtually effective divisor $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v})$ with $\deg(\boldsymbol{x}) < \mu$ and a blowup F of T at $\boldsymbol{x}$ such that $V = F \cap T$ contains and is equivalent to U, while (V, F) is a maximal order pair.*

(2) *There exist a g-divisible algebra W with $U \subseteq W \subseteq V$ such that U is a right subidealiser inside W and W is a left subidealiser inside V. In more detail,*
   (a) *There exists a graded g-divisible left ideal L of V such either $L = V$ or else V/L is 2-pure, and a g-divisible ideal K of $X = \mathbb{I}(L)$ such that $K \subseteq W \subseteq X$ and $\operatorname{GKdim}_X(X/K) \leq 1$;*
   (b) *V is a finitely generated left W-module, while X/K is a finitely generated $\Bbbk[g]$-module and so X is finitely generated over W on both sides;*
   (c) *the properties given for $W \subseteq V$ also hold for the pair $U \subseteq W$, but with left and right interchanged.*

Proof. (1) Use Theorem 7.4(1–2).

(2) By (1), $aVb \subseteq U$ for some $a, b \in U \smallsetminus \{0\}$. Set $W' = U + Vb$ and $W = \widehat{W'}$. By Lemma 2.13(1), $aW = \widehat{aW'} \subseteq \widehat{U} = U$. By Proposition 2.9, W is noetherian and so (modulo a shift) $V \cong Vb$ is a finitely generated left W-module. Similarly, W is a finitely generated right U-module. We will now just prove parts (2a) and (2b), leaving the reader to check that the same argument does indeed work for the pair (U, W).

Write $V = \sum_{i=1}^{v} We_i$ for some e_i. Then the right annihilator

$$K = \text{r-ann}_W(V/W) = \bigcap \text{r-ann}(e_i)$$

is nonzero. Let L/K be the largest left V-submodule of V/K with $\operatorname{GKdim}(L/K) \leq 1$. Then either $L = V$, or else V/L is 2-pure. For $a \in W$, the module $(La + K)/K$ is a homomorphic image of La/Ka and hence of L/K. Thus $\operatorname{GKdim}((La+K)/K) \leq 1$ and $La \subseteq L$; in other words, L is still a (V, W)-bimodule.

As $W = \widehat{W}$, it is routine to see that K is g-divisible, but since we use the argument several times we give the details. So, suppose that $\theta g \in K$ for some $\theta \in V$. Then $(V\theta)g \subseteq \widehat{W} = W$, whence $V\theta \subseteq W$ and $\theta \in K$, as required. It follows that L/K is g-torsionfree and so, by Lemma 2.14, L/K is a finitely generated right $\Bbbk[g]$-module. Thus, by [Krause and Lenagan 1985, Lemma 5.3], $I = \ell\text{-ann}_V(L/K)$ satisfies $\operatorname{GKdim}_V(V/I) = \operatorname{GKdim}(L/K) \leq 1$. Again, I is g-divisible. Also, if $\theta \in V$ has $\theta g \in L$ then $(I\theta)g \subseteq K$ and so $I\theta \subseteq K$. Hence $\operatorname{GKdim}(V\theta + K)/K \leq \operatorname{GKdim}(V/I) \leq 1$ and $\theta \in L$. So L is also g-divisible.

Finally, let $X = \mathbb{I}_V(L) = \{x \in V : Lx \subseteq L\}$. As usual, X is g-divisible. Clearly IL is an ideal of X, and since I and L are g-divisible, $\operatorname{GKdim}(X/IL) \leq 1$, by Lemma 2.15(4). Since $X \supset K \supseteq IL$, it follows that $\operatorname{GKdim} X/K \leq 1$. Finally, since X/K is g-torsionfree of GK-dimension 1, it must be a finitely generated $\Bbbk[g]$-module by Lemma 2.14; in particular, X/W and hence X are finitely generated as right W-modules. □

There is a close correspondence between subalgebras A of the function skewfield $D = D_{\mathrm{gr}}(T)$ and g-divisible subalgebras of $T_{(g)}$, and so we end the section by studying the consequences of our earlier results for such an algebra A.

For a cg subalgebra $R \subseteq T_{(g)}$ with $g \in R$, define

$$R^\circ = R[g^{-1}]_0 = \bigcup_{n \geq 0} R_n g^{-n} \subseteq D = D_{\mathrm{gr}}(T).$$

Conversely, given an algebra $A \subseteq T^\circ$, define

$$\Omega A = \bigoplus_{m \geq 0} (\Omega A)_m \quad \text{for } (\Omega A)_m = \{a \in T_m : ag^{-m} \in A\}.$$

Clearly ΩA is g-divisible with $(\Omega A)^\circ = A$ and, if $R \subseteq T$, then $\Omega(R^\circ) = \widehat{R}$; thus we obtain a one-to-one correspondence between cg g-divisible subalgebras of T and subalgebras of T°.

Given a left ideal I of R or a left ideal J of A we define I° and ΩJ by the same formulæ. If R is g-divisible, the map $I \mapsto I^\circ$ gives a one-to-one correspondence between g-divisible left ideals of R and left ideals of R°, with analogous results for two-sided ideals (see [Artin et al. 1991, Proposition 7.5]).

An algebra $A \subseteq T^\circ$ is filtered by $A = \bigcup \Gamma^n A$ for $\Gamma^n A = (\Omega A)_n g^{-n}$. By [RSS 2014, Lemmas 2.1 and 2.2],

$$\operatorname{gr}_\Gamma A = \bigoplus \Gamma^n A / \Gamma^{n-1} A \cong \Omega A / g\Omega A, \tag{7.7}$$

where the isomorphism is induced by the map

$$\Gamma^n A \smallsetminus \Gamma^{n-1} A \to \Omega A, \quad x = rg^{-n} \mapsto r.$$

Lemma 7.8. *Let A, A' be orders in T°. Then A and A' are equivalent orders if and only if ΩA and $\Omega A'$ are equivalent orders in $Q_{\mathrm{gr}}(T)$.*

Proof. Let $0 \neq a \in \Gamma_m A'$ and $0 \neq b \in \Gamma_n A'$. To prove the lemma, it suffices to show that $aAb \subseteq A'$ if and only if $ag^m(\Omega A)bg^n \subseteq \Omega A'$. However, if $0 \neq \alpha \in \Omega A$, write

$\alpha = xg^k$ for some k and $x \in A$. Then

$$axb \in A' \iff axb \in \Gamma_{m+n+k}A' \iff ag^m(xg^k)bg^n \in \Omega A',$$

as desired. □

Corollary 7.9. *A subalgebra $A \subseteq T$ is a maximal T°-order if and only if $\Omega(A)$ is a maximal T-order.* □

By [RSS 2014, Theorem 1.1], every subalgebra of T° is finitely generated and noetherian; these subalgebras thus give a rich supply of noetherian domains of GK-dimension 2. Our earlier results about cg maximal T-orders translate easily to results about maximal T°-orders. An ideal I of a $\Bbbk$-algebra A is called *cofinite* if $\dim_{\Bbbk}(A/I) < \infty$.

Corollary 7.10. *Let A be a subalgebra of T° with $Q(A) = Q(T^\circ)$.*

(1) *There exists a maximal order pair (V, F), where F is a blowup of T at some virtually effective divisor $\boldsymbol{x}$, such that A is contained in and equivalent to the maximal T-order V°.*

(2) *In part* (1), *F° is a maximal order in $Q(T^\circ) = D_{\mathrm{gr}}(T)$.*

(3) *The algebras V° and F° have a cofinite ideal K° in common. Also, we have* $\mathrm{gr}_\Gamma V \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$.

(4) *Suppose that all nonzero ideals I of $T(\boldsymbol{d})^\circ$ generate cofinite right ideals of T°* (*in particular, this happens if $T(\boldsymbol{d})^\circ$ is simple*) *and that A is a maximal T°-order. Then A is a maximal order.*

Proof. (1) By Theorem 7.4(2), ΩA is contained in and equivalent to some such V. Now use Lemma 7.8 and Corollary 7.9.

(2) Since F need not be contained in T, this does not follow directly from the above discussion. However, it does follow from Lemma 6.2 combined with the fact that, in the notation of Corollary 6.6,

$$F^\circ = \mathrm{End}_{T(\boldsymbol{d})^\circ}((M^{**})^\circ) = \mathrm{End}_{T(\boldsymbol{d})^\circ}((M^\circ)^{**}).$$

(3) By definition and Lemma 2.14, V and F have an ideal K in common such that F/K is finitely generated as a $\Bbbk[g]$-module. Consequently F°/K° and V°/K° are finite-dimensional. The final assertion follows from Theorem 7.4(1c).

(4) Use Corollary 6.6(3). □

We also have a converse to Corollary 7.10(3).

Corollary 7.11. *Let $\boldsymbol{x}$ be a virtually effective divisor on E with* $\deg \boldsymbol{x} < \mu$. *Then there exists a maximal T°-order A with* $\mathrm{gr}_\Gamma A \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$.

Proof. Let U be the g-divisible maximal T-order given by Theorem 7.4(3); thus $\overline{U} \doteq B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$ by part (1c) of that result. By (7.7), $A = U^\circ$ satisfies the conclusion of this corollary. □

Del Pezzo surfaces. The blowup of T at ≤ 8 points on E can be thought of as a noncommutative del Pezzo surface. More carefully, it should be thought of as the anticanonical ring of a noncommutative del Pezzo surface; this corresponds to the fact that the central element g is in degree 1. Let U be a blowup of T at a virtually effective divisor $\boldsymbol{d}'$ of degree ≤ 8. By analogy, we should think of U as a (new type of) noncommutative del Pezzo surface, and the localisation U° as a particular kind of noncommutative affine surface. Corollary 7.10(3) can then be reinterpreted as saying that any maximal order $A \subseteq T^\circ$ is the coordinate ring of just such a noncommutative affine surface.

In [Etingof and Ginzburg 2010], the authors study noncommutative affine surfaces which are deformations of the commutative symplectic affine surfaces obtained from removing an anticanonical divisor from $\mathbb{P}^2$. These surfaces are related to ours but not the same; for example, we consider $A = T^\circ \cong T/(g-1)$, but the algebra $A' = S/(g-1)$ is considered in [ibid.]. The algebra A' is a rank 3 A-module, so "Spec A'" is a triple cover of "Spec A" (inasmuch as these terms make sense in a noncommutative context).

8. Sporadic ideals and g-divisible hulls

One of the main results in [Rogalski 2011] showed that the algebras considered there have minimal sporadic ideals, in a sense we define momentarily. In this section we show that, under minor assumptions, this generalises to cg subalgebras $U \subseteq T$ with $g \in U$ (see Corollary 8.8 for the precise statement). The significance of this result is that it provides a tight connection between the algebra U and its g-divisible hull $\widehat{U}$ and provides the final step in the proof of Theorem 1.2, that maximal orders are noetherian blowups of T (see Theorem 8.11).

Recall that a graded ideal I of a cg graded algebra R is called sporadic if $\operatorname{GKdim}(R/I) = 1$.

Definition 8.1. An ideal I of a cg algebra R is called a *minimal sporadic ideal* if $\operatorname{GKdim}(R/I) \leq 1$ and, for all sporadic ideals J, we have $\dim_{\Bbbk} I/(J \cap I) < \infty$.

Note that one can make the minimal sporadic ideal I unique by demanding that it be saturated, but we will not do so since this causes extra complications.

Beginning in this section, we need to strengthen our hypothesis on the ring T.

Assumption 8.2. In addition to Assumption 2.1, we assume that T has a minimal sporadic ideal and that there exists an uncountable algebraically closed field extension $K \supseteq \Bbbk$ such that, in the notation of [RSS 2015, Definition 7.2], $\mathrm{Div}(T \otimes_{\Bbbk} K)$ is countable.

We emphasise that, by [RSS 2015, Theorem 8.8 and Proposition 8.7], these extra assumptions do hold both for the algebras T from Examples 2.2(1–2) and for their blowups $T(\boldsymbol{d})$ at effective divisors $\boldsymbol{d}$ with $\deg \boldsymbol{d} < \mu$.

For the rest of this section we assume that our algebras T satisfy Assumptions 2.1 and 8.2. We do not know if Assumption 8.2 holds for Stephenson's algebras from Examples 2.2(3). By a routine exercise, Examples 2.2(4) does not have a minimal sporadic ideal, so Assumption 8.2 is strictly stronger than Assumption 2.1.

As noted above, the blowups $T(\boldsymbol{d})$ with $\deg \boldsymbol{d} < \mu$ have a minimal sporadic ideal, and the first goal of this section is to extend this to more general subalgebras of $T_{(g)}$. We start with the case of g-divisible algebras.

Lemma 8.3. *Let (V, F) be a maximal order pair, in the sense of Definition 6.5. Then both F and V have a minimal sporadic ideal.*

Proof. By Corollary 6.6, there exists an effective divisor $\boldsymbol{d}$ with $\deg \boldsymbol{d} < \mu$ and a right R-module $M \supseteq R$, where $R = T(\boldsymbol{d})$, such that

$$F = \mathrm{End}_R(M^{**}) \supseteq F \cap T = V = \mathrm{End}_R(M).$$

We will use a minimal sporadic ideal of R to construct such an ideal for F and for V.

Set $J = M^* = M^{***} \subseteq R$; thus $F = \mathrm{End}_R(J)$ as well. Also, write $X = JJ^*$, a nonzero ideal of R, and $W = J^*J$, a nonzero ideal of F. By Lemma 2.13(3), J and $J^* = M^{**}$ are g-divisible; in particular, $J \not\subseteq gT_{(g)}$ and $J^* \not\subseteq gT_{(g)}$. Thus, by Lemma 2.15(4), $\mathrm{GKdim}(R/X) \le 1$ and $\mathrm{GKdim}(F/W) \le 1$. By Assumption 8.2 and [RSS 2015, Proposition 8.7] we can choose a minimal sporadic ideal X' of R such that $X' \subseteq X$. Let $I = J^*X'J$. Since $\mathrm{GKdim}(X') \le 1$ and R is g-divisible, $\mathrm{GKdim}\, R/gR = 2$ and so $X' \not\subseteq gT_{(g)}$ also. Thus I is an ideal of F with $\mathrm{GKdim}(F/I) \le 1$ by Lemma 2.15(4).

Now consider an arbitrary sporadic ideal L of F, if such an ideal exists. Since F is g-divisible, $L \not\subseteq gT_{(g)}$ and so, just as in the previous paragraph, JLJ^* is an ideal of R satisfying $\mathrm{GKdim}_R(R/JLJ^*) \le 1$. Hence $JLJ^* \supseteq X'H$ for an ideal H of R with $\dim_{\Bbbk}(R/H) < \infty$. Now, $L \supseteq (J^*J)L(J^*J) \supseteq J^*X'HJ$, and [Krause and Lenagan 1985, Proposition 5.6] implies that $\dim_{\Bbbk}(J^*X'J)/(J^*X'HJ) < \infty$. Thus I is a minimal sporadic ideal of F.

In conclusion, F and V have a common ideal K with $\mathrm{GKdim}(F/K) \le 1$ (see Proposition 6.4). Thus KIK is a minimal sporadic ideal for F that lies in V and so it is also a minimal sporadic ideal for V. □

Proposition 8.4. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let $U \subseteq T$ be a g-divisible graded algebra with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. Then U has a minimal sporadic ideal.*

Proof. By Theorem 7.4(2), U is contained in and equivalent to some g-divisible maximal T-order V, say with $aVb \subseteq U$ for some nonzero homogeneous $a, b \in U$. Set $U' = U + UaV \subseteq V$, and $W = \widehat{U'}$. Thus $aV \subseteq U' \subseteq W$ and $U'b \subseteq U$. By Lemma 2.13(1), $Wb = \widehat{U'b} \subseteq \widehat{U} = U$. Set $J = \ell\text{-ann}_W\, V/W$, noticing that J is a nonzero ideal of W (since $a \in J$) and a right ideal of V. Also, as W is g-divisible, it follows that J is g-divisible. Thus, by Lemma 2.15(3), $\operatorname{GKdim} W/J \leq 1$.

If K is a minimal sporadic ideal in V given by Lemma 8.3, we claim that JK is a minimal sporadic ideal in W. To see this, let L be any ideal of W with $\operatorname{GKdim} W/L \leq 1$. Then $I = VLJ$ is an ideal of V. Since none of V, L, or J is contained in $gT_{(g)}$, $\operatorname{GKdim} V/I \leq 1$ by Lemma 2.15(4). Hence $I \supseteq KM$ for some ideal M of V with $\dim_{\Bbbk}(V/M) < \infty$ and so $L \supseteq JVLJ \supseteq JKM$. This implies that JK is a minimal sporadic ideal for W. Finally, a symmetric argument, using the fact that W is g-divisible with a minimal sporadic ideal, proves that U has such an ideal. □

As in Section 7, results on g-divisible rings have close analogues for subalgebras of T°.

Corollary 8.5. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let A be a subalgebra of T° with $Q(A) = Q(T^\circ)$. Then A has a unique minimal nonzero ideal I, and $\dim_{\Bbbk} A/I < \infty$. Further, A has DCC on ideals and finitely many primes.*

Proof. Recall from Section 7 that there is a one-to-one correspondence between g-divisible ideals of ΩA and ideals of A. Since every nonzero g-divisible ideal of ΩA is sporadic, when combined with Proposition 8.4 this gives the existence of I as described. Since A/I is artinian it has finitely many prime ideals and DCC on ideals. Thus the same holds for A. □

We now turn to a more general subalgebra U of T, with the aim of controlling its sporadic ideals also. We achieve this by relating U to its *g-divisible hull* $\widehat{U}$ and we begin with a straightforward lemma on subalgebras of TCRs. Recall that, for any subalgebra $U \subseteq T_{(g)}$, we write $\overline{U} = U + gT_{(g)}/T_{(g)}$.

Lemma 8.6. *Let $B = B(E, \mathcal{M}, \tau)$ for some smooth elliptic curve E, invertible sheaf $\mathcal{M}$ of degree $d > 0$ and τ of infinite order. Then for any $0 \neq x \in B_k$ we have $B_n x + x B_n = B_{n+k}$ for $n \gg 0$.*

In particular, if A is a graded subalgebra of B such that $A \neq \Bbbk$, then B is a noetherian (A, A)-bimodule.

Proof. By [Artin and Van den Bergh 1990, Theorem 1.3] and its left-right analogue, there exist effective divisors $\boldsymbol{x}$ and $\boldsymbol{x}'$ such that

$$xB_{\geq n_0} = \bigoplus_{n\geq n_0} H^0(E, \mathcal{M}_{n+k}(-\boldsymbol{x})) \quad \text{and} \quad (B_{\geq n_0})x = \bigoplus_{n\geq n_0} H^0(E, \mathcal{M}_{n+k}(-\tau^{-n}\boldsymbol{x}'))$$

(With a little thought one can see that this holds with $n_0 = 0$ and $\boldsymbol{x} = \boldsymbol{x}'$, but that is not relevant here.) Since $|\tau| = \infty$, we may choose n_0 so that $\boldsymbol{x} \cap \tau^{-n}\boldsymbol{x}' = \varnothing$ for all $n \geq n_0$. For such n there is an exact sequence

$$0 \to \mathcal{O}_E(-\boldsymbol{x} - \tau^{-n}\boldsymbol{x}') \to \mathcal{O}_E(-\boldsymbol{x}) \oplus \mathcal{O}_E(-\tau^{-n}\boldsymbol{x}') \to \mathcal{O}_E \to 0.$$

Tensoring with $\mathcal{M}_{n+k}$ and taking global sections gives a long exact sequence that reads, in part,

$$\begin{tikzcd} H^0(E, \mathcal{M}_{n+k}(-\boldsymbol{x})) \oplus H^0(E, \mathcal{M}_{n+k}(-\tau^{-n}\boldsymbol{x}')) \arrow[r] \arrow[d, equal] & H^0(E, \mathcal{M}_{n+k}) \arrow[r] \arrow[d, equal] & H \\ xB_n \oplus B_n x \arrow[r, "\theta"] & B_{n+k} & \end{tikzcd}$$

for $H = H^1(E, \mathcal{M}_{n+k}(-\boldsymbol{x} - \tau^{-n}\boldsymbol{x}'))$ and θ the natural map. Since

$$\deg(\mathcal{M}_{n+k}(-\boldsymbol{x} - \tau^{-n}\boldsymbol{x})) > 0 \quad \text{for } n \gg 0,$$

Riemann–Roch ensures that $H = 0$ and hence that θ is surjective for such n.

This implies that B is a noetherian $(\Bbbk\langle x\rangle, \Bbbk\langle x\rangle)$-bimodule, which certainly suffices to prove the final assertion of the lemma. □

We now show that, under mild hypotheses, $\widehat{U}$ is equivalent to U. In this result the hypothesis that $\overline{U} \neq \Bbbk$ is annoying but necessary (see Example 10.8) but, as will be shown in Section 9, there are ways of circumventing it.

Proposition 8.7. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let U be a cg subalgebra of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$, $g \in U$ and $\overline{U} \neq \Bbbk$.*

(1) *There exists $n \geq 0$ such that $U \cap Tg^m = \widehat{U} \cap Tg^m = g^m\widehat{U}$ for all $m \geq n$. Thus U and $\widehat{U}$ are equivalent orders.*

(2) *If U is right noetherian then $\widehat{U}$ is a finitely generated right U-module.*

Proof. (1) Let $V = \widehat{U}$. Since T is g-divisible, $V \subseteq T$. Working inside $Q_{\mathrm{gr}}(T)$, we get

$$\{x \in T : xg^k \in U\} = g^{-k}U \cap T,$$

and hence $V = \bigcup_{k\geq 0} g^{-k}U \cap T$. Now define $Q^{(k)} = (g^{-k}U \cap T + gT)/gT \subseteq \overline{T}$. Then, since $g \in U$,

$$\overline{U} = Q^{(0)} \subseteq Q^{(1)} \subseteq \cdots \subseteq \bigcup_k Q^{(k)} = \overline{V}.$$

Each $Q^{(i)}$ is an $\overline{U}$-subbimodule of $\overline{T}$ and so, by Lemma 8.6, $Q^{(n)} = \overline{V}$ for some n.

We claim that $U \cap Tg^m = V \cap Tg^m$ for all $m \geq n$. If not, there exists $y = xg^m \in V \cap Tg^m \smallsetminus U$ for some such m. Choose x of minimal degree with this property. This ensures that $y \notin g^{m+1}T$, since otherwise one could write $y = g^{m+1}x'$ with $\deg(x') = \deg(x) - 1$. Since $\bar{x} = [x + gT] \in \overline{V} = Q^{(n)}$, we have $\bar{x} = \overline{w}$, where $wg^n \in U$. Thus $wg^n - xg^n \in V \cap Tg^{n+1}$ and so $w - x = vg$, where $vg^{n+1} \in V \cap Tg^{n+1}$. Since $\deg v < \deg x$, the minimality of $\deg x$ ensures that $vg^{n+1} \in U$. Then $xg^n = wg^n - vg^{n+1} \in U$, and so $y = xg^n(g^{m-n}) \in U$, a contradiction. Thus $U \cap Tg^m = V \cap Tg^m$ as claimed. Finally, as $gV = V \cap gT$, an easy induction shows that $V \cap Tg^m = g^m V$.

(2) This is immediate from part (1). □

In the next result, we construct an ideal with a property that is slightly weaker than being a minimal sporadic ideal. However, it will have the same consequences.

Corollary 8.8. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let C be a cg subalgebra of T with $Q_{\mathrm{gr}}(C) = Q_{\mathrm{gr}}(T)$. Assume that $g \in C$ and $\overline{C} \neq \Bbbk$. Then C has a sporadic ideal K (possibly $K = C$) that is minimal among sporadic ideals I for which C/I is g-torsionfree.*

Proof. Note that $\widehat{C}$ is noetherian by Proposition 2.9 and has a minimal sporadic ideal, say J, by Proposition 8.4. By Lemma 2.15(2), $\widehat{J}$ is also a minimal sporadic ideal of $\widehat{C}$. Thus, replacing J by $\widehat{J}$, we can assume that $\widehat{C}/J$ is g-torsionfree.

We will show that $K = J \cap C$ satisfies the conclusion of the corollary. So, let I be a sporadic ideal of C such that C/I is g-torsionfree (if such an ideal exists). We first show that $J \cap C \subseteq I$. By Proposition 8.7, $H = g^n\widehat{C} \subseteq C$ for some $n \geq 1$ and so $I \supseteq HIH = g^{2n}\widehat{C}I\widehat{C}$. By Lemma 2.15(3), $HIH = g^r L$ for some r and ideal L of $\widehat{C}$ with $\mathrm{GKdim}(\widehat{C}/L) \leq 1$. As J is sporadic, $\dim_{\Bbbk} J/(J \cap L) < \infty$ and so $L \cap J \supseteq J_{\geq s} \supseteq g^s J$ for some integer s. Combining these observations shows that $I \supseteq g^t J$ for some integer t. Pick u minimal such that $I \supseteq g^u(J \cap C)$. If $u \neq 0$, then

$$\frac{I + g^{u-1}(J \cap C)}{I} = \frac{I + g^{u-1}(J \cap C)}{I + g^u(J \cap C)}$$

is g-torsion, and hence zero since C/I is g-torsionfree by assumption. Hence $u = 0$ and $I \supseteq J \cap C$.

It remains to show that $\mathrm{GKdim}\, C/(C \cap J) \leq 1$. Since $\widehat{C}/J$ is g-torsionfree, Lemma 2.14 implies that $M = \widehat{C}/J$ is a finitely generated $\Bbbk[g]$-module. Then the C-submodule $(C + J)/J \cong C/(J \cap C)$ is also. Therefore, by [Krause and Lenagan 1985, Corollary 5.4],

$$\mathrm{GKdim}_C(C/(C \cap J)) = \mathrm{GKdim}_{\Bbbk[g]}(C/(C \cap J)) \leq 1.$$

Thus $K = J \cap C$ satisfies the conclusions of the corollary. □

Lemma 8.9. *The set of orders $\{C \subseteq T$ with $\overline{C} \neq \Bbbk$ and $g \in C\}$ satisfies ACC.*

Proof. By Zorn's lemma, it suffices to prove that any such ring C is finitely generated as an algebra; equivalently, that $C_{\geq 1}$ is finitely generated as a right ideal.

We first show that, for any $m \geq 1$, $C/(g^m T \cap C)$ is finitely generated as an algebra. The result holds for $m = 1$ by [RSS 2014, Theorem 1.1(1)]. By induction, choose $a_1, \dots, a_k \in C_{\geq 1}$ whose images generate $C/(g^{m-1} T \cap C)$ as an algebra. Set $X = (g^{m-1} T \cap C)/(g^m T \cap C)$. Then, up to shifts,

$$X \cong \frac{T \cap g^{1-m} C}{gT \cap g^{1-m} C} \cong \frac{(T \cap g^{1-m} C) + gT}{gT} \subseteq T/gT$$

as $\overline{C}$-bimodules. Thus, by Lemma 8.6, X is a finitely generated $\overline{C}$-bimodule, say by the images of $b_1, \dots, b_n \in g^{m-1} T \cap C$. Then $\{a_1, \dots, a_k, b_1, \dots, b_n\}$ generate $C/(g^m T \cap C)$, completing the induction.

By Proposition 8.7, there exists $\ell \in \mathbb{N}$ such that $g^m \widehat{C} = C \cap g^m T \subseteq C$ for all $m \geq \ell$. By the above, choose $c_1, \dots, c_N \in C_{\geq 1}$ whose images generate $C/g^{\ell+1}\widehat{C}$ as an algebra. Then for any $f \in C_{\geq 1}$, there exists $x \in \sum c_i C$ so that $f - x \in g^{\ell+1}\widehat{C} = g(g^{\ell}\widehat{C})$; thus $f \in gC + \sum c_i C$. Therefore, $C_{\geq 1}$ is generated as a right ideal by $g, c_1, \dots, c_N$. □

Proposition 8.10. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let U be a cg subalgebra of T with $\overline{U} \neq \Bbbk$ and $D_{\mathrm{gr}}(U) = D_{\mathrm{gr}}(T)$. Then there exists a nonzero ideal of $C = U\langle g\rangle$ that is finitely generated as both a left and a right U-module.*

Proof. By Lemma 8.9, there is a finitely generated cg subalgebra W of U with $C = U\langle g\rangle = W\langle g\rangle$. Note that $Q_{\mathrm{gr}}(C) = Q_{\mathrm{gr}}(T)$ as $g \in C$.

Fix $n \in \mathbb{N}$. Observe that $CW_{\geq n} = \sum_m W_{\geq n} g^m = W_{\geq n} C$ is an ideal of C. Moreover, $C/CW_{\geq n}$ is a homomorphic image of the polynomial ring $(W/W_{\geq n})[g]$. Since $\dim_{\Bbbk}(W/W_{\geq n}) < \infty$, it follows that $C/CW_{\geq n}$ is a finitely generated $\Bbbk[g]$-module. In particular, by [Krause and Lenagan 1985, Corollary 5.4],

$$\mathrm{GKdim}_C(C/CW_{\geq n}) = \mathrm{GKdim}_{\Bbbk[g]}(C/CW_{\geq n}) \leq 1.$$

Moreover, $K_n = \mathrm{tors}_g(C/CW_{\geq n})$ is finite-dimensional.

Let $Z_n = C \cap \widehat{CW_{\geq n}}$; thus $Z_n/CW_{\geq n} = K_n$. Note that C/Z_n is a finitely generated torsion-free, hence free, $\Bbbk[g]$-module. Therefore, if d_n denotes the rank of that free module, then

$$\begin{aligned} d_n &= \dim_{\Bbbk}(C/Z_n)_m && \text{for } m \gg 0 \\ &= \dim_{\Bbbk}(C/CW_{\geq n})_m && \text{for } m \gg 0. \end{aligned}$$

Also, $CW_{\geq n} \supseteq CW_{\geq n+1}$, whence $Z_n \supseteq Z_{n+1}$ and $d_n \leq d_{n+1}$.

Let J be a minimal sporadic ideal of $\widehat{C}$ such that $\widehat{C}/J$ is g-torsionfree; thus, by Corollary 8.8 and its proof, $C \cap J$ is minimal among sporadic ideals of C such that the factor is g-torsionfree. By construction, each Z_n is either sporadic or equal to C; in either case $Z_n \supseteq C \cap J$. Now $C/(C \cap J) \hookrightarrow \widehat{C}/J$, which, by Lemma 2.14, has an eventually constant Hilbert series; say $\dim_{\Bbbk}(\widehat{C}/J)_m = N$ for all $m \gg 0$. Hence $\dim(C/Z_n)_m \leq N$ for all such m and, in particular, $d_n \leq N$. Since $d_n \leq d_{n+1}$, it follows that $d_n = d_{n+1}$ for all $n \gg 0$; say for all $n \geq n_0$. Thus, by the last display, $CW_{\geq n} \doteq CW_{\geq n_0}$ for all $n \geq n_0$.

Finally, if W is generated as an algebra by elements of degree at most e, then $CW_{\geq n_0}W_{\geq 1} \supseteq CW_{\geq n_0+e}$. By the last paragraph, $\dim_{\Bbbk}(CW_{\geq n_0}/CW_{\geq n_0+e}) < \infty$, and so $\dim_{\Bbbk}(CW_{\geq n_0}/CW_{\geq n_0}W_{\geq 1}) < \infty$. Thus, by the graded Nakayama's lemma, $CW_{\geq n_0} = W_{\geq n_0}C = W_{\geq n_0}U\langle g\rangle$ is finitely generated as a right W-module, and hence as a right U-module. $\square$

Finally, we can reap the benefits of the last few results.

Theorem 8.11. *Suppose that T satisfies Assumptions 2.1 and 8.2. For some $n \geq 1$, let U be a cg maximal $T^{(n)}$-order with $\overline{U} \neq \Bbbk$. Then U is strongly noetherian; in particular, noetherian and finitely generated as an algebra. Moreover:*

(1) *If $n = 1$, so $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$, then U is g-divisible and $U = F \cap T$, where F is a blowup of T at a virtually effective divisor $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v} + \tau^{-1}(\boldsymbol{v})$ of degree $< \mu$.*

(2) *If $Q_{\mathrm{gr}}(U) \neq Q_{\mathrm{gr}}(T)$, then there is a virtually effective divisor $\boldsymbol{x}$ of degree $< \mu$ and a blowup F of T at $\boldsymbol{x}$ so that $U = (F \cap T)^{(n)}$.*

Proof. (1) Let $C = U\langle g\rangle$; thus $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(\widehat{C}) = Q_{\mathrm{gr}}(T)$. By Proposition 8.10, there exists an ideal X of C that is finitely generated as a right U-module. In particular, as U is a right Ore domain and $X \subseteq Q_{\mathrm{gr}}(U)$, we can clear denominators from the left to find $q \in Q_{\mathrm{gr}}(U)$ such that $X \subseteq qU$. As X is an ideal of C, we have $pC \subseteq X$ for any $0 \neq p \in X$ and hence $C \subseteq p^{-1}qU$. Thus C and U are equivalent orders. By Proposition 8.7 it follows that U and $\widehat{C}$ are equivalent orders and hence $U = \widehat{C}$. Now apply Proposition 2.9 and Theorem 7.4.

(2) Keep C and X as above. In this case, as $Q_{\mathrm{gr}}(U) = \Bbbk(E)[g^n, g^{-n}, \tau^n]$, clearly U and $C' = U\langle g^n\rangle$ have the same graded quotient ring and, moreover, $C' = C^{(n)}$. Therefore $X^{(n)}$ is an ideal of $C^{(n)}$ which, since it is a U-module summand of X, is also finitely generated as a right (and left) U-module. The argument used in (1) therefore implies that U and $C^{(n)}$ are equivalent orders and hence that $U = C^{(n)}$. In particular, $C = \sum_{i=0}^{n-1} g^i C^{(n)}$ is a finitely generated right U-module.

Consider $\widehat{C}$. As $g \in C$, we have $Q_{\mathrm{gr}}(\widehat{C}) = Q_{\mathrm{gr}}(T)$ and so, by Corollary 7.6(1), there exists a cg maximal T-order $V = \widehat{V} \subseteq T$ containing and equivalent to $\widehat{C}$. By

Proposition 8.7, V is equivalent to C. Further, $V = F \cap T$ where F is a blowup of T at some virtually effective divisor $\boldsymbol{x}$ on E with $\deg \boldsymbol{x} < \mu$.

Now, $aVb \subseteq C$ for some $a, b \in C \smallsetminus \{0\}$. By multiplying by further elements of C we may suppose that $a, b \in C^{(n)} = U$ and hence that $aV^{(n)}b \subseteq U$. As U is a maximal $T^{(n)}$-order, and certainly $V^{(n)} \subseteq T^{(n)}$, it follows that $U = V^{(n)}$. □

One consequence of the theorem is that maximal $T^{(n)}$-orders have a number of pleasant properties, as we next illustrate. The undefined terms in the following corollary can be found in [Rogalski 2011, §2] and [Van den Bergh 1997].

Corollary 8.12. *Suppose that T satisfies Assumptions 2.1 and 8.2. For some $n \geq 1$, let U be a cg maximal $T^{(n)}$-order with $\overline{U} \neq \Bbbk$. Then* qgr-U *has cohomological dimension ≤ 2, while U has a balanced dualising complex and satisfies the Artin–Zhang χ conditions.*

Proof. By Theorem 8.11, $U = V^{(n)}$ for a g-divisible maximal T-order V. Hence $\overline{V} \doteq B(E, \mathcal{N}, \tau)$, by Theorem 6.7. Thus [Rogalski 2011, Lemma 2.2] and [Artin and Zhang 1994, Lemma 8.2(5)] imply that qgr-$\overline{V}$ has cohomological dimension one, and that $\overline{V}$ satisfies χ. The fact that V satisfies χ and that qgr-V has cohomological dimension ≤ 2 then follow from [ibid., Theorem 8.8]. By [Artin and Stafford 1995, Lemma 4.10(3)], V is a noetherian U-module and so, by [Artin and Zhang 1994, Proposition 8.7(2)], these properties then descend to U. (With a little more work one can show that qgr-V and qgr-U have cohomological dimension exactly 2.) Finally, by [Van den Bergh 1997, Theorem 6.3], this implies the existence of a balanced dualising complex. □

Let U be a maximal order in T with $\overline{U} \neq \Bbbk$. Theorem 8.11 also allows us to determine the simple objects in qgr-U, although we do not formalise their geometric structure.

Corollary 8.13. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let U be a cg maximal T-order with $\overline{U} \neq \Bbbk$. Then the simple objects in* qgr-U *are in one-to-one correspondence with the closed points of the elliptic curve E together with a (possibly empty) finite set.*

Proof. A simple object in qgr-U equals $\pi(M)$ for a cyclic critical right U-module M with the property that every proper factor of M is finite-dimensional. Suppose first that M is g-torsion; thus $Mg = 0$ by Lemma 3.8. Hence, by Theorems 8.11 and 7.4, $\pi(M) \in$ qgr-B for some TCR $B = B(E, \mathcal{N}, \tau)$. Thus, under the equivalence of categories qgr-$B \simeq \operatorname{coh}(E)$, $\pi(M)$ corresponds to a closed point of E.

On the other hand, if M is not annihilated by g, then Lemma 3.8 implies that M is g-torsionfree. By comparing Hilbert series, it follows that $\operatorname{GKdim}(M/Mg) = \operatorname{GKdim}(M) - 1$ and so, as $\dim_{\Bbbk} M/Mg < \infty$ by construction, $\operatorname{GKdim}(M) = 1$. In particular, $M' = M[g^{-1}]_0$ is then a finite-dimensional simple U°-module and hence is annihilated by the minimal nonzero ideal of U° (see Corollary 8.5). Pulling back to U, this says that M is killed by the minimal sporadic ideal K of U. Thus, by Lemma 3.8, $P = \text{r-ann}(M)$ is one of the finitely many prime ideals P minimal over K.

In order to complete the proof we need to show that $\pi(M)$ is uniquely determined by P. Note that, as $\dim_{\Bbbk}(M/Mg) < \infty$, we have $\pi(M) \cong \pi(Mg) = \pi(M[-1])$ in qgr-U, and so we do not need to worry about shifts. Next, as $\operatorname{GKdim}(M) = \operatorname{GKdim}(U/P)$, M is a (Goldie) torsion-free U/P-module and hence is isomorphic to (a shift of) a uniform right ideal J of U/P. However, given a second uniform right ideal $J' \subseteq U/P$, then J' is isomorphic to (a shift of) a submodule $L \subseteq J$ (use the proof of [McConnell and Robson 2001, Corollary 3.3.3]). Once again, $\dim_{\Bbbk}(J/L) < \infty$ and so $\pi(J) \cong \pi(J')$, as required. □

Corollary 8.14. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let $U \subseteq T$ be a noetherian cg algebra with $D_{\mathrm{gr}}(U) = D_{\mathrm{gr}}(T)$ and $\overline{U} \neq \Bbbk$. Then $C = U\langle g\rangle$ and $\widehat{C}$ are both finitely generated right (and left) U-modules.*

Proof. Again, let $X = CU_{\geq n_0}$ be the ideal of C that is finitely generated as a right U-module given by Proposition 8.10. In this case, X is a noetherian right U-module and hence so is $C \cong xC[n]$ for any $0 \neq x \in X_n$. The rest of the result follows from Proposition 8.7. □

9. Arbitrary orders

The assumption $\overline{U} \neq \Bbbk$ that appeared in most of the results from Section 8 is annoying but, as Example 10.8 shows, necessary. Fortunately one can bypass the problem, although at the cost of passing to a Veronese ring. In this section we explain the trick and apply it to describe arbitrary cg orders in T.

Up to now graded homomorphisms of algebras have been degree-preserving, but this will not be the case for the next few results, and so we make the following definition. A homomorphism $A \to B$ between $\mathbb{N}$-graded algebras is called *graded of degree t* if $\phi(A_n) \subseteq B_{nt}$ for all n. The map ϕ is called *semigraded* if it is graded of degree t for some t.

Proposition 9.1. *Suppose that T satisfies Assumption 2.1 and that U is a cg noetherian subalgebra of T with $U \not\subseteq \Bbbk[g]$. Then there exist $N, M \in \mathbb{N}$ and an injective graded homomorphism $\phi : U^{(N)} \to T$ of degree M such that $U' = \phi(U^{(N)}) \not\subseteq \Bbbk + gT$. In addition, $D_{\mathrm{gr}}(U) = D_{\mathrm{gr}}(U') \subseteq D_{\mathrm{gr}}(T)$.*

Proof. For $n \geq 0$, define $f : \mathbb{N} \to \mathbb{N} \cup \{-\infty\}$ by

$$f(n) = \min\{i : U_n \subseteq g^{n-i}T\}, \quad \text{with } f(n) = -\infty \text{ if } U_n = 0.$$

Trivially, $f(n) \in \{0, 1, \dots, n\} \cup \{-\infty\}$ for all $n \geq 0$, and $f(n) = 0$ if and only if $U_n = \Bbbk g^n$.

We first claim that $f(n) + f(m) \leq f(n+m)$ for all $m, n \geq 0$. As A is a domain, this is clear if one of the terms equals $-\infty$, and so we may assume that $f(r) \geq 0$ for $r = n, m, n+m$. Write $U_r = X_r g^{r-f(r)}$ for such r; thus $X_r \subseteq T$ but $X_r \not\subseteq gT$. Since gT is a completely prime ideal, $X_n X_m \subseteq T$ but $X_n X_m \not\subseteq gT$. In other words, $U_n U_m \not\subseteq Y = g^{(n-f(n)+m-f(m)+1)}T$. Since $U_n U_m \subseteq U_{n+m}$ it follows that $U_{n+m} \not\subseteq Y$ and hence that $f(n+m) \geq f(n) + f(m)$, as claimed.

A noetherian cg algebra is finitely generated by the graded Nakayama's lemma, so suppose that U is generated in degrees $\leq r$. Then $U_n = \sum_{i=1}^{r} U_i U_{n-i}$ for all $n > r$. Arguing as in the previous paragraph shows that

$$f(n) = \max\{f(n-i) + f(i) : 1 \leq i \leq r\} \quad \text{for } n > r, \tag{9.2}$$

with the obvious conventions if any of these numbers equals $-\infty$.

We claim that there exists N with $f(N) > 0$ such that $f(nN) = nf(N)$ for all $n \geq 1$. This follows by exactly the same proof as in [Artin and Stafford 1995, Lemma 2.7]. Namely, choose $1 \leq N \leq r$ such that $\lambda = f(N)/N$ is as large as possible; by induction using (9.2) it follows that $f(n) \leq \lambda n$ for all $n \geq 0$, and this forces $f(nN) = nf(N)$ for all $n \geq 0$, as claimed.

Let $M = f(N)$ and note that $M > 0$ since $U \not\subseteq \Bbbk[g]$. Thus, for each $n \geq 0$ we have $U_{nN} \subseteq g^{nN-nM}T$ but $U_{nN} \not\subseteq g^{nN-nM+1}T$. Therefore the function $U_{nN} \to T_{nM}$ given by $x \mapsto xg^{n(M-N)}$ is well-defined, and it defines an injective vector space homomorphism $\theta : U^{(N)} \to T$ with $\theta(U^{(N)}) \not\subseteq \Bbbk + gT$. It is routine to see that θ is an algebra homomorphism which is graded of degree M. The final claim of the proposition is clear because $D_{\mathrm{gr}}(U) = D_{\mathrm{gr}}(U^{(N)}) = D_{\mathrm{gr}}(U')$. □

Corollary 9.3. (1) *Suppose T satisfies Assumption 2.1 and that U is a noetherian subring of T generated in a single degree N, with $U \neq \Bbbk[g^N]$. Then up to a semigraded isomorphism we may assume that $U \not\subseteq \Bbbk + gT$.*

(2) *Suppose also that T satisfies Assumption 8.2. If U is a noetherian maximal $T^{(N)}$-order generated in degree N then, again up to a semigraded isomorphism, $U \cong V^{(M)}$, where (V, F) is a maximal order pair and $M \leq N$.*

Proof. In this proof, Veronese rings are unregraded; that is, they are given the grading induced from T.

(1) Pick $M \in \mathbb{N}$ minimal such that $U_N \subseteq g^{N-M}T$. Necessarily, $M \leq N$. Then, either directly or by Proposition 9.1, there is a semigraded monomorphism

$$\phi : U = U^{(N)} \to T \quad \text{given by} \quad u \mapsto g^{M-N}u \text{ for } u \in U_N.$$

Hence $U \cong \phi(U)$, and $\phi(U) \not\subseteq \Bbbk + gT$ by the choice of M.

(2) As U is an order in $T^{(N)}$, certainly $\phi(U)$ is an order in $T^{(M)}$. So, suppose that $\phi(U) \subseteq W \subseteq T^{(M)}$ for some equivalent order W; say with $aWb \subseteq \phi(U)$, for $a, b \in \phi(U)$. Since $M \leq N$, the map ϕ^{-1} extends to give a well-defined semigraded homomorphism $\psi : T^{(M)} \to T^{(N)}$ defined by $\gamma \mapsto g^{N-M}\gamma$ for all $\gamma \in T_M$. Therefore, $\psi(a)U\psi(b) \subseteq U \subseteq \psi(W) \subseteq T^{(N)}$ and hence $U = \psi(W)$. Thus, $\phi(U) = W$ is a maximal order in $T^{(M)}$ with $\phi(U) \not\subseteq \Bbbk + gT$. Now apply Theorem 8.11(2). □

One question we have been unable to answer is the following.

Question 9.4. Suppose that $U \subseteq T$ is a cg maximal T-order or, indeed, a maximal order. Then is each Veronese ring $U^{(n)}$ also a maximal $T^{(n)}$-order? The question is open even when U is noetherian.

If this question has a positive answer, one can mimic the proof of Corollary 9.3 for any noetherian maximal order U to get a precise description of some Veronese ring $U^{(N)}$. However, the best we can do at the moment is to use the much less precise result given by the next corollary, which also describes arbitrary noetherian cg subalgebras of T.

Corollary 9.5. *Suppose that T satisfies Assumptions 2.1 and 8.2. Let $U \subseteq T$ be a noetherian algebra with $D_{\mathrm{gr}}(U) = D_{\mathrm{gr}}(T)$. Then, up to taking Veronese subrings, U is an iterated subidealiser inside a virtual blowup of T. More precisely, the following hold.*

(1) *There is a semigraded isomorphism of Veronese rings $U^{(N)} \cong U'$, where $U' \subseteq T$ is a noetherian algebra such that $D_{\mathrm{gr}}(U') = D_{\mathrm{gr}}(T)$ and $U' \not\subseteq \Bbbk + gT$.*

(2) *If $C = U'\langle g\rangle$ and $Z = \widehat{C}$, then Z is a finitely generated (left and right) U'-module and Z is a noetherian algebra with $Q_{\mathrm{gr}}(Z) = Q_{\mathrm{gr}}(T)$. The g-divisible algebra Z is described by Corollary 7.6.*

Proof. By [Artin and Zhang 1994, Proposition 5.10], the Veronese ring $U^{(N)}$ is noetherian and so part (1) follows from Proposition 9.1. Part (2) then follows from Corollary 8.14 (and Corollary 7.6). □

10. Examples

We end the paper with several examples that illustrate some of the subtleties involved here. For simplicity, these examples will all be constructed from $T = S^{(3)}$ for the standard Sklyanin algebra S of Examples 2.2(1); thus $\mu = \deg \mathcal{M} = 9$.

We first construct a g-divisible, maximal T-order U that is not a maximal order in $Q_{\mathrm{gr}}(U)$, as promised in Section 6. This shows, in particular, that the concept of maximal order pairs is indeed necessary in that section. In order to construct the example, we need the following notation.

Notation 10.1. Fix $0 \neq x \in S_1$ and let $\boldsymbol{c} = p+q+r$ be the hyperplane section of E where x vanishes. We can and will assume that no two of p, q, r lie on the same σ-orbit on E, where $S/gS \cong B(E, \mathcal{L}, \sigma)$. Set $R = T(\boldsymbol{c})$. By [Rogalski 2011, Example 11.3], R has a sporadic ideal $I = xS_2R$. Write $N = xT_1x^{-1}R$ and $M = xS_5R + R$. Finally, set $\boldsymbol{d} = \sigma^{-2}(\boldsymbol{c}) = \sigma^{-2}(p) + \sigma^{-2}(q) + \sigma^{-2}(r)$ and hence $\boldsymbol{d}^\tau = \sigma^{-5}(\boldsymbol{c})$.

As we will see, $U = \operatorname{End}_R(M)$ will (essentially) be the required maximal T-order with equivalent maximal order being $F = \operatorname{End}_R(N)$. The proof will require some detailed computations, which form the content of the next lemma. We note that for subspaces of homogeneous pieces of S we use the grading on S, but for subspaces that live naturally in T we use the T-grading. For example, we write $T_1S_2 = S_5$.

Lemma 10.2. *Keep the data from Notation 10.1.*

(1) $NI = xS_5R \subseteq M$ *and* $M_{\geq 1} \subseteq N$. *Hence* $N^{**} = M^{**} = (\widehat{M})^{**} = \widehat{M^{**}}$.

(2) $U' = \operatorname{End}_R(\widehat{M}) \subseteq T$, *but*

(3) $F = \operatorname{End}_R(M^{**}) = \operatorname{End}_R(NI) = xT(\boldsymbol{d}^\tau)x^{-1}$. *Moreover*, $F \not\subseteq T$.

Proof. (1) Clearly

$$NI = xT_1S_2R = xS_5R \subseteq M = xS_5R + R.$$

By [Rogalski 2011, Example 11.3], $R_1 = xS_2 + \Bbbk g$ and so $R_1x \subseteq xT$. Equivalently, $R_1 \subseteq xT_1x^{-1} \subseteq N$. As $R = T(\boldsymbol{c})$ is generated in degree one by Proposition 4.10(2), $R \subseteq xTx^{-1}$. In particular, $M_{\geq 1} = xS_5R + R_{\geq 1} \subseteq N$. As I is a sporadic ideal, it follows from Proposition 4.10 and Lemma 4.11(1) that $N^{**} = (NI)^{**}$ and hence that $M^{**} = N^{**}$.

Now consider $\widehat{M}$. Since $1 \in M$, certainly $MT = T$ and so $M^{**} = (\widehat{M})^{**} = \widehat{M^{**}}$ by Lemma 2.13(3).

(2) Since $MT = T$ we have $\widehat{M}T = T$, from which the result follows.

(3) We will first prove that $\operatorname{End}_R(N) = xT(\boldsymbol{d}^\tau)x^{-1}$. As in (1), $R_1 = xS_2 + \Bbbk g$. Equivalently, $(x^{-1}Rx)_1 = S_2x + \Bbbk g$ is a 7-dimensional subspace of T_1 that vanishes at the points $\sigma^{-2}(p)$, $\sigma^{-2}(q)$ and $\sigma^{-2}(r)$. Now, $T(\boldsymbol{d})_1$ is also 7-dimensional by [Rogalski 2011, Theorem 1.1(1)]. Consequently, $(x^{-1}Rx)_1 = T(\boldsymbol{d})_1$ and so $x^{-1}Rx = T(\boldsymbol{d})$, since both algebras are generated in degree 1 by Proposition 4.10(2). Therefore,

$$x^{-1}Nx = T_1(x^{-1}Rx) = T_1T(\boldsymbol{d}) = T(\boldsymbol{d}^\tau)T_1,$$

where the final equality follows from [RSS 2015, Corollary 4.14]. Thus

$$xT(\boldsymbol{d}^\tau)T_1x^{-1} = N$$

and so $\mathrm{End}_R(N) \supseteq G = xT(\boldsymbol{d}^\tau)x^{-1}$. Since $N = GxT_1x^{-1}$, Lemma 6.3 implies that $\mathrm{End}_R(N)$ is a finitely generated left G-module. But G is a maximal order by [Rogalski 2011, Theorem 1.1(2)], and so $\mathrm{End}_R(N) = G$. Thus, by part (1) and Lemma 6.2, $\mathrm{End}_R(N) = \mathrm{End}_R(N^{**}) = \mathrm{End}(M^{**})$. Moreover, $\mathrm{End}_R(N) \subseteq \mathrm{End}_R(NI)$ and we again have equality by Lemma 6.3.

It remains to prove that $xT(\boldsymbol{d}^\tau)x^{-1} \not\subseteq T$. This will follow if we show that $\bar{x}\overline{X}\bar{x}^{-1} \not\subseteq \overline{T}$, where $X = T(\boldsymbol{d}^\tau)$ and $\overline{X} = (X + gT_{(g)})/T_{(g)}$. So, assume that $\bar{x}\overline{X}\bar{x}^{-1} \subseteq \overline{T}$. Then $\bar{x}\overline{X}_1 \subseteq \overline{T}_1\bar{x} = \bar{S}_3\bar{x}$. However, inside $\bar{S}_4$,

$$\bar{x}\overline{X}_1 \subseteq H^0\big(E, \mathscr{L}_4(-p-q-r-\sigma^{-6}(p)-\sigma^{-6}(q)-\sigma^{-6}(r))\big),$$

and, since both are 6-dimensional, they are equal. On the other hand,

$$\bar{S}_3\bar{x} = H^0\big(E, \mathscr{L}_4(-\sigma^{-3}(p)-\sigma^{-3}(q)-\sigma^{-3}(r))\big).$$

Inside $\bar{S}_4$, vanishing conditions at ≤ 12 distinct points give independent conditions. So there exists z that vanishes at the first 6 points $p, \dots, \sigma^{-6}(r)$ but not at the points $\sigma^{-3}(p), \sigma^{-3}(q), \sigma^{-3}(r)$. This implies that $\bar{x}\overline{X}_1 \not\subseteq \overline{T}_1\bar{x}$, and completes the proof of the lemma. □

We are now able to give the desired example.

Proposition 10.3. *There exists a maximal order pair (V, F) with $V \neq F$. In particular, V is a maximal T-order that is not a maximal order.*

In more detail, and using the data from Notation 10.1, $F = \mathrm{End}_R((\widehat{M})^{**}) = xT(\boldsymbol{d}^\tau)x^{-1}$ is a blowup of T at $\boldsymbol{x} = \boldsymbol{c} - \tau^{-1}(\boldsymbol{c}) + \tau^{-2}(\boldsymbol{c})$. The algebra F is also Auslander–Gorenstein and CM.

Proof. As $1 \in M$, Theorem 6.7 and Lemma 10.2 imply that $F = \mathrm{End}_R((\widehat{M})^{**}) = xT(\boldsymbol{d}^\tau)x^{-1}$ is a maximal order with $F \not\subseteq T$. By Theorem 6.7, again, $V = T \cap F$ is a g-divisible maximal T-order, but V is not a maximal order as $V \neq F$. That F is Auslander–Gorenstein and CM follows from Proposition 4.10.

Theorem 6.7 also implies that F is a blowup of T at some virtual divisor $\boldsymbol{y}$, so it remains to check that $\boldsymbol{y} = \boldsymbol{x}$. By Lemma 10.2, $F = \mathrm{End}_R(NI) = \mathrm{End}_R(xS_5R)$ and hence $\overline{F} \subseteq \mathrm{End}_{\overline{R}}(\overline{xS_5R})$. Now, for any $n \geq 2$, one has

$$\bar{R}_{n-2} = H^0\big(E, \mathcal{M}(-\boldsymbol{c} - \boldsymbol{c}^\tau - \cdots - \boldsymbol{c}^{\tau^{n-3}})\big),$$

and so

$$(\overline{xS_5}\,\overline{R})_n = H^0(E, \mathcal{M}(-\boldsymbol{c} - \boldsymbol{c}^{\tau^2} - \boldsymbol{c}^{\tau^3} - \cdots - \boldsymbol{c}^{\tau^{n-1}}) = H^0(E, \mathcal{O}(\boldsymbol{c}^\tau)\mathcal{M}(-\boldsymbol{c})_n).$$

Hence

$$\overline{F} \subseteq \mathrm{End}_{\overline{R}}(\overline{xS_5R}) = \mathrm{End}_{\overline{R}}\bigg(\bigoplus_{n\geq 2} H^0(E, \mathcal{O}(\boldsymbol{c}^\tau)\mathcal{M}(-\boldsymbol{c})_n)\bigg).$$

Therefore, by Lemma 6.14(1), $\overline{F} \overset{\bullet}{=} B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$. By [Rogalski 2011, Theorem 1.1(2)] and Riemann–Roch, $\dim \overline{F}_n = 6n = \dim B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$, for $n \geq 1$, and hence $\overline{F} = B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$, as required. □

When $\boldsymbol{y}$ is effective, the blowup $T(\boldsymbol{y})$ is both Auslander–Gorenstein and CM (see Proposition 4.10), as is the blowup of T at $\boldsymbol{x}$ from Proposition 10.3. Despite this example, neither the Auslander–Gorenstein nor the CM condition is automatic for a blowup of T at virtually effective divisors.

Example 10.4. Let $\boldsymbol{x} = p - \tau(p) + \tau^2(p)$ for a closed point $p \in E$ and let U be a blowup of T at $\boldsymbol{x}$. Then U is a maximal order contained in T that is neither Auslander–Gorenstein nor AS Gorenstein nor CM.

Proof. By Definition 6.9 and Corollary 6.6(2), $U = \mathrm{End}_{T(q)}(M)$, where $M = M^{**}$ satisfies $MT = T$ and q is a closed point that is τ-equivalent to $\boldsymbol{x}$ and hence to p. By [RSS 2015, Example 9.5], $T(q)$ has no sporadic ideals and so, by Corollary 6.6(3), U is a g-divisible maximal order contained in T.

Now consider $\overline{U} = U/gU$. By Theorem 6.7, $\overline{U} \overset{\bullet}{=} B = B(E, \mathcal{M}(-\boldsymbol{x}), \tau)$. We emphasise that we always identify $\mathcal{M}(-\boldsymbol{x})$ and $\mathcal{M}$ with the appropriate subsheaves of the field $\Bbbk(E)$ and B with the corresponding subring of the Ore extension $T_{(g)}/gT_{(g)} \cong \Bbbk(E)[z, z^{-1}; \tau]$. We first want to show that $\overline{U} \neq B$. Since

$$\deg(\mathcal{M}(-\boldsymbol{x})) = \deg \mathcal{M} - \deg \boldsymbol{x} = 8,$$

[Hartshorne 1977, Corollary IV.3.2] implies that $\mathcal{M}(-\boldsymbol{x})$ is very ample and generated by its sections $B_1 = H^0(E, \mathcal{M}(-\boldsymbol{x}))$. On the other hand, the inclusion $U \subseteq T$ forces $\overline{U} \subseteq \overline{T} = B(E, \mathcal{M}, \tau)$ and again $\overline{T}_1$ generates $\mathcal{M}$. Therefore, if $\overline{U} = B$ or even if $\overline{U}_1 = B_1$ then $\mathcal{M}(-\boldsymbol{x}) \subseteq \mathcal{M}$. Since $\boldsymbol{x}$ is not effective, this is impossible and so $\overline{U} \neq B$, as claimed.

We now turn to the homological questions. By [Levasseur 1992, Theorem 5.10], U is Auslander–Gorenstein, AS Gorenstein or CM if and only if the same holds for $\overline{U}$. Thus we can concentrate on $\overline{U}$. Since $B/\overline{U}$ is a nonzero, finite-dimensional vector space, and B is a domain, certainly $\mathrm{Ext}^1_{\overline{U}}(\Bbbk, \overline{U}) \neq 0$ (on either side). Since $\mathrm{GKdim}\, \overline{U} = \mathrm{GKdim}\, B = 2$ this certainly implies that $\overline{U}$ is not CM. Moreover, if we can prove that $\mathrm{Ext}^2_{\overline{U}}(\Bbbk, \overline{U}) \neq 0$ on either side, then $\overline{U}$ will be neither AS Gorenstein nor Auslander–Gorenstein.

By [Levasseur 1992, Proposition 6.5], $\mathrm{Ext}^i_B(\Bbbk, B) = \delta_{i,2}\Bbbk$, up to a shift in degree. Therefore [Rotman 2009, Corollary 10.65], with $A = \Bbbk$, $B = S$ and $R = C = \overline{U}$, gives

$$\mathrm{Ext}^2_{\overline{U}}(\Bbbk, \overline{U}) = \mathrm{Ext}^2_{\overline{U}}(B \otimes_B \Bbbk, \overline{U}) = \mathrm{Ext}^2_B(\Bbbk, J) \quad \text{for } J = \mathrm{Hom}_{\overline{U}}(B, \overline{U}). \tag{10.5}$$

Since $\overline{U} \overset{\bullet}{=} B$, clearly $L = B/J$ is also a nonzero finite-dimensional $\Bbbk$-vector space. We claim that the same is true of $\mathrm{Ext}^2_B(\Bbbk, J)$. As $\mathrm{Ext}^1_B(\Bbbk, B) = 0$, we have an exact sequence

$$0 \longrightarrow \operatorname{Ext}^1_B(\Bbbk, L) \longrightarrow \operatorname{Ext}^2_B(\Bbbk, J) \longrightarrow \operatorname{Ext}^2_B(\Bbbk, B) \longrightarrow \cdots . \tag{10.6}$$

Since $\dim_\Bbbk \operatorname{Ext}^1_B(\Bbbk, L) < \infty$, the claim will follow once we show that $\operatorname{Ext}^1_B(\Bbbk, L) \neq 0$.

As in [Artin and Zhang 1994, (7.1.2)], let $I(L)$ denote the largest essential extension of L by locally finite-dimensional modules. If $\operatorname{soc}(L)$ denotes the socle of L, then L and $\operatorname{soc}(L)$ have the same injective hulls and hence the same torsion-injective hulls $I(L) = I(\operatorname{soc}(L))$. By [Rogalski 2011, Lemma 2.2(2)], B satisfies χ in the sense of [Artin and Zhang 1994, Definition 3.2] and so, by [ibid., Proposition 7.7], $I(L)$ is a direct sum of copies of shifts of the vector space dual B^*. Since this is strictly larger than L, $\operatorname{Ext}^1_B(\Bbbk, L) \neq 0$ and the claim follows.

In conclusion, by (10.6) we know that $0 < \dim_\Bbbk \operatorname{Ext}^2_B(\Bbbk, J) < \infty$ and hence by (10.5) it follows that (up to a shift) $\Bbbk \hookrightarrow \operatorname{Ext}^2_{\overline{U}}(\Bbbk, \overline{U})$ as left $\overline{U}$-modules. As noted earlier, this shows that both Gorenstein conditions fail. □

Remark 10.7. (1) By expanding upon the above proof one can in fact show that U from Example 10.4 will have infinite injective dimension.

(2) Explicit computation shows that U is not uniquely determined by $\boldsymbol{x}$ as a subalgebra of T, although the factor $\overline{U}$ is determined in large degree. We do not know whether U is unique up to isomorphism.

Let U be a noetherian subring of T with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$. In Proposition 8.7, we had to assume that $U \not\subseteq \Bbbk + gT$ in order to find a g-divisible, equivalent order and this meant that the same assumption was needed for the rest of Section 8. In our next example we show that the conclusions of Proposition 8.7 can fail without this assumption, as does Theorem 8.11. Thus Proposition 9.1 is necessary for Section 9.

In order to define the ring, pick algebra generators of T in degree 1; say $T = \Bbbk\langle a_1, \dots, a_r\rangle$, set $T^g = \Bbbk\langle ga_1, \dots, ga_r\rangle$ and write $U = T^g\langle g\rangle \subset T$ for the subring of T generated by T^g and g.

Example 10.8. Keep T^g and $U = T^g\langle g\rangle$ as above. Then:

(1) There is a semigraded isomorphism $T^g \cong T$. Thus U is noetherian and there is a semigraded isomorphism $T[x]/(x^2 - g) \cong U$ mapping x to g. Moreover, $U^{(2)} = T^g$ and so $U^\circ = (T^g)^\circ \cong T^\circ$.

(2) $U \subseteq \Bbbk + gT$ and so $\overline{U} = \Bbbk$.

(3) gU is a prime ideal of U such that there is a semigraded isomorphism $U/gU \cong B = T/gT$.

(4) $\widehat{U} = T$ but T is not finitely generated as a right (or left) U-module.

(5) U is a maximal order with $Q_{\mathrm{gr}}(U) = Q_{\mathrm{gr}}(T)$.

Proof. (1–2) These are routine computations.

(3) Under the identification $U = T[x]/(x^2 - g)$, clearly $U/xU = T/gT$.

(4) For any $\theta \in T_n$ one has $g^n\theta \in T^g \subseteq U$ and hence $\widehat{U} = \widehat{T^g} = T$. If T were finitely generated as a (right) U-module then the factor $B = T/gT$ would be finitely generated as a module over the image $(U + gT)/gT = \overline{U}$ of U in B. This contradicts (2).

(5) Write $U = T[x]/(x^2 - g)$; thus $x \in U_1$ but the grading of T is shifted. If U is not a maximal order then there exists a cg ring $U \subsetneq V \subset Q(U)$ such that either $aV \subseteq U$ or $Va \subseteq U$ for some $0 \neq a \in U$. By symmetry we may assume the former, in which case $IV = I$ for the nonzero ideal $I = UaV$ of U. Thus $I^{(2)}V^{(2)} = I^{(2)}$, and $I^{(2)} \neq 0$ since U is a domain. Since $U^{(2)} = T$ is a maximal order by Proposition 4.10(4), it follows that $V^{(2)} = U^{(2)} = T$. Let $f \in V \smallsetminus U$ be homogeneous. Then f appears in odd degree and so $fx \in V^{(2)} = U^{(2)} = T$ and $f = tx^{-1}$ for some $t \in T$. However, $T = V^{(2)} \ni f^2 = (tx^{-1})^2 = t^2 g^{-1}$. Hence $t^2 \in gT$ which, since T/gT is a domain, forces $t = gt_1 \in gT$. But this implies that $f = tx^{-1} = xt_1 \in U$, a contradiction. Thus U is indeed a maximal order. Moreover, as $g \in U$, clearly each a_i lies in $Q_{\mathrm{gr}}(U)$ and hence $Q_{\mathrm{gr}}(T) = Q_{\mathrm{gr}}(U)$. □

In this paper we have only been concerned with two-sided noetherian rings, since we believe that this is the appropriate context for noncommutative geometry. For one-sided noetherian rings there are further examples that can appear, as is illustrated by the following example.

Example 10.9. Let J be a right ideal of T such that $g \in J$ and $\mathrm{GKdim}(T/J) = 1$. Then the idealiser $A = \mathbb{I}(J)$ is right but not left noetherian.

Proof. Let $\overline{J} = J/gJ$. Since $B = T/gT$ is just infinite [Rogalski 2011, Lemma 3.2], $\dim_{\Bbbk} T/TJ < \infty$. Since $TJ = \sum_{i=1}^{m} t_i J$ for some t_j, it follows that TJ and hence T are finitely generated right A-modules. Thus, by the proof of [Stafford and Zhang 1994, Theorem 3.2], A is right noetherian. On the other hand, B is not a finitely generated left A/gT-module, and so gT is an ideal of A that cannot be finitely generated as a left A-module. □

Acknowledgements

Part of this material is based upon work supported by the National Science Foundation under Grant No. 0932078 000, while the authors were in residence at the Mathematical Science Research Institute (MSRI) in Berkeley, California, during the spring semester of 2013. During this trip, Sierra was also partially supported by the Edinburgh Research Partnership in Engineering and Mathematics, and Stafford was partially supported by the Clay Mathematics Institute and Simons Foundation. The authors gratefully acknowledge the support of all these organisations.

Index of notation

References

[Artin 1997] M. Artin, "Some problems on three-dimensional graded domains", pp. 1–19 in *Representation theory and algebraic geometry* (Waltham, MA, 1995), edited by A. Martsinkovsky and G. Todorov, London Math. Soc. Lecture Note Ser. **238**, Cambridge Univ. Press, 1997. MR 99a:16037 Zbl 0888.16025

[Artin and Stafford 1995] M. Artin and J. T. Stafford, "Noncommutative graded domains with quadratic growth", *Invent. Math.* **122**:2 (1995), 231–276. MR 96g:16027 Zbl 0849.16022

[Artin and Van den Bergh 1990] M. Artin and M. Van den Bergh, "Twisted homogeneous coordinate rings", *J. Algebra* **133**:2 (1990), 249–271. MR 91k:14003 Zbl 0717.14001

[Artin and Zhang 1994] M. Artin and J. J. Zhang, "Noncommutative projective schemes", *Adv. Math.* **109**:2 (1994), 228–287. MR 96a:14004 Zbl 0833.14002

[Artin et al. 1990] M. Artin, J. Tate, and M. Van den Bergh, "Some algebras associated to automorphisms of elliptic curves", pp. 33–85 in *The Grothendieck Festschrift, I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR 92e:14002 Zbl 0744.14024

[Artin et al. 1991] M. Artin, J. Tate, and M. Van den Bergh, "Modules over regular algebras of dimension 3", *Invent. Math.* **106**:2 (1991), 335–388. MR 93e:16055 Zbl 0763.14001

[Artin et al. 1999] M. Artin, L. W. Small, and J. J. Zhang, "Generic flatness for strongly Noetherian algebras", *J. Algebra* **221**:2 (1999), 579–610. MR 2001a:16006 Zbl 0958.16024

[Björk and Ekström 1990] J.-E. Björk and E. K. Ekström, "Filtered Auslander–Gorenstein rings", pp. 425–448 in *Operator algebras, unitary representations, enveloping algebras, and invariant theory* (Paris, 1989), edited by A. Connes et al., Progr. Math. **92**, Birkhäuser, Boston, 1990. MR 92d:16017 Zbl 0733.16016

[Cozzens 1976] J. H. Cozzens, "Maximal orders and reflexive modules", *Trans. Amer. Math. Soc.* **219** (1976), 323–336. MR 54 #7524 Zbl 0334.16009

[Etingof and Ginzburg 2010] P. Etingof and V. Ginzburg, "Noncommutative del Pezzo surfaces and Calabi–Yau algebras", *J. Eur. Math. Soc.* (*JEMS*) **12**:6 (2010), 1371–1416. MR 2012d:16082 Zbl 1204.14004

[Goodearl and Warfield 1989] K. R. Goodearl and R. B. Warfield, Jr., *An introduction to noncommutative Noetherian rings*, London Math. Soc. Student Texts **16**, Cambridge Univ. Press, 1989. MR 91c:16001 Zbl 0679.16001

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

[Keeler et al. 2005] D. S. Keeler, D. Rogalski, and J. T. Stafford, "Naïve noncommutative blowing up", *Duke Math. J.* **126**:3 (2005), 491–546. MR 2006g:14005 Zbl 1082.14003

[Krause and Lenagan 1985] G. R. Krause and T. H. Lenagan, *Growth of algebras and Gelfand–Kirillov dimension*, Research Notes in Mathematics **116**, Pitman, Boston, 1985. MR 86g:16001 Zbl 0564.16001

[Levasseur 1992] T. Levasseur, "Some properties of noncommutative regular graded rings", *Glasgow Math. J.* **34**:3 (1992), 277–300. MR 93k:16045 Zbl 0824.16032

[McConnell and Robson 2001] J. C. McConnell and J. C. Robson, *Noncommutative Noetherian rings*, Revised ed., Graduate Studies in Mathematics **30**, American Mathematical Society, Providence, RI, 2001. MR 2001i:16039 Zbl 0980.16019

[Năstăsescu and van Oystaeyen 1982] C. Năstăsescu and F. van Oystaeyen, *Graded ring theory*, North-Holland Mathematical Library **28**, North-Holland Publishing Co., Amsterdam, 1982. MR 84i:16002 Zbl 0494.16001

[Rogalski 2011] D. Rogalski, "Blowup subalgebras of the Sklyanin algebra", *Adv. Math.* **226**:2 (2011), 1433–1473. MR 2012d:16083 Zbl 1207.14004

[Rogalski and Sierra 2012] D. Rogalski and S. J. Sierra, "Some projective surfaces of GK-dimension 4", *Compos. Math.* **148**:4 (2012), 1195–1237. MR 2956041 Zbl 1262.16024

[Rogalski and Stafford 2009] D. Rogalski and J. T. Stafford, "A class of noncommutative projective surfaces", *Proc. Lond. Math. Soc.* (3) **99**:1 (2009), 100–144. MR 2010j:14007 Zbl 1173.14005

[Rotman 2009] J. J. Rotman, *An introduction to homological algebra*, 2nd ed., Springer, New York, 2009. MR 2009i:18011 Zbl 1157.18001

[RSS 2014] D. Rogalski, S. J. Sierra, and J. T. Stafford, "Algebras in which every subalgebra is Noetherian", *Proc. Amer. Math. Soc.* **142**:9 (2014), 2983–2990. MR 3223353 Zbl 1311.16015

[RSS 2015] D. Rogalski, S. J. Sierra, and J. T. Stafford, "Noncommutative blowups of elliptic algebras", *Algebr. Represent. Theory* **18**:2 (2015), 491–529. MR 3336351 Zbl 06445654

[Sierra 2011] S. J. Sierra, "Classifying birationally commutative projective surfaces", *Proc. Lond. Math. Soc.* (3) **103**:1 (2011), 139–196. MR 2012i:16054 Zbl 1226.14004

[Stafford and Van den Bergh 2001] J. T. Stafford and M. Van den Bergh, "Noncommutative curves and noncommutative surfaces", *Bull. Amer. Math. Soc.* (*N.S.*) **38**:2 (2001), 171–216. MR 2002d:16036 Zbl 1042.16016

[Stafford and Zhang 1994] J. T. Stafford and J. J. Zhang, "Examples in non-commutative projective geometry", *Math. Proc. Cambridge Philos. Soc.* **116**:3 (1994), 415–433. MR 95h:14001 Zbl 0821.16026

[Stephenson 1997] D. R. Stephenson, "Algebras associated to elliptic curves", *Trans. Amer. Math. Soc.* **349**:6 (1997), 2317–2340. MR 97m:16080 Zbl 0868.16028

[Van den Bergh 1997] M. Van den Bergh, "Existence theorems for dualizing complexes over non-commutative graded and filtered rings", *J. Algebra* **195**:2 (1997), 662–679. MR 99b:16010 Zbl 0894.16020

[Van den Bergh 2001] M. Van den Bergh, *Blowing up of non-commutative smooth surfaces*, Mem. Amer. Math. Soc. **734**, Amer. Math. Soc., Providence, 2001. MR 2002k:16057 Zbl 0998.14002

[Van den Bergh 2011] M. Van den Bergh, "Noncommutative quadrics", *Int. Math. Res. Not.* **2011**:17 (2011), 3983–4026. MR 2012m:14004 Zbl 1311.14003

[Van den Bergh 2012] M. Van den Bergh, "Non-commutative $\mathbb{P}^1$-bundles over commutative schemes", *Trans. Amer. Math. Soc.* **364**:12 (2012), 6279–6313. MR 2958936 Zbl 06199600

[Zhang 1998] J. J. Zhang, "On lower transcendence degree", *Adv. Math.* **139**:2 (1998), 157–193. MR 2000a:16046 Zbl 0924.16015

Communicated by Efim Zelmanov
Received 2014-04-08 Revised 2015-03-10 Accepted 2015-09-03

drogalsk@math.ucsd.edu — *Department of Mathematics, University of California, San Diego, 9500 Gilman Drive #0112, La Jolla, CA 92093-0112, United States*

s.sierra@ed.ac.uk — *School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom*

toby.stafford@manchester.ac.uk — *School of Mathematics, University of Manchester, Alan Turing Building, Oxford Road, Manchester, M13 9PL, United Kingdom*

ALGEBRA AND NUMBER THEORY 9:9 (2015)
dx.doi.org/10.2140/ant.2015.9.2121

Congruence property in conformal field theory

Chongying Dong, Xingjun Lin and Siu-Hung Ng

The congruence subgroup property is established for the modular representations associated to any modular tensor category. This result is used to prove that the kernel of the representation of the modular group on the conformal blocks of any rational, C_2-cofinite vertex operator algebra is a congruence subgroup. In particular, the q-character of each irreducible module is a modular function on the same congruence subgroup. The Galois symmetry of the modular representations is obtained and the order of the anomaly for those modular categories satisfying some integrality conditions is determined.

Introduction

Modular invariance of characters of a rational conformal field theory (RCFT) has been known since the work of Cardy [1986], and it was proved by Zhu [1996] for rational and C_2-cofinite vertex operator algebras (VOA), which constitute a mathematical formalization of RCFT. The associated matrix representation of $\mathrm{SL}_2(\mathbb{Z})$ relative to the distinguished basis, formed by the trace functions of the irreducible modules or primary fields, is a powerful tool in the study of vertex operator algebras and conformal field theory. This matrix representation conceives many intriguing arithmetic properties, and the Verlinde formula [1988] is certainly a notable example. Moreover, it has been shown that these matrices representing the modular group are defined over a certain cyclotomic field [de Boer and Goeree 1991].

An important characteristic of the modular representation ρ associated with a RCFT is its kernel. It has been conjectured by many authors that the kernel is a congruence subgroup of a certain level n (see [Moore 1987; Eholzer 1995; Eholzer and Skoruppa 1995; Dong and Mason 1996; Bauer et al. 1997]). Eholzer further conjectured that this representation is defined over the n-th cyclotomic field $\mathbb{Q}_n$. In this case, the Galois group $\mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q})$ acts on the representation ρ by

Dong and Ng were partially supported by NSF grants.
MSC2010: primary 17B69; secondary 18D10, 20H05, 81R05.
Keywords: Frobenius–Schur indicator, modular tensor category, modular group, vertex operator algebra.

its entrywise action. Coste and Gannon [1994] proved that ρ determines a signed permutation matrix G_σ for each automorphism σ of $\mathbb{Q}_n$. They also conjectured that the representation $\sigma^2\rho$ is equivalent to ρ under the intertwining operator G_σ. These conjectural properties were summarized as the congruence property of the modular data associated with RCFT in [Coste and Gannon 1999; Gannon 2006]. These remarkable properties of RCFT were established by Bantay [2003] under certain assumptions, and by Coste and Gannon [1994] under the condition that the order of the Dehn twist is odd. In the formalization of RCFT through conformal nets, the congruence property was proved by Xu [2006].

In this paper we give a positive answer to the conjecture on the congruence property for a rational and C_2-cofinite vertex operator algebra V. Such a V has only finitely many irreducible modules [Dong et al. 1998a] $M^0, \dots, M^p$ up to isomorphism and there exist $\lambda_i \in \mathbb{C}$ for $i = 0, \dots, p$ such that

$$M^i = \bigoplus_{n=0}^{\infty} M^i_{\lambda_i+n}$$

where $M^i_{\lambda_i} \neq 0$ and $L(0)|_{M^i_{\lambda_i+n}} = \lambda_i + n$ for any $n \in \mathbb{Z}$. Moreover, λ_i and the central charge c are rational numbers (see [Dong et al. 2000]).

The trace function for $v \in V_k$ on M^i is defined as

$$Z_i(v, q) = q^{\lambda_i - c/24} \sum_{n=0}^{\infty} (\operatorname{tr}_{M^i_{\lambda_i+n}} o(v)) q^n$$

where $o(v) = v_{k-1}$ is the $(k-1)$-st component operator of $Y(v, z) = \sum_{n\in\mathbb{Z}} v_n z^{-n-1}$ which maps each homogeneous subspace of M^i to itself. If $v = \mathbb{1}$ is the vacuum vector we get the q-character $\chi_i(q)$ of M^i. It is proved in [Zhu 1996] that if V is C_2-cofinite then $Z_i(v, q)$ converges to a holomorphic function on the upper half-plane in variable τ where $q = e^{2\pi i\tau}$. By abusing the notation we also denote this holomorphic function by $Z_i(v, \tau)$. There is another vertex operator algebra structure on V [Zhu 1996] with grading $V = \oplus_{n\in\mathbb{Z}} V_{[n]}$. We will write $\mathrm{wt}[v] = n$ if $v \in V_{[n]}$. Then there is a representation ρ_V of the modular group $\mathrm{SL}_2(\mathbb{Z})$ on the space spanned by $\{Z_i(v, \tau) \mid i = 0, \dots, p\}$:

$$Z_i(v, \gamma\tau) = (c\tau + d)^{\mathrm{wt}[v]} \sum_{j=0}^{p} \gamma_{ij} Z_j(v, \tau)$$

where $\gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z})$ and $\rho_V(\gamma) = [\gamma_{ij}]$ [Zhu 1996].

Theorem I. *Let V be a rational, C_2-cofinite, self-dual simple vertex operator algebra. Then each $Z_i(v, \tau)$ is a modular form of weight* $\mathrm{wt}[v]$ *on a congruence subgroup of* $\mathrm{SL}_2(\mathbb{Z})$ *of level n, which is the smallest positive integer such that*

$n(\lambda_i - c/24)$ is an integer for all i. In particular, each q-character χ_i is a modular function on the same congruence subgroup.

We should remark that the modularity of the q-characters of irreducible modules for some known vertex operator algebras such as those associated to the highest weight unitary representations for Kac–Moody algebras [Kac and Peterson 1984; Kac 1990] and the Virasoro algebra [Rocha-Caridi 1985] were previously known. The readers are referred to [Dong et al. 2001] for the modularity of $Z_i(v, \tau)$ when V is a vertex operator algebra associated to a positive definite even lattice.

According to [Huang 2008a; 2008b], the category $\mathcal{C}_V$ of modules of a rational and C_2-cofinite vertex operator algebra V under the tensor product defined in [Huang and Lepowsky 1995a; 1995b; 1995c; Huang 1995] is a modular tensor category over $\mathbb{C}$. To establish this theorem we have to turn our attention to general modular tensor categories.

Modular tensor categories, or simply called modular categories, play an integral role in the Reshetikhin–Turaev TQFT invariant of 3-manifolds [Turaev 2010] and topological quantum computation [Wang 2010]. They also constitute another formalization of RCFT [Moore and Seiberg 1990; Bakalov and Kirillov 2001].

Parallel to a rational conformal field theory, associated to a modular category $\mathcal{A}$ are the invertible matrices $\tilde{s}$ and $\tilde{t}$ indexed by the set Π of isomorphism classes of simple objects of $\mathcal{A}$. These matrices define a projective representation $\bar{\rho}_{\mathcal{A}}$ of $\mathrm{SL}_2(\mathbb{Z})$ by the assignment

$$\mathfrak{s} := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mapsto \tilde{s} \quad \text{and} \quad \mathfrak{t} := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \mapsto \tilde{t},$$

and the well-known presentation $\mathrm{SL}_2(\mathbb{Z}) = \langle \mathfrak{s}, \mathfrak{t} \mid \mathfrak{s}^4 = 1,\ (\mathfrak{s}\mathfrak{t})^3 = \mathfrak{s}^2 \rangle$ of the modular group. It was proved by Ng and Schauenburg [2010] that the kernel of this projective representation of $\mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup of level N, where N is the order of $\tilde{t}$. Moreover, both $\tilde{s}$ and $\tilde{t}$ are matrices over $\mathbb{Q}_N$. For factorizable semisimple Hopf algebras, the corresponding result was proved previously by Sommerhäuser and Zhu [2012].

The projective representation $\bar{\rho}_{\mathcal{A}}$ can be lifted to an ordinary representation of $\mathrm{SL}_2(\mathbb{Z})$ which is called a *modular representation of* $\mathcal{A}$ in [Ng and Schauenburg 2010]. There are only finitely many modular representations of $\mathcal{A}$ but, in general, none of them is a canonical choice. However, if $\mathcal{A}$ is the Drinfeld center of a spherical fusion category, then $\mathcal{A}$ is modular (see [Müger 2003b]) and it admits a canonical modular representation defined over $\mathbb{Q}_N$ whose kernel is a congruence subgroup of level N (see [Ng and Schauenburg 2010]). The canonical modular representation of the module category over the Drinfeld double of a semisimple Hopf algebra was shown to have a congruence kernel as well as Galois symmetry (see Theorem II (iii) and (iv)) in [Sommerhäuser and Zhu 2012].

The second main theorem of this paper is to prove that the congruence property and Galois symmetry holds for all modular representations of any modular category.

Theorem II. *Let $\mathcal{A}$ be a modular category over any algebraically closed field $\Bbbk$ of characteristic zero with the set of isomorphism classes of simple objects Π and Frobenius–Schur exponent N. Suppose $\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_\Pi(\Bbbk)$ is a modular representation of $\mathcal{A}$ where $\mathrm{GL}_\Pi(\Bbbk)$ denotes the group of invertible matrices over $\Bbbk$ indexed by Π. Set $s = \rho(\mathfrak{s})$ and $t = \rho(\mathfrak{t})$. Then*:

(i) $\ker \rho$ *is a congruence subgroup of level n where $n = \mathrm{ord}(t)$ and*, *moreover*, $N \mid n \mid 12N$.

(ii) ρ *is $\mathbb{Q}_n$-rational, i.e.*, $\mathrm{im}\,\rho \le \mathrm{GL}_\Pi(\mathbb{Q}_n)$, *where $\mathbb{Q}_n = \mathbb{Q}(\zeta_n)$ for some primitive n-th root of unity $\zeta_n \in \Bbbk$.*

(iii) *For $\sigma \in \mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q})$, the matrix $G_\sigma = \sigma(s)s^{-1}$ is a signed permutation matrix, and*
$$\sigma^2(\rho(\gamma)) = G_\sigma \rho(\gamma) G_\sigma^{-1}$$
for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. In particular, if $(G_\sigma)_{ij} = \epsilon_\sigma(i)\delta_{\hat\sigma(i)j}$ for some sign function ϵ_σ and permutation $\hat\sigma$ on Π, then $\sigma^2(t_{ii}) = t_{\hat\sigma(i)\hat\sigma(i)}$ for all $i \in \Pi$.

(iv) *Let a be an integer relatively prime to n with an inverse b modulo n. For the automorphism σ_a of $\mathbb{Q}_n$ given by $\zeta_n \mapsto \zeta_n^a$,*
$$G_{\sigma_a} = t^a s t^b s t^a s^{-1}.$$

We return to the modular tensor category $\mathcal{C}_V$ associated to a rational, C_2-cofinite and self-dual vertex operator algebra V. This yields a projective representation of $\mathrm{SL}_2(\mathbb{Z})$ on space spanned by the equivalence classes of irreducible V-modules. We show in Theorem 3.10 that the representation ρ_V of $\mathrm{SL}_2(\mathbb{Z})$ is a modular representation of $\mathcal{C}_V$. This implies that the kernel of ρ_V is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$.

Although the congruence property proved in Theorem II is motivated by solving the congruence property conjecture on the trace functions of vertex operator algebras, the result has its own importance. We will discuss this in the rest of the introduction.

It was also shown in [Sommerhäuser and Zhu 2012] that the (unnormalized) T-matrix $\tilde{t}$ of the module category over a factorizable Hopf algebra also enjoys the Galois symmetry $\sigma^2(\tilde{t}) = G_\sigma \tilde{t} G_\sigma^{-1}$ for any $\sigma \in \mathrm{Gal}(\mathbb{Q}_N/\mathbb{Q})$. However, this extra symmetry does not hold for a general modular category $\mathcal{A}$ (see Example 4.6). This condition is, in fact, related to the order of the quotient of the Gauss sums, called the *anomaly*, of $\mathcal{A}$. It is proved in Proposition 4.7 that Galois symmetry of the T-matrix is equivalent to the condition that the anomaly is a fourth root of unity. We will prove in Proposition 6.7 that the anomaly of any *integral* modular category is always a fourth root of unity. Therefore, the T-matrix of any integral modular category enjoys the Galois symmetry. For a *weakly integral* modular category, such as the Ising model, the anomaly is always an eighth root of unity (Theorem 6.10).

Using Theorem II, we uncover some relations among the global dimension $\dim \mathcal{A}$, the Frobenius–Schur exponent N and the order of the anomaly α of a modular category $\mathcal{A}$. We define

$$J_{\mathcal{A}} = (-1)^{1+\operatorname{ord}\alpha}$$

to record the parity of the order of the anomaly. If N is not a multiple of 4, then $J_{\mathcal{A}} \dim \mathcal{A}$ has a square root in $\mathbb{Q}_N$. If, in addition, $\dim \mathcal{A}$ is an odd integer, then $J_{\mathcal{A}}$ coincides with the Jacobi symbol $\left(\frac{-1}{\dim \mathcal{A}}\right)$. The consequence of this observation is a result closely related to the Cauchy theorem of integral fusion category.

The organization of this paper is as follows: Section 1 covers some basic definitions, conventions and preliminary results on spherical fusion categories and modular categories. In Section 2, we prove the congruence property, Theorem II (i) and (ii), by proving a lifting theorem of modular projective representations with congruence kernels. In Section 3, we prove the associated representation of modular invariance of trace functions of a rational, C_2-cofinite vertex operator algebra V is a modular representation of its module category. Using Theorem II (i) and (ii), we prove Theorem I: the trace functions of V are modular forms. In Section 4, we assume the technical Lemma 4.2 to prove the Galois symmetry of modular categories as well as RCFTs, Theorem II (iii) and (iv). Section 5 is devoted to the proof of Lemma 4.2 by using generalized Frobenius–Schur indicators. In Section 6, we use the congruence property and Galois symmetry of modular categories (Theorem II) to uncover some arithmetic relations among the global dimension, the Frobenius–Schur exponent and the anomaly of a modular category. In particular, we determine the order of the anomaly of a modular category satisfying certain integrality conditions.

1. Basics of modular tensor categories

In this section, we will collect some conventions and preliminary results on spherical fusion categories and modular categories. Most of these results are quite well-known, and the readers are referred to [Turaev 2010; Bakalov and Kirillov 2001; Ng and Schauenburg 2007a; 2007b; 2008; 2010] and the references therein.

Throughout this paper, $\Bbbk$ is always assumed to be an algebraically closed field of characteristic zero. The group of invertible matrices over a commutative ring K indexed by Π is denoted by $\mathrm{GL}_\Pi(K)$, and we will write $\mathrm{PGL}_\Pi(K)$ for its associated projective linear group. If $\Pi = \{1, \dots, r\}$ for some positive integer r, then $\mathrm{GL}_\Pi(K)$ (resp. $\mathrm{PGL}_\Pi(K)$) will be denoted by the standard notation $\mathrm{GL}_r(K)$ (resp. $\mathrm{PGL}_r(K)$) instead.

For any primitive n-th root of unity $\zeta_n \in \Bbbk$, we let $\mathbb{Q}_n := \mathbb{Q}(\zeta_n)$ be the smallest subfield of $\Bbbk$ containing all the n-th roots of unity in $\Bbbk$. Recall that $\mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q})$ is isomorphic to $U(\mathbb{Z}_n)$, the group of units of $\mathbb{Z}_n$. Let a be an integer relatively prime to n. The associated $\sigma_a \in \mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q})$ is defined by

$$\sigma_a(\zeta_n) = \zeta_n^a.$$

Define $\mathbb{Q}_{\mathrm{ab}} = \bigcup_{n\in\mathbb{N}} \mathbb{Q}_n$, the abelian closure of $\mathbb{Q}$ in $\Bbbk$. Since $\mathbb{Q}_n$ is Galois over $\mathbb{Q}$, we have $\sigma(\mathbb{Q}_n) = \mathbb{Q}_n$ for all automorphisms σ of $\mathbb{Q}_{\mathrm{ab}}$. Moreover, the restriction map $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}}) \xrightarrow{\mathrm{res}} \mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q})$ is surjective for all positive integers n. Thus, for any integer a relatively prime to n, there exists a $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ such that $\sigma|_{\mathbb{Q}_n} = \sigma_a$.

1.1. ***Spherical fusion categories.*** In a left rigid monoidal category $\mathcal{C}$ with tensor product $\otimes$ and unit object $\mathbf{1}$, we denote a left dual $V^\vee$ of $V \in \mathcal{C}$ with morphisms $\mathrm{db}_V : \mathbf{1} \to V \otimes V^\vee$ and $\mathrm{ev}_V : V^\vee \otimes V \to \mathbf{1}$ by the triple $(V^\vee, \mathrm{db}_V, \mathrm{ev}_V)$. The left duality can be extended to a monoidal functor $(-)^\vee : \mathcal{C} \to \mathcal{C}^{\mathrm{op}}$, and so $(-)^{\vee\vee} : \mathcal{C} \to \mathcal{C}$ defines a monoidal equivalence. Moreover we can choose $\mathbf{1}^\vee = \mathbf{1}$. A *pivotal structure* of $\mathcal{C}$ is an isomorphism $j : \mathrm{Id}_{\mathcal{C}} \to (-)^{\vee\vee}$ of monoidal functors. One can respectively define the left and the right pivotal traces of an endomorphism $f : V \to V$ in $\mathcal{C}$ as

$$\underline{\mathrm{ptr}}^\ell(f) = \left(\mathbf{1} \xrightarrow{\mathrm{db}_{V^\vee}} V^\vee \otimes V^{\vee\vee} \xrightarrow{\mathrm{id}\otimes j_V^{-1}} V^\vee \otimes V \xrightarrow{\mathrm{id}\otimes f} V^\vee \otimes V \xrightarrow{\mathrm{ev}_V} \mathbf{1}\right)$$

and

$$\underline{\mathrm{ptr}}^r(f) = \left(\mathbf{1} \xrightarrow{\mathrm{db}_V} V \otimes V^\vee \xrightarrow{f\otimes \mathrm{id}} V \otimes V^\vee \xrightarrow{j_V\otimes \mathrm{id}} V^{\vee\vee} \otimes V^\vee \xrightarrow{\mathrm{ev}_{V^\vee}} \mathbf{1}\right).$$

The pivotal structure is called *spherical* if the two pivotal traces coincide for all endomorphisms f in $\mathcal{C}$.

A *pivotal* (resp. *spherical*) *category* $(\mathcal{C}, j)$ is a left rigid monoidal category $\mathcal{C}$ equipped with a pivotal (resp. spherical) structure j. We will simply denote the pair $(\mathcal{C}, j)$ by $\mathcal{C}$ when there is no ambiguity. The left and the right pivotal dimensions of $V \in \mathcal{C}$ are defined as $d_\ell(V) = \underline{\mathrm{ptr}}^\ell(\mathrm{id}_V)$ and $d_r(V) = \underline{\mathrm{ptr}}^r(\mathrm{id}_V)$ respectively. In a spherical category, the pivotal traces and dimensions will be denoted by $\underline{\mathrm{ptr}}(f)$ and $d(V)$ (or $\dim V$), respectively.

A *fusion category* $\mathcal{C}$ over the field $\Bbbk$ is an abelian $\Bbbk$-linear semisimple (left) rigid monoidal category with a simple unit object $\mathbf{1}$, finite-dimensional morphism spaces and finitely many isomorphism classes of simple objects (see [Etingof et al. 2005]). We will denote by $\Pi_{\mathcal{C}}$ the set of isomorphism classes of simple objects of $\mathcal{C}$, and by 0 the isomorphism class of $\mathbf{1}$, unless stated otherwise. If $i \in \Pi_{\mathcal{C}}$, we write i^* for the (left) dual of the isomorphism class i. Moreover, $i \mapsto i^*$ defines a permutation of order at most 2 on $\Pi_{\mathcal{C}}$.

In a spherical fusion category $\mathcal{C}$ over $\Bbbk$, $d(V)$ can be identified with a scalar in $\Bbbk$ for $V \in \mathcal{C}$. We use the abbreviation $d_i \in \Bbbk$ for the pivotal dimension of $i \in \Pi_{\mathcal{C}}$. By

[Müger 2003a, Lemma 2.8], $d_i = d_{i^*}$ for all $i \in \Pi_{\mathcal{C}}$. The global dimension $\dim \mathcal{C}$ of $\mathcal{C}$ is defined by

$$\dim \mathcal{C} = \sum_{i \in \Pi_{\mathcal{C}}} d_i^2.$$

A pivotal category $(\mathcal{C}, j)$ is said to be *strict* if $\mathcal{C}$ is a strict monoidal category and if the pivotal structure j and the canonical isomorphism $(V \otimes W)^\vee(1/2) \to W^\vee \otimes V^\vee$ are identities. It has been proved in [Ng and Schauenburg 2007b, Theorem 2.2] that every pivotal category is *pivotally equivalent* to a strict pivotal category.

1.2. ***Representations of the modular group.*** The modular group $\mathrm{SL}_2(\mathbb{Z})$ is the group of 2×2 integral matrices with determinant 1. It is well-known that the modular group is generated by

$$\mathfrak{s} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathfrak{t} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{with defining relations } (\mathfrak{s}\mathfrak{t})^3 = \mathfrak{s}^2,\ \mathfrak{s}^4 = \mathrm{id}. \tag{1-1}$$

We denote by $\Gamma(n)$ the kernel of the reduction modulo n epimorphism $\pi_n : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}_n)$. A subgroup L of $\mathrm{SL}_2(\mathbb{Z})$ is called a *congruence subgroup of level n* if n is the least positive integer for which $\Gamma(n) \le L$.

For any pair of matrices $A, B \in \mathrm{GL}_r(\Bbbk)$, with $r \in \mathbb{N}$, satisfying the conditions

$$A^4 = \mathrm{id} \quad \text{and} \quad (AB)^3 = A^2,$$

one can define a representation $\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_r(\Bbbk)$ such that $\rho(\mathfrak{s}) = A$ and $\rho(\mathfrak{t}) = B$ via the presentation (1-1) of $\mathrm{SL}_2(\mathbb{Z})$.

Suppose $\bar{\rho} : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{PGL}_r(\Bbbk)$ is a projective representation of $\mathrm{SL}_2(\mathbb{Z})$. A *lifting* of $\bar{\rho}$ is an *ordinary* representation $\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_r(\Bbbk)$ such that $\eta \circ \rho = \bar{\rho}$, where $\eta : \mathrm{GL}_r(\Bbbk) \to \mathrm{PGL}_r(\Bbbk)$ is the natural surjection map. One can always lift $\bar{\rho}$ to a representation $\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_r(\Bbbk)$ as follows: let $\hat{A}, \hat{B} \in \mathrm{GL}_r(\Bbbk)$ such that $\bar{\rho}(\mathfrak{s}) = \eta(\hat{A})$ and $\bar{\rho}(\mathfrak{t}) = \eta(\hat{B})$. Then

$$\hat{A}^4 = \mu_s \,\mathrm{id} \quad \text{and} \quad (\hat{A}\hat{B})^3 = \mu_t \hat{A}^2$$

for some scalars $\mu_s, \mu_t \in \Bbbk^\times$. Take $\lambda, \zeta \in \Bbbk$ such that $\lambda^4 = \mu_s$ and $\zeta^3 = \mu_t/\lambda$, and set $A = \hat{A}/\lambda$ and $B = \hat{B}/\zeta$. Then we have

$$A^4 = \mathrm{id} \quad \text{and} \quad (AB)^3 = A^2.$$

Therefore, the assignment $\rho : \mathfrak{s} \mapsto A,\ \mathfrak{t} \mapsto B$ defines a lifting of $\bar{\rho}$.

Let ρ be a lifting of $\bar{\rho}$. Suppose $x \in \Bbbk$ is a 12-th root of unity. Then the assignment

$$\rho_x : \mathfrak{s} \mapsto \frac{1}{x^3}\rho(\mathfrak{s}), \quad \mathfrak{t} \mapsto x\rho(\mathfrak{t}) \tag{1-2}$$

also defines a lifting of $\bar{\rho}$. If $\rho' : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_r(\Bbbk)$ is another lifting of $\bar{\rho}$, then

$$\rho'(\mathfrak{s}) = a\rho(\mathfrak{s}) \quad \text{and} \quad \rho'(\mathfrak{t}) = b\rho(\mathfrak{t})$$

for some $a, b \in \Bbbk^\times$. It follows immediately from (1-1) that $a^4 = 1$ and $(ab)^3 = a^2$. This implies $b^{12} = 1$ and $b^{-3} = a$. Therefore, we have $\rho' = \rho_b$ and so $\bar{\rho}$ has at most 12 liftings.

For any 12-th root of unity $x \in \Bbbk$, the assignment $\chi_x : \mathfrak{s} \mapsto x^{-3}$, $\mathfrak{t} \mapsto x$ defines a linear character of $\mathrm{SL}_2(\mathbb{Z})$. It is straightforward to check that $\chi_x \otimes \rho$ is isomorphic to ρ_x as representations of $\mathrm{SL}_2(\mathbb{Z})$. Therefore, the lifting of $\bar{\rho}$ is *unique* up to a linear character of $\mathrm{SL}_2(\mathbb{Z})$.

1.3. ***Modular categories.*** Following [Kassel 1995], a *twist* (or *ribbon structure*) of a left rigid braided monoidal category $\mathcal{C}$ with a braiding c is an automorphism θ of the identity functor $\mathrm{Id}_{\mathcal{C}}$ satisfying

$$\theta_{V\otimes W} = (\theta_V \otimes \theta_W) \circ c_{W,V} \circ c_{V,W}, \quad \theta_V^\vee = \theta_{V^\vee}$$

for $V, W \in \mathcal{C}$. Associated to the braiding c is the Drinfeld isomorphism $u : \mathrm{Id}_{\mathcal{C}} \to (-)^{\vee\vee}$. When $\mathcal{C}$ is a braided fusion category over $\Bbbk$, there is a one-to-one correspondence between twists θ and spherical structures j of $\mathcal{C}$ given by $\theta = u^{-1}j$ (see [Ng and Schauenburg 2007a, p. 38] for more details).

A *modular tensor category* over $\Bbbk$ (see [Turaev 2010; Bakalov and Kirillov 2001]), also called a modular category, is a braided spherical fusion category $\mathcal{A}$ over $\Bbbk$ such that the S-matrix of $\mathcal{A}$ defined by

$$\tilde{s}_{ij} = \underline{\mathrm{ptr}}(c_{V_j,V_{i^*}} \circ c_{V_{i^*},V_j})$$

is nonsingular, where V_j denotes an object in the class $j \in \Pi_{\mathcal{A}}$. In this case, the associated ribbon structure θ is of finite order N (see [Vafa 1988; Bakalov and Kirillov 2001]). Let $\theta_{V_i} = \theta_i \,\mathrm{id}_{V_i}$ for some $\theta_i \in \Bbbk$. Since $\theta_{\mathbf{1}} = \mathrm{id}_{\mathbf{1}}$, we have $\theta_0 = 1$. The *T-matrix* $\tilde{t}$ of $\mathcal{A}$ is defined by $\tilde{t}_{ij} = \delta_{ij}\theta_j$ for $i, j \in \Pi_{\mathcal{A}}$. It is immediate to see that $\mathrm{ord}(\tilde{t}) = N$, which is called the *Frobenius–Schur exponent* of $\mathcal{A}$ and denoted by $\mathrm{FSexp}(\mathcal{A})$, is finite (see [Ng and Schauenburg 2007a, Theorem 7.7]).

The matrices $\tilde{s}, \tilde{t}$ of a modular category $\mathcal{A}$ satisfy the conditions

$$(\tilde{s}\tilde{t})^3 = p_{\mathcal{A}}^+ \tilde{s}^2, \quad \tilde{s}^2 = p_{\mathcal{A}}^+ p_{\mathcal{A}}^- C, \quad C\tilde{t} = \tilde{t}C, \quad C^2 = \mathrm{id}, \tag{1-3}$$

where $p_{\mathcal{A}}^\pm = \sum_{i\in\Pi_{\mathcal{A}}} d_i^2 \theta_i^{\pm 1}$ are called the *Gauss sums*, and $C = [\delta_{ij^*}]_{i,j\in\Pi_{\mathcal{A}}}$ is called the *charge conjugation matrix* of $\mathcal{A}$. The quotient $p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ is a root of unity (see [Bakalov and Kirillov 2001, Theorem 3.1.19] or [Vafa 1988]), and

$$p_{\mathcal{A}}^+ p_{\mathcal{A}}^- = \dim \mathcal{A} \neq 0. \tag{1-4}$$

Moreover, $\tilde{s}$ satisfies

$$\tilde{s}_{ij} = \tilde{s}_{ji} \quad \text{and} \quad \tilde{s}_{ij^*} = \tilde{s}_{i^*j} \tag{1-5}$$

for all $i, j \in \Pi_{\mathcal{A}}$.

The relations (1-3) imply that

$$\bar{\rho}_{\mathcal{A}} : \mathfrak{s} \mapsto \eta(\tilde{s}), \quad \mathfrak{t} \mapsto \eta(\tilde{t}) \tag{1-6}$$

defines a projective representation of $\mathrm{SL}_2(\mathbb{Z})$, where $\eta : \mathrm{GL}_{\Pi_{\mathcal{A}}}(\Bbbk) \to \mathrm{PGL}_{\Pi_{\mathcal{A}}}(\Bbbk)$ is the natural surjection. By [Ng and Schauenburg 2010, Theorem 6.8], $\ker \bar{\rho}_{\mathcal{A}}$ is a congruence subgroup of level N.

It is well-known that $\bar{\rho}_{\mathcal{A}}$ can be lifted to an ordinary representation (see [Bakalov and Kirillov 2001, Remark 3.1.9] or Section 1.2). Following [Ng and Schauenburg 2010], a lifting ρ of $\bar{\rho}_{\mathcal{A}}$ is called a *modular representation* of $\mathcal{A}$. By (1-4), for any 6-th root $\zeta \in \Bbbk$ of $p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$, we have that $(p_{\mathcal{A}}^+/\zeta^3)^2 = \dim \mathcal{A}$. It follows from (1-3) that the assignment

$$\rho^{\zeta} : \mathfrak{s} \mapsto \frac{\zeta^3}{p_{\mathcal{A}}^+}\tilde{s}, \quad \mathfrak{t} \mapsto \frac{1}{\zeta}\tilde{t} \tag{1-7}$$

defines a modular representation of $\mathcal{A}$.

Thus, if ρ is a modular representation of $\mathcal{A}$, it follows from Section 1.2 that $\rho = \rho_x^{\zeta}$ for some 12-th root of unity $x \in \Bbbk$. Thus $\rho(\mathfrak{s})^2 = \pm C$. More precisely, $\rho(\mathfrak{s})^2 = x^6 C$.

A modular category $\mathcal{A}$ is called *anomaly-free* if the quotient $p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ equals 1. The terminology addresses the associated anomaly-free TQFT with such a modular category [Turaev 2010]. In this spirit, we will simply call the quotient $\alpha_{\mathcal{A}} := p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ the *anomaly* of $\mathcal{A}$.

If $\mathcal{A}$ is an anomaly-free modular category, then $p_{\mathcal{A}}^+$ is a *canonical* choice of square root of $\dim \mathcal{A}$, and hence a *canonical* modular representation of $\mathcal{A}$ is determined by the assignment

$$\rho_{\mathcal{A}} : \mathfrak{s} \mapsto \frac{1}{p_{\mathcal{A}}^+}\tilde{s}, \quad \mathfrak{t} \mapsto \tilde{t}. \tag{1-8}$$

For any modular category $\mathcal{A}$ over $\mathbb{C}$, we have that $\dim \mathcal{A} > 0$ (see [Etingof et al. 2005]). The *central charge* $\boldsymbol{c}$ of $\mathcal{A}$ is a rational number modulo 8 defined by $\exp(\pi i \boldsymbol{c}/4) = p_{\mathcal{A}}^+/\sqrt{\dim \mathcal{A}}$ where $\sqrt{\dim \mathcal{A}}$ denotes the positive square root of $\dim \mathcal{A}$, and so the anomaly α of $\mathcal{A}$ is given by

$$\alpha = \exp\left(\frac{\pi i \boldsymbol{c}}{2}\right). \tag{1-9}$$

We will show in Theorem 3.10 that the central charge $\boldsymbol{c}$ of the modular category $\mathcal{C}_V$ is equal to central charge c of V modulo 4.

Remark. The S- and T-matrices of a modular category are preserved by equivalence of braided pivotal categories over $\Bbbk$, and so are the dimensions of simple objects, the global dimension, the Gauss sums and the anomaly. By the last paragraph of Section 1.1, without loss of generality, we may assume that the underlying pivotal category of a modular category over $\Bbbk$ is *strict*.

1.4. ***Quantum doubles of spherical fusion categories.*** Let $\mathcal{C}$ be a strict monoidal category. The left Drinfeld center $Z(\mathcal{C})$ of $\mathcal{C}$ is a category whose objects are pairs of the form $\boldsymbol{X} = (X, \sigma_X)$, where X is an object of $\mathcal{C}$, and the *half-braiding* $\sigma_X(-) : X \otimes (-) \to (-) \otimes X$ is a natural isomorphism satisfying the properties $\sigma_X(\mathbf{1}) = \mathrm{id}_X$ and

$$(\mathrm{id}_V \otimes \sigma_X(W)) \circ (\sigma_X(V) \otimes \mathrm{id}_W) = \sigma_X(V \otimes W)$$

for all $V, W \in \mathcal{C}$. It is well-known that $Z(\mathcal{C})$ is a braided strict monoidal category (see [Kassel 1995]) with unit object $(\mathbf{1}, \sigma_{\mathbf{1}})$ and tensor product $(X, \sigma_X) \otimes (Y, \sigma_Y) := (X \otimes Y, \sigma_{X\otimes Y})$, where

$$\sigma_{X\otimes Y}(V) = (\sigma_X(V) \otimes \mathrm{id}_Y) \circ (\mathrm{id}_X \otimes \sigma_Y(V)), \quad \sigma_{\mathbf{1}}(V) = \mathrm{id}_V$$

for $V \in \mathcal{C}$. The forgetful functor $Z(\mathcal{C}) \to \mathcal{C}$, $\boldsymbol{X} = (X, \sigma_X) \mapsto X$, is a strict monoidal functor.

When $\mathcal{C}$ is a (strict) spherical fusion category over $\Bbbk$, by Müger's result [2003b], the center $Z(\mathcal{C})$ is a modular category over $\Bbbk$ with the inherited spherical structure from $\mathcal{C}$. In addition,

$$p^+_{Z(\mathcal{C})} = \dim \mathcal{C} = p^-_{Z(\mathcal{C})}.$$

Therefore, $Z(\mathcal{C})$ is anomaly-free and it admits a canonical modular representation $\rho_{Z(\mathcal{C})}$ described in (1-8). In particular,

$$\rho_{Z(\mathcal{C})}(\mathfrak{t}) = \tilde{t} \quad \text{and} \quad \rho_{Z(\mathcal{C})}(\mathfrak{s}) = \frac{1}{\dim \mathcal{C}}\tilde{s}, \tag{1-10}$$

which is called the *canonical normalization* of the S-matrix of $Z(\mathcal{C})$. By [Ng and Schauenburg 2010, Theorems 6.7 and 7.1], $\ker \rho_{Z(\mathcal{C})}$ is a congruence subgroup of level N, and $\operatorname{im} \rho_{Z(\mathcal{C})} \le \mathrm{GL}_{\Pi_{Z(\mathcal{C})}}(\mathbb{Q}_N)$, where $N = \operatorname{ord}(\tilde{t})$.

2. Rationality and kernels of modular representations

In this section, we prove the congruence property given in (i) and (ii) of Theorem II. Recall that a projective representation $\bar{\rho} : G \to \mathrm{PGL}_r(\Bbbk)$ of a group G determines a cohomology class $\kappa_{\bar{\rho}} \in H^2(G, \Bbbk^\times)$. For any section $\iota : \mathrm{PGL}_r(\Bbbk) \to \mathrm{GL}_r(\Bbbk)$ of the natural surjection $\eta : \mathrm{GL}_r(\Bbbk) \to \mathrm{PGL}_r(\Bbbk)$, the function $\gamma_\iota : G \times G \to \Bbbk^\times$ given by

$$\rho_\iota(ab) = \gamma_\iota(a, b)\rho_\iota(a)\rho_\iota(b)$$

determines a 2-cocycle in $\kappa_{\bar\rho}$, where $\rho_\iota = \iota \circ \bar\rho$. The cohomology class $\kappa_{\bar\rho}$ is trivial if and only if $\bar\rho$ can be *lifted* to a linear representation $\rho : G \to \mathrm{GL}_r(\Bbbk)$, i.e., $\eta \circ \rho = \bar\rho$ (see [Karpilovsky 1985, p. 72]).

Let $\pi : L \to G$ be a group homomorphism. For any 2-cocycle $\gamma \in Z^2(G, \Bbbk^\times)$, we have $\gamma \circ (\pi \times \pi) \in Z^2(L, \Bbbk^\times)$. The assignment $\gamma \mapsto \gamma \circ (\pi \times \pi)$ of 2-cocycles induces the group homomorphism $\pi^* : H^2(G, \Bbbk^\times) \to H^2(L, \Bbbk^\times)$, which is called the inflation map along π. In particular, $\pi^*\kappa_{\bar\rho} \in H^2(L, \Bbbk^\times)$ is associated with the projective representation $\bar\rho \circ \pi : L \to \mathrm{PGL}_r(\Bbbk)$.

The homology group $H_2(G, \mathbb{Z})$ is often called the *Schur multiplier* of G [Weibel 1994]. Since $\Bbbk^\times$ is a divisible abelian group, $H^2(G, \Bbbk^\times)$ is naturally isomorphic to $\mathrm{Hom}(H_2(G, \mathbb{Z}), \Bbbk^\times)$ for any group G. This natural isomorphism allows us to summarize the result of Beyl [1986, Theorem 3.9 and Corollary 3.10] on the Schur multiplier of $\mathrm{SL}_2(\mathbb{Z}_m)$ as the following theorem. A proof of the statement is provided for the sake of completeness. The case for odd integers m was originally proved by Mennicke [1967].

Theorem 2.1. *Let $\Bbbk$ be an algebraically closed field of characteristic zero and let m be an integer greater than* 1. *Then $H^2(\mathrm{SL}_2(\mathbb{Z}_m), \Bbbk^\times)$ is isomorphic to $\mathbb{Z}_2$ when $4 \mid m$ and is trivial otherwise. Moreover, the image of the inflation map $\pi^* : H^2(\mathrm{SL}_2(\mathbb{Z}_m), \Bbbk^\times) \to H^2(\mathrm{SL}_2(\mathbb{Z}_{2m}), \Bbbk^\times)$ along the natural reduction map $\pi : \mathrm{SL}_2(\mathbb{Z}_{2m}) \to \mathrm{SL}_2(\mathbb{Z}_m)$ is always trivial.*

Proof. The first statement is a direct consequence of [Beyl 1986, Theorem 3.9]. For the second statement, it suffices to consider the case $m = 2^a q$ with $a \geq 2$ and q odd. Then, by the Chinese Remainder Theorem, there are split surjections $p : \mathrm{SL}_2(\mathbb{Z}_m) \to \mathrm{SL}_2(\mathbb{Z}_{2^a})$ and $p' : \mathrm{SL}_2(\mathbb{Z}_{2m}) \to \mathrm{SL}_2(\mathbb{Z}_{2^{a+1}})$ such that the following diagram of group homomorphisms commutes, where π' is the reduction map:

$$\begin{tikzcd}
\mathrm{SL}_2(\mathbb{Z}_{2m}) \arrow[r, "p'"] \arrow[d, "\pi"'] & \mathrm{SL}_2(\mathbb{Z}_{2^{a+1}}) \arrow[d, "\pi'"] \\
\mathrm{SL}_2(\mathbb{Z}_m) \arrow[r, "p"] & \mathrm{SL}_2(\mathbb{Z}_{2^a})
\end{tikzcd}$$

Applying the functor $H^2(-, \Bbbk^\times)$ to this commutative diagram, we obtain the following commutative diagram of abelian groups:

$$\begin{tikzcd}
H^2(\mathrm{SL}_2(\mathbb{Z}_{2m}), \Bbbk^\times) & H^2(\mathrm{SL}_2(\mathbb{Z}_{2^{a+1}}), \Bbbk^\times) \arrow[l, "(p')^*"'] \\
H^2(\mathrm{SL}_2(\mathbb{Z}_m), \Bbbk^\times) \arrow[u, "\pi^*"] & H^2(\mathrm{SL}_2(\mathbb{Z}_{2^a}), \Bbbk^\times) \arrow[l, "p^*"] \arrow[u, "(\pi')^*"']
\end{tikzcd}$$

Since p and p' are split surjections, both p^* and $(p')^*$ are injective. Hence, by the first statement, they are isomorphisms. By [Beyl 1986, Corollary 3.10], $(\pi')^*$ is trivial, and so is π^*. □

Theorem 2.1 is essential to the proof of the following lifting lemma for projective representations of $\mathrm{SL}_2(\mathbb{Z})$.

Lemma 2.2. *Suppose* $\bar\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{PGL}_r(\Bbbk)$ *is a projective representation for some positive integer r such that* $\ker\bar\rho$ *is a congruence subgroup of level n. Let* $\bar\rho_n : \mathrm{SL}_2(\mathbb{Z}_n) \to \mathrm{PGL}_r(\Bbbk)$ *be the projective representation which satisfies* $\bar\rho = \bar\rho_n \circ \pi_n$, *where* $\pi_n : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}_n)$ *is the reduction modulo n map and κ denotes the associated second cohomology class in* $H^2(\mathrm{SL}_2(\mathbb{Z}_n), \Bbbk^\times)$. *Then*:

(i) *The class κ is trivial if and only if $\bar\rho$ admits a lifting whose kernel is a congruence subgroup of level n.*

(ii) *If κ is not trivial, then $4\,|\,n$ and $\bar\rho$ admits a lifting whose kernel is a congruence subgroup of level $2n$.*

In particular, there exists a lifting ρ of $\bar\rho$ such that $\ker\rho$ *is a congruence subgroup containing* $\Gamma(2n)$.

Proof. (i) If κ is trivial, there exists a linear representation $\rho_n : \mathrm{SL}_2(\mathbb{Z}_n) \to \mathrm{GL}_r(\Bbbk)$ such that $\eta \circ \rho_n = \bar\rho_n$. Then $\rho := \rho_n \circ \pi_n$ is a lifting of $\bar\rho$ since

$$\eta \circ \rho = \eta \circ \rho_n \circ \pi_n = \bar\rho_n \circ \pi_n = \bar\rho.$$

In particular, $\ker\rho$ is a congruence subgroup of level at most n. Obviously, $\ker\rho \le \ker\bar\rho$. Since $\ker\bar\rho$ is of level n, the level of $\ker\rho$ is at least n. Therefore, $\ker\rho$ is of level n.

Conversely, assume $\rho : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_r(\Bbbk)$ is a representation whose kernel is a congruence subgroup of level n and assume $\bar\rho = \eta \circ \rho$. Then ρ factors through $\mathrm{SL}_2(\mathbb{Z}_n)$ and so there exists a linear representation $\rho_n : \mathrm{SL}_2(\mathbb{Z}_n) \to \mathrm{GL}_r(\Bbbk)$ such that $\rho = \rho_n \circ \pi_n$. Since

$$\eta \circ \rho_n \circ \pi_n = \eta \circ \rho = \bar\rho = \bar\rho_n \circ \pi_n,$$

we have $\eta \circ \rho_n = \bar\rho_n$. Therefore, ρ_n is a lifting of $\bar\rho_n$ and hence κ is trivial.

(ii) Now we consider the case when κ is not trivial. By Theorem 2.1, 4 divides n and $\pi^*(\kappa) \in H^2(\mathrm{SL}_2(\mathbb{Z}_{2n}), \Bbbk^\times)$ is trivial, where $\pi : \mathrm{SL}_2(\mathbb{Z}_{2n}) \to \mathrm{SL}_2(\mathbb{Z}_n)$ is the natural surjection (reduction) map. The composition $\bar\rho_n \circ \pi : \mathrm{SL}_2(\mathbb{Z}_{2n}) \to \mathrm{PGL}_r(\Bbbk)$ defines a projective representation of $\mathrm{SL}_2(\mathbb{Z}_{2n})$, and its associated class in $H^2(\mathrm{SL}_2(\mathbb{Z}_{2n}), \Bbbk^\times)$ is $\pi^*(\kappa)$. Since $\pi^*(\kappa)$ is trivial, $\bar\rho_n \circ \pi$ can be lifted to a linear representation $f : \mathrm{SL}_2(\mathbb{Z}_{2n}) \to \mathrm{GL}_r(\Bbbk)$, i.e., $\eta \circ f = \bar\rho_n \circ \pi$. Thus, we have the following commutative diagram:

$$\begin{tikzcd}
\mathrm{SL}_2(\mathbb{Z}_{2n}) \arrow[rrd, bend left, "f"] \arrow[rdd, bend right, "\pi"'] & & \\
& \mathrm{SL}_2(\mathbb{Z}) \arrow[lu, "\pi_{2n}"'] \arrow[d, "\pi_n"'] \arrow[rd, "\bar{\rho}"] & \mathrm{GL}_r(\Bbbk) \arrow[d, "\eta"] \\
& \mathrm{SL}_2(\mathbb{Z}_n) \arrow[r, "\bar{\rho}_n"'] & \mathrm{PGL}_r(\Bbbk)
\end{tikzcd}$$

The commutativity of the upper quadrangle is given by

$$\eta \circ f \circ \pi_{2n} = \bar{\rho}_n \circ \pi \circ \pi_{2n} = \bar{\rho}_n \circ \pi_n = \bar{\rho}.$$

Set $\rho = f \circ \pi_{2n}$. Then $\eta \circ \rho = \bar{\rho}$ and so $\Gamma(2n) \leq \ker \rho$. Suppose $\Gamma(m) \leq \ker \rho$ for some positive integer $m < 2n$ and suppose $m \mid 2n$. Then $\Gamma(m) \leq \ker \rho \leq \ker \bar{\rho}$. Since $\ker \bar{\rho}$ is of level n, we have that $n \mid m$. Thus, $m = n$, and hence $\ker \rho$ is a congruence subgroup of level n. It follows from (i) that κ is trivial, a contradiction. Therefore, $\ker \rho$ is of level $2n$. □

Now we can prove the following lifting theorem for projective representations of $\mathrm{SL}_2(\mathbb{Z})$ with congruence kernels.

Theorem 2.3. *Suppose $\bar{\rho} : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{PGL}_r(\Bbbk)$ is a projective representation for some positive integer r such that $\ker \bar{\rho}$ is a congruence subgroup of level n. Then the kernel of any lifting of $\bar{\rho}$ is a congruence subgroup of level m where $n \mid m \mid 12n$.*

Proof. By Lemma 2.2, $\bar{\rho}$ admits a lifting ξ such that $\ker \xi$ is congruence subgroup containing $\Gamma(2n)$. Let ρ be a lifting of $\bar{\rho}$. By Section 1.2, $\rho = \xi_x \cong \chi_x \otimes \xi$ for some 12-th root of unity $x \in \Bbbk$. Note that $\mathrm{SL}_2(\mathbb{Z})/\,\mathrm{SL}_2(\mathbb{Z})' \cong \mathbb{Z}_{12}$ and $\Gamma(12) \leq \mathrm{SL}_2(\mathbb{Z})'$; see for example [Beyl 1986, Lemma 1.13]. Therefore, $\Gamma(12) \leq \ker \chi_x$ and hence

$$\ker(\chi_x \otimes \xi) \supseteq \mathrm{SL}_2(\mathbb{Z})' \cap \Gamma(2n) \supseteq \Gamma(12) \cap \Gamma(2n) \supseteq \Gamma(12n).$$

Therefore, ρ has a congruence kernel containing $\Gamma(12n)$ and so $m \mid 12n$. Since $\Gamma(m) \leq \ker \rho \leq \ker \bar{\rho}$ and $\ker \bar{\rho}$ is of level n, we have $n \mid m$. □

A consequence of Theorem 2.3 is a proof for the statements (i) and (ii) of Theorem II.

Proof of Theorem II (i) and (ii). By [Ng and Schauenburg 2010, Theorem 6.8], the projective modular representation $\bar{\rho}_{\mathcal{A}}$ of a modular category $\mathcal{A}$ over $\Bbbk$ has a congruence kernel of level N where N is the order of the T-matrix of $\mathcal{A}$. It follows immediately from Theorem 2.3 that every modular representation ρ has a congruence kernel of level n where $N \mid n \mid 12N$. By Lemma A.1, $\mathrm{ord}(\rho(\mathfrak{t})) = n$.

Now the statement Theorem II (ii) follows directly from [Ng and Schauenburg 2010, Theorem 7.1]. $\square$

The congruence property, Theorem II (i) and (ii), is essential to the proof of Theorem I and to the Galois symmetry of modular categories in Sections 4 and 5.

Definition 2.4. Let $\mathcal{A}$ be a modular category over $\Bbbk$ with $\mathrm{FSexp}(\mathcal{A}) = N$.

(i) By virtue of Theorem II (i), a modular representation ρ of $\mathcal{A}$ is said to be of level n if $\mathrm{ord}(\rho(\mathfrak{t})) = n$.

(ii) The projective modular representation $\bar{\rho}_{\mathcal{A}}$ of $\mathcal{A}$ factors through a projective representation $\bar{\rho}_{\mathcal{A},N}$ of $\mathrm{SL}_2(\mathbb{Z}_N)$. We denote by $\kappa_{\mathcal{A}}$ the cohomology class in $H^2(\mathrm{SL}_2(\mathbb{Z}_N), \Bbbk^{\times})$ associated with $\bar{\rho}_{\mathcal{A},N}$.

By Theorem 2.1, the order of $\kappa_{\mathcal{A}}$ is at most 2. If $4 \nmid \mathrm{FSexp}(\mathcal{A})$, then $\kappa_{\mathcal{A}}$ is trivial. However, if $4 \mid \mathrm{FSexp}(\mathcal{A})$, Lemma 2.2 provides the following criterion to decide the order of $\kappa_{\mathcal{A}}$.

Corollary 2.5. *Let $\mathcal{A}$ be a modular category over $\Bbbk$. Suppose $N = \mathrm{FSexp}(\mathcal{A})$ and suppose $\zeta \in \Bbbk$ is a 6-th root of the anomaly of $\mathcal{A}$. Then $\kappa_{\mathcal{A}}$ is trivial if and only if $(x/\zeta)^N = 1$ for some 12-th root of unity $x \in \Bbbk$. In this case, $x^3 p_{\mathcal{A}}^+/\zeta^3 \in \mathbb{Q}_N$. In particular, if $4 \nmid N$, then there exists a 12-th root of unity $x \in \Bbbk$ such that*

$$(x/\zeta)^N = 1 \quad \textit{and} \quad x^3 p_{\mathcal{A}}^+/\zeta^3 \in \mathbb{Q}_N.$$

Proof. By (1-7), ζ determines the modular representation ρ^{ζ} of $\mathcal{A}$ given by

$$\rho^{\zeta} : \mathfrak{s} \mapsto \frac{\zeta^3}{p_{\mathcal{A}}^+}\tilde{s}, \qquad \mathfrak{t} \mapsto \frac{1}{\zeta}\tilde{t}.$$

By Lemma 2.2 (i) and the last two paragraphs of Section 1.2, $\kappa_{\mathcal{A}}$ is trivial if and only if there exists a 12-th root of unity $x \in \Bbbk$ such that ρ_x^{ζ} is a level N modular representation of $\mathcal{A}$. By Theorem II (i), this is equivalent to $\mathrm{id} = (x\tilde{t}/\zeta)^N$ or $(x/\zeta)^N = 1$. In this case, Theorem II (ii) implies $\zeta^3/(x^3 p_{\mathcal{A}}^+)\tilde{s} \in \mathrm{GL}_{\Pi_{\mathcal{A}}}(\mathbb{Q}_N)$ and hence $\zeta^3/(x^3 p_{\mathcal{A}}^+) \in \mathbb{Q}_N$. The last statement follows immediately from Theorem 2.1. $\square$

The corollary implies some arithmetic relations among the Frobenius–Schur exponent, the global dimension and the anomaly of a modular category. These arithmetic consequences will be discussed in Section 6.

3. Modularity of trace functions for rational vertex operator algebras

In this section we prove that the trace functions of a rational, C_2-cofinite vertex operator algebra V are modular forms on some congruence subgroup by showing that the representation ρ_V of $\mathrm{SL}_2(\mathbb{Z})$, defined by modular transformation of the trace functions of V, is a modular representation of $\mathcal{C}_V$. The congruence subgroup property obtained in Section 2 is then applied to ρ_V to conclude the modularity of the trace functions of V.

Preliminaries. In this subsection we briefly review some basics of vertex operator algebras following [Frenkel et al. 1988; Frenkel et al. 1993; Dong et al. 1997; 1998a; Lepowsky and Li 2004; Zhu 1996].

Let $V = (V, Y, \mathbb{1}, \omega)$ be a vertex operator algebra. Then V is C_2-cofinite if the subspace $C_2(V)$ of V spanned by all elements of type $a_{-2}b$ for $a, b \in V$ has finite codimension in V. Recall from [Dong et al. 1998a] that V is rational if any admissible module is completely reducible. The component operator $L(n)$ of $Y(\omega, z) = \sum_{n\in\mathbb{Z}} L(n)z^{-n-2}$ will be used frequently. It is proved in [Dong et al. 1998a] that if V is rational then V has only finitely many irreducible admissible modules $M^0, \dots, M^p$ up to isomorphism and there exist $\lambda_i \in \mathbb{C}$ for $i = 0, \dots, p$ such that

$$M^i = \bigoplus_{n=0}^{\infty} M^i_{\lambda_i+n}$$

where $M^i_{\lambda_i} \neq 0$ and $L(0)|_{M^i_{\lambda_i+n}} = \lambda_i + n$ for any $n \in \mathbb{Z}$. Moreover, if V is also assumed to be C_2-cofinite, then λ_i and the central charge c of V are rational numbers (see [Dong et al. 2000]). In this paper we always assume that V is simple and we take M^0 to be V.

Another important concept is the contragredient module. Let $M = \bigoplus_{\lambda\in\mathbb{C}} M_\lambda$ be a V-module. Let $M' = \bigoplus_{\lambda\in\mathbb{C}} M^*_\lambda$ be the restricted dual of M. It is proved in [Frenkel et al. 1993] that $M' = (M', Y')$ is naturally a V-module such that

$$\langle Y'(a, z)u', v\rangle = \langle u', Y(e^{zL(1)}(-z^{-2})^{L(0)}a, z^{-1})v\rangle,$$

for $a \in V$, $u' \in M'$ and $v \in M$, and that $(M')' \simeq M$. Moreover, if M is irreducible, so is M'. A V-module M is said to be self-dual if M and M' are isomorphic. In this paper, we'll always assume that the vertex operator algebra V satisfies the following assumptions:

(V1) $V = \bigoplus_{n\geq 0} V_n$ with $\dim V_0 = 1$ is simple and self-dual.

(V2) V is C_2-cofinite and rational.

The assumption (V2) is equivalent to the regularity [Dong et al. 1997]. That is, any weak module is completely reducible.

We now recall the notion of intertwining operators and fusion rules from [Frenkel et al. 1993]. Let $W^i = (W^i, Y_{W^i})$ for $i = 1, 2, 3$ be weak V-modules. Then an intertwining operator $\mathcal{Y}(\,\cdot\,, z)$ of type $\binom{W^3}{W^1\,W^2}$ is a linear map

$$\mathcal{Y}(\,\cdot\,, z) : W^1 \to \operatorname{Hom}(W^2, W^3)\{z\}, \quad v^1 \mapsto \mathcal{Y}(v^1, z) = \sum_{n\in\mathbb{C}} v^1_n z^{-n-1}$$

satisfying the following conditions:

(i) For any $v^1 \in W^1$, $v^2 \in W^2$ and $\lambda \in \mathbb{C}$, we have $v^1_{n+\lambda}v^2 = 0$ for $n \in \mathbb{Z}$ sufficiently large.

(ii) For any $a \in V$, $v^1 \in W^1$, we have

$$z_0^{-1}\delta\left(\frac{z_1 - z_2}{z_0}\right)Y_{W^3}(a, z_1)\mathcal{Y}(v^1, z_2) - z_0^{-1}\delta\left(\frac{z_1 - z_2}{-z_0}\right)\mathcal{Y}(v^1, z_2)Y_{W^2}(a, z_1) = z_2^{-1}\delta\left(\frac{z_1 - z_0}{z_2}\right)\mathcal{Y}(Y_{W^1}(a, z_0)v^1, z_2).$$

(iii) For $v^1 \in W^1$, we have $\frac{d}{dz}\mathcal{Y}(v^1, z) = \mathcal{Y}(L(-1)v^1, z)$.

The sum in the definition of intertwining operator in [Frenkel et al. 1993] is over rational numbers. For a rational vertex operator algebra, this is true. In general, the sum should be over complex numbers. All of the intertwining operators of type $\binom{W^3}{W^1\,W^2}$ form a vector space denoted by $I_V\binom{W^3}{W^1\,W^2}$. The dimension of $I_V\binom{W^3}{W^1\,W^2}$ is called the fusion rule of type $\binom{W^3}{W^1\,W^2}$ for V, which is denoted by $N^{W^3}_{W^1,W^2}$.

The following properties of the fusion rule are well-known (see [Frenkel et al. 1993]).

Proposition 3.1. *Let V be a vertex operator algebra, and let M^i, M^j, M^k be three irreducible V-modules. Then*:

(i) $N^i_{j,k} = N^{k^*}_{j,i^*}$, *where we use W^{i^*} to denote $(W^i)'$ and where $N^i_{j,k} = N^{M^i}_{M^j,M^k}$.*

(ii) $N^i_{j,k} = N^i_{k,j}$.

Let M^1 and M^2 be two V-modules. A tensor product for the ordered pair (M^1, M^2) is a pair $(M, F(\cdot, z))$, which consists of a V-module M and an intertwining operator $F(\cdot, z)$ of type $\binom{M}{M^1\,M^2}$, such that the following universal property holds: for any V-module X and any intertwining operator $I(\cdot, z)$ of type $\binom{X}{M^1\,M^2}$, there exists a unique V-homomorphism ϕ from M to X such that $I(\cdot, z) = \phi \circ F(\cdot, z)$. Note that if there is a tensor product, then it is unique by the universal mapping property. In this case we will denote it by $M^1 \boxtimes M^2$.

In a series of papers [Huang and Lepowsky 1995a; 1995b; 1995c; Huang 1995; 2008a; 2008b], the tensor product $\boxtimes$ of the modules for a vertex operator algebra V has been defined and studied extensively. We have the following result (see [Abe et al. 2004, Corollary 10] and [Huang and Lepowsky 1995a, Proposition 4.13]).

Theorem 3.2. *Let V be a rational and C_2-cofinite vertex operator algebra, and let M^i, M^j, M^k be any three irreducible modules of V. Then*:

(i) *The fusion rules $N^k_{i,j}$ are finite.*

(ii) *The tensor product $M^i \boxtimes M^j$ of M^i and M^j exists and is equal to $\sum_k N^k_{i,j}M^k$.*

We finally review some facts about the modular transformation of trace functions of irreducible modules of a vertex operator algebra from [Zhu 1996]. Let V be a rational and C_2-cofinite vertex operator algebra, and let $M^0, \dots, M^p$ be the irreducible V-modules as before. There is another VOA structure on V, given by $(V, Y[\,\cdot\,, z], \mathbb{1}, \omega - c/24)$ and introduced in [Zhu 1996]. In particular,

$$V = \bigoplus_{n\geq 0} V_{[n]}.$$

We will write $wt[v] = n$ if $v \in V_{[n]}$. For each $v \in V_n$, we denote v_{n-1} by $o(v)$ and extend to V linearly. Recall that $M^i = \bigoplus_{n=0}^{\infty} M^i_{\lambda_i+n}$. For $v \in V$ we set

$$Z_i(v, q) = \operatorname{tr}_{M^i} o(v) q^{L(0)-c/24} = \sum_{n\geq 0} (\operatorname{tr}_{M^i_{\lambda_i+n}} o(v)) q^{\lambda_i+n-c/24},$$

which is a formal power series in variable q. The constant c here is the central charge of V, and $Z_i(\mathbb{1}, q)$ is sometimes called the q-character of M^i. Then $Z_i(v, q)$ converges to a holomorphic function in $0 < |q| < 1$ [Zhu 1996]. As usual we let $\mathfrak{h} = \{\tau \in \mathbb{C} \mid \operatorname{im} \tau > 0\}$ and $q = e^{2\pi i \tau}$ with $\tau \in \mathfrak{h}$. We also denote the holomorphic function $Z_i(v, q)$ by $Z_i(v, \tau)$ when we discuss modular transformations of these functions.

The full modular group $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathfrak{h}$ by

$$\gamma : \tau \mapsto \frac{a\tau + b}{c\tau + d}, \quad \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

The following theorem was established in [Zhu 1996].

Theorem 3.3. *Let V be a rational and C_2-cofinite vertex operator algebra, and let $M^0, \dots, M^p$ be the irreducible V-modules. Then for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ there exists a $\rho_V(\gamma) = [\gamma_{ij}]_{i,j=0,\dots,p} \in \mathrm{GL}_{p+1}(\mathbb{C})$ such that, for any $0 \leq i \leq p$ and $v \in V_{[n]}$,*

$$Z_i(v, \gamma\tau) = (c\tau + d)^n \sum_{j=0}^{p} \gamma_{ij} Z_j(v, \tau).$$

Theorem 3.3, in fact, gives a group homomorphism $\rho_V : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_{p+1}(\mathbb{C})$. We call $\rho_V(\gamma)$ the genus one modular matrices. In particular,

$$S = \rho_V\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) \quad \text{and} \quad T = \rho_V\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right).$$

are respectively called the *genus one* S- and T-matrices of V. It is immediate to see that $T_{jk} = \delta_{jk} e^{2\pi i (\lambda_j - c/24)}$.

One of our main goals is to show that the kernel of ρ_V is a congruence subgroup.

We need the following results on the Verlinde formula [1988] from [Huang 2008a; 2008b] (also see [Moore and Seiberg 1990]).

Theorem 3.4. *Let V be a vertex operator algebra satisfying (V1) and (V2). Then the genus one S-matrix of V defined above has the following properties*:

(i) *S is symmetric and $S^2 = C$, where $C_{ij} = \delta_{ij^*}$. In particular, C has order at most 2 and is also symmetric.*

(ii) $S_{ij}^{-1} = S_{i^*j} = S_{ij^*}$.

(iii) (*Verlinde formula*) *For any $i, j, k \in \{0, \dots, p\}$,*

$$N_{i,j}^k = \sum_{q=0}^{p} \frac{S_{iq}S_{jq}S_{k^*q}}{S_{0q}}.$$

Unitarity of S. In this subsection, we will prove that the genus one S-matrix of V defined on page 2137 is unitary and consequences of this fact. Our approach is slightly different from that given in [Etingof et al. 2005] for the unitarity of a normalized S-matrix of a modular category. Recall that $S^2 = C$. In fact, this equality holds for any symmetric matrix satisfying the Verlinde formula as follows:

Lemma 3.5. *Let $\mathcal{C}$ be a fusion category over $\mathbb{C}$ with commutative Grothendieck ring. Suppose A is a complex symmetric matrix indexed by $\Pi_{\mathcal{C}}$ such that $A_{0r} \neq 0$ for all $r \in \Pi_{\mathcal{C}}$ and suppose A satisfies the Verlinde formula in the sense that*

$$N_{ij}^k = \sum_{r\in\Pi_{\mathcal{C}}} \frac{A_{ir}A_{jr}A_{k^*r}}{A_{0r}} \tag{3-1}$$

for all $i, j, k \in \Pi_{\mathcal{C}}$, where N_{ij}^k is the fusion rule of $\mathcal{C}$. Then we have $A_{0r} \in \mathbb{R}$ and $A^2 = C$, and we have that A is unitary, where $C_{ij} = \delta_{ij^*}$ for $i, j \in \Pi_{\mathcal{C}}$.

Proof. By the Verlinde formula (3-1), $\sum_{r\in\Pi_{\mathcal{C}}} A_{ir}A_{jr} = N_{ij}^0 = \delta_{ij^*}$ for any $i, j \in \Pi_{\mathcal{C}}$. This implies A is invertible and $(A^{-1})_{ij} = A_{ij^*} = A_{i^*j}$ for $i, j \in \Pi_{\mathcal{C}}$. Hence, we have $A_{i^*j^*} = A_{ij}$ and $A_{0j} = A_{0j^*}$ for all $i, j \in \Pi_{\mathcal{C}}$. Let $\mathcal{K}_0(\mathcal{C})$ be the Grothendieck ring of $\mathcal{C}$ and let $\mathcal{K}_{\mathbb{C}}(\mathcal{C}) = \mathcal{K}_0(\mathcal{C}) \otimes_{\mathbb{Z}} \mathbb{C}$. Note that $\mathcal{K}_{\mathbb{C}}(\mathcal{C})$ is commutative $\mathbb{C}$-algebra. For $b \in \Pi_{\mathcal{C}}$, let

$$e_b = A_{0b} \sum_{a\in\Pi_{\mathcal{C}}} A_{ab}\, a \in \mathcal{K}_{\mathbb{C}}(\mathcal{C}),$$

and $E = \{e_b \mid b \in \Pi_{\mathcal{C}}\}$. Then

$$\begin{aligned} e_a e_b &= A_{0a}A_{0b} \sum_{c,d} A_{ac}A_{bd}cd = A_{0a}A_{0b} \sum_{c,d,r} A_{ac}A_{bd}N_{cd}^r r \\ &= A_{0a}A_{0b} \sum_{c,d,r,z} A_{ac}A_{bd} \frac{A_{cz}A_{dz}A_{r^*z}}{A_{0z}} r = A_{0a}A_{0b} \sum_{r,z} \frac{\delta_{az^*}\delta_{bz^*}A_{r^*z}}{A_{0z}} r \\ &= \delta_{ab}A_{0a}^2 \sum_r \frac{A_{r^*a^*}}{A_{0a^*}} r = \delta_{ab}A_{0a} \sum_r A_{ra} r = \delta_{ab} e_a. \end{aligned}$$

Hence, E is the set of all primitive idempotents of $\mathcal{K}_{\mathbb{C}}(\mathcal{C})$.

The duality permutation defined on $\Pi_{\mathcal{C}}$ can be extended to a sesquilinear linear map $\dagger$ on $\mathcal{K}_{\mathbb{C}}(\mathcal{C})$, i.e.,

$$\left(\sum_{x\in\Pi_{\mathcal{C}}} \alpha_x x\right)^{\dagger} = \sum_{x\in\Pi_{\mathcal{C}}} \bar{\alpha}_x x^*$$

for $\alpha_x \in \mathbb{C}$. Moreover, $\dagger$ is an $\mathbb{R}$-algebra automorphism of $\mathcal{K}_{\mathbb{C}}(\mathcal{C})$, but $\dagger$ is not $\mathbb{C}$-linear. In particular, $e_b^{\dagger}$ is in E and hence $\dagger$ defines a permutation on E.

For $x \in \mathcal{K}_{\mathbb{C}}(\mathcal{C})$, denote by $\epsilon(x)$ the coefficient of the unit object 0 in x. Then

$$\epsilon(ab) = N_{ab}^0 = \delta_{ab^*} \quad \text{for } a, b \in \Pi_{\mathcal{C}}.$$

We now define the sesquilinear form $(\cdot\,,\cdot)$ on $\mathcal{K}_{\mathbb{C}}(\mathcal{C})$ by

$$(x, y) = \epsilon(xy^{\dagger}).$$

Note that $(x, x) > 0$ for $x \neq 0$. Thus

$$0 < (e_b, e_b) = \epsilon(e_b e_b^{\dagger}).$$

Therefore, $e_b^{\dagger} = e_b$ and so $(e_b, e_b) = A_{0b}^2 > 0$ and $A_{0b}A_{ab} = \overline{A_{0b}A_{a^*b}}$ for all $a, b \in \Pi_{\mathcal{C}}$. The former implies $A_{0b} \in \mathbb{R}$ and hence $A_{ab} = \overline{A_{a^*b}}$ for all $a, b \in \Pi_{\mathcal{C}}$. Therefore, A is unitary. □

The following corollary is an immediate consequence of Lemma 3.5 and the modularity of $\mathcal{C}_V$ presented in Theorem 3.9.

Corollary 3.6. *Let V be a vertex operator algebra satisfying (V1) and (V2). Then the genus one S-matrix of V defined on page 2137 is unitary and satisfies $\bar{S} = SC$.*

The following result can be proved easily by using Corollary 3.6.

Corollary 3.7. *Let V be a vertex operator algebra satisfying (V1) and (V2). For any $u \in V_{[m]},\ v \in V_{[n]},\ \gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z})$ and $\tau_1, \tau_2 \in \mathfrak{h}$ we have*

$$\sum_i Z_i(u, \gamma\tau_1)\overline{Z_i(v, \gamma\tau_2)} = (c\tau_1 + d)^m\,\overline{(c\tau_2 + d)^n} \sum_i Z_i(u, \tau_1)\overline{Z_i(v, \tau_2)}.$$

In particular, $\sum_{0\le i\le p} |\chi_i(\tau)|^2$ is invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$.

Proof. Note that T is a diagonal matrix with diagonal entries $e^{2\pi i(\lambda_j - c/24)}$ for $j = 0, \ldots, p$ which is clearly a unitary matrix, as λ_j and c are rational numbers. It follows from Corollary 3.6 that the representation ρ is unitary. Set

$$f(\tau_1, \tau_2) = \sum_i Z_i(u, \tau_1)\overline{Z_i(v, \tau_2)}.$$

Then

$$\begin{aligned} f(\gamma\tau_1,\gamma\tau_2) &= \sum_i Z_i(u,\gamma\tau_1)\overline{Z_i(v,\gamma\tau_2)} \\ &= (c\tau_1+d)^m\overline{(c\tau_2+d)^n}\sum_{i,j,k}\gamma_{ij}Z_j(u,\tau_1)\overline{\gamma_{ik}}\,\overline{Z_k(v,\tau_2)} \\ &= (c\tau_1+d)^m\overline{(c\tau_2+d)^n}\sum_i Z_i(u,\tau_1)\overline{Z_i(v,\tau_2)}. \end{aligned}$$

□

Here we use Corollary 3.7 to study the extensions of vertex operator algebras. As before we assume that V is a vertex operator algebra satisfying (V1) and (V2). We also assume that U is an extension of V satisfying (V1) and (V2). Then $U=\sum_i n_i M^i$ as a V-module, where n_i is nonnegative and $n_0=1$, as the vacuum vector is unique. The main goal is to determine the possible values of n_i. There have been a lot of discussions on this in the literature using the modular invariance of the characters (see, for example, [Cappelli et al. 1987a; 1987b; Gannon 2005]). It seems that using the characters of irreducible modules is not good enough, as the characters of irreducible modules are not linearly independent in general. In this section we use the conformal blocks instead of the characters to approach the problem.

For $u, v\in V$, we set

$$f_V(u,v,\tau_1,\tau_2)=\sum_{i=0}^{p} Z_i(u,\tau_1)\overline{Z_i(v,\tau_2)}$$

(see Corollary 3.7). Similarly we can define

$$f_U(u,v,\tau_1,\tau_2)=\sum_{M} Z_M(u,\tau_1)\overline{Z_M(v,\tau_2)}$$

for $u, v\in U$ where M ranges through the equivalent classes of irreducible U-modules. Since each irreducible U-module M is a direct sum of irreducible V-modules, we see that, for $u, v\in V$,

$$f_U(u,v,\tau_1,\tau_2)=\sum_{i,j=0}^{p} X_{ij}Z_i(u,\tau_1)\overline{Z_j(v,\tau_2)}$$

for some $X_{ij}\in\mathbb{Z}_+$ and all i, j. If $u=v=\mathbb{1}$ and $\tau_1=\tau_2=\tau$, then $f_U(\mathbb{1},\mathbb{1},\tau,\tau)$, which is the sum of square norms of the irreducible characters of U, is $\mathrm{SL}_2(\mathbb{Z})$-invariant. We now determine the matrix $X=[X_{ij}]$. It will be clear from our proof below that the $\mathrm{SL}_2(\mathbb{Z})$-invariance of $f_U(\mathbb{1},\mathbb{1},\tau,\tau)$ is not good enough to determine the matrix X.

Proposition 3.8. *The matrix X satisfies $X_{00}=1$ and $X\gamma=\gamma X$, where $\gamma\in SL_2(\mathbb{Z})$, and is identified with the modular transformation matrix $\rho_V(\gamma)$.*

Proof. For any $u \in V_{[m]}$, let

$$\boldsymbol{Z}(u,\tau)=\begin{bmatrix} Z_0(u,\tau)\\ \vdots \\ Z_p(u,\tau)\end{bmatrix}.$$

Then

$$\boldsymbol{Z}(u,\gamma\tau)=(c\tau+d)^m\gamma\boldsymbol{Z}(u,\tau)\quad\text{and}\quad f_U(u,v,\tau_1,\tau_2)=\boldsymbol{Z}(u,\tau_1)^T X\overline{\boldsymbol{Z}(v,\tau_2)}.$$

By Corollary 3.7,

$$\begin{aligned}(c\tau_1+d)^m\overline{(c\tau_2+d)^n}\boldsymbol{Z}(u,\tau_1)^T X\overline{\boldsymbol{Z}(v,\tau_2)}&\\ &= f_U(u,v,\gamma\tau_1,\gamma\tau_2)=\boldsymbol{Z}(u,\gamma\tau_1)^T X\overline{\boldsymbol{Z}(v,\gamma\tau_2)}\\ &=(c\tau_1+d)^m\overline{(c\tau_2+d)^n}\boldsymbol{Z}(u,\tau_1)^T\gamma^T X\bar{\gamma}\overline{\boldsymbol{Z}(v,\tau_2)}.\end{aligned}$$

This implies that

$$\boldsymbol{Z}(u,\tau_1)^T X\overline{\boldsymbol{Z}(v,\tau_2)}=\boldsymbol{Z}(u,\tau_1)^T\gamma^T X\bar{\gamma}\overline{\boldsymbol{Z}(v,\tau_2)}$$

for all u, v. Since γ is unitary, it is enough to show that if $\boldsymbol{Z}(u,\tau_1)^T A\overline{\boldsymbol{Z}(v,\tau_2)}=0$ for all $u, v \in V$ where $A=[a_{ij}]$ is a fixed matrix, then $A=0$.

Next note the equality $\boldsymbol{Z}(u,\tau_1)^T A\overline{\boldsymbol{Z}(v,\tau_2)}=\sum_{ij} a_{ij} Z_i(u,\tau_1)\overline{Z_j(v,\tau_2)}$. For simplicity, set $q_j=e^{2\pi i\tau_j}$ for $j=1,2$. Then

$$\begin{aligned}0&=\boldsymbol{Z}(u,\tau_1)^T A\overline{\boldsymbol{Z}(v,\tau_2)}\\ &=\sum_{i,j}\sum_{m_i,n_j\geq 0} a_{ij}(\mathrm{tr}_{M^i_{\lambda_i+m_i}}o(u)\overline{\mathrm{tr}_{M^j_{\lambda_j+n_j}}o(v)})q_1^{\lambda_i+m_i-c/24}\overline{q_2^{\lambda_j+n_j-c/24}}.\end{aligned}$$

This implies that each coefficient of $q_1^m\overline{q_2^n}$ for any rational numbers m, n must be zero. We now prove that $a_{ij}=0$ for all i, j. Fix i and j. Then the coefficient of $q_1^{\lambda_i-c/24}\bar{q}_2^{\lambda_j-c/24}$ in $\boldsymbol{Z}(u,\tau_1)^T A\overline{\boldsymbol{Z}(v,\tau_2)}$ is

$$\sum_{k,l} a_{kl}\,\mathrm{tr}_{M^k_{\lambda_k+m_k}}o(u)\overline{\mathrm{tr}_{M^l_{\lambda_l+n_l}}o(v)}$$

where $k, l \in \{0,\ldots,p\}$ satisfy $m_k+\lambda_k=\lambda_i$, $n_l+\lambda_l=\lambda_j$. Fix $n\geq 0$ such that $n\geq m_k$ and $n\geq n_l$ for all k, l occurring in the summation above. Recall from [Dong et al. 1998b] that there is a finite dimensional semisimple associative algebra $A_n(V)$ such that $M^k_{m_k+\lambda_k}$, $M^l_{n_l+\lambda_l}$ are the inequivalent simple modules of $A_n(V)$. As a result we can choose $u, v \in V$ such that $o(u)=1$ on $M^i_{\lambda_i}$ and $o(u)=0$ on all other $M^k_{\lambda_k+m_k}$, and such that $o(v)=1$ on $M^j_{\lambda_j}$ and $o(v)=0$ on all other $M^l_{\lambda_l+n_l}$. Therefore, for this u and v, we have that the coefficient of $q_1^{\lambda_i-c/24}\bar{q}_2^{\lambda_j-c/24}$ in $\boldsymbol{Z}(u,\tau_1)^T A\overline{\boldsymbol{Z}(v,\tau_2)}$ is a nonzero multiple of a_{ij}. This forces $a_{ij}=0$, completing the proof. □

The congruence property theorem. Now we come back to the theories of vertex operator algebras. Let V be a rational and C_2-cofinite vertex operator algebra. For any V-module M, set $\theta_M = e^{2\pi i L(0)}$. The following result from [Huang 2008a, Theorem 4.1] is important in this paper.

Theorem 3.9. *Let V be a vertex operator algebra satisfying (V1) and (V2). Then the V-module category $\mathcal{C}_V$ with the dual M' (M a V-module), braiding σ which is denoted by $\mathcal{C}$ in* [Huang 2008a, p. 877] *and twist θ is a modular tensor category over $\mathbb{C}$.*

Note that $\mathrm{End}_V(M^i) = \mathbb{C}$, $0 \le i \le p$. Recall from discussions in Sections 1.1 and 1.3 that the pivotal dimension d_i of the simple V-module is a nonzero real number and the global dimension $\dim \mathcal{C}_V = \sum_{i=0}^{p} d_i^2$ is at least 1. Let $\tilde{s}$ and $\tilde{t}$ be the S- and T-matrices of $\mathcal{C}_V$, and $D = \sqrt{\dim \mathcal{C}_V}$ the positive square root of $\dim \mathcal{C}_V$, and $\boldsymbol{c}$ the central charge of $\mathcal{C}_V$. We fix the normalization $s = \tilde{s}/D$, and simply call s the *normalized S-matrix of $\mathcal{C}_V$*. We will prove in Theorem 3.10 that s is identical to the genus one S-matrix of V up to a sign.

Theorem 3.10. *Let V be a vertex operator algebra satisfying (V1) and (V2). Then*:

(i) *The normalized S-matrix s of $\mathcal{C}_V$ and the genus one S-matrix of V are identical up to a sign.*

(ii) *The representation ρ_V defined by modular transformation of trace functions is a modular representation of $\mathcal{C}_V$. In particular*, $\ker \rho_V$ *is a congruence subgroup of level n where n is the order of the genus one T-matrix of V, and ρ_V is $\mathbb{Q}_n$-rational.*

(iii) *The central charge $\boldsymbol{c}$ of $\mathcal{C}_V$ is equal to the central charge c of V modulo* 4.

Proof. Let

$$\sigma_{M^i M^j} : M^i \boxtimes M^j \to M^j \boxtimes M^i$$

be the braiding of $\mathcal{C}_V$. It is proved in [Huang 2008a] that the pivotal trace of $\sigma_{M^{i*} M^j} \sigma_{M^j M^{i*}}$ on $M^j \boxtimes M^{i^*}$ equals S_{ij}/S_{00}. This implies that $S = \lambda s$ where $\lambda = S_{00}/s_{00}$. Using the unitarity of s and S, we conclude that λ is a complex number of norm 1. This forces $\lambda = \pm 1$, which proves the first statement.

It follows from Theorem 3.9 that the T-matrix of $\mathcal{C}_V$ is given by $\tilde{t} = [\delta_{ij}\theta_i]_{i,j=0,\dots,p}$ and $\theta_j = e^{2\pi i \lambda_j}$. Therefore, that genus one T-matrix of V is given by $T = \tilde{t}\, e^{-2\pi i c/24}$, where c is the central charge of V. In particular, ρ_V is a modular representation of $\mathcal{C}_V$. The second part of the second statement is an immediate consequence of Theorem II (i) and (ii).

By (i), (1-3) and Theorem 3.4 we see that

$$C = (ST)^3 = \pm (s\tilde{t} e^{-2\pi i c/24})^3 = \pm \frac{p^+}{D} e^{-6\pi i c/24} C,$$

where p^+ is the Gauss sum of $\mathcal{C}_V$. This implies that $\pm 1 = (p^+/D)e^{-\pi i c/4}$ or $p^+/D = \pm e^{\pi i c/4}$. In particular, $\boldsymbol{c} = c \bmod 4$. $\square$

Theorem I now follows from Theorem 3.10 immediately.

We next discuss two different definitions of dimension of modules of rational and C_2-cofinite vertex operator algebras given in [Dong et al. 2013; Bakalov and Kirillov 2001]. As before we assume that V is a vertex operator algebra satisfying (V1) and (V2). Recall the following definition of quantum dimension from [Dong et al. 2013]. Let M be a V-module. Set $Z_M(\tau) = \operatorname{ch}_q M = Z_M(\mathbb{1}, \tau)$. The quantum dimension of M over V is defined as

$$\operatorname{qdim}_V M = \lim_{y \to 0} \frac{Z_M(iy)}{Z_V(iy)}$$

where y is real and positive. It is shown in [Dong et al. 2013] that if V is a vertex operator algebra satisfying (V1) and (V2) with the irreducibles M^i for $i = 0, \ldots, p$ such that $\lambda_i > 0$ for $i \neq 0$, then

$$\operatorname{qdim}_V M^i = \frac{S_{i0}}{S_{00}}. \tag{3-2}$$

On the other hand, because V is a vertex operator algebra satisfying (V1) and (V2), the tensor category $\mathcal{C}_V$ of V-modules is modular by Theorem 3.9. The pivotal dimension $d_i = \dim M^i$ of M^i is also defined in the modular tensor category $\mathcal{C}_V$. We now prove that these two dimensions coincide.

Proposition 3.11. *Let V be a vertex operator algebra satisfying (V1) and (V2), and suppose $\lambda_i > 0$ for $i \neq 0$. Then for any irreducible V-module M^i, we have* $\dim M^i = \operatorname{qdim}_V M^i$.

Proof. Since $\dim M^i = d_i = s_{0i}/s_{00}$, the result follows from Theorem 3.10 and (3-2) immediately. $\square$

The modular transformation property on the conformal blocks has been used extensively in the study of rational vertex operator algebras. The modular transformation property gives an estimation of the growth conditions on the dimensions of homogeneous subspaces as the q-character of an irreducible module is a component of a vector-valued modular function [Knopp and Mason 2003]. The growth condition helps us to show that a rational and C_2-cofinite vertex operator algebra with central charge less than one is an extension of the Virasoro vertex operator algebra associated to the discrete series [Dong and Zhang 2008], and to characterize vertex operator algebra $L(1/2, 0) \otimes L(1/2, 0)$ [Zhang and Dong 2009; Dong and Jiang 2010]. The congruence subgroup property of the action of the modular group on the conformal block is expected to play an important role in the classification of rational vertex operator algebras. Since the q-character of an irreducible module is a modular function on a congruence subgroup and the sum of the square norms of the q-characters of the irreducible modules is invariant under $\mathrm{SL}_2(\mathbb{Z})$, this gives a lot of information on the dimensions of homogeneous subspaces of vertex operator algebras. For example, one can use these properties to determine the possible characters of the rational vertex operator algebras of central charge 1 [Kiritsis 1989]. This will avoid some difficult work in [Dong and Jiang 2011; 2013] of determining the dimensions of homogenous subspaces of small weights when characterizing certain classes of rational vertex operator algebras of central charge one.

4. Galois symmetry of modular representations

It was conjectured by Coste and Gannon that the representation of $\mathrm{SL}_2(\mathbb{Z})$ associated with a RCFT admits a Galois symmetry (see [Coste and Gannon 1999, Conjecture 3; Gannon 2006, Conjecture 6.1.7]). Under certain assumptions, the Galois symmetry of these representations of $\mathrm{SL}_2(\mathbb{Z})$ was established by Coste and Gannon [1999] and by Bantay [2003].

In this section, we will prove that such Galois symmetry holds for all modular representations of a modular category as stated in Theorem II (iii) and (iv). It will follow from Theorem 3.10 that this Galois symmetry holds for the representation ρ_V defined by modular transformation of the trace functions of any VOA V satisfying (V1) and (V2).

The Galois symmetry for the canonical modular representation of the Drinfeld center of a spherical fusion category (Lemma 4.2) plays a crucial for the general case, and we will provide its proof in the next section.

Galois action on a normalized S-matrix. Let $\mathcal{A}$ be a modular category over $\Bbbk$ with Frobenius–Schur exponent N, and let ρ be a level n modular representation of $\mathcal{A}$. By virtue of Theorem II (i) and (ii), $N \mid n \mid 12N$ and $\rho(\mathrm{SL}_2(\mathbb{Z})) \le \mathrm{GL}_\Pi(\mathbb{Q}_n)$, where $\Pi_\mathcal{A}$ is simply abbreviated as Π.

A fixed 6-th root ζ of the anomaly of $\mathcal{A}$ determines the modular representation ρ^ζ of $\mathcal{A}$ (see (1-7)). It follows from Section 1.2 that $\rho = \rho_x^\zeta$ for some 12-th root of unity $x \in \Bbbk$. Let

$$s = \rho(\mathfrak{s}) \quad \text{and} \quad t = \rho(\mathfrak{t}).$$

Then

$$s = \frac{\zeta^3}{x^3 p_\mathcal{A}^+}\tilde{s},\ t = \frac{x}{\zeta}\tilde{t} \in \mathrm{GL}_\Pi(\mathbb{Q}_n). \tag{4-1}$$

Thus $s^2 = x^6 C = \pm C$, where C is the charge conjugation matrix $[\delta_{ij^*}]_{i,j\in\Pi}$. Set $\mathrm{sgn}(s) = x^6$.

Following [de Boer and Goeree 1991, Appendix B], [Coste and Gannon 1994] or [Etingof et al. 2005, Appendix], for each $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$, there exists a unique permutation, denoted by $\hat{\sigma}$, on Π such that

$$\sigma\left(\frac{s_{ij}}{s_{0j}}\right) = \frac{s_{i\hat{\sigma}(j)}}{s_{0\hat{\sigma}(j)}} \quad \text{for all } i, j \in \Pi. \tag{4-2}$$

Moreover, there exists a function $\epsilon_\sigma : \Pi \to \{\pm 1\}$ such that

$$\sigma(s_{ij}) = \epsilon_\sigma(i) s_{\hat{\sigma}(i)j} = \epsilon_\sigma(j) s_{i\hat{\sigma}(j)} \quad \text{for all } i, j \in \Pi. \tag{4-3}$$

Define $G_\sigma \in \mathrm{GL}_\Pi(\mathbb{Z})$ by $(G_\sigma)_{ij} = \epsilon_\sigma(i)\delta_{\hat{\sigma}(i)j}$. Then (4-3) can be rewritten as

$$\sigma(s) = G_\sigma s = s G_\sigma^{-1} \tag{4-4}$$

where $(\sigma(y))_{ij} = \sigma(y_{ij})$ for $y \in \mathrm{GL}_\Pi(\mathbb{Q}_n)$. Since $G_\sigma \in \mathrm{GL}_\Pi(\mathbb{Z})$, this equation implies that the assignment,

$$\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}}) \to \mathrm{GL}_\Pi(\mathbb{Z}), \quad \sigma \mapsto G_\sigma$$

defines a representation of the group $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ (see [Coste and Gannon 1994]). Moreover,

$$\sigma^2(s) = G_\sigma s G_\sigma^{-1}, \tag{4-5}$$

$$G_\sigma = \sigma(s)s^{-1} = \sigma(s^{-1})s. \tag{4-6}$$

Note that the permutation $\hat{\sigma}$ on Π depends only on the modular category $\mathcal{A}$, as $s_{ij}/s_{0j} = \tilde{s}_{ij}/\tilde{s}_{0j}$ in (4-2). However, the matrix G_σ does depend on s, and hence the representation ρ.

Suppose $\tilde{t} = [\delta_{ij}\theta_j]_{i,j\in\Pi}$. Then $t = x\tilde{t}/\zeta$ is a diagonal matrix of order n. If $\sigma|_{\mathbb{Q}_n} = \sigma_a$ for some integer a relatively prime to n, then

$$\sigma(t) = \sigma_a(t) = t^a.$$

By virtue of (4-5), to prove Theorem II (iii), it suffices to show that

$$\sigma^2(t) = G_\sigma t G_\sigma^{-1}. \tag{4-7}$$

We first establish the following simple observation.

Lemma 4.1. *For any integers a, b such that $ab \equiv 1 \pmod{n}$, we have*

$$s^2 = (t^a s t^b s t^a)^2.$$

Proof. It follows from direct computation that

$$\mathfrak{s}^2 \equiv \begin{bmatrix} 0 & -a \\ b & 0 \end{bmatrix}^2 \equiv (\mathfrak{t}^a \mathfrak{s} \mathfrak{t}^b \mathfrak{s} \mathfrak{t}^a)^2 \pmod{n}.$$

By Theorem II (i), ρ factors through $\mathrm{SL}_2(\mathbb{Z}_n)$ and so we obtain the equality. $\square$

Galois symmetry of Drinfeld doubles. Before we return to prove the Galois symmetry for general modular categories, we need to settle the special case, stated in the following lemma, when $\mathcal{A}$ is the Drinfeld center of a spherical fusion category over $\Bbbk$, and ρ is the canonical modular representation of $\mathcal{A}$.

Lemma 4.2. *Let $\mathcal{C}$ be a spherical fusion category over $\Bbbk$, and take $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$. Suppose G_σ is the signed permutation matrix determined by the canonical normalization $s = \tilde{s}/\dim\mathcal{C}$ of the S-matrix of the center $Z(\mathcal{C})$, i.e., $G_\sigma = \sigma(s)s^{-1}$. Then the T-matrix $\tilde{t}$ of $Z(\mathcal{C})$ satisfies*

$$\sigma^2(\tilde{t}) = G_\sigma \tilde{t} G_\sigma^{-1}. \tag{4-8}$$

In particular, if $(G_\sigma)_{ij} = \epsilon_\sigma(i)\delta_{\hat{\sigma}(i)j}$ for some sign function ϵ_σ and permutation $\hat{\sigma}$ on $\Pi_{Z(\mathcal{C})}$, then $\sigma^2(\tilde{t}_{ii}) = \tilde{t}_{\hat{\sigma}(i)\hat{\sigma}(i)}$ for all $i \in \Pi_{Z(\mathcal{C})}$. Moreover, for any integers a, b relatively prime to $N = \mathrm{ord}(\tilde{t})$ such that $\sigma|_{\mathbb{Q}_N} = \sigma_a$ and $ab \equiv 1 \pmod N$,

$$G_\sigma = \tilde{t}^a s \tilde{t}^b s \tilde{t}^a s^{-1}.$$

The proof of this lemma, which requires the machinery of generalized Frobenius–Schur indicators, will be developed independently in Section 5.

Galois symmetry of general modular categories. Let c be the braiding of the modular category $\mathcal{A}$. Without loss of generality, we further assume the underlying pivotal category of $\mathcal{A}$ is *strict*. We set

$$\sigma_{X\otimes Y}(V) = (c_{X,V} \otimes Y) \circ (X \otimes c^{-1}_{V,Y})$$

for any $X, Y, V \in \mathcal{A}$. Then $(X \otimes Y, \sigma_{X\otimes Y})$ is a simple object of $Z(\mathcal{A})$ if X, Y are simple objects of $\mathcal{A}$. Moreover, if V_i denotes a representative of $i \in \Pi$, then

$$\{(V_i \otimes V_j, \sigma_{V_i\otimes V_j}) \mid i, j \in \Pi\}$$

forms a complete set of representatives of simple objects in $Z(\mathcal{A})$ (see [Müger 2003b, Section 7]). Let $(i, j) \in \Pi \times \Pi$ denote the isomorphism class of $(V_i \otimes V_j, \sigma_{V_i\otimes V_j})$ in $Z(\mathcal{A})$. Then we have $\Pi_{Z(\mathcal{A})} = \Pi \times \Pi$ and the isomorphism class of the unit object of $Z(\mathcal{A})$ is $(0, 0) \in \Pi_{Z(\mathcal{A})}$.

Let $\tilde{s}$ and $\tilde{t} = [\delta_{ij}\theta_i]_{i,j\in\Pi}$ be the S- and T-matrices of $\mathcal{A}$ respectively. Then the S- and T-matrices of the center $Z(\mathcal{A})$, denoted by $\tilde{\boldsymbol{s}}$ and $\tilde{\boldsymbol{t}}$ respectively, are indexed by $\Pi \times \Pi$. By [Ng and Schauenburg 2010, Section 6],

$$\tilde{\boldsymbol{s}}_{ij,kl} = \tilde{s}_{ik}\tilde{s}_{jl^*}, \quad \tilde{\boldsymbol{t}}_{ij,kl} = \delta_{ik}\delta_{jl}\frac{\theta_i}{\theta_j}.$$

Thus $\mathrm{FSexp}(\mathcal{A}) = \mathrm{ord}(\tilde{\boldsymbol{t}}) = \mathrm{ord}(\tilde{t}) = N$.

Proof of Theorem II (iii) and (iv). The canonical normalization $\boldsymbol{s}$ of $\tilde{s}$ is

$$\boldsymbol{s}_{ij,kl} = \frac{1}{\dim \mathcal{A}} \tilde{s}_{ik}\tilde{s}_{jl^*} = \operatorname{sgn}(s) s_{ik} s_{jl^*},$$

where $\operatorname{sgn}(s) = \pm 1$ is given by $s^2 = \operatorname{sgn}(s) C$ (see (4-1)). Moreover, $\boldsymbol{s} \in \mathrm{GL}_{\Pi\times\Pi}(\mathbb{Q}_N)$.

For $\sigma \in \operatorname{Aut}(\mathbb{Q}_{\mathrm{ab}})$, we have

$$\sigma(\boldsymbol{s}_{ij,kl}) = \operatorname{sgn}(s)\epsilon_\sigma(i)\epsilon_\sigma(j) s_{\hat{\sigma}(i)k} s_{\hat{\sigma}(j)l^*} = \epsilon_\sigma(i)\epsilon_\sigma(j)\boldsymbol{s}_{\hat{\sigma}(i)\hat{\sigma}(j),kl} = \boldsymbol{\epsilon}_\sigma(i,j)\boldsymbol{s}_{\hat{\boldsymbol{\sigma}}(i,j),kl},$$

where $\boldsymbol{\epsilon}_\sigma$ and $\hat{\boldsymbol{\sigma}}$ are respectively the associated sign function and permutation on $\Pi \times \Pi$. Thus,

$$\boldsymbol{\epsilon}_\sigma(i,j) = \epsilon_\sigma(i)\epsilon_\sigma(j), \quad \hat{\boldsymbol{\sigma}}(i,j) = (\hat{\sigma}(i), \hat{\sigma}(j))$$

and so

$$(\boldsymbol{G}_\sigma)_{ij,kl} = \epsilon_\sigma(i)\epsilon_\sigma(j)\delta_{\hat{\sigma}(i)k}\delta_{\hat{\sigma}(j)l}$$

where $\boldsymbol{G}_\sigma$ is the associated signed permutation matrix of σ on $\boldsymbol{s}$. By Lemma 4.2, we find

$$\sigma^2\left(\frac{\theta_i}{\theta_j}\right) = \sigma^2(\tilde{\boldsymbol{t}}_{ij,ij}) = \tilde{\boldsymbol{t}}_{\hat{\boldsymbol{\sigma}}(i,j),\hat{\boldsymbol{\sigma}}(i,j)} = \tilde{\boldsymbol{t}}_{\hat{\sigma}(i)\hat{\sigma}(j),\hat{\sigma}(i)\hat{\sigma}(j)} = \frac{\theta_{\hat{\sigma}(i)}}{\theta_{\hat{\sigma}(j)}}$$

for all $i, j \in \Pi$. Since $\theta_0 = 1$,

$$\frac{\theta_{\hat{\sigma}(i)}}{\sigma^2(\theta_i)} = \frac{\theta_{\hat{\sigma}(0)}}{\sigma^2(\theta_0)} = \theta_{\hat{\sigma}(0)}$$

for all $i \in \Pi$. By (4-1), $t = \tilde{\zeta}^{-1}\tilde{t}$ where $\tilde{\zeta} = \zeta/x$. Then

$$t_{\hat{\sigma}(i)\hat{\sigma}(i)} = \frac{\theta_{\hat{\sigma}(i)}}{\tilde{\zeta}} = \frac{\sigma^2(\theta_i)\theta_{\hat{\sigma}(0)}}{\tilde{\zeta}} = \sigma^2(t_{ii})\beta \tag{4-9}$$

for all $i \in \Pi$, where $\beta = t_{\hat{\sigma}(0)\hat{\sigma}(0)} \cdot \sigma^2(\tilde{\zeta}) \in \Bbbk^\times$. Suppose $\sigma|_{\mathbb{Q}_n} = \sigma_a$ for some integer a relatively prime to n. Then (4-9) is equivalent to the equalities

$$G_\sigma t G_\sigma^{-1} = \beta t^{a^2} \quad \text{or} \quad G_\sigma^{-1} t^{a^2} G_\sigma = \beta^{-1} t. \tag{4-10}$$

Now it suffices to show that $\beta = 1$.

Apply σ^2 to the equation $(s^{-1}t)^3 = \mathrm{id}$. It follows from (4-10) that

$$\mathrm{id} = G_\sigma s^{-1} G_\sigma^{-1} t^{a^2} G_\sigma s^{-1} G_\sigma^{-1} t^{a^2} G_\sigma s^{-1} G_\sigma^{-1} t^{a^2} = \beta^{-2}(G_\sigma s^{-1} t s^{-1} t s^{-1} G_\sigma^{-1} t^{a^2}).$$

This implies

$$\mathrm{id} = \beta^{-2}(s^{-1} t s^{-1} t s^{-1} G_\sigma^{-1} t^{a^2} G_\sigma) = \beta^{-3}(s^{-1} t s^{-1} t s^{-1} t) = \beta^{-3}\,\mathrm{id}.$$

Therefore, $\beta^3 = 1$.

Apply σ^{-1} to the equality $sts = t^{-1}st^{-1}$. Since $\sigma^{-1}|_{\mathbb{Q}_n} = \sigma_b$ where b is an inverse of a modulo n, we have

$$G_\sigma^{-1} st^b s G_\sigma = t^{-b} s G_\sigma t^{-b} \quad \text{or} \quad st^b s = G_\sigma t^{-b} s G_\sigma t^{-b} G_\sigma^{-1}.$$

This implies

$$\begin{aligned} G_\sigma^{-1} t^a st^b st^a G_\sigma &= G_\sigma^{-1} t^a G_\sigma t^{-b} s G_\sigma t^{-b} G_\sigma^{-1} t^a G_\sigma \\ &= \sigma^{-1}(G_\sigma^{-1} t^{a^2} G_\sigma) t^{-b} s G_\sigma t^{-b} \sigma^{-1}(G_\sigma^{-1} t^{a^2} G_\sigma) \\ &= \sigma^{-1}(\beta^{-1}) t^b t^{-b} s G_\sigma t^{-b} \sigma^{-1}(\beta^{-1}) t^b \\ &= \sigma^{-1}(\beta^{-2}) s G_\sigma. \end{aligned}$$

Therefore,

$$t^a st^b st^a = \sigma^{-1}(\beta^{-2}) G_\sigma s. \tag{4-11}$$

Note that

$$(G_\sigma s)^2 = G_\sigma s G_\sigma s = s G_\sigma^{-1} G_\sigma s = s^2.$$

Square both sides of (4-11) and apply Lemma 4.1. We obtain

$$s^2 = \sigma^{-1}(\beta^{-4}) s^2.$$

Consequently, $\sigma^{-1}(\beta^{-4}) = 1$ and this is equivalent to $\beta^4 = 1$. Now we can conclude that $\beta = 1$ and so

$$G_\sigma t G_\sigma^{-1} = t^{a^2}.$$

By (4-11), we also have $G_\sigma = t^a st^b st^a s^{-1}$. □

Remark 4.3. For the case $\mathcal{A} = \mathrm{Rep}(D(H))$, where H is a semisimple Hopf algebra, the T-matrix $\tilde{t}$ of $\mathcal{A}$ was proven to satisfy $\sigma^2(\tilde{t}_{ii}) = \tilde{t}_{\hat{\sigma}(i)\hat{\sigma}(i)}$ in [Sommerhäuser and Zhu 2012, Proposition 12.1]. The underlying modular representation of $\mathcal{A}$, in the context of Theorem II (iii) and (iv), is the canonical modular representation of $\mathcal{A}$ described in Section 1.4.

We can now establish the Galois symmetry of RCFT as a corollary.

Corollary 4.4. *Let V be a vertex operator algebra satisfying (V1) and (V2) with simple V-modules $M^0, \ldots, M^p$. Then the genus one S- and T-matrices of V admit the Galois symmetry: for $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$, there exists a signed permutation matrix $G_\sigma \in \mathrm{GL}_{p+1}(\mathbb{C})$ such that*

$$\sigma(S) = G_\sigma S = S G_\sigma \quad \text{and} \quad \sigma^2(T) = G_\sigma T G_\sigma^{-1}$$

where the associated permutation $\hat{\sigma} \in S_{p+1}$ of G_σ is determined by

$$\sigma\left(\frac{S_{ij}}{S_{0j}}\right) = \frac{S_{i\hat{\sigma}(j)}}{S_{0\hat{\sigma}(j)}} \quad \text{for all } i, j = 0, \ldots, p.$$

In particular, $\sigma^2(T_{ii}) = T_{\hat\sigma(i)\hat\sigma(i)}$. *If* $n = \operatorname{ord}(T)$ *and* $\sigma|_{\mathbb{Q}_n} = \sigma_a$ *for some integer* a *relatively prime to* n, *then*

$$G_\sigma = T^a S T^b S T^a S^{-1}$$

where b *is an inverse of* a *modulo* n.

Proof. The result is an immediate consequence of Theorem 3.10 and Theorem II (iii) and (iv). □

Remark 4.5. The modular representation ρ factors through a representation given by $\rho_n : \mathrm{SL}_2(\mathbb{Z}_n) \to \mathrm{GL}_\Pi(\Bbbk)$. For any integers a, b such that $ab \equiv 1 \pmod n$, the matrix

$$\mathfrak{d}_a = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \equiv \mathfrak{t}^a \mathfrak{s} \mathfrak{t}^b \mathfrak{s} \mathfrak{t}^a \mathfrak{s}^{-1} \pmod n$$

is uniquely determined in $\mathrm{SL}_2(\mathbb{Z}_n)$ by the coset $a + n\mathbb{Z}$ of $\mathbb{Z}$. Moreover, the assignment $u : \mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q}) \to \mathrm{SL}_2(\mathbb{Z}_n)$, $\sigma_a \mapsto \mathfrak{d}_a$, defines a group monomorphism. Theorem II (iv) implies that the representation $\phi_\rho : \mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q}) \to \mathrm{GL}_\Pi(\mathbb{Z})$, $\sigma \mapsto G_\sigma$, associated with ρ, also factors through ρ_n, satisfying the following commutative diagram:

$$\begin{tikzcd}
\mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q}) \arrow[r, "\phi_\rho"] \arrow[d, "u"'] & \mathrm{GL}_\Pi(\Bbbk) \\
\mathrm{SL}_2(\mathbb{Z}_n) \arrow[ur, "\rho_n"] & \mathrm{SL}_2(\mathbb{Z}) \arrow[l, "\pi_n"] \arrow[u, "\rho"']
\end{tikzcd}$$

The Galois symmetry enjoyed by the T-matrix of the Drinfeld center of a spherical fusion category (Lemma 4.2) does not hold for a general modular category, as demonstrated in the following example.

Example 4.6. Consider the Fibonacci modular category $\mathcal{A}$ over $\mathbb{C}$ which has only one isomorphism class of non-unit simple objects. We abbreviate this non-unit class by 1 (see [Rowell et al. 2009, Section 5.3.2]). Thus, $\Pi_\mathcal{A} = \{0, 1\}$. The S- and T-matrices are given by

$$\tilde{s} = \begin{bmatrix} 1 & \varphi \\ \varphi & -1 \end{bmatrix}, \quad \tilde{t} = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{4\pi i}{5}} \end{bmatrix}.$$

where $\varphi = (1+\sqrt{5})/2$. The central charge is $\boldsymbol{c} = 14/5$ and the global dimension is $\dim \mathcal{A} = 2 + \varphi$. Therefore, $\alpha = e^{7\pi i/5}$ is the anomaly of $\mathcal{A}$ and $\zeta = e^{7\pi i/30}$ is a 6-th root of α (see (1-9)). Thus

$$s = \rho^\zeta(\mathfrak{s}) = \frac{1}{\sqrt{2+\varphi}} \tilde{s}, \quad t = \rho^\zeta(\mathfrak{t}) = \begin{bmatrix} e^{\frac{-7\pi i}{30}} & 0 \\ 0 & e^{\frac{17\pi i}{30}} \end{bmatrix}$$

and ρ^ζ is a level 60 modular representation of $\mathcal{A}$ by Theorem II. In $\mathrm{Gal}(\mathbb{Q}_{60}/\mathbb{Q})$, the unique nontrivial square is σ_{49}. Since $\sigma_7(\sqrt{5}) = -\sqrt{5}$, we have $\sigma_7(\tilde{s}_{i0}/\tilde{s}_{00}) = \tilde{s}_{i1}/\tilde{s}_{01}$. Therefore, $\hat{\sigma}_7$ is the transposition $(0, 1)$ on $\Pi_{\mathcal{A}}$, and

$$\sigma_7^2(t) = \sigma_{49}(t) = \begin{bmatrix} e^{\frac{17\pi i}{30}} & 0 \\ 0 & e^{\frac{-7\pi i}{30}} \end{bmatrix} = \begin{bmatrix} t_{11} & 0 \\ 0 & t_{00} \end{bmatrix}.$$

However, the Galois symmetry does not hold for $\tilde{t}$, as

$$\sigma_7^2(\tilde{t}) = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{6\pi i}{5}} \end{bmatrix} \neq \begin{bmatrix} \tilde{t}_{11} & 0 \\ 0 & \tilde{t}_{00} \end{bmatrix}.$$

We close this section with the following proposition which provides a necessary and sufficient condition for such Galois symmetry of the T-matrix $\tilde{t}$ of a modular category.

Proposition 4.7. *Suppose $\mathcal{A}$ is a modular category over $\Bbbk$ with Frobenius–Schur exponent N and T-matrix $\tilde{t} = [\delta_{ij}\theta_i]_{i,j\in\Pi_{\mathcal{A}}}$. Let $\zeta \in \Bbbk$ be a 6-th root of the anomaly $\alpha = p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ of $\mathcal{A}$. Then, for any $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ and $i \in \Pi_{\mathcal{A}}$,*

$$\frac{\theta_{\hat{\sigma}(i)}}{\sigma^2(\theta_i)} = \theta_{\hat{\sigma}(0)} = \frac{\zeta}{\sigma^2(\zeta)}. \tag{4-12}$$

Moreover, the following statements are equivalent:

(i) $\theta_{\hat{\sigma}(0)} = 1$ *for all* $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$.

(ii) $\sigma^2(\theta_i) = \theta_{\hat{\sigma}(i)}$ *for all* $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$.

(iii) $(p_{\mathcal{A}}^+/p_{\mathcal{A}}^-)^4 = 1$.

Proof. By (1-7), the assignment

$$\rho^\zeta(\mathfrak{s}) = s = \lambda^{-1}\tilde{s}, \quad \rho^\zeta(\mathfrak{t}) = t = \zeta^{-1}\tilde{t}$$

defines a modular representation of $\mathcal{A}$ where $\lambda = p_{\mathcal{A}}^+/\zeta^3$. For $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ and $i \in \Pi_{\mathcal{A}}$, Theorem II (iii) implies that

$$\sigma^2\left(\frac{\theta_i}{\zeta}\right) = \sigma^2(t_{ii}) = t_{\hat{\sigma}(i)\hat{\sigma}(i)} = \frac{\theta_{\hat{\sigma}(i)}}{\zeta}.$$

Thus (4-12) follows, as $\theta_0 = 1$.

By (4-12), the equivalence of (i) and (ii) is obvious. Statement (i) is equivalent to

$$\sigma^2(\zeta) = \zeta \quad \text{for all } \sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}}). \tag{4-13}$$

Since the anomaly α is a root of unity, so is ζ. By Lemma A.2, (4-13) holds if and only if $\zeta^{24} = 1$ or $\alpha^4 = 1$. $\square$

Remark 4.8. For a modular category $\mathcal{A}$ over $\mathbb{C}$, it follows from (1-9) that the anomaly of $\mathcal{A}$ is a fourth root of unity if and only if its central charge $\boldsymbol{c}$ is an integer modulo 8.

5. Galois symmetry of quantum doubles

In this section, we provide a proof for Lemma 4.2 which is a special case of Theorem II (iii) and (iv), but which is also crucial to the proof of the theorem. We will invoke the machinery of *generalized Frobenius–Schur indicators* for spherical fusion categories introduced in [Ng and Schauenburg 2010].

Generalized Frobenius–Schur indicators. Frobenius–Schur indicators for group representations have been recently generalized to the representations of Hopf algebras [Linchenko and Montgomery 2000] and quasi-Hopf algebras [Mason and Ng 2005; Schauenburg 2004; Ng and Schauenburg 2008]. A version of the second Frobenius–Schur indicator was introduced in conformal field theory [Bantay 1997], and some categorical versions were studied in [Fuchs et al. 1999; Fuchs and Schweigert 2003]. All these different contexts of indicators are specializations of the Frobenius–Schur indicators for pivotal categories introduced in [Ng and Schauenburg 2007b].

The most recent introduction of the equivariant Frobenius–Schur indicators for semisimple Hopf algebras by [Sommerhäuser and Zhu 2012] has motivated the discovery of generalized Frobenius–Schur indicators for pivotal categories [Ng and Schauenburg 2010]. The specialization of these generalized Frobenius–Schur indicators to spherical fusion categories carries a natural action of $\mathrm{SL}_2(\mathbb{Z})$. This modular group action has played a crucial role for the congruence subgroup theorem [Ng and Schauenburg 2010, Theorem 6.8] of the projective representation of $\mathrm{SL}_2(\mathbb{Z})$ associated with a modular category. These indicators also admit a natural action of $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ which will be employed to prove the Galois symmetry of quantum doubles in this section. For the purpose of this paper, we will only provide relevant details of generalized Frobenius–Schur indicators for our proof to be presented here. The readers are referred to [Ng and Schauenburg 2010] for more details.

Suppose $\mathcal{C}$ is a strict spherical fusion category over $\Bbbk$ with Frobenius–Schur exponent N. For any pair (m, l) of integers, $V \in \mathcal{C}$ and $\boldsymbol{X} = (X, \sigma_X) \in Z(\mathcal{C})$, there is a naturally defined $\Bbbk$-linear operator $E_{\boldsymbol{X},V}^{(m,l)}$ on the finite-dimensional $\Bbbk$-space $\mathcal{C}(X, V^m)$ (see [Ng and Schauenburg 2010, Section 2]). Here, $V^0 = \mathbf{1}$; $V^m = (V^\vee)^{-m}$ if $m < 0$; and V^m is the m-fold tensor product of V if $m > 0$. The (m, l)-th *generalized Frobenius–Schur indicator* for $\boldsymbol{X} \in Z(\mathcal{C})$ and $V \in \mathcal{C}$ is

$$\nu_{m,l}^{\boldsymbol{X}}(V) := \mathrm{tr}\big(E_{\boldsymbol{X},V}^{(m,l)}\big) \tag{5-1}$$

where tr denotes the ordinary trace map. In particular, for $m > 0$ and $f \in \mathcal{C}(X, V^m)$, the operator $E^{(m,1)}_{X,V}(f)$ is the following composition:

$$X \xrightarrow{X\otimes \mathrm{db}_{V^\vee}} X\otimes V^\vee \otimes V \xrightarrow{\sigma_X(V^\vee)\otimes V} V^\vee \otimes X \otimes V \xrightarrow{V^\vee \otimes f \otimes V} V^\vee \otimes V^m \otimes V \xrightarrow{\mathrm{ev}_V \otimes V^m} V^m.$$

It can be shown by graphical calculus that, for $m, l \in \mathbb{Z}$ with $m \neq 0$,

$$E^{(m,l)}_{\boldsymbol{X},V} = \big(E^{(m,1)}_{\boldsymbol{X},V}\big)^l \quad \text{and} \quad \big(E^{(m,1)}_{\boldsymbol{X},V}\big)^{mN} = \mathrm{id} \tag{5-2}$$

(see [Ng and Schauenburg 2010, Lemmas 2.5 and 2.7]). Hence, for $m \neq 0$, we have

$$\nu^{\boldsymbol{X}}_{m,l}(V) = \mathrm{tr}\Big(\big(E^{(m,1)}_{\boldsymbol{X},V}\big)^l\Big). \tag{5-3}$$

Note that $\nu^{\mathbf{1}}_{m,1}(V)$ coincides with the Frobenius–Schur indicator $\nu_m(V)$ of $V \in \mathcal{C}$ introduced in [Ng and Schauenburg 2007b].

Galois group action on generalized Frobenius–Schur indicators. Let $\mathcal{K}(Z(\mathcal{C}))$ denote the Grothendieck ring of $Z(\mathcal{C})$ and let $\mathcal{K}_{\Bbbk}(Z(\mathcal{C})) = \mathcal{K}(Z(\mathcal{C})) \otimes_{\mathbb{Z}} \Bbbk$. For any matrix $y \in \mathrm{GL}_\Pi(\Bbbk)$, we define the linear operator $F(y)$ on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ by

$$F(y)(j) = \sum_{i\in\Pi} y_{ij} i \quad \text{for all } \ j \in \Pi,$$

where $\Pi = \Pi_{Z(\mathcal{C})}$. Then $F : \mathrm{GL}_\Pi(\Bbbk) \to \mathrm{Aut}_{\Bbbk}(\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ is a group isomorphism. In particular, every representation $\rho : G \to \mathrm{GL}_\Pi(\Bbbk)$ of a group G can be considered as a G-action on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ through F. More precisely, for $g \in G$, we define

$$gj = F(\rho(g))(j) \quad \text{for all } \ j \in \Pi.$$

Let $\tilde{s}$ and $\tilde{t}$ be the S- and T-matrices of $Z(\mathcal{C})$. The $\mathrm{SL}_2(\mathbb{Z})$-action on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ associated with the canonical modular representation $\rho_{Z(\mathcal{C})}$ of $Z(\mathcal{C})$ is then given by

$$\mathfrak{s} j = \sum_{i\in\Pi} s_{ij} i \quad \text{and} \quad \mathfrak{t} j = \theta_j j, \tag{5-4}$$

where $\tilde{t} = [\delta_{ij}\theta_j]_{i,j\in\Pi}$ and $s = \tilde{s}/\dim \mathcal{C}$ (see (1-10)). Note that $s \in \mathrm{GL}_\Pi(\mathbb{Q}_N)$ by Theorem II (ii), since $N = \mathrm{ord}(\tilde{t})$.

Now we extend the generalized indicator $\nu^{\boldsymbol{X}}_{m,l}(V)$ linearly via the basis Π to a functional $I_V((m,l), -)$ on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$. Let $V \in \mathcal{C}$ and $(m,l) \in \mathbb{Z}^2$, and let $z = \sum_{i\in\Pi} \alpha_i i \in \mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ for some $\alpha_i \in \Bbbk$. Then we define

$$I_V((m,l), z) = \sum_{i\in\Pi} \alpha_i \nu^{\boldsymbol{X}_i}_{m,l}(V)$$

where $\boldsymbol{X}_i$ denotes an arbitrary object in the isomorphism class i. The $\mathrm{SL}_2(\mathbb{Z})$-actions on $\mathbb{Z}^2$ and on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ are related by these functionals on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$. In the following theorem, we summarize some results on these generalized indicators relevant to the proof of Lemma 4.2 (see Section 5 of [Ng and Schauenburg 2010]).

Theorem 5.1. *Let $\mathcal{C}$ be a spherical fusion category $\mathcal{C}$ over $\Bbbk$ with Frobenius–Schur exponent N. Suppose $z \in \mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$, $\boldsymbol{X} = (X, \sigma_X) \in Z(\mathcal{C})$, $V \in \mathcal{C}$, $(m, l) \in \mathbb{Z}^2$ and $J = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$. Then*:

(i) $\nu^{\boldsymbol{X}}_{m,l}(V) \in \mathbb{Q}_N$.

(ii) $\nu^{\boldsymbol{X}}_{1,0}(V) = \dim_{\Bbbk} \mathcal{C}(X, V)$.

(iii) $I_V((m,l)\gamma, z) = I_V((m,l), \gamma^J z)$ *for* $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, *where* $\gamma^J = J\gamma J$.

In particular, $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ *acts on the generalized Frobenius–Schur indicators* $\nu^{\boldsymbol{X}}_{m,l}(V)$. □

For $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$, the matrix $G_\sigma = \sigma(s)s^{-1}$ is also given by

$$(G_\sigma)_{ij} = \epsilon_\sigma(i)\delta_{\hat{\sigma}(i)j}$$

for some sign function ϵ_σ and permutation $\hat{\sigma}$ on Π (see (4-2), (4-3) and (4-4)). Define $\mathfrak{f}_\sigma = F(G_\sigma)$. Then

$$\mathfrak{f}_\sigma j = \epsilon_\sigma(\hat{\sigma}^{-1}(j))\hat{\sigma}^{-1}(j) \quad \text{for } j \in \Pi. \tag{5-5}$$

Since the assignment $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}}) \to \mathrm{GL}_\Pi(\mathbb{Z})$, $\sigma \mapsto G_\sigma$ is a representation of $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$,

$$\mathfrak{f}_\sigma \mathfrak{f}_\tau = \mathfrak{f}_{\sigma\tau} \quad \text{for all } \sigma, \tau \in \mathrm{Gal}(\mathbb{Q}_N/\mathbb{Q}).$$

Therefore, by direct computation,

$$\mathfrak{f}_{\sigma^{-1}} j = \mathfrak{f}_\sigma^{-1} j = \epsilon_\sigma(j)\hat{\sigma}(j) \quad \text{for } j \in \Pi.$$

Remark 5.2. Since $s \in \mathrm{GL}_\Pi(\mathbb{Q}_N)$, if $\sigma, \sigma' \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ such that $\sigma|_{\mathbb{Q}_N} = \sigma'|_{\mathbb{Q}_N}$, then

$$G_\sigma = G_{\sigma'} \quad \text{and so} \quad \mathfrak{f}_\sigma = \mathfrak{f}_{\sigma'}.$$

Now we can establish the following lemma which describes a relation between the $\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$-action on $\mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$ and the $\mathrm{SL}_2(\mathbb{Z})$-action in terms of the functionals $I_V((m,l), -)$.

Lemma 5.3. *Take $V \in \mathcal{C}$ and let a, l be nonzero integers such that a is relatively prime to lN. Suppose $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ satisfies $\sigma|_{\mathbb{Q}_N} = \sigma_a$. Then, for all $z \in \mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$,*

$$I_V((a,l), z) = I_V((1,0), \mathfrak{t}^{-al}\mathfrak{f}_\sigma z).$$

Proof. Let $\boldsymbol{X}_j$ be a representative of $j \in \Pi$. By (5-2), (5-3) and Theorem 5.1 (i), for any nonzero integer m, there is a linear operator $E_m = E^{(m,1)}_{\boldsymbol{X}_j, V}$ on a finite-dimensional space such that $(E_m)^{mN} = \mathrm{id}$ and

$$\nu^{\boldsymbol{X}_j}_{m,k}(V) = \mathrm{tr}(E^k_m) \in \mathbb{Q}_N$$

for all integers k. In particular, the eigenvalues of E_m are $|mN|$-th roots of unity.

Suppose $\tau \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ such that $\tau|_{\mathbb{Q}_{|lN|}} = \sigma_a$. Then $\tau|_{\mathbb{Q}_N} = \sigma_a = \sigma|_{\mathbb{Q}_N}$. Therefore,

$$\sigma(\nu_{l,-1}^{X_j}(V)) = \tau(\mathrm{tr}(E_l^{-1})) = \mathrm{tr}(E_l^{-a}) = \nu_{l,-a}^{X_j}(V) = I_V((l,-a), j) \tag{5-6}$$

and

$$\begin{aligned}\sigma(\nu_{1,l}^{X_j}(V)) &= \sigma_a(\mathrm{tr}(E_1^l)) = \mathrm{tr}(E_1^{la}) = \nu_{1,la}^{X_j}(V)\\ &= I_V((1,la), j) = I_V((1,0)\mathfrak{t}^{la}, j) = I_V((1,0), \mathfrak{t}^{-la} j).\end{aligned} \tag{5-7}$$

Here, the last equality follows from Theorem 5.1 (iii).

On the other hand, by Theorem 5.1 (iii), we have

$$\nu_{1,l}^{X_j}(V) = I_V((1,l), j) = I_V((l,-1)\mathfrak{s}^{-1}, j) = I_V((l,-1), \mathfrak{s} j) = \sum_{i\in\Pi} s_{ij}\nu_{l,-1}^{X_i}(V).$$

Therefore, (5-6) and Theorem 5.1 (iii) imply

$$\begin{aligned}\sigma(\nu_{1,l}^{X_j}(V)) &= \sigma\Big(\sum_{i\in\Pi} s_{ij}\nu_{l,-1}^{X_i}(V)\Big) = \sum_{i\in\Pi}\epsilon_\sigma(j)s_{i\hat\sigma(j)}\sigma(\nu_{l,-1}^{X_i}(V))\\ &= \sum_{i\in\Pi}\epsilon_\sigma(j)s_{i\hat\sigma(j)} I_V((l,-a), i) = I_V((l,-a), \epsilon_\sigma(j)\mathfrak{s}\,\hat\sigma(j))\\ &= I_V((l,-a), \mathfrak{s}(\mathfrak{f}_{\sigma^{-1}} j)) = I_V((l,-a)\mathfrak{s}^{-1}, \mathfrak{f}_{\sigma^{-1}} j) = I_V((a,l), \mathfrak{f}_{\sigma^{-1}} j).\end{aligned}$$

It follows from (5-7) that, for all $j \in \Pi$,

$$I_V((a,l), \mathfrak{f}_{\sigma^{-1}} j) = I_V((1,0), \mathfrak{t}^{-la} j)$$

and so

$$I_V((a,l), \mathfrak{f}_{\sigma^{-1}} z) = I_V((1,0), \mathfrak{t}^{-la} z)$$

for all $z \in \mathcal{K}_{\Bbbk}(Z(\mathcal{C}))$. The assertion follows by replacing z with $\mathfrak{f}_\sigma z$. □

Remark 5.4. Some related equalities for the representation categories of semisimple Hopf algebras were obtained in [Sommerhäuser and Zhu 2012, Corollary 12.4] with a similar strategy. Because of the conceptual differences of the definitions of generalized Frobenius–Schur indicators for spherical fusion categories and the counterpart for semisimple Hopf algebras introduced in that paper, their approach generally cannot be adapted in fusion categories.

Proof of Lemma 4.2. Let $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$ and let $\sigma|_{\mathbb{Q}_N} = \sigma_a$ for some integer a relatively prime to N. Then $\sigma^{-1}|_{\mathbb{Q}_N} = \sigma_b$ where b is an inverse of a modulo N. By Dirichlet's theorem on primes in arithmetic progressions, there exists a prime q

such that $q \equiv b \pmod N$ and $q \nmid a$. By Lemma 5.3 and Theorem 5.1 (iii), for $j \in \Pi$,

$$\begin{aligned}
I_V((1,0), \mathfrak{t}^{-1}\mathfrak{f}_\sigma \mathfrak{t}^q \mathfrak{f}_{\sigma^{-1}} j)& \\
&= I_V((1,0), \mathfrak{t}^{-aq}\mathfrak{f}_\sigma \mathfrak{t}^q \mathfrak{f}_{\sigma^{-1}} j) = I_V((a,q), \mathfrak{t}^q \mathfrak{f}_{\sigma^{-1}} j) \\
&= I_V((a,q)\mathfrak{t}^{-q}, \mathfrak{f}_{\sigma^{-1}} j) = I_V((a, q-aq), \mathfrak{f}_{\sigma^{-1}} j) \\
&= I_V((1,0), \mathfrak{t}^{-aq+a^2 q}\mathfrak{f}_\sigma \mathfrak{f}_{\sigma^{-1}} j) = I_V((1,0), \mathfrak{t}^{-1+a} j). \qquad (5\text{-}8)
\end{aligned}$$

Using (5-4) and (5-5), we can compute directly the two sides of (5-8). This implies

$$\theta_j^{-1}\theta_{\hat\sigma(j)}^q \nu_{1,0}^{X_j}(V) = \theta_j^{a-1}\nu_{1,0}^{X_j}(V)$$

for all $V \in \mathcal{C}$. Take $V = X_j$ to be the underlying $\mathcal{C}$-object of $\boldsymbol{X}_j$. We then have $\nu_{1,0}^{\boldsymbol{X}_j}(X_j) = \dim_{\Bbbk} \mathcal{C}(X_j, X_j) \geq 1$. Therefore, we have $\theta_j^{-1}\theta_{\hat\sigma(j)}^q = \theta_j^{a-1}$, and hence

$$\theta_{\hat\sigma(j)}^q = \theta_j^a \quad \text{or} \quad \theta_{\hat\sigma(j)} = \theta_j^{a^2}.$$

This is equivalent to the equality

$$\sigma^2(\tilde t) = G_\sigma \tilde t G_\sigma^{-1}.$$

Since $\tilde t s \tilde t s \tilde t = s$, we find that

$$\begin{aligned}
G_\sigma s = \sigma(s) &= \sigma(\tilde t s \tilde t s \tilde t) = \tilde t^a s G_\sigma^{-1} \tilde t^a G_\sigma s \tilde t^a \\
&= \tilde t^a s G_\sigma^{-1} \tilde t^{a^2 b} G_\sigma s \tilde t^a = \tilde t^a s (G_\sigma^{-1} \tilde t^{a^2} G_\sigma)^b s \tilde t^a = \tilde t^a s \tilde t^b s \tilde t^a. \qquad (5\text{-}9)
\end{aligned}$$

Therefore,

$$G_\sigma = \tilde t^a s \tilde t^b s \tilde t^a s^{-1}. \qquad \square$$

6. Anomaly of modular categories

In this section, we apply the congruence property and Galois symmetry of a modular category (Theorem II) to deduce some arithmetic relations among the global dimension, the Frobenius–Schur exponent and the order of the anomaly.

Let $\mathcal{A}$ be a modular category over $\Bbbk$ with Frobenius–Schur exponent N. Since $d(V) \in \mathbb{Q}_N$ for $V \in \mathcal{A}$ (see [Ng and Schauenburg 2010, Proposition 5.7]), the anomaly $\alpha = p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ of $\mathcal{A}$ is a root of unity in $\mathbb{Q}_N$. Therefore, $\alpha^N = 1$ if N is even, and $\alpha^{2N} = 1$ if N is odd.

Let us define $J_{\mathcal{A}} = (-1)^{1+\operatorname{ord}\alpha}$ to record the parity of the order of the anomaly α of $\mathcal{A}$. Note that $J_{\mathcal{A}}$ is intrinsically defined by $\mathcal{A}$. It will become clear that $J_{\mathcal{A}}$ is closely related to the Jacobi symbol $\left(\frac{*}{*}\right)$ in number theory. When $4 \nmid N$, the quantity $J_{\mathcal{A}}$ determines whether $\dim \mathcal{A}$ has a square root in $\mathbb{Q}_N$.

Theorem 6.1. *Let $\mathcal{A}$ be a modular category over $\Bbbk$ with Frobenius–Schur exponent N such that $4 \nmid N$. Then $J_{\mathcal{A}} \dim \mathcal{A}$ has a square root in $\mathbb{Q}_N$ and $-J_{\mathcal{A}} \dim \mathcal{A}$ does not have any square root in $\mathbb{Q}_N$.*

Proof. Let $\zeta \in \Bbbk$ be a 6-th root of the anomaly $\alpha = p_{\mathcal{A}}^{+}/p_{\mathcal{A}}^{-}$ of $\mathcal{A}$. By Corollary 2.5, there exists a 12-th root of unity $x \in \Bbbk$ such that

$$\left(\frac{x}{\zeta}\right)^{N} = 1 \quad \text{and} \quad \frac{x^3 p_{\mathcal{A}}^{+}}{\zeta^3} \in \mathbb{Q}_N.$$

Note that $(p_{\mathcal{A}}^{+}/\zeta^3)^2 = \dim \mathcal{A}$.

Set $N' = N$ if N is odd and $N' = N/2$ if N is even. In particular, N' is odd. Then $(x/\zeta)^{N'} = \pm 1$ and so

$$\alpha^{N'} = \zeta^{6N'} = x^{6N'} = x^6.$$

By straightforward verification, one can show that $x^6 = J_{\mathcal{A}}$. Therefore,

$$\left(\frac{x^3 p_{\mathcal{A}}^{+}}{\zeta^3}\right)^2 = x^6 \dim \mathcal{A} = J_{\mathcal{A}} \dim \mathcal{A}.$$

Suppose $-J_{\mathcal{A}} \dim \mathcal{A}$ also has a square root in $\mathbb{Q}_N$. Since $J_{\mathcal{A}} \dim \mathcal{A}$ has a square root in $\mathbb{Q}_N$, so does -1. Therefore, $4 \mid N$, a contradiction. $\square$

When $\dim \mathcal{A}$ is an odd integer, we will show that $J_{\mathcal{A}} = \left(\frac{-1}{\dim \mathcal{A}}\right)$. Let us fix our convention in the following definition for the remainder of this paper.

Definition 6.2. Let $\mathcal{A}$ be a modular category over $\Bbbk$.

(i) $\mathcal{A}$ is called *mock integral* if its global dimension $\dim \mathcal{A}$ is an integer.

(ii) $\mathcal{A}$ is called *integral* if $d(V) \in \mathbb{Z}$ for all $V \in \mathcal{A}$.

Remark 6.3. The standard definition of integral fusion categories is defined in terms of Frobenius–Perron dimensions. Following [Etingof et al. 2005], a fusion category $\mathcal{C}$ is called integral (resp. weakly integral) if $\operatorname{FPdim} V \in \mathbb{Z}$ for all $V \in \mathcal{C}$ (resp. $\operatorname{FPdim} \mathcal{C} \in \mathbb{Z}$). Moreover, any weakly integral spherical fusion category $\mathcal{C}$ satisfies the pseudounitary condition: $\operatorname{FPdim} \mathcal{C} = \dim \mathcal{C}$. Therefore, weakly integral modular categories are obviously mock integral. The Deligne product of the Fibonacci modular category (see [Rowell et al. 2009, Section 5.3.2]) with its Galois conjugate is a mock integral modular category but not weakly integral.

It follows from [Hong and Rowell 2010, Lemma A.1] and [Etingof et al. 2005, Proposition 8.24] that $d(V) \in \mathbb{Z}$ for all objects V in a modular category $\mathcal{A}$ if and only if $\operatorname{FPdim} V \in \mathbb{Z}$ for all $V \in \mathcal{A}$. Therefore, these two definitions of *integral* modular categories are equivalent. A weakly integral modular category can also be characterized by the integrality of $d(V)^2$ as in the following lemma.

Lemma 6.4. *A modular category $\mathcal{A}$ over $\Bbbk$ is weakly integral if and only if $d(V)^2$ is an integer for any simple object $V \in \mathcal{A}$.*

Proof. By the modularity of $\mathcal{A}$, we have that $\operatorname{FPdim}\mathcal{A} = \dim\mathcal{A}/d(U)^2$ for some simple object U. If $d(V)^2 \in \mathbb{Z}$ for all simple objects $V \in \mathcal{A}$, then $\dim\mathcal{A} \in \mathbb{Z}$ and hence $\operatorname{FPdim}\mathcal{A} \in \mathbb{Z}$. Conversely, if $\operatorname{FPdim}\mathcal{A} \in \mathbb{Z}$, then $\operatorname{FPdim}\mathcal{A} = \dim\mathcal{A}$ and $(\operatorname{FPdim} V)^2 \in \mathbb{Z}$ for all simple objects $V \in \mathcal{A}$ by [Etingof et al. 2005, Propositions 8.24 and 8.27]. Since $d(V)^2 \le (\operatorname{FPdim} V)^2$, the pseudounitarity of $\mathcal{A}$ implies $d(V)^2 = (\operatorname{FPdim} V)^2 \in \mathbb{Z}$. □

Proposition 6.5. *Let $\mathcal{A}$ be a mock integral modular category over $\Bbbk$ with Frobenius–Schur exponent N and odd global dimension* $\dim\mathcal{A}$. *Then $J_{\mathcal{A}} = \left(\frac{-1}{\dim\mathcal{A}}\right)$. In particular*,

$$J_{\mathcal{A}} = \begin{cases} 1 & \textit{if } \dim\mathcal{A} \equiv 1 \pmod 4, \\ -1 & \textit{if } \dim\mathcal{A} \equiv 3 \pmod 4. \end{cases}$$

Moreover, the square-free part of $\dim\mathcal{A}$ *is a divisor of N.*

Proof. We may simply assume $\mathcal{A}$ contains a non-unit simple object. By [Etingof 2002, Theorem 5.1], N divides $(\dim\mathcal{A})^3$. In particular, N is odd. Let $\varphi : \mathbb{Q}_N \to \mathbb{C}$ be any embedding. It follows from the proof of [Etingof et al. 2005, Proposition 2.9] that $\varphi(d_i)$ is real for $i \in \Pi_{\mathcal{A}}$, and so $\varphi(\dim\mathcal{A}) > 1$. We can identify $\mathbb{Q}_N$ with $\varphi(\mathbb{Q}_N)$.

If $\dim\mathcal{A}$ is the square of an integer, then $J_{\mathcal{A}} = 1$ by Theorem 6.1, and we have $\left(\frac{-1}{\dim\mathcal{A}}\right) = 1$. In this case, the last statement is trivial. Suppose $\dim\mathcal{A}$ is not the square of any integer. It follows from Theorem 6.1 that $\mathbb{Q}(\sqrt{J_{\mathcal{A}}\dim\mathcal{A}})$ is a quadratic subfield of $\mathbb{Q}_N$. Note that $\mathbb{Q}(\sqrt{p^*})$ is the unique quadratic subfield of $\mathbb{Q}_{p^\ell}$ for any odd prime p and positive integer ℓ (see [Washington 1997]), where $p^* = \left(\frac{-1}{p}\right)p$, and that $\mathbb{Q}(\sqrt{m}) \ne \mathbb{Q}(\sqrt{m'})$ for any two distinct square-free integers m, m'. Let $p_1, \dots, p_k$ be the distinct prime factors of N. By counting the order 2 elements of $\operatorname{Gal}(\mathbb{Q}_N/\mathbb{Q})$, the quadratic subfields of $\mathbb{Q}_N$ are of the form $\mathbb{Q}(\sqrt{d^*})$ where d is a positive divisor of $p_1 \cdots p_k$ and where $d^* = \left(\frac{-1}{d}\right)d$.

Let a be the square-free part of $\dim\mathcal{A}$. Then we have that $\left(\frac{-1}{\dim\mathcal{A}}\right) = \left(\frac{-1}{a}\right)$ and $\mathbb{Q}(\sqrt{J_{\mathcal{A}}a}) = \mathbb{Q}(\sqrt{J_{\mathcal{A}}\dim\mathcal{A}})$. By the preceding paragraph, $a \mid p_1 \cdots p_k$ and $J_{\mathcal{A}} = \left(\frac{-1}{a}\right)$. □

Remark 6.6. In [Sommerhäuser and Zhu 2009], integral modular categories with the special Galois property

$$\sigma(\tilde{s}_{ij}) = \tilde{s}_{\hat{\sigma}(i)j} \tag{6-1}$$

were discussed. These conditions are not satisfied by some common modular categories such as the Ising and Fibonacci modular categories. However, for semisimple quasi-Hopf algebras with modular module categories, the first statement of the preceding proposition was proved in [Sommerhäuser and Zhu 2009, Theorem 5.3].

A number of new results appear in the serious revision [Sommerhäuser and Zhu 2013] of [Sommerhäuser and Zhu 2009]. In Theorem 2.6 and Proposition 3.5 of these papers, the same statement was established for integral modular categories satisfying (6-1) by considering the quadratic subfields of $\mathbb{Q}_N$ but using a different approach.

The following proposition on modular categories is a slight variation of [Coste and Gannon 1999, Proposition 3], and it was essentially proved [loc. cit.] under the assumption of Galois symmetry which has been proved in the previous sections.

Proposition 6.7. *Let $\mathcal{A}$ be a modular category over $\Bbbk$, and let ρ be a modular representation of $\mathcal{A}$. Set $s=\rho(\mathfrak{s})$, $t=[\delta_{ij}t_i]_{i,j\in\Pi_{\mathcal{A}}}=\rho(\mathfrak{t})$, $n=\operatorname{ord}(t)$ and*

$$\mathbb{K}_b=\mathbb{Q}(s_{ib}/s_{0b}\mid i\in\Pi_{\mathcal{A}})\quad \textit{for } b\in\Pi_{\mathcal{A}}.$$

(i) *Then $\sigma^2(t_b)=t_b$ for $\sigma\in\operatorname{Gal}(\mathbb{Q}_n/\mathbb{K}_b)$.*

(ii) *If $\mathcal{A}$ is integral, then the anomaly $\alpha=p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ of $\mathcal{A}$ is a 4-th root of unity.*

(iii) *Let $\mathbb{K}=\mathbb{Q}(s_{ib}/s_{0b}\mid i,b\in\Pi_{\mathcal{A}})$, and let k be the conductor of $\mathbb{K}$, i.e., the smallest positive integer k such that $\mathbb{K}\subseteq\mathbb{Q}_k$. Then $\operatorname{Gal}(\mathbb{Q}_n/\mathbb{K})$ is an elementary 2-group, and $|\operatorname{Gal}(\mathbb{Q}_n/\mathbb{Q}_k)|$ is a divisor of 8. Moreover, n/k is a divisor of 24, and $\gcd(n/k,k)$ divides 2.*

Proof. (i) For $\sigma\in\operatorname{Gal}(\mathbb{Q}_n/\mathbb{K}_b)$, let ϵ_σ be the sign function determined by s (see (4-3)). Suppose $s^2=\operatorname{sgn}(s)C$ where $\operatorname{sgn}(s)=\pm1$. Then, by (4-2),

$$\frac{\operatorname{sgn}(s)}{s_{0b}^2}=\sum_{i\in\Pi_{\mathcal{A}}}\frac{s_{ib}s_{ib^*}}{s_{0b}^2}=\sum_{i\in\Pi_{\mathcal{A}}}\left(\frac{s_{ib}}{s_{0b}}\right)\left(\frac{s_{ib^*}}{s_{0b}}\right)=\sum_{i\in\Pi_{\mathcal{A}}}\left(\frac{s_{ib}}{s_{0b}}\right)\left(\frac{s_{i^*b}}{s_{0b}}\right)\in\mathbb{K}_b.$$

Therefore, $s_{0b}^2\in\mathbb{K}_b$ and so $\sigma(s_{0b}^2)=s_{0b}^2$. Since $\sigma(s_{0b})=\epsilon_\sigma(b)s_{0\hat\sigma(b)}$, we have $s_{0\hat\sigma(b)}=\epsilon s_{0b}$ for some sign ϵ. Now, for $i\in\Pi_{\mathcal{A}}$,

$$\frac{s_{ib}}{s_{0b}}=\sigma\left(\frac{s_{ib}}{s_{0b}}\right)=\frac{s_{i\hat\sigma(b)}}{s_{0\hat\sigma(b)}}=\frac{\epsilon s_{i\hat\sigma(b)}}{s_{0b}}.$$

Thus, $s_{ib}=\epsilon s_{i\hat\sigma(b)}$ for all $i\in\Pi_{\mathcal{A}}$. If $\hat\sigma(b)\neq b$, then the b-th and the $\hat\sigma(b)$-th columns of s are linearly dependent but this contradicts the invertibility of s. Therefore, $\hat\sigma(b)=b$ and hence, by Theorem II (iii), $\sigma^2(t_b)=t_{\hat\sigma(b)}=t_b$.

(ii) If $\mathcal{A}$ is integral, then $\mathbb{K}_0=\mathbb{Q}$ and hence $\sigma^2(t_0)=t_0$ for all $\sigma\in\operatorname{Gal}(\mathbb{Q}_n/\mathbb{Q})$. Recall from Section 1.3 that $t_0=x/\zeta$ for some 6-th root ζ of α and some 12-th root of unity $x\in\Bbbk$. By Lemma A.2, x/ζ is a 24-th root of unity. Therefore,

$$\alpha^4=\zeta^{24}=(\zeta/x)^{24}=1.$$

(iii) By (i), for $\sigma\in\operatorname{Gal}(\mathbb{Q}_n/\mathbb{K})$, we have $\sigma^2(t_b)=t_b$ for all $b\in\Pi_{\mathcal{A}}$. Since $\mathbb{Q}_n$ is generated by t_b ($b\in\Pi_{\mathcal{A}}$), we have $\sigma^2=\mathrm{id}$. Therefore, $\operatorname{Gal}(\mathbb{Q}_n/\mathbb{K})$ is an elementary 2-group, and so is $\operatorname{Gal}(\mathbb{Q}_n/\mathbb{Q}_k)$. Thus, for any integer a relatively prime to n such that $a\equiv1 \pmod{k}$, we have $a^2\equiv1 \pmod{n}$. By Lemma A.3, we have that n/k is a divisor of 24 and that $\gcd(n/k, k)\,|\,2$. Moreover, $|\mathrm{Gal}(\mathbb{Q}_n/\mathbb{Q}_k)| = \phi(n)/\phi(k)$ is a divisor of 8. $\square$

Remark 6.8. The proof of the preceding proposition is a mere adaptation of [Coste and Gannon 1999, Proposition 3]. For integral modular categories satisfying (6-1) (see Remark 6.6), Proposition 6.7 (ii) and (iii) also appear in the final version of [Sommerhäuser and Zhu 2013, Theorems 2.3.2 and 3.4] with similar ideas. The following corollary was also established for factorizable quasi-Hopf algebras in Theorem 4.3 of [Sommerhäuser and Zhu 2009; 2013] with a different approach.

Corollary 6.9. *Let $\mathcal{A}$ be an integral modular category with anomaly $\alpha = p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$. If $\dim\mathcal{A}$ is odd, then $\alpha = \left(\frac{-1}{\dim\mathcal{A}}\right)$.*

Proof. If $\dim\mathcal{A}$ is odd, then so is the Frobenius–Schur exponent N of $\mathcal{A}$, as $N \mid (\dim\mathcal{A})^3$. Since $\alpha \in \mathbb{Q}_N$ and $\alpha^4 = 1$, we have $\alpha^2 = 1$. It follows from Proposition 6.5 that

$$\alpha = (-1)^{1+\mathrm{ord}\,\alpha} = J_{\mathcal{A}} = \left(\frac{-1}{\dim\mathcal{A}}\right). \qquad \square$$

The Ising modular category is an example of a weakly integral modular category (see [Rowell et al. 2009, Section 5.3.4]) and its central charge is $\boldsymbol{c} = 1/2$. Therefore, its anomaly is $e^{\pi i/4}$, an eighth root of unity, and this holds for every weakly integral modular category.

Theorem 6.10. *The anomaly $\alpha = p_{\mathcal{A}}^+/p_{\mathcal{A}}^-$ of any weakly integral modular category $\mathcal{A}$ is an eighth root of unity.*

Proof. Suppose $\zeta \in \Bbbk$ is a 6-th root of the anomaly α of a weakly integral modular category $\mathcal{A}$. Then $\lambda = p_{\mathcal{A}}^+/\zeta^3$ is a square root of $\dim\mathcal{A}$. Consider the modular representation ρ^ζ of $\mathcal{A}$ given by

$$\rho^\zeta : \mathfrak{s} \mapsto s := \frac{1}{\lambda}\tilde{s}, \quad \mathfrak{t} \mapsto t := \frac{1}{\zeta}\tilde{t}.$$

Let $\tilde{t} = [\delta_{ij}\theta_i]_{i,j\in\Pi_{\mathcal{A}}}$ be the T-matrix of $\mathcal{A}$. Since $s_{0i}^2 = d_i^2/\dim\mathcal{A} \in \mathbb{Q}$, we have, for $\sigma \in \mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$,

$$s_{0i}^2 = \sigma(s_{0i}^2) = s_{0\hat{\sigma}(i)}^2$$

or $d_i^2 = d_{\hat{\sigma}(i)}^2$ for all $i \in \Pi_{\mathcal{A}}$. By Theorem II (iii),

$$\sigma^2\left(\sum_{i\in\Pi_{\mathcal{A}}} d_i^2\frac{\theta_i}{\zeta}\right) = \sum_{i\in\Pi_{\mathcal{A}}} d_i^2\frac{\theta_{\hat{\sigma}(i)}}{\zeta} = \sum_{i\in\Pi_{\mathcal{A}}} d_{\hat{\sigma}(i)}^2\frac{\theta_{\hat{\sigma}(i)}}{\zeta} = \sum_{i\in\Pi_{\mathcal{A}}} d_i^2\frac{\theta_i}{\zeta}.$$

Thus, we have

$$\frac{\sigma^2(p_{\mathcal{A}}^+)}{p_{\mathcal{A}}^+} = \frac{\sigma^2(\zeta)}{\zeta}.$$

Since $\dim\mathcal{A}$ is a positive integer, $\sigma^2(\lambda)=\lambda$ and so

$$\frac{\sigma^2(\zeta^3)}{\zeta^3}=\frac{\sigma^2(p_{\mathcal{A}}^+/\lambda)}{p_{\mathcal{A}}^+/\lambda}=\frac{\sigma^2(p_{\mathcal{A}}^+)}{p_{\mathcal{A}}^+}=\frac{\sigma^2(\zeta)}{\zeta}.$$

Therefore, we find $\sigma^2(\zeta^2)/\zeta^2=1$ for all $\sigma\in\mathrm{Aut}(\mathbb{Q}_{\mathrm{ab}})$. It follows from Lemma A.2 that $\zeta^{48}=1$ and so $\alpha^8=1$. □

Corollary 6.9 and the Cauchy theorem for Hopf algebras [Kashina et al. 2006] as well as quasi-Hopf algebras [Ng and Schauenburg 2007a] suggest a more general version of the Cauchy theorem may hold for spherical fusion categories or modular categories over $\Bbbk$. We finish this paper with two equivalent questions.

Question 6.11. Let $\mathcal{C}$ be a spherical fusion category over $\Bbbk$ with Frobenius–Schur exponent N. Let $\mathcal{O}$ denote the ring of integers of $\mathbb{Q}_N$. Must the principal ideals $\mathcal{O}(\dim\mathcal{C})$ and $\mathcal{O}N$ of $\mathcal{O}$ have the same prime ideal factors?

Since $Z(\mathcal{C})$ is a modular category over $\Bbbk$ and $(\dim\mathcal{C})^2=\dim Z(\mathcal{C})$, the preceding question is equivalent to the following:

Question 6.12. Let $\mathcal{A}$ be a modular category over $\Bbbk$ with Frobenius–Schur exponent N. Let $\mathcal{O}$ denote the ring of integers of $\mathbb{Q}_N$. Must the principal ideals $\mathcal{O}(\dim\mathcal{A})$ and $\mathcal{O}N$ of $\mathcal{O}$ have the same prime ideal factors?

By [Etingof 2002], $(\dim\mathcal{A})^3/N\in\mathcal{O}$. Therefore, the prime ideal factors of $\mathcal{O}N$ are a subset of $\mathcal{O}\dim\mathcal{A}$. The converse is only known to be true for the representation categories of semisimple quasi-Hopf algebras, by [Ng and Schauenburg 2007a, Theorem 8.4]. Question 6.11 was originally raised for semisimple Hopf algebras in [Etingof and Gelaki 1999, Question 5.1], which had been solved in [Kashina et al. 2006, Theorem 3.4].

Appendix

The first lemma in this appendix could be known to some experts. An analogous result for $\mathrm{PSL}_2(\mathbb{Z})$ was proved by Wohlfahrt [1964, Theorem 2] (see also Newman's proof [1972, Theorem VIII.8]). However, we do not see the lemma as an immediate consequence of Wohlfahrt's theorem for $\mathrm{PSL}_2(\mathbb{Z})$.

Lemma A.1. *Let H be a congruence normal subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Then the level of H is equal to the order of $\mathfrak{t}H$ in $\mathrm{SL}_2(\mathbb{Z})/H$.*

Proof. Let m be the level of H and let $n=\operatorname{ord}\mathfrak{t}H$. Since $\mathfrak{t}^m\in\Gamma(m)\le H$, we have $\mathfrak{t}^m\in H$ and hence $n\,|\,m$.

Suppose $\gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \Gamma(n)$. Since $ad - bc = 1$, by Dirichlet's theorem, there exists a prime $p \nmid m$ such that $p = d + kc$ for some integer k. Then

$$\mathfrak{t}^{-k}\gamma\mathfrak{t}^{k} = \begin{bmatrix} a' & b' \\ c & p \end{bmatrix} \in \Gamma(n)$$

for some integers a', b'. In particular,

$$a'p - b'c = 1, \quad p \equiv a' \equiv 1 \pmod{n} \quad \text{and} \quad c \equiv b' \equiv 0 \pmod{n}.$$

Since $p \nmid m$, there exists an integer q such that $pq \equiv 1 \pmod{m}$. Thus we have $pq \equiv 1 \pmod{n}$ and so $q \equiv 1 \pmod{n}$. One can verify directly that

$$\begin{bmatrix} a' & b' \\ c & p \end{bmatrix} \equiv \mathfrak{t}^{b'q}\mathfrak{s}^{-1}\mathfrak{t}^{(-c+1)p}\mathfrak{s}\mathfrak{t}^{q}\mathfrak{s}\mathfrak{t}^{p} \pmod{m}.$$

Therefore,

$$\mathfrak{t}^{-k}\gamma\mathfrak{t}^{k}H = \mathfrak{t}^{b'q}\mathfrak{s}^{-1}\mathfrak{t}^{(-c+1)p}\mathfrak{s}\mathfrak{t}^{q}\mathfrak{s}\mathfrak{t}^{p}H = \mathfrak{s}^{-1}\mathfrak{t}\mathfrak{s}\mathfrak{t}\mathfrak{s}\mathfrak{t}H = \mathfrak{s}^{-1}\mathfrak{s}H = H.$$

This implies $\mathfrak{t}^{-k}\gamma\mathfrak{t}^{k} \in H$, and hence $\gamma \in H$. Therefore, $\Gamma(n) \leq H$ and so $m \mid n$. □

The following fact should be well-known. We include the proof here for the convenience of the reader.

Lemma A.2. *Let ζ be a root of unity in $\mathbb{k}$. Then $\sigma^2(\zeta) = \zeta$ for all $\sigma \in \operatorname{Aut}(\mathbb{Q}_{\text{ab}})$ if and only if $\zeta^{24} = 1$.*

Proof. Let m be the order of ζ. Then $\operatorname{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q}) \cong U(\mathbb{Z}_m)$. Note that the group $U(\mathbb{Z}_m)$ has exponent at most 2 if and only if $m \mid 24$. Since $\mathbb{Q}(\zeta)$ is a Galois extension over $\mathbb{Q}$, the restriction map $\operatorname{Aut}(\mathbb{Q}_{\text{ab}}) \xrightarrow{\text{res}} \operatorname{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$ is surjective. Thus, if $\sigma^2(\zeta) = \zeta$ for all $\sigma \in \operatorname{Aut}(\mathbb{Q}_{\text{ab}})$, then the exponent of $\operatorname{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$ is at most 2, and hence $m \mid 24$. Conversely, if $m \mid 24$, then the exponent of $\operatorname{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$ is at most 2, and so $\sigma^2(\zeta) = \zeta$ for all $\sigma \in \operatorname{Aut}(\mathbb{Q}_{\text{ab}})$. □

The next lemma is a variation of the argument used in the proof of [Coste and Gannon 1999, Proposition 3].

Lemma A.3. *Let k be a positive divisor of a positive integer n. Suppose that, for any integer a relatively prime to n such that $a \equiv 1 \pmod{k}$, we have $a^2 \equiv 1 \pmod{n}$. Then $\gcd(n/k, k)$ divides 2 and n/k is a divisor of 24. Moreover, $\phi(n)/\phi(k)$ is a divisor of 8.*

Proof. Let $\pi : U(\mathbb{Z}_n) \to U(\mathbb{Z}_k)$ be the reduction map. The assumption implies that $\ker\pi$ is an elementary 2-group. It follows from the exact sequence

$$0 \to \ker\pi \to U(\mathbb{Z}_n) \xrightarrow{\pi} U(\mathbb{Z}_k) \to 0$$

that $\phi(n)/\phi(k)$ is a power of 2, and so is $\gcd(n/k, k)$. Thus, if $2 \nmid \gcd(n/k, k)$, then $\gcd(n/k, k) = 1$. By the Chinese remainder theorem, for any integer y relatively prime to n/k, there exists an integer a such that $a \equiv y \pmod{n}/k$ and $a \equiv 1 \pmod{k}$. Thus, $a^2 \equiv 1 \pmod{n}$, and hence $y^2 \equiv 1 \pmod{n}/k$. This implies the exponent of $U(\mathbb{Z}_{n/k})$ is at most 2, and therefore $n/k \mid 24$. Moreover, $\phi(n)/\phi(k) = \phi(n/k)$ is a factor of 8.

Suppose $2 \mid \gcd(n/k, k)$. Then $k = 2^u k'$ for some positive integer u and odd integer k'. The aforementioned conclusion implies $n = 2^v n' k'$ where $v > u$ and $\gcd(n', 2^v k') = 1$. By the Chinese remainder theorem, the given condition implies the kernel of the reduction map $U(\mathbb{Z}_{2^v}) \to U(\mathbb{Z}_{2^u})$ is an elementary 2-group. Therefore, $2 \le v \le 3$ if $u = 1$, and $v = u + 1$ if $u > 1$. In both cases, $\gcd(n/k, k) = 2$ and $\phi(2^v)/\phi(2^u)$ is a divisor of 4. By the aforementioned argument, for any integer y relatively prime to n', we have $y^2 \equiv 1 \pmod{n}'$. Therefore, $n' \mid 24$ and hence $n' \mid 3$. Thus, $n/k = n' 2^{v-u} \mid 12$, and

$$\frac{\phi(n)}{\phi(k)} = \phi(n')\,\frac{\phi(2^v)}{\phi(2^u)}$$

is also a divisor of 8. □

Acknowledgements

Part of this paper was carried out while Ng was visiting the National Center for Theoretical Sciences and Shanghai University. He would like to thank these institutes for their generous hospitality, especially Ching-Hung Lam, Wen-Ching Li and Xiuyun Guo for being wonderful hosts. He particularly thanks Ling Long for many invaluable discussions, and Eric Rowell for his suggestions and stimulative discussions after the first version of this paper was posted on the arXiv.

References

[Abe et al. 2004] T. Abe, G. Buhl, and C. Dong, "Rationality, regularity, and C_2-cofiniteness", *Trans. Amer. Math. Soc.* **356**:8 (2004), 3391–3402. MR 2005c:17041 Zbl 1070.17011

[Bakalov and Kirillov 2001] B. Bakalov and A. Kirillov, Jr., *Lectures on tensor categories and modular functors*, University Lecture Series **21**, American Mathematical Society, Providence, RI, 2001. MR 2002d:18003 Zbl 0965.18002

[Bantay 1997] P. Bantay, "The Frobenius–Schur indicator in conformal field theory", *Phys. Lett. B* **394**:1–2 (1997), 87–88. MR 98c:81195 Zbl 0925.81331

[Bantay 2003] P. Bantay, "The kernel of the modular representation and the Galois action in RCFT", *Comm. Math. Phys.* **233**:3 (2003), 423–438. MR 2004g:81240 Zbl 1035.81055

[Bauer et al. 1997] M. Bauer, A. Coste, C. Itzykson, and P. Ruelle, "Comments on the links between su(3) modular invariants, simple factors in the Jacobian of Fermat curves, and rational triangular billiards", *J. Geom. Phys.* **22**:2 (1997), 134–189. MR 98g:81189 Zbl 0895.14009

[Beyl 1986] F. R. Beyl, "The Schur multiplicator of $\mathrm{SL}(2, \mathbb{Z}/m\mathbb{Z})$ and the congruence subgroup property", *Math. Z.* **191**:1 (1986), 23–42. MR 87b:20071 Zbl 0581.20050

[de Boer and Goeree 1991] J. de Boer and J. Goeree, "Markov traces and II_1 factors in conformal field theory", *Comm. Math. Phys.* **139**:2 (1991), 267–304. MR 93i:81211 Zbl 0760.57002

[Cappelli et al. 1987a] A. Cappelli, C. Itzykson, and J.-B. Zuber, "Modular invariant partition functions in two dimensions", *Nuclear Phys. B* **280**:3 (1987), 445–465. MR 88i:81132 Zbl 0661.17017

[Cappelli et al. 1987b] A. Cappelli, C. Itzykson, and J.-B. Zuber, "The A-D-E classification of minimal and $A_1^{(1)}$ conformal invariant theories", *Comm. Math. Phys.* **113**:1 (1987), 1–26. MR 89b:81178 Zbl 0639.17008

[Cardy 1986] J. L. Cardy, "Operator content of two-dimensional conformally invariant theories", *Nuclear Phys. B* **270**:2 (1986), 186–204. MR 87k:17017 Zbl 0689.17016

[Coste and Gannon 1994] A. Coste and T. Gannon, "Remarks on Galois symmetry in rational conformal field theories", *Phys. Lett. B* **323**:3–4 (1994), 316–321. MR 95h:81031

[Coste and Gannon 1999] A. Coste and T. Gannon, "Congruence subgroups and rational conformal field theory", preprint, 1999. arXiv math/9909080

[Dong and Jiang 2010] C. Dong and C. Jiang, "A characterization of vertex operator algebra $L(1/2, 0)\otimes L(1/2, 0)$", *Comm. Math. Phys.* **296**:1 (2010), 69–88. MR 2011c:17049 Zbl 1207.17034

[Dong and Jiang 2011] C. Dong and C. Jiang, "A characterization of vertex operator algebras $V_{\mathbb{Z}\alpha}^{+}$, I", 2011. To appear in *J. Reine Angew. Math.* arXiv 1110.1882

[Dong and Jiang 2013] C. Dong and C. Jiang, "A characterization of the rational vertex operator algebra $V_{\mathbb{Z}\alpha}^{+}$, II", *Adv. Math.* **247** (2013), 41–70. MR 3096793 Zbl 1292.17021

[Dong and Mason 1996] C. Dong and G. Mason, "Vertex operator algebras and Moonshine: a survey", pp. 101–136 in *Progress in algebraic combinatorics* (Fukuoka, 1993), edited by E. Bannai and A. Munemasa, Adv. Stud. Pure Math. **24**, Math. Soc. Japan, Tokyo, 1996. MR 97h:17027 Zbl 0861.17018

[Dong and Zhang 2008] C. Dong and W. Zhang, "On classification of rational vertex operator algebras with central charges less than 1", *J. Algebra* **320**:1 (2008), 86–93. MR 2009c:17045 Zbl 1163.17029

[Dong et al. 1997] C. Dong, H. Li, and G. Mason, "Regularity of rational vertex operator algebras", *Adv. Math.* **132**:1 (1997), 148–166. MR 98m:17037 Zbl 0902.17014

[Dong et al. 1998a] C. Dong, H. Li, and G. Mason, "Twisted representations of vertex operator algebras", *Math. Ann.* **310**:3 (1998), 571–600. MR 99d:17030 Zbl 0890.17029

[Dong et al. 1998b] C. Dong, H. Li, and G. Mason, "Vertex operator algebras and associative algebras", *J. Algebra* **206**:1 (1998), 67–96. MR 99i:17029 Zbl 0911.17017

[Dong et al. 2000] C. Dong, H. Li, and G. Mason, "Modular-invariance of trace functions in orbifold theory and generalized Moonshine", *Comm. Math. Phys.* **214**:1 (2000), 1–56. MR 2001k:17043 Zbl 1061.17025

[Dong et al. 2001] C. Dong, G. Mason, and K. Nagatomo, "Quasi-modular forms and trace functions associated to free boson and lattice vertex operator algebras", *Internat. Math. Res. Notices* 8 (2001), 409–427. MR 2002b:11060 Zbl 0990.17022

[Dong et al. 2013] C. Dong, X. Jiao, and F. Xu, "Quantum dimensions and quantum Galois theory", *Trans. Amer. Math. Soc.* **365**:12 (2013), 6441–6469. MR 3105758 Zbl 06218191

[Eholzer 1995] W. Eholzer, "On the classification of modular fusion algebras", *Comm. Math. Phys.* **172**:3 (1995), 623–659. MR 96e:11060 Zbl 0884.17018

[Eholzer and Skoruppa 1995] W. Eholzer and N.-P. Skoruppa, "Modular invariance and uniqueness of conformal characters", *Comm. Math. Phys.* **174**:1 (1995), 117–136. MR 96k:11052 Zbl 0884.17019

[Etingof 2002] P. Etingof, "On Vafa's theorem for tensor categories", *Math. Res. Lett.* **9**:5–6 (2002), 651–657. MR 2003i:18009 Zbl 1035.18004

[Etingof and Gelaki 1999] P. Etingof and S. Gelaki, "On the exponent of finite-dimensional Hopf algebras", *Math. Res. Lett.* **6**:2 (1999), 131–140. MR 2000f:16045 Zbl 0954.16028

[Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, "On fusion categories", *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR 2006m:16051 Zbl 1125.16025

[Frenkel et al. 1988] I. Frenkel, J. Lepowsky, and A. Meurman, *Vertex operator algebras and the Monster*, Pure and Applied Mathematics **134**, Academic Press, Boston, 1988. MR 90h:17026 Zbl 0674.17001

[Frenkel et al. 1993] I. B. Frenkel, Y.-Z. Huang, and J. Lepowsky, "On axiomatic approaches to vertex operator algebras and modules", *Mem. Amer. Math. Soc.* **104**:494 (1993), viii+64. MR 94a:17007 Zbl 0789.17022

[Fuchs and Schweigert 2003] J. Fuchs and C. Schweigert, "Category theory for conformal boundary conditions", pp. 25–70 in *Vertex operator algebras in mathematics and physics* (Toronto, ON, 2000), edited by S. Berman et al., Fields Inst. Commun. **39**, Amer. Math. Soc., Providence, RI, 2003. MR 2005b:17056 Zbl 1084.17012

[Fuchs et al. 1999] J. Fuchs, A. C. Ganchev, K. Szlachányi, and P. Vecsernyés, "S_4 symmetry of $6j$ symbols and Frobenius–Schur indicators in rigid monoidal C^* categories", *J. Math. Phys.* **40**:1 (1999), 408–426. MR 99k:81111 Zbl 0986.81044

[Gannon 2005] T. Gannon, "Modular data: the algebraic combinatorics of conformal field theory", *J. Algebraic Combin.* **22**:2 (2005), 211–250. MR 2006h:81254 Zbl 1103.17007

[Gannon 2006] T. Gannon, *Moonshine beyond the Monster*, Cambridge University Press, 2006. MR 2008a:17032 Zbl 1146.11026

[Hong and Rowell 2010] S.-m. Hong and E. Rowell, "On the classification of the Grothendieck rings of non-self-dual modular categories", *J. Algebra* **324**:5 (2010), 1000–1015. MR 2011k:18018 Zbl 1210.18006

[Huang 1995] Y.-Z. Huang, "A theory of tensor products for module categories for a vertex operator algebra, IV", *J. Pure Appl. Algebra* **100**:1–3 (1995), 173–216. MR 98a:17050 Zbl 0841.17015

[Huang 2008a] Y.-Z. Huang, "Rigidity and modularity of vertex tensor categories", *Commun. Contemp. Math.* **10**:suppl. 1 (2008), 871–911. MR 2009i:17041 Zbl 1169.17019

[Huang 2008b] Y.-Z. Huang, "Vertex operator algebras and the Verlinde conjecture", *Commun. Contemp. Math.* **10**:1 (2008), 103–154. MR 2009e:17056 Zbl 1180.17008

[Huang and Lepowsky 1995a] Y.-Z. Huang and J. Lepowsky, "A theory of tensor products for module categories for a vertex operator algebra, I", *Selecta Math.* (*N.S.*) **1**:4 (1995), 699–756. MR 98a:17047 Zbl 0854.17032

[Huang and Lepowsky 1995b] Y.-Z. Huang and J. Lepowsky, "A theory of tensor products for module categories for a vertex operator algebra, II", *Selecta Math.* (*N.S.*) **1**:4 (1995), 757–786. MR 98a:17047 Zbl 0854.17033

[Huang and Lepowsky 1995c] Y.-Z. Huang and J. Lepowsky, "A theory of tensor products for module categories for a vertex operator algebra, III", *J. Pure Appl. Algebra* **100**:1–3 (1995), 141–171. MR 98a:17049 Zbl 0841.17014

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, 1990. MR 92k:17038 Zbl 0716.17022

[Kac and Peterson 1984] V. G. Kac and D. H. Peterson, "Infinite-dimensional Lie algebras, theta functions and modular forms", *Adv. in Math.* **53**:2 (1984), 125–264. MR 86a:17007 Zbl 0584.17007

[Karpilovsky 1985] G. Karpilovsky, *Projective representations of finite groups*, Monographs and Textbooks in Pure and Applied Mathematics **94**, Marcel Dekker, New York, 1985. MR 86m:20014 Zbl 0568.20016

[Kashina et al. 2006] Y. Kashina, Y. Sommerhäuser, and Y. Zhu, "On higher Frobenius–Schur indicators", *Mem. Amer. Math. Soc.* **181**:855 (2006), viii+65. MR 2007k:16071 Zbl 1163.16029

[Kassel 1995] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, New York, 1995. MR 96e:17041 Zbl 0808.17003

[Kiritsis 1989] E. B. Kiritsis, "Proof of the completeness of the classification of rational conformal theories with $c = 1$", *Phys. Lett. B* **217**:4 (1989), 427–430. MR 89k:81150

[Knopp and Mason 2003] M. Knopp and G. Mason, "On vector-valued modular forms and their Fourier coefficients", *Acta Arith.* **110**:2 (2003), 117–124. MR 2004j:11043 Zbl 1044.11020

[Lepowsky and Li 2004] J. Lepowsky and H. Li, *Introduction to vertex operator algebras and their representations*, Progress in Mathematics **227**, Birkhäuser, Boston, 2004. MR 2004k:17050 Zbl 1055.17001

[Linchenko and Montgomery 2000] V. Linchenko and S. Montgomery, "A Frobenius–Schur theorem for Hopf algebras", *Algebr. Represent. Theory* **3**:4 (2000), 347–355. MR 2001k:16073 Zbl 0971.16018

[Mason and Ng 2005] G. Mason and S.-H. Ng, "Central invariants and Frobenius–Schur indicators for semisimple quasi-Hopf algebras", *Adv. Math.* **190**:1 (2005), 161–195. MR 2005h:16066 Zbl 1100.16033

[Mennicke 1967] J. Mennicke, "On Ihara's modular group", *Invent. Math.* **4** (1967), 202–228. MR 37 #1485 Zbl 0189.02504

[Moore 1987] G. Moore, "Atkin–Lehner symmetry", *Nuclear Phys. B* **293**:1 (1987), 139–188. MR 89c:81151a

[Moore and Seiberg 1990] G. Moore and N. Seiberg, "Lectures on RCFT", pp. 263–361 in *Physics, geometry, and topology* (Banff, AB, 1989), edited by H. C. Lee, NATO Adv. Sci. Inst. Ser. B Phys. **238**, Plenum, New York, 1990. MR 93m:81133b Zbl 0728.57012

[Müger 2003a] M. Müger, "From subfactors to categories and topology, I: Frobenius algebras in and Morita equivalence of tensor categories", *J. Pure Appl. Algebra* **180**:1–2 (2003), 81–157. MR 2004f:18013 Zbl 1033.18002

[Müger 2003b] M. Müger, "From subfactors to categories and topology, II: The quantum double of tensor categories and subfactors", *J. Pure Appl. Algebra* **180**:1–2 (2003), 159–219. MR 2004f:18014 Zbl 1033.18003

[Newman 1972] M. Newman, *Integral matrices*, Pure and Applied Mathematics **45**, Academic Press, New York-London, 1972. MR 49 #5038 Zbl 0254.15009

[Ng and Schauenburg 2007a] S.-H. Ng and P. Schauenburg, "Frobenius–Schur indicators and exponents of spherical categories", *Adv. Math.* **211**:1 (2007), 34–71. MR 2008b:16067 Zbl 1138.16017

[Ng and Schauenburg 2007b] S.-H. Ng and P. Schauenburg, "Higher Frobenius–Schur indicators for pivotal categories", pp. 63–90 in *Hopf algebras and generalizations*, edited by L. H. Kauffman et al., Contemp. Math. **441**, American Mathematical Society, Providence, RI, 2007. MR 2008m:18015 Zbl 1153.18008

[Ng and Schauenburg 2008] S.-H. Ng and P. Schauenburg, "Central invariants and higher indicators for semisimple quasi-Hopf algebras", *Trans. Amer. Math. Soc.* **360**:4 (2008), 1839–1860. MR 2009d:16065 Zbl 1141.16028

[Ng and Schauenburg 2010] S.-H. Ng and P. Schauenburg, "Congruence subgroups and generalized Frobenius–Schur indicators", *Comm. Math. Phys.* **300**:1 (2010), 1–46. MR 2012g:81193 Zbl 1206.18007

[Rocha-Caridi 1985] A. Rocha-Caridi, "Vacuum vector representations of the Virasoro algebra", pp. 451–473 in *Vertex operators in mathematics and physics* (Berkeley, Calif., 1983), edited by J. Lepowsky et al., Math. Sci. Res. Inst. Publ. **3**, Springer, New York, 1985. MR 87b:17011 Zbl 0557.17005

[Rowell et al. 2009] E. Rowell, R. Stong, and Z. Wang, "On classification of modular tensor categories", *Comm. Math. Phys.* **292**:2 (2009), 343–389. MR 2011b:18013 Zbl 1186.18005

[Schauenburg 2004] P. Schauenburg, "On the Frobenius–Schur indicators for quasi-Hopf algebras", *J. Algebra* **282**:1 (2004), 129–139. MR 2005h:16068 Zbl 1071.16037

[Sommerhäuser and Zhu 2009] Y. Sommerhäuser and Y. Zhu, "On the central charge of a factorizable Hopf algebra", preprint, 2009. arXiv 0906.3471

[Sommerhäuser and Zhu 2012] Y. Sommerhäuser and Y. Zhu, "Hopf algebras and congruence subgroups", *Mem. Amer. Math. Soc.* **219**:1028 (2012), vi+134. MR 2985696 Zbl 1314.16021

[Sommerhäuser and Zhu 2013] Y. Sommerhäuser and Y. Zhu, "On the central charge of a factorizable Hopf algebra", *Adv. Math.* **236** (2013), 158–223. MR 3019720 Zbl 1280.16028

[Turaev 2010] V. G. Turaev, *Quantum invariants of knots and 3-manifolds*, 2nd ed., de Gruyter Studies in Mathematics **18**, Walter de Gruyter & Co., Berlin, 2010. MR 2011f:57023 Zbl 1213.57002

[Vafa 1988] C. Vafa, "Toward classification of conformal theories", *Phys. Lett. B* **206**:3 (1988), 421–426. MR 89k:81178

[Verlinde 1988] E. Verlinde, "Fusion rules and modular transformations in 2D conformal field theory", *Nuclear Phys. B* **300**:3 (1988), 360–376. MR 89h:81238 Zbl 1180.81120

[Wang 2010] Z. Wang, *Topological quantum computation*, CBMS Regional Conference Series in Mathematics **112**, American Mathematical Society, Providence, RI, 2010. MR 2011f:81054 Zbl 1239.81005

[Washington 1997] L. C. Washington, *Introduction to cyclotomic fields*, 2nd ed., Graduate Texts in Mathematics **83**, Springer, New York, 1997. MR 97h:11130 Zbl 0966.11047

[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

[Wohlfahrt 1964] K. Wohlfahrt, "An extension of F. Klein's level concept", *Illinois J. Math.* **8** (1964), 529–535. MR 29 #4805 Zbl 0135.29101

[Xu 2006] F. Xu, "Some computations in the cyclic permutations of completely rational nets", *Comm. Math. Phys.* **267**:3 (2006), 757–782. MR 2007h:81199 Zbl 1138.81510

[Zhang and Dong 2009] W. Zhang and C. Dong, "W-algebra $W(2,2)$ and the vertex operator algebra $L(1/2,0)\otimes L(1/2,0)$", *Comm. Math. Phys.* **285**:3 (2009), 991–1004. MR 2009k:17048 Zbl 1194.17015

[Zhu 1996] Y. Zhu, "Modular invariance of characters of vertex operator algebras", *J. Amer. Math. Soc.* **9**:1 (1996), 237–302. MR 96c:17042 Zbl 0854.17034

Communicated by Susan Montgomery
Received 2015-03-05 Revised 2015-07-20 Accepted 2015-08-19

dong@ucsc.edu — *Department of Mathematics, UC Santa Cruz, 194 Baskin Engineering, Santa Cruz, CA 95064, United States*

xingjunlin88@gmail.com — *Department of Mathematics, Sichuan University, Chengdu, 610064, China*

rng@math.lsu.edu — *Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803, United States*

ALGEBRA AND NUMBER THEORY 9:9 (2015)
dx.doi.org/10.2140/ant.2015.9.2167

An averaged form of Chowla's conjecture

Kaisa Matomäki, Maksym Radziwiłł and Terence Tao

Let λ denote the Liouville function. A well-known conjecture of Chowla asserts that, for any distinct natural numbers $h_1, \dots, h_k$, one has

$$\sum_{1 \le n \le X} \lambda(n+h_1) \cdots \lambda(n+h_k) = o(X)$$

as $X \to \infty$. This conjecture remains unproven for any $h_1, \dots, h_k$ with $k \ge 2$. Using the recent results of Matomäki and Radziwiłł on mean values of multiplicative functions in short intervals, combined with an argument of Kátai and Bourgain, Sarnak, and Ziegler, we establish an averaged version of this conjecture, namely

$$\sum_{h_1,\dots,h_k \le H} \left| \sum_{1 \le n \le X} \lambda(n+h_1) \cdots \lambda(n+h_k) \right| = o(H^k X)$$

as $X \to \infty$, whenever $H = H(X) \le X$ goes to infinity as $X \to \infty$ and k is fixed. Related to this, we give the exponential sum estimate

$$\int_0^X \left| \sum_{x \le n \le x+H} \lambda(n) e(\alpha n) \right| dx = o(HX)$$

as $X \to \infty$ uniformly for all $\alpha \in \mathbb{R}$, with H as before. Our arguments in fact give quantitative bounds on the decay rate (roughly on the order of $\log\log H / \log H$) and extend to more general bounded multiplicative functions than the Liouville function, yielding an averaged form of a (corrected) conjecture of Elliott.

1. Introduction

Let $\lambda : \mathbb{N} \to \{-1, +1\}$ be the Liouville function, that is to say, the completely multiplicative function such that $\lambda(p) = -1$ for all primes p. The prime number theorem implies that[1]

$$\sum_{1 \le n \le X} \lambda(n) = o(X)$$

MSC2010: 11P32.
Keywords: multiplicative functions, Hardy–Littlewood circle method, Chowla conjecture.

[1] See page 2174 for our asymptotic notation conventions.

as $X \to \infty$. More generally, a famous conjecture of Chowla [1965] asserts that, for any distinct natural numbers $h_1, \dots, h_k$, one has

$$\sum_{1 \le n \le X} \lambda(n+h_1) \cdots \lambda(n+h_k) = o(X) \tag{1-1}$$

as $X \to \infty$.

Chowla's conjecture remains open for any $h_1, \dots, h_k$ with $k \ge 2$. Our first main theorem establishes an averaged form of this conjecture:

Theorem 1.1 (Chowla's conjecture on average). *For any natural number k, and any $10 \le H \le X$, we have*

$$\sum_{1 \le h_1, \dots, h_k \le H} \left| \sum_{1 \le n \le X} \lambda(n+h_1) \cdots \lambda(n+h_k) \right| \ll k \left(\frac{\log \log H}{\log H} + \frac{1}{\log^{1/3000} X} \right) H^k X. \tag{1-2}$$

In fact, we have the slightly stronger bound

$$\sum_{1 \le h_2, \dots, h_k \le H} \left| \sum_{1 \le n \le X} \lambda(n) \lambda(n+h_2) \cdots \lambda(n+h_k) \right| \ll k \left(\frac{\log \log H}{\log H} + \frac{1}{\log^{1/3000} X} \right) H^{k-1} X. \tag{1-3}$$

In the case $k = 2$ our result implies that

$$\sum_{1 \le h \le H} \left| \sum_{1 \le n \le X} \lambda(n) \lambda(n+h) \right| = o(HX)$$

provided that $H \to \infty$ arbitrarily slowly with $X \to \infty$ (and $H \le X$). Note that the $k = 2$ case of Chowla's conjecture is equivalent to the above asymptotic holding in the case that H is bounded rather than going to infinity.

In fact, we have a more precise bound than (1-2) (or (1-3)) that gives more control on the exceptional tuples $(h_1, \dots, h_k)$ for which the sums of the form $\sum_{1 \le n \le X} \lambda(n+h_1) \cdots \lambda(n+h_k)$ are large; see Remark 5.2. In particular, in the special case $k = 2$ we get the following result.

Theorem 1.2. *Let $\delta \in (0, 1]$ be fixed. There is a large but fixed $H = H(\delta)$ such that, for all large enough X,*

$$\left| \sum_{1 \le n \le X} \lambda(n) \lambda(n+h) \right| \le \delta X \tag{1-4}$$

for all but at most $H^{1-\delta/5000}$ integers $|h| \le H$.

One can also replace the ranges $1 \le h_j \le H$ in Theorem 1.1 by $b_j+1 \le h_j \le b_j+H$ for any $b_j = O(X)$; see Theorem 1.6.

The exponents $1/3000$ and $1/5000$ in the above theorems may certainly be improved, but we did not attempt to optimize the constants here. However, our methods cannot produce a gain much larger than $1/\log H$, as one would then have to somehow control λ on numbers that are not divisible by any prime less than H, at which point we are no longer able to exploit the averaging in the $h_1, \dots, h_k$ parameters. It would be of particular interest to obtain a gain of more than $1/\log X$, as one could then potentially localize λ to primes and obtain some version of the prime tuples conjecture when the $h_1, \dots, h_k$ parameters are averaged over short intervals, but this is well beyond the capability of our methods. (If instead one is allowed to average the $h_1, \dots, h_k$ over long intervals of scale comparable to X, one can obtain various averaged forms of the prime tuples conjecture and its relatives, by rather different methods than those used here; see [Balog 1990; Mikawa 1992; Kawada 1993; 1995; Green and Tao 2010].)

Theorem 1.1 is closely related to the following averaged short exponential sum estimate, which may be of independent interest.

Theorem 1.3 (exponential sum estimate). *For any* $10 \le H \le X$, *one has*

$$\sup_{\alpha\in\mathbb{R}} \int_0^X \left| \sum_{x\le n\le x+H} \lambda(n) e(\alpha n) \right| dx \ll \left(\frac{\log\log H}{\log H} + \frac{1}{\log^{1/700} X} \right) HX.$$

Actually, for technical reasons it is convenient to prove a sharper version of Theorem 1.3 in which the Liouville function has been restricted to those numbers that have "typical" factorization; see Theorem 2.3. This sharper version will then be used to establish Theorem 1.1.

The relationship between Theorems 1.1 and 1.3 stems from the following Fourier-analytic identity:

Lemma 1.4 (Fourier identity). *For* $H > 0$, *if* $f : \mathbb{Z} \to \mathbb{C}$ *is a function supported on a finite set*, *then*

$$\int_{\mathbb{T}} \left(\int_{\mathbb{R}} \left| \sum_{x\le n\le x+H} f(n) e(\alpha n) \right|^2 dx \right)^2 d\alpha = \sum_{|h|\le H} (H-|h|)^2 \left| \sum_n f(n) \bar{f}(n+h) \right|^2.$$

Proof. Using the Fourier identity $\int_{\mathbb{T}} e(n\alpha)\, d\alpha = 1_{n=0}$, we can expand the left-hand side as

$$\sum_{n,n',m,m'} f(n)\bar{f}(n')f(m)\bar{f}(m') 1_{n+m-n'-m'=0} \\ \times \left(\int_{\mathbb{R}} 1_{x\le n,n'\le x+H}\, dx \right) \left(\int_{\mathbb{R}} 1_{y\le m,m'\le y+H}\, dy \right).$$

Writing $n' = n + h$, we see that both integrals are equal to $H - |h|$ if $|h| \leq H$ and vanish otherwise. The claim follows. $\square$

Theorem 1.3 may be compared with the classical estimate

$$\sup_{\alpha \in \mathbb{R}} \left| \sum_{1 \leq n \leq X} \lambda(n) e(\alpha n) \right| \ll_A X \log^{-A} X$$

of Davenport [1937], valid for any $A > 0$. Indeed, one can view Theorem 1.3 as asserting that a weak form of Davenport's estimate holds on average in short intervals. It would be of interest to also obtain nontrivial bounds on the larger quantity

$$\int_0^X \sup_{\alpha \in \mathbb{R}} \left| \sum_{x \leq n \leq x+H} \lambda(n) e(\alpha n) \right| dx \tag{1-5}$$

but this appears difficult to establish with our methods.

As with other applications of the circle method, our proof of Theorem 1.3 splits into two cases, depending on whether the quantity α is on "major arc" or on "minor arc". In the "major arc" case we are able to use the recent results of Matomäki and Radziwiłł [2015] on the average size of mean values of multiplicative functions on short intervals. Actually, in order to handle the presence of complex Dirichlet characters, we need to extend the results in [Matomäki and Radziwiłł 2015] to complex-valued multiplicative functions rather than real-valued ones; this is accomplished in an appendix to this paper (Appendix A). In the "minor arc" case we use a variant of the arguments of Kátai [1986] and Bourgain, Sarnak, and Ziegler [Bourgain et al. 2013] (see also the earlier works of Montgomery and Vaughan [1977] and Daboussi and Delange [1982]) to obtain the required cancellation. One innovation here is to rely on a combinatorial identity of Ramaré (also used in [Matomäki and Radziwiłł 2015]) as a substitute for the Turan–Kubilius inequality, as this leads to superior quantitative estimates (particularly if one first restricts the variable n to have a "typical" prime factorization).

Extension to more general multiplicative functions. Define a 1-*bounded multiplicative function* to be a multiplicative function $f : \mathbb{N} \to \mathbb{C}$ such that $|f(n)| \leq 1$ for all $n \in \mathbb{N}$. Given two 1-bounded multiplicative functions f, g and a parameter $X \geq 1$, we define the distance $\mathbb{D}(f, g; X) \in [0, +\infty)$ by the formula

$$\mathbb{D}(f, g; X) := \left(\sum_{p \leq X} \frac{1 - \operatorname{Re}(f(p)\overline{g(p)})}{p} \right)^{1/2}.$$

This is known to give a (pseudo)metric on 1-bounded multiplicative functions; see [Granville and Soundararajan 2007, Lemma 3.1]. We also define the asymptotic counterpart $\mathbb{D}(f, g; \infty) \in [0, +\infty]$ by the formula

$$\mathbb{D}(f, g; \infty) := \left(\sum_p \frac{1 - \mathrm{Re}(f(p)\overline{g(p)})}{p}\right)^{1/2}.$$

We informally say that f *pretends* to be g if $\mathbb{D}(f, g; X)$ (or $\mathbb{D}(f, g; \infty)$) is small (or finite).

For any 1-bounded multiplicative function g and real number $X > 1$, we introduce the quantity

$$M(g; X) := \inf_{|t| \leq X} \mathbb{D}(g, n \mapsto n^{it}; X)^2, \tag{1-6}$$

and then the more general quantity

$$M(g; X, Q) := \inf_{q \leq Q; \chi(q)} M(g\overline{\chi}; X) = \inf_{|t| \leq X; q \leq Q; \chi(q)} \mathbb{D}(g, n \mapsto \chi(n)n^{it}; X)^2,$$

where χ ranges over all Dirichlet characters of modulus $q \leq Q$. Informally, $M(g; X)$ is small when g pretends to be like a multiplicative character $n \mapsto n^{it}$, and $M(g; X, Q)$ is small when g pretends to be like a twisted Dirichlet character of modulus at most Q and twist of height at most X. We also define the asymptotic counterpart

$$M(g; \infty, \infty) = \inf_{\chi, t} \mathbb{D}(g, n \mapsto \chi(n)n^{it}; \infty)^2$$

where χ now ranges over all Dirichlet characters and t ranges over all real numbers.

Elliott proposed in [1992, Conjecture II] the following more general form of Chowla's conjecture, which we phrase here in contrapositive form.

Conjecture 1.5 (Elliott's conjecture). *Let $g_1, \ldots, g_k : \mathbb{N} \to \mathbb{C}$ be 1-bounded multiplicative functions, and let $a_1, \ldots, a_k,\ b_1, \ldots, b_k$ be natural numbers such that any two of the pairs $(a_1, b_1), \ldots, (a_k, b_k)$ are linearly independent in $\mathbb{Q}^2$. Suppose that there is an index $1 \leq j_0 \leq k$ such that*

$$M(g_{j_0}; \infty, \infty) = \infty. \tag{1-7}$$

Then

$$\sum_{1 \leq n \leq X} \prod_{j=1}^{k} g_j(a_j n + b_j) = o(X) \tag{1-8}$$

as $X \to \infty$.

Informally, this conjecture asserts that for pairwise linearly independent pairs $(a_1, b_1), \ldots, (a_k, b_k)$ and any 1-bounded multiplicative $g_1, \ldots, g_k$, one has the asymptotic (1-8) as $X \to \infty$, unless each of the g_j pretends to be a twisted Dirichlet character $n \mapsto \chi_j(n)n^{it_j}$. Note that some condition of this form is necessary, since if $g(n)$ is equal to $\chi(n)n^{it}$ then $g(n)\overline{g(n+h)}$ will be biased to be positive for large n, if h is fixed and divisible by the modulus q of χ; one also expects some bias when

h is not divisible by this modulus since the sums $\sum_{n\in\mathbb{Z}/q\mathbb{Z}} \chi(n)\overline{\chi(n+h)}$ do not vanish in general. From the prime number theorem in arithmetic progressions it follows that

$$M(\lambda; \infty, \infty) = \infty,$$

so Elliott's conjecture implies Chowla's conjecture (1-1).

When one allows the functions g_j to be complex-valued rather than real-valued, Elliott's conjecture turns out to be false on a technicality; one can choose 1-bounded multiplicative functions g_j which are arbitrarily close at various scales to a sequence of functions of the form $n \mapsto n^{it_m}$ (which allows one to violate (1-8)) without *globally* pretending to be n^{it} (or $\chi(n)n^{it}$) for any *fixed* t; we present this counterexample in Appendix B. However, this counterexample can be removed by replacing (1-7) with the stronger condition that

$$M(g_{j_0}; X, Q) \to \infty \tag{1-9}$$

as $X \to \infty$ for each fixed Q. In the real-valued case, (1-9) and (1-7) are equivalent by a triangle inequality argument of Granville and Soundararajan which we give in Appendix C.

As evidence for the corrected form of Conjecture 1.5 (in both the real-valued and complex-valued cases), we present the following averaged form of that conjecture:

Theorem 1.6 (Elliott's conjecture on average). *Let $10 \le H \le X$ and $A \ge 1$. Let $g_1, \dots, g_k : \mathbb{N} \to \mathbb{C}$ be 1-bounded functions, and let $a_1, \dots, a_k$, $b_1, \dots, b_k$ be natural numbers with $a_j \le A$ and $b_j \le AX$ for $j = 1, \dots, k$. Let $1 \le j_0 \le k$, and suppose that g_{j_0} is multiplicative. Then one has*

$$\sum_{1\le h_1,\dots,h_k\le H} \left| \sum_{1\le n\le X} \prod_{j=1}^{k} g_j(a_j n + b_j + h_j) \right| \ll A^2 k \left(\exp(-M/80) + \frac{\log\log H}{\log H} + \frac{1}{\log^{1/3000} X} \right) H^k X \tag{1-10}$$

where

$$M := M(g_{j_0}; 10AX, Q) \quad \text{and} \quad Q := \min(\log^{1/125} X, \log^{20} H).$$

In fact, we have the slightly stronger bound

$$\sum_{1\le h_2,\dots,h_k\le H} \left| \sum_{1\le n\le X} g_1(a_1 n + b_1) \prod_{j=2}^{k} g_j(a_j n + b_j + h_j) \right| \ll A^2 k \left(\exp(-M/80) + \frac{\log\log H}{\log H} + \frac{1}{\log^{1/3000} X} \right) H^{k-1} X. \tag{1-11}$$

Note that if $a_1, \dots, a_k,\ b_1, \dots, b_k$ are fixed, g_{j_0} is independent of X and obeys the condition (1-9) for any fixed Q, and $H = H(X)$ is chosen to go to infinity arbitrarily slowly as $X \to \infty$, then the quantity M in the above theorem goes to infinity (note that $M(g; X, Q)$ is nondecreasing in Q), and (1-11) then implies an averaged form of the asymptotic (1-8). Thus Theorem 1.6 is indeed an averaged form of the corrected form of Conjecture 1.5. (We discovered the counterexample in Appendix B while trying to interpret Theorem 1.6 as an averaged version of the original form of Conjecture 1.5.) Interestingly, only one of the functions $g_1, \dots, g_k$ in Theorem 1.6 is required to be multiplicative;[2] one can use a van der Corput argument to reduce matters to obtaining cancellation for a sum roughly of the form $\sum_{h \le H} \bigl| \sum_{1 \le n \le X} g_{j_0}(n) \overline{g_{j_0}(n+h)} \bigr|^2$, which can then be treated using Lemma 1.4.

For $g(n) = \lambda(n)$ and X, Q, M as in the above theorem, one obtains, for every $\varepsilon > 0$, the bound

$$\begin{aligned} M &\ge \inf_{|t| \le X; q \le Q; \chi(q)} \sum_{\exp((\log X)^{2/3+\varepsilon}) \le p \le X} \frac{1 + \operatorname{Re} \chi(p) p^{it}}{p} \\ &\ge \left(\frac{1}{3} - \varepsilon \right) \log \log X + O(1). \end{aligned} \tag{1-12}$$

The last inequality is established via standard methods from the Vinogradov–Korobov type zero-free region

$$\left\{ \sigma + it : \sigma > 1 - \frac{c}{\max\{\log q, (\log(3+|t|))^{2/3} (\log\log(3+|t|))^{1/3}\}} \right\}$$

for $L(s, \chi)$ and some absolute constant $c > 0$, which applies since χ has conductor $q \le (\log X)^{1/125}$ (so that there are no exceptional zeros); see [Montgomery 1994, §9.5]. Hence Theorem 1.6 implies Theorem 1.1. The same argument gives Theorem 1.1 when the Liouville function λ is replaced by the Möbius function μ. We remark that, as our arguments make no use of exceptional zeroes, all the implied constants in our theorems are effective.

We also have a generalized form of Theorem 1.3:

Theorem 1.7 (exponential sum estimate). *Let $X \ge H \ge 10$ and let g be a 1-bounded multiplicative function. Then*

$$\sup_{\alpha \in \mathbb{T}} \int_0^X \Bigl| \sum_{x \le n \le x+H} g(n) e(\alpha n) \Bigr| \, dx \ll \left(\exp(-M(g; X, Q)/20) + \frac{\log \log H}{\log H} + \frac{1}{\log^{1/700} X} \right) HX$$

[2] We thank the referee for observing this fact. In a previous version of this paper, all of the g_j were required to be multiplicative.

where

$$Q := \min(\log^{1/125} X, \log^5 H).$$

By (1-12), Theorem 1.7 implies Theorem 1.3.

Remark 1.8. In the recent preprint [Frantzikinakis and Host 2015], a different averaged form of Elliott's conjecture is established, in which one uses fewer averaging parameters h_i than in Theorem 1.6 (indeed, one can average over just a single such parameter, provided that the linear parts of the forms are independent), but the averaging parameters range over a long range (comparable to X) rather than on the short range given here. The methods of proof are rather different (in particular, the arguments in [Frantzikinakis and Host 2015] rely on higher order Fourier analysis). In the long-range averaged situation considered in [Frantzikinakis and Host 2015], the counterexample in Appendix B does not apply, and one can use the original form of Elliott's conjecture in place of the corrected version. It may be possible to combine the results here with those in [Frantzikinakis and Host 2015] to obtain an averaged version of Chowla's or Elliott's conjecture in which the number of averaging parameters is small, and the averaging is over a short range, but this seems to require nontrivial estimates on quantities such as (1-5), which we are currently unable to handle.

Notation. Our asymptotic notation conventions are as follows. We use $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the estimate $|X| \leq CY$ for some absolute constant C. If x is a parameter going to infinity, we use $X = o(Y)$ to denote the claim that $|X| \leq c(x)Y$ for some quantity $c(x)$ that goes to zero as $x \to \infty$ (holding all other parameters fixed).

Unless otherwise specified, all sums are over the integers, except for sums over the variable p (or p_1, p_2, etc.) which are understood to be over primes.

We use $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ to denote the standard unit circle and let $e : \mathbb{T} \to \mathbb{C}$ be the standard character $e(x) := e^{2\pi i x}$.

We use 1_S to denote the indicator of a predicate S, thus $1_S = 1$ when S is true and $1_S = 0$ when S is false. If A is a set, we write $1_{A(n)}$ for $1_{n \in A}$, so that 1_A is the indicator function of A.

2. Restricting to numbers with typical factorization

To prove Theorem 1.6 and Theorem 1.7 (and hence Theorem 1.1 and Theorem 1.3), it is technically convenient (as in the previous paper [Matomäki and Radziwiłł 2015]) to restrict the support of the multiplicative functions to a certain dense set $\mathcal{S}$ of natural numbers that have a "typical" prime factorization in a certain specific sense, in order to fully exploit a useful combinatorial identity of Ramaré (see (3-2)).

This will lead to improved quantitative estimates in the arguments in subsequent sections of the paper.

More precisely, we introduce the following sets $\mathcal{S}$ of numbers with typical prime factorization, which previously appeared in [Matomäki and Radziwiłł 2015].

Definition 2.1. Let $10 < P_1 < Q_1 \le X$ and $\sqrt{X} \le X_0 \le X$ be quantities such that $Q_1 \le \exp(\sqrt{\log X_0})$. We then define P_j, Q_j for $j > 1$ by the formulas

$$P_j := \exp(j^{4j}(\log Q_1)^{j-1}\log P_1), \quad Q_j := \exp(j^{4j+2}(\log Q_1)^j)$$

for $j > 1$. Note that the intervals $[P_j, Q_j]$ are disjoint and increase to infinity; indeed, one easily verifies that

$$P_1 < Q_1 < \exp(2^8 \log Q_1 \log P_1) = P_2$$

and

$$P_j < \exp(j^{4j}(\log Q_1)^j) < Q_j < \exp((j+1)^{4(j+1)}(\log Q_1)^j) < P_{j+1}$$

for all $j > 1$. Let J be the largest index such that $Q_J \le \exp(\sqrt{\log X_0})$. Then we define $\mathcal{S}_{P_1,Q_1,X_0,X}$ to be the set of all the numbers $1 \le n \le X$ which have at least one prime factor in the interval $[P_j, Q_j]$ for each $1 \le j \le J$.

In practice, X will be taken to be slightly smaller than X_0^2. The need to have two parameters X, X_0 instead of one is technical (we need to have the freedom later in the argument to replace X with a slightly smaller quantity X/d without altering J), but the reader may wish to pretend that $X_0 = \sqrt{X}$ for most of the argument.

This set is fairly dense if P_1 and Q_1 are widely separated:

Lemma 2.2. *Let* $10 < P_1 < Q_1 \le X$ *and* $\sqrt{X} \le X_0 \le X$ *be quantities such that* $Q_1 \le \exp(\sqrt{\log X_0})$. *Then, for every large enough* X,

$$\#\{1 \le n \le X : n \notin \mathcal{S}_{P_1,Q_1,X_0,X}\} \ll \frac{\log P_1}{\log Q_1} \cdot X.$$

Proof. From the fundamental lemma of sieve theory (see, e.g., [Friedlander and Iwaniec 2010, Theorem 6.17]) we know that, for any $1 \le j \le J$ and large enough X, the number of $1 \le n \le X$ that are not divisible by any prime in $[P_j, Q_j]$ is at most

$$\ll X \prod_{P_j \le p \le Q_j}\left(1 - \frac{1}{p}\right) \ll \frac{\log P_j}{\log Q_j} X = \frac{1}{j^2}\frac{\log P_1}{\log Q_1} X.$$

Summing over j, we obtain the claim. □

Both Theorems 1.6 and 1.7 will be deduced from the following claim.

Theorem 2.3 (key exponential sum estimate). *Let* $X, H, W \ge 10$ *be such that*

$$(\log H)^5 \le W \le \min\{H^{1/250}, (\log X)^{1/125}\}$$

and let g be a 1*-bounded multiplicative function such that*

$$W \le \exp(M(g; X, Q)/3). \tag{2-1}$$

Set

$$\mathcal{S} := \mathcal{S}_{P_1, Q_1, \sqrt{X}, X} \quad \text{where } P_1 := W^{200},\ Q_1 := H/W^3.$$

Then, for any $\alpha \in \mathbb{T}$, one has

$$\int_{\mathbb{R}} \Bigl| \sum_{x \le n \le x+H} 1_{\mathcal{S}}(n) g(n) e(\alpha n) \Bigr| \, dx \ll \frac{(\log H)^{1/4} \log\log H}{W^{1/4}} HX. \tag{2-2}$$

In Section 5 we will show how this theorem implies Theorem 1.6. For now, let us at least see how it implies Theorem 1.7:

Proof of Theorem 1.7 assuming Theorem 2.3. We may assume that X, H, and $M(g; X, Q)$ are larger than any specified absolute constant, since if one of these expressions is bounded, then so is W. The claim (2-2) is then trivial with a suitable choice of implied constant (discarding the $(\log H)^{1/4} \log\log H$ factor).

Choose H_0 such that

$$\log H_0 := \min(\log^{1/700} X \log\log X, \exp(M(g; X, Q)/20) M(g; X, Q)).$$

We divide into two cases: $H \le H_0$ and $H > H_0$.

First suppose that $H \le H_0$. Then if we set $W := \log^5 H$, one verifies that all the hypotheses of Theorem 2.3 hold, and hence

$$\int_0^X \Bigl| \sum_{x \le n \le x+H} 1_{\mathcal{S}}(n) g(n) e(\alpha n) \Bigr| \, dx \ll \frac{\log\log H}{\log H} HX.$$

On the other hand, from Lemma 2.2, the choice of W, P_1, Q_1, and the bound on H, we see that

$$\#\{1 \le n \le X + H : n \notin \mathcal{S}\} \ll \frac{\log\log H}{\log H} X$$

and thus, by Fubini's theorem and the triangle inequality,

$$\int_0^X \Bigl| \sum_{x \le n \le x+H} (1 - 1_{\mathcal{S}}(n)) g(n) e(\alpha n) \Bigr| \, dx \ll \frac{\log\log H}{\log H} HX.$$

Summing, we obtain Theorem 1.7 in this case.

Now suppose that $H > H_0$. Covering $[0, H]$ by $O(H/H_0)$ intervals of length H_0, we see that

$$\int_0^X \Bigl| \sum_{x \le n \le x+H} g(n) e(\alpha n) \Bigr| \, dx \ll \frac{H}{H_0} \int_0^{X+H} \Bigl| \sum_{x \le n \le x+H_0} g(n) e(\alpha n) \Bigr| \, dx.$$

Also, observe from the choice of H_0 that the quantity

$$\exp(-M(g; X, Q)/20) + \frac{\log\log H}{\log H} + \frac{1}{\log^{1/700} X}$$

is unchanged up to multiplicative constants if one reduces H to H_0. Finally, from Mertens' theorem we see that $M(g; X+H, Q) = M(g; X, Q) + O(1)$. The claim then follows from the $H = H_0$ case (after performing the minor alteration of replacing X with $X + H$). □

We now begin the proof of Theorem 2.3. The first step is to reduce to the case where g is completely multiplicative rather than multiplicative. More precisely, we will deduce Theorem 2.3 from the following proposition.

Proposition 2.4 (completely multiplicative exponential sum estimate). *Assume* $X, H, W \geq 10$ *are such that*

$$(\log H)^5 \leq W \leq \min\{H^{1/250}, (\log X)^{1/125}\},$$

and let g *be a* 1*-bounded completely multiplicative function such that*

$$W \leq \exp(M(g; X, W)/3). \tag{2-3}$$

Let d *be a natural number with* $d < W$. *Set*

$$\mathcal{S} := \mathcal{S}_{P_1, Q_1, \sqrt{X}, X/d} \quad \textit{where } P_1 := W^{200},\ Q_1 := H/W^3.$$

Then for any $\alpha \in \mathbb{T}$ *one has*

$$\int_{\mathbb{R}} \Bigl| \sum_{x/d \leq n \leq x/d + H/d} 1_{\mathcal{S}}(n) g(n) e(\alpha n) \Bigr| \, dx \ll \frac{1}{d^{3/4}} \frac{(\log H)^{1/4} \log\log H}{W^{1/4}} HX. \tag{2-4}$$

Let us explain why Theorem 2.3 follows from Proposition 2.4. Let the hypotheses and notation be as in Theorem 2.3. The function g is not necessarily completely multiplicative, but we may approximate it by the 1-bounded completely multiplicative function $g_1 : \mathbb{N} \to \mathbb{C}$, defined as the completely multiplicative function with $g_1(p) = g(p)$ for all primes p. By Möbius inversion we may then write $g = g_1 * h$ where $*$ denotes Dirichlet convolution and h is the multiplicative function $h = g * \mu g_1$. Observe that, for all primes p, we have $h(p) = 0$ and $|h(p^j)| \leq 2$ for $j \geq 2$. We now write

$$\sum_{x \leq n \leq x+H} 1_{\mathcal{S}_{P_1, Q_1, \sqrt{X}, X}} g(n) e(\alpha n) = \sum_{d=1}^{\infty} h(d) \sum_{x/d \leq m \leq x/d + H/d} 1_{\mathcal{S}_{P_1, Q_1, \sqrt{X}, X}}(dm) g_1(m) e(d\alpha m)$$

and so by the triangle inequality we may upper bound the left-hand side of (2-2) by

$$\sum_{d=1}^{\infty} |h(d)| \int_{\mathbb{R}} \left| \sum_{x/d \le m \le x/d+H/d} 1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X}}(dm) g_1(m) e(d\alpha m) \right| dx.$$

Let us first dispose of the contribution where $d \ge W$. Here we trivially bound this contribution by

$$\sum_{d \ge W} |h(d)| \sum_{m \le (2X+H)/d} O(H)$$

(after moving the absolute values inside the m summation and then performing the integration on x first). We can bound this in turn by

$$\ll HX \frac{1}{W^{1/4}} \sum_{d=1}^{\infty} \frac{|h(d)|}{d^{3/4}}.$$

From Euler products we see that $\sum_{d=1}^{\infty} |h(d)|/d^{3/4} = O(1)$, so the contribution of this case is acceptable.

Now we consider the contribution $d < W < P_1$. In this case we may reduce

$$1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X}}(dm) = 1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X/d}}(m)$$

and so this contribution to (2-2) can be upper bounded by

$$\sum_{1 \le d < W} |h(d)| \int_{\mathbb{R}} \left| \sum_{x/d \le m \le x/d+H/d} 1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X/d}}(m) g_1(m) e(d\alpha m) \right| dx.$$

By Proposition 2.4, this is bounded by

$$\sum_{d=1}^{\infty} \frac{|h(d)|}{d^{3/4}} \frac{(\log H)^{1/4} \log\log H}{W^{1/4}} HX.$$

As before, we have $\sum_{d=1}^{\infty} |h(d)|/d^{3/4} = O(1)$, and Theorem 2.3 follows.

It remains to prove Proposition 2.4. For any $\alpha \in \mathbb{T}$, we know from the Dirichlet approximation theorem that there exists a rational number a/q with $(a, q) = 1$ and $1 \le q \le H/W$ such that

$$\left| \alpha - \frac{a}{q} \right| \le \frac{W}{qH} \le \frac{1}{q^2}.$$

In the next two sections, we will apply separate arguments to prove Proposition 2.4 in the *minor arc* case $q > W$ and the *major arc* case $q \le W$.

3. Proof of minor arc estimate

We now prove Proposition 2.4 in the minor arc case $q > W$. It suffices to show that

$$\int_{\mathbb{R}} \theta(x) \sum_{x/d \le n \le x/d + H/d} 1_{\mathcal{S}}(n) g(n) e(\alpha n)\, dx \ll \frac{1}{d^{3/4}} \frac{(\log H)^{1/4} \log\log H}{W^{1/4}} HX \tag{3-1}$$

whenever $\theta : \mathbb{R} \to \mathbb{C}$ is measurable, with $|\theta(x)|$ at most 1 for all x and supported on $[0, X]$. We will now use a variant of an idea of Bourgain, Sarnak, and Ziegler [Bourgain et al. 2013] (building on earlier works of Kátai [1986], Montgomery and Vaughan [1977] and Daboussi and Delange [1982]).

Let $\mathcal{P}$ be the set consisting of the primes lying between P_1 and Q_1. Then notice that each $n \in \mathcal{S}$ has at least one prime factor from $\mathcal{P}$. This leads to the following variant of Ramaré's identity (see [Friedlander and Iwaniec 2010, Section 17.3]):

$$1_{\mathcal{S}}(n) = \sum_{p \in \mathcal{P}, m : mp = n} \frac{1_{\mathcal{S}'}(mp)}{1 + \#\{q | m : q \in \mathcal{P}\}}, \tag{3-2}$$

where $\mathcal{S}'$ is the set of all $1 \le n \le X/d$ that have at least one prime factor in each of the intervals $[P_j, Q_j]$ for $j \ge 2$; the constraint $n \le X/d$ arises from the corresponding constraint in the definition of $\mathcal{S}$.

Using this identity, we may write the left-hand side of (3-1) as

$$\sum_{p \in \mathcal{P}} \sum_m \frac{1_{\mathcal{S}'}(mp) g(mp) e(mp\alpha)}{1 + \#\{q | m : q \in \mathcal{P}\}} \int_{\mathbb{R}} \theta(x) 1_{x/d \le mp \le (x+H)/d}\, dx.$$

As g is completely multiplicative, $g(mp) = g(m) g(p)$. Thus it suffices to show that

$$\sum_{p \in \mathcal{P}} \sum_m \frac{1_{\mathcal{S}'}(mp) g(m) g(p) e(mp\alpha)}{1 + \#\{q | m : q \in \mathcal{P}\}} \int_{\mathbb{R}} \theta(x) 1_{x/d \le mp \le (x+H)/d}\, dx \ll \frac{(\log H)^{1/4} \log\log H}{d^{3/4} W^{1/4}} HX.$$

We can cover $\mathcal{P}$ by intervals $[P, 2P]$ with $P_1 \ll P \ll Q_1$ and P a power of two, and observe that

$$\sum_{\substack{P_1 \ll P \ll Q_1 \\ P = 2^j}} \frac{1}{\log P} \ll \log\log Q_1 - \log\log P_1 \ll \log\log H,$$

so by the triangle inequality it suffices to show that

$$\sum_{\substack{p \in \mathcal{P} \\ P \le p \le 2P}} \sum_m \frac{1_{\mathcal{S}'}(mp) g(m) g(p) e(mp\alpha)}{1 + \#\{q | m : q \in \mathcal{P}\}} \int_{\mathbb{R}} \theta(x) 1_{x/d \le mp \le (x+H)/d}\, dx \ll \frac{(\log H)^{1/4}}{d^{3/4} W^{1/4} \log P} HX$$

for each such P. We can rearrange the left-hand side as

$$\sum_{m\in\mathcal{S}'}\frac{g(m)}{1+\#\{q|m:q\in\mathcal{P}\}}\sum_{\substack{p\in\mathcal{P}\\ P\le p\le 2P}}1_{mp\le X/d}\,g(p)e(mp\alpha)\int_{\mathbb{R}}\theta(x)1_{x/d\le mp\le(x+H)/d}\,dx.$$

Observe that the summand vanishes unless we have $m\le X/dP$. Crudely bounding[3] $g(m)/(1+\#\{q|m:q\in\mathcal{P}\})$ in magnitude by 1 and applying Hölder's inequality, we may bound the previous expression in magnitude by

$$\left(\frac{X}{dP}\right)^{3/4}\left(\sum_{m\le X/dP}\left|\sum_{\substack{p\in\mathcal{P}\\ P\le p\le 2P}}1_{mp\le X/d}\,g(p)e(mp\alpha)\int_{\mathbb{R}}\theta(x)1_{x/d\le mp\le(x+H)/d}\right|^4dx\right)^{1/4}.$$

It thus suffices to show that

$$\sum_{m\le X/dP}\left|\sum_{\substack{p\in\mathcal{P}\\ P\le p\le 2P}}1_{mp\le X/d}\,g(p)e(mp\alpha)\int_{\mathbb{R}}\theta(x)1_{x/d\le mp\le(x+H)/d}\,dx\right|^4 \ll \frac{\log H}{W\log^4 P}H^4XP^3.$$

The left-hand side may be expanded as

$$\sum_{\substack{p_1,p_2,p_3,p_4\in\mathcal{P}\\ P\le p_1,p_2,p_3,p_4\le 2P}}\int\cdots\int g(p_1)g(p_2)\overline{g(p_3)}\,\overline{g(p_4)}\theta(x_1)\theta(x_2)\overline{\theta(x_3)}\,\overline{\theta(x_4)}$$
$$\times\sum_{\substack{m\le X/(dp_i)\\ x_i/(dp_i)\le m\le(x_i+H)/(dp_i)\\ \forall i=1,2,3,4}}e(m(p_1+p_2-p_3-p_4)\alpha)\,dx_1\,dx_2\,dx_3\,dx_4.$$

From summing the geometric series, we observe that the summation over m is $O(\min(H/P,1/\|(p_1+p_2-p_3-p_4)\alpha\|))$, where $\|z\|$ denotes the distance from z to the nearest integer. Also, the sum vanishes unless we have $x_1=O(X)$ and $x_i=x_1p_i/p_1+O(H)$ for $i=2,3,4$, so there are only $O(XH^3)$ quadruples (x_1,x_2,x_3,x_4) which contribute. Thus we may bound the previous expression by

$$O\left(XH^3\sum_{p_1,p_2,p_3,p_4\le 2P}\min\left(\frac{H}{P},\frac{1}{\|(p_1+p_2-p_3-p_4)\alpha\|}\right)\right)$$

and so we reduce to showing that

$$\sum_{p_1,p_2,p_3,p_4\le 2P}\min\left(\frac{H}{P},\frac{1}{\|(p_1+p_2-p_3-p_4)\alpha\|}\right)\ll\log H\frac{HP^3}{W\log^4 P}. \tag{3-3}$$

[3]By using the Turan–Kubilius inequality here one could save a factor of $\log\log H$, but such a gain will not make a significant impact on our final estimates.

The quantity $p_1+p_2-p_3-p_4$ is clearly of size $O(P)$. Conversely, from a standard upper bound sieve,[4] the number of representations of an integer $n=O(P)$ of the form $p_1+p_2-p_3-p_4$ with $p_1, p_2, p_3, p_4 \le 2P$ prime is $O(P^3/\log^4 P)$. Thus it suffices to show that

$$\sum_{n=O(P)} \min\left(\frac{H}{P}, \frac{1}{\|n\alpha\|}\right) \ll \frac{\log H}{W} H.$$

But from the Vinogradov lemma (see, e.g., [Iwaniec and Kowalski 2004, page 346]), the left-hand side is bounded by

$$O\left(\left(\frac{P}{q}+1\right)\left(\frac{H}{P}+q\log q\right)\right) \ll \frac{H}{q}+P\log q+\frac{H}{P}+q\log q$$

which, since

$$W^{200}=P_1 \ll P \ll Q_1 = H/W^3 \quad \text{and} \quad W \le q \le H/W,$$

is bounded by $O(\log H/WH)$ as required. □

4. Proof of major arc estimate

We now prove Proposition 2.4 in the major arc case $q \le W$. We will discard the factor $d^{1/4}(\log H)^{1/4} \log\log H$ and prove the stronger bound

$$\int_{\mathbb{R}} \left| \sum_{x/d \le n \le (x+H)/d} 1_{\mathcal{S}}(n) g(n) e(\alpha n) \right| dx \ll \frac{HX}{dW^{1/4}}. \tag{4-1}$$

By hypothesis we have $\alpha = a/q+\theta$ with $q \le W$ and $\theta = O(W/(Hq))$. Integrating by parts we see that

$$\begin{aligned}\left| \sum_{x/d \le n \le (x+H)/d} 1_{\mathcal{S}}(n) g(n) e(\alpha n) \right| &\ll \left| \sum_{x/d \le n \le (x+H)/d} 1_{\mathcal{S}}(n) g(n) e(an/q) \right| \\ &\quad + \frac{W}{Hq} \int_0^{H/d} \left| \sum_{x/d \le n \le x/d+H'} 1_{\mathcal{S}}(n) g(n) e(an/q) \right| dH'. \end{aligned} \tag{4-2}$$

[4] For instance, from [Montgomery and Vaughan 2007, Theorem 3.13] one sees that any number $N=O(P)$ has $O((N/\phi(N))(P/\log^2 P))$ representations as the sum of two primes; since $\sum_{N=O(P)} N^2/\phi(N)^2 = O(P)$ (see, e.g., [Montgomery and Vaughan 2007, Exercise 2.1.14]), the claim then follows from the Cauchy–Schwarz inequality.

Thus let us focus on bounding

$$\int_{\mathbb{R}} \Bigl| \sum_{x/d \le n \le x/d+H'} 1_{\mathcal{S}}(n) g(n) e(an/q) \Bigr| \, dx \tag{4-3}$$

with $0 \le H' \le H/d$. Splitting into residues classes we see that (4-3) is

$$\le \sum_{b \pmod q} \int_{\mathbb{R}} \Bigl| \sum_{\substack{x/d \le n \le x/d+H' \\ n \equiv b \pmod q}} 1_{\mathcal{S}}(n) g(n) \Bigr| \, dx.$$

For $n \equiv b \pmod q$ we have $d_0 := (b, q) | n$. Therefore let us write $b = d_0 b_0$, $q = d_0 q_0$ and $n = d_0 m$, so that the condition $n \equiv b \pmod q$ simplifies to $m \equiv b_0 \pmod{q_0}$. In addition, since g is completely multiplicative and since $d_0 \le q \le W \le P_1$, we have

$$1_{\mathcal{S}}(n) g(n) = g(d_0) \cdot 1_{\mathcal{S}_{P_1, Q_1, \sqrt{X}, X/(dd_0)}}(m) g(m).$$

Finally we express $m \equiv b_0 \pmod{q_0}$ in terms of Dirichlet characters noting that

$$1_{m \equiv b_0 \pmod{q_0}}(m) = \frac{1}{\varphi(q_0)} \sum_{\chi \pmod{q_0}} \chi(b_0) \overline{\chi(m)}.$$

Putting everything together we see that (4-3) is less than

$$\sum_{b \pmod q} \frac{1}{\varphi(q_0)} \sum_{\chi \pmod{q_0}} \int_{\mathbb{R}} \Bigl| \sum_{x/(dd_0) \le m \le x/(dd_0) + H'/d_0} 1_{\mathcal{S}_{P_1, Q_1, \sqrt{X}, X/(dd_0)}} g(m) \overline{\chi(m)} \Bigr| \, dx.$$

In the integral we make the linear change of variable $y = x/(dd_0)$, so that the above expression becomes

$$d \sum_{b \pmod q} \frac{d_0}{\varphi(q_0)} \sum_{\chi \pmod{q_0}} \int_{\mathbb{R}} \Bigl| \sum_{y \le m \le y + H'/d_0} 1_{\mathcal{S}_{P_1, Q_1, \sqrt{X}, X/(dd_0)}} g(m) \overline{\chi(m)} \Bigr| \, dy. \tag{4-4}$$

We bound the part of the integral with $y \le X/W^{10}$ trivially. This produces in (4-3) an error which is

$$\ll dq \cdot \frac{X}{W^{10}} \cdot H' \le \frac{HX}{W^9} \ll \frac{HX}{dW^3}$$

since $q, d \le W$ and $H' \le H/d$. We split the remaining range $X/W^{10} \le y \le 2X/(dd_0)$ into dyadic blocks $X/W^{10} \le X' \le X/(dd_0)$ with X' running through powers of two. Thus the previous expression is

$$\ll d \sum_{X'} \sum_{b \pmod q} \frac{d_0}{\varphi(q_0)} \sum_{\chi \pmod{q_0}} \int_{X'}^{2X'} \Bigl| \sum_{y \le m \le y + H'/d_0} 1_{\mathcal{S}_{P_1, Q_1, \sqrt{X}, X/(dd_0)}} g(m) \overline{\chi(m)} \Bigr| \, dy + \frac{HX}{dW^3}.$$

At this point we apply Theorem A.2 with $\eta = 1/20$ (note that $P_1 \geq (\log Q_1)^{40/\eta}$) to conclude that

$$\int_{X'}^{2X'} \Bigg| \sum_{y \leq m \leq y+H'/d_0} 1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X/(dd_0)}}(m) g(m) \overline{\chi}(m) \Bigg|^2 dy$$
$$\ll \left(\exp(-M(g\overline{\chi}; X')) M(g\overline{\chi}; X') + \frac{(\log H'/d_0)^{1/3}}{P_1^{1/6-1/20}} + \frac{1}{(\log X')^{1/50}} \right) \frac{H'^2}{d_0^2} X'.$$

Since $P_1 = W^{200}$ and $H'/d_0 \leq H$ and $W \geq \log^5 H$, we have

$$\frac{(\log H'/d_0)^{1/3}}{P_1^{1/6-1/20}} \leq \frac{(\log H)^{1/3}}{P_1^{1/6-1/20}} \ll \frac{1}{W^{5/2}}$$

and certainly

$$\frac{1}{(\log X')^{1/50}} \ll \frac{1}{(\log X)^{1/50}} \ll \frac{1}{W^{5/2}}.$$

From Mertens' theorem and definition of $M(g, X, W)$,

$$M(g\overline{\chi}; X') \geq M(g\overline{\chi}; X) - O(1) \geq M(g, X, W) - O(1)$$

and thus, by (2-3),

$$\exp(-M(g\overline{\chi}; X')) M(g\overline{\chi}; X') \ll \frac{1}{W^{5/2}}.$$

Putting all this together, we obtain

$$\int_{X'}^{2X'} \Bigg| \sum_{y \leq m \leq y+H'/d_0} 1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X/(dd_0)}}(m) g(m) \overline{\chi}(m) \Bigg|^2 dy \ll \frac{1}{W^{5/2}} \frac{H'^2}{d_0^2} X'.$$

It follows from Cauchy–Schwarz that

$$\int_{X'}^{2X'} \Bigg| \sum_{y \leq m \leq y+H'/d_0} 1_{\mathcal{S}_{P_1,Q_1,\sqrt{X},X/(dd_0)}}(m) g(m) \overline{\chi(m)} \Bigg| dy \ll W^{-5/4} \cdot \frac{H'X'}{d_0}.$$

Inserting this bound into (4-4) we see that (4-3) is bounded by

$$\ll dq \cdot \frac{1}{W^{5/4}} \cdot \frac{H}{d} \cdot \frac{X}{d} \ll \frac{qHX}{dW^{5/4}}.$$

Therefore using (4-2) and using $q \leq W$ we see that (4-1) is

$$\ll \frac{qHX}{dW^{5/4}} \cdot \left(1 + \frac{W}{Hq} \cdot \frac{H}{d}\right) \ll \frac{HX}{dW^{1/4}}. \qquad \square$$

5. Elliott's conjecture on the average

In this section we use Theorem 2.3 to prove Theorem 1.6, which will be deduced from the following result (compare also with Theorem 2.3 and deduction of Theorem 1.7 from it). For brevity, we write $1_{\mathcal{S}}g$ for the function $n \mapsto 1_{\mathcal{S}}(n)g(n)$.

Proposition 5.1 (truncated Elliott on the average). *Let $X, H, W, A \geq 10$ be such that*

$$\log^{20} H \leq W \leq \min\{H^{1/500}, (\log X)^{1/125}\}.$$

Let $g_1, \ldots, g_k : \mathbb{N} \to \mathbb{C}$ be 1-bounded multiplicative functions, and let $a_1, \ldots, a_k$, $b_1, \ldots, b_k$ be natural numbers with $a_j \leq A$ and $b_j \leq 3AX$ for $j = 1, \ldots, k$. Let $1 \leq j_0 \leq k$ be such that

$$W \leq \exp(M(g_{j_0}; 10AX, Q)/3).$$

Set

$$\mathcal{S} = \mathcal{S}_{P_1, Q_1, \sqrt{10AX}, 10AX} \quad \textit{where } P_1 := W^{200},\ Q_1 := H^{1/2}/W^3.$$

Then

$$\sum_{1 \leq h_2, \ldots, h_k \leq H} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \ll \frac{kA^2}{W^{1/20}} H^{k-1} X. \quad (5\text{-}1)$$

Proof of Theorem 1.6 assuming Proposition 5.1. We may assume that X, H, and M are larger than any specified absolute constant, as the claim is trivial otherwise. We first make some initial reductions. The first estimate (1-10) of Theorem 1.6 follows from the second (1-11) after shifting b_1 by h_1 in (1-11) and averaging, provided that we relax the hypotheses $b_j \leq AX$ slightly to $b_j \leq 2AX$. Thus it suffices to prove (1-11) under the relaxed hypotheses $b_j \leq 2AX$.

Let H_0 be such that

$$\log H_0 = \min\{\log^{1/3000} X \log\log X, \exp(M(g_{j_0}; 10AX, Q)/80) M(g_{j_0}; 10AX, Q)\}. \quad (5\text{-}2)$$

If $H \leq H_0$ we take $W = \log^{20} H$ and let $\mathcal{S}$ be as in Proposition 5.1. All the assumptions of Proposition 5.1 hold and thus

$$\sum_{1 \leq h_2, \ldots, h_k \leq H} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \ll \frac{kA^2}{\log H} H^{k-1} X.$$

Furthermore, from Lemma 2.2 we have

$$\sum_{\substack{n \leq 10AX \\ n \notin \mathcal{S}}} 1 \ll AX \frac{\log W}{\log H}. \quad (5\text{-}3)$$

From this and the triangle inequality, we have

$$\sum_{1\le n\le X} g_1(a_1n+b_1)\prod_{j=2}^{k} g_j(a_jn+b_j+h_j)$$
$$= \sum_{1\le n\le X} 1_{\mathcal{S}}g_1(a_1n+b_1)\prod_{j=2}^{k} 1_{\mathcal{S}}g_j(a_jn+b_j+h_j) + O\left(kAX\frac{\log W}{\log H}\right). \tag{5-4}$$

Hence the claim follows in the case when $H \le H_0$.

If $H > H_0$, one can cover the summation over the h_j indices by intervals of length H_0 and apply Theorem 1.6 to each subinterval (shifting the b_j by at most AX when doing so), and then sum, noting that the quantity

$$\exp(-M(g_{j_0};\, 10AX,\, Q)/80) + \frac{\log\log H}{\log H} + \frac{1}{\log^{1/3000} X}$$

is essentially unchanged after replacing H with H_0. □

Remark 5.2. By using larger choices of W, one can obtain more refined information on the large values of the correlations $\sum_{1\le n\le X} g_1(a_1n+b_1)\prod_{j=2}^{k} g_j(a_jn+b_j+h_j)$. For instance, if we take $W = H^{\delta}$ for some H, δ such that $10 \le H \le H_0$ and $20\log\log H/\log H \le \delta \le 1/500$, we see from Proposition 5.1, (5-4), and Markov's inequality that

$$\sum_{1\le n\le X} g_1(a_1n+b_1)\prod_{j=2}^{k} g_j(a_jn+b_j+h_j) \ll kA^2\delta X$$

for all but at most $O(H^{k-1}/\delta H^{\delta/20})$ tuples $(h_1, \dots, h_{k-1})$ with $1 \le h_j \le H$ for $j = 2, \dots, k$. Thus we can obtain a power saving in the number of exceptional tuples, at the cost of only obtaining a weak bound on the individual correlations $\sum_{1\le n\le X} g_1(a_1n+b_1)\prod_{j=2}^{k} g_j(a_jn+b_j+h_j)$.

It remains to prove Proposition 5.1. We start by proving the following simpler case to which the general case will be reduced.

Proposition 5.3. *Let $X, H, W \ge 10$ be such that*

$$\log^{20} H \le W \le \min\{H^{1/250}, (\log X)^{1/125}\}.$$

Let $g : \mathbb{N} \to \mathbb{C}$ be a 1-bounded multiplicative function such that

$$W \le \exp(M(g;\, X,\, W)/3).$$

Set

$$\mathcal{S} = \mathcal{S}_{P_1,Q_1,\sqrt{X},X} \quad \text{where } P_1 := W^{200},\ Q_1 := H/W^3.$$

Then

$$\sum_{1\le h\le H}\Big|\sum_{1\le n\le X} 1_{\mathcal{S}}g(n)1_{\mathcal{S}}\bar{g}(n+h)\Big|^2 \ll \frac{HX^2}{W^{1/5}}. \tag{5-5}$$

To deduce Theorem 1.2 we let $\mathcal{S}$ be as in this proposition with $W := H^{\delta/900}$. The argument of Lemma 2.2 actually gives $\#\{1\le n\le X : n\notin\mathcal{S}\} \le 2X\log P_1/\log Q_1$ in this case, and thus the numbers n with $n\notin\mathcal{S}$ or $n+h\notin\mathcal{S}$ contribute to the left-hand side of (1-4) at most $9\delta/10$. Hence, recalling (1-12), the claim follows from the previous proposition and Markov's inequality.

Proof of Proposition 5.3. The claim follows once we have shown

$$\sum_{|h|\le 2H}(2H-|h|)^2\cdot\Big|\sum_n 1_{\mathcal{S}}g(n)1_{\mathcal{S}}\bar{g}(n+h)\Big|^2 \ll \frac{1}{W^{1/5}}H^3X^2.$$

Applying Lemma 1.4, it will suffice to show that

$$\int_{\mathbb{T}}\Big(\int_{\mathbb{R}}\Big|\sum_{x\le n\le x+2H}1_{\mathcal{S}}g(n)e(\alpha n)\Big|^2dx\Big)^2d\alpha \ll \frac{1}{W^{1/5}}H^3X^2.$$

From the Parseval identity we have

$$\int_{\mathbb{T}}\int_{\mathbb{R}}\Big|\sum_{x\le n\le x+2H}1_{\mathcal{S}}g(n)e(\alpha n)\Big|^2dx\,d\alpha = \int_{\mathbb{R}}\sum_{x\le n\le x+2H}|1_{\mathcal{S}}g(n)|^2\,dx \ll HX$$

so it suffices to show that

$$\sup_\alpha\int_{\mathbb{R}}\Big|\sum_{x\le n\le x+2H}1_{\mathcal{S}}g(n)e(\alpha n)\Big|^2dx \ll \frac{1}{W^{1/5}}H^2X.$$

Using the trivial bound

$$\Big|\sum_{x\le n\le x+2H}1_{\mathcal{S}}g(n)e(\alpha n)\Big| \ll H$$

we thus reduce to showing

$$\sup_\alpha\int_{\mathbb{R}}\Big|\sum_{x\le n\le x+2H}1_{\mathcal{S}}g(n)e(\alpha n)\Big|\,dx \ll \frac{HX}{W^{1/5}}. \tag{5-6}$$

This follows from Theorem 2.3 (using the lower bound $W\ge\log^{20}H$ in the hypotheses of Proposition 5.3 to absorb the $\log^{1/4}H\log\log H$ factors in Theorem 2.3). $\square$

Proof of Proposition 5.1. We first remove the special treatment afforded to the g_1 factor in (5-1). Note that we may assume

$$W^{1/20}\ge kA^2 \tag{5-7}$$

and thus

$$H \geq W^{500} \geq (kA^2)^{10000}$$

since the claim is trivial otherwise.

Set $H' := \sqrt{H}$. For any $1 \leq h_1 \leq H'/A$, we may shift n by h_1 and conclude that

$$\sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \\ = \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1 + a_1 h_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j + a_j h_1) + O(H')$$

and thus we may write the left-hand side of (5-1) as

$$\sum_{1 \leq h_2, \dots, h_k \leq H} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1 + a_1 h_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j + a_j h_1) \right| \\ + O(H^{k-1} H').$$

If one shifts each of the h_j for $j = 2, \dots, k$ in turn by $a_j h_1 = O(H')$, we may rewrite this as

$$\sum_{1 \leq h_2, \dots, h_k \leq H} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1 + a_1 h_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \\ + O(H^{k-1} H') + O(k H^{k-2} H' X).$$

Averaging in h_1, and replacing h_1 by $a_1 h_1$ (crudely dropping the constraint that $a_1 h_1$ is divisible by a_1), we may thus bound the left-hand side of (5-1) by

$$\ll \frac{A}{H'} \sum_{1 \leq h_1 \leq H'} \sum_{1 \leq h_2, \dots, h_k \leq H} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1 + h_1) \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \\ + H^{k-1} H' + k H^{k-2} H' X.$$

The g_1 term may now be combined with the product over the remaining g_j terms to form $\prod_{j=1}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j)$. The error term $H^{k-1} H' + k H^{k-2} H' X$ is certainly of size $O((kA^2/W^{1/20}) H^{k-1} X)$, so it suffices to show that

$$\sum_{1 \leq h_1 \leq H'} \sum_{1 \leq h_2, \dots, h_k \leq H} \left| \sum_{1 \leq n \leq X} \prod_{j=1}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \ll \frac{A}{W^{1/20}} H^{k-1} H' X.$$

By covering the ranges $1 \leq h_j \leq H$ by intervals of length H' and averaging, it suffices (after relaxing the conditions $b_j \leq 3AX$ to $b_j \leq 4AX$) to prove that

$$\sum_{1 \leq h_1, h_2, \ldots, h_k \leq H'} \left| \sum_{1 \leq n \leq X} \prod_{j=1}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \ll \frac{A}{W^{1/20}} (H')^k X.$$

The situation is now symmetric with respect to permuting the indices $1, \ldots, k$, so we may assume that the index j_0 in Proposition 5.1 is equal to 1. By the triangle inequality in $h_2, \ldots, h_k$, it suffices to show that

$$\sum_{1 \leq h_1 \leq H'} \left| \sum_{1 \leq n \leq X} \prod_{j=1}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j) \right| \ll \frac{A}{W^{1/20}} H' X$$

for all $h_2, \ldots, h_k$. Writing $G(n) := \prod_{j=2}^{k} 1_{\mathcal{S}} g_j(a_j n + b_j + h_j)$, it thus suffices to show that

$$\sum_{1 \leq h_1 \leq H'} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1 + h_1) G(n) \right| \ll \frac{A}{W^{1/20}} H' X$$

for any 1-bounded function $G : \mathbb{Z} \to \mathbb{C}$.

We use a standard van der Corput argument. By the Cauchy–Schwarz inequality, it suffices to show that

$$\sum_{1 \leq h_1 \leq H'} \left| \sum_{1 \leq n \leq X} 1_{\mathcal{S}} g_1(a_1 n + b_1 + h_1) G(n) \right|^2 \ll \frac{A^2}{W^{1/10}} (H')^2 X^2.$$

The left-hand side may be rewritten as

$$\sum_{n, n' \leq X} G(n) \overline{G(n')} \sum_{1 \leq h_1 \leq H'} 1_{\mathcal{S}} g_1(a_1 n + b_1 + h_1) 1_{\mathcal{S}} \overline{g_j}(a_1 n' + b_1 + h_1).$$

By the triangle inequality, it thus suffices to show that

$$\sum_{n, n' \leq X} \left| \sum_{1 \leq h_1 \leq H'} 1_{\mathcal{S}} g_1(a_1 n + b_1 + h_1) 1_{\mathcal{S}} \overline{g_1}(a_1 n' + b_1 + h_1) \right| \ll \frac{A^2}{W^{1/10}} H' X^2.$$

To abbreviate notation we now write $h = h_1,\ g = g_1,\ a = a_1,\ b = b_1$. By the Cauchy–Schwarz inequality, it suffices to show that

$$\sum_{n, n' \leq X} \left| \sum_{1 \leq h \leq H'} 1_{\mathcal{S}} g(an + b + h) 1_{\mathcal{S}} \bar{g}(an' + b + h) \right|^2 \ll \frac{A^4}{W^{1/5}} (H')^2 X^2.$$

Replacing n, n' by $an+b$, $an'+b$ respectively, it suffices to show that

$$\sum_{n,n'}\Bigl|\sum_{1\le h\le H'} 1_{\mathcal{S}}g(n+h)1_{\mathcal{S}}\bar{g}(n'+h)\Bigr|^2 \ll \frac{A^4}{W^{1/5}}(H')^2X^2$$

where we have extended $1_{\mathcal{S}}g$ by zero to the negative integers. The left-hand side can be rewritten as

$$\sum_{|h|<H'}(\lfloor H'\rfloor-|h|)\Bigl|\sum_{n} 1_{\mathcal{S}}g(n)1_{\mathcal{S}}\bar{g}(n+h)\Bigr|^2,$$

and the claim follows from Proposition 5.3. □

Appendix A: Mean values of complex multiplicative functions in short intervals

In this section we prove a complex variant of results in [Matomäki and Radziwiłł 2015] in the case that f is not p^{it} pretentious. In particular, we show that the mean value of a 1-bounded nonpretentious multiplicative function is small for most short intervals:

Theorem A.1. *Let f be a 1-bounded multiplicative function and let $M(f;X)$ be as in (1-6). Then, for $X\ge h\ge 10$,*

$$\frac{1}{X}\int_X^{2X}\Bigl|\frac{1}{h}\sum_{x\le n\le x+h} f(n)\Bigr|^2dx \ll \exp(-M(f;X))M(f;X)+\frac{(\log\log h)^2}{(\log h)^2}+\frac{1}{(\log X)^{1/50}}.$$

Actually, as in [Matomäki and Radziwiłł 2015] and earlier in this paper, one gets better quantitative results if one first restricts to a subset of n with a typical factorization. Let us first define such a subset $\mathcal{S}$ in this setting.

Let $\eta\in(0,1/6)$, and let X_0 be a quantity with $\sqrt{X}\le X_0\le X$. (The results in [Matomäki and Radziwiłł 2015] used the choice $X_0=X$, but for technical reasons we will need a more flexible choice of this parameter.) Consider a sequence of increasing intervals $[P_j,Q_j]$, $j\ge 1$ such that:

- $Q_1\le \exp(\sqrt{\log X_0})$.
- The intervals are not too far from each other; precisely, for all $j\ge 2$,

$$\frac{\log\log Q_j}{\log P_{j-1}-1}\le \frac{\eta}{4j^2}. \tag{A-1}$$

- The intervals are not too close to each other; precisely, for all $j\ge 2$,

$$\frac{\eta}{j^2}\log P_j \ge 8\log Q_{j-1}+16\log j. \tag{A-2}$$

For example, given $0 < \eta < 1/6$, the sequence of intervals $[P_j, Q_j]$ defined in Definition 2.1 can be verified to obey the above estimates if

$$\exp(\sqrt{\log X_0}) \geq Q_1 \geq P_1 \geq (\log Q_1)^{40/\eta}$$

and if P_1 is sufficiently large.

Let $\mathcal{S}$ be the set of integers $X \leq n \leq 2X$ having at least one prime factor in each of the intervals $[P_j, Q_j]$ for $j \leq J$, where J is chosen to be the largest index j such that $Q_j \leq \exp((\log X_0)^{1/2})$. We will establish the following variant of [Matomäki and Radziwiłł 2015, Theorem 3].

Theorem A.2. *Let f be a 1-bounded multiplicative function. Let $\mathcal{S}$ be as above with $\eta \in (0, 1/6)$. If $[P_1, Q_1] \subset [1, h]$, then for all $X > X(\eta)$ large enough and all $h \geq 3$,*

$$\frac{1}{X}\int_X^{2X} \Bigg|\frac{1}{h}\sum_{\substack{x \leq n \leq x+h \\ n \in \mathcal{S}}} f(n)\Bigg|^2 dx \ll \exp(-M(f; X))M(f; X) + \frac{(\log h)^{1/3}}{P_1^{1/6-\eta}} + \frac{1}{(\log X)^{1/50}}.$$

The proof of Theorem A.2 proceeds as the proof of [Matomäki and Radziwiłł 2015, Theorem 3]. The first step is a Parseval bound:

$$\frac{1}{X}\int_X^{2X} \Bigg|\frac{1}{h}\sum_{\substack{x \leq n \leq x+h \\ n \in \mathcal{S}}} f(n)\Bigg|^2 dx \ll \int_1^{1+iX/h_1} |F(s)|^2\,|ds| + \max_{T \geq X/h_1} \frac{X/h_1}{T}\int_{1+iT}^{1+i2T} |F(s)|^2\,|ds|.$$

This follows exactly in the same way as [Matomäki and Radziwiłł 2015, Lemma 14] but there is no need to split the integral into two parts, and one can just work as for $V(x)$ there. Theorem A.2 now follows immediately from the following variant of [Matomäki and Radziwiłł 2015, Proposition 1].

Proposition A.3. *Let f be a 1-bounded multiplicative function. Let $\mathcal{S}$ be as above, and let*

$$F(s) = \sum_{\substack{X \leq n \leq 2X \\ n \in \mathcal{S}}} \frac{f(n)}{n^s}.$$

Then, for any T,

$$\int_0^T |F(1+it)|^2\,dt \ll \left(\frac{T}{X/Q_1} + 1\right)\left(\frac{(\log Q_1)^{1/3}}{P_1^{1/6-\eta}} + \exp(-M(f; X))M(f; X) + \frac{1}{(\log X)^{1/50}}\right).$$

Proof. Since the mean value theorem gives the bound $O(T/X + 1)$, we can assume $T \leq X/2$ and $M(f; X) \geq 1$.

Now let t_1 be the value of t which attains the minimum in

$$M(f;X) = \inf_{|t|\le X} \mathbb{D}(g, n \mapsto n^{it}; X)^2.$$

We split the integration into three ranges:

$$\mathcal{T}_0 = \{0 \le t \le T : |t-t_1| \le \exp(M(f;X))/M(f;X)\},$$
$$\mathcal{T}_1 = \{0 \le t \le T : \exp(M(f;X))/M(f;X) \le |t-t_1| \le (\log X)^{1/16}\},$$
$$\mathcal{T}_2 = \{0 \le t \le T : |t-t_1| \ge (\log X)^{1/16}\}.$$

Notice that by the definition of t_1, the triangle inequality and arguing as in (1-12), for any $|t| \le X$ with $|t-t_1| \ge 1$, and any $\varepsilon > 0$,

$$\begin{aligned} 2\mathbb{D}(f, p^{it}; X) &\ge \mathbb{D}(f, p^{it}; X) + \mathbb{D}(f, p^{it_1}; X) \ge \mathbb{D}(1, p^{i(t-t_1)}) \\ &\ge \left(\frac{1}{\sqrt{3}} - \varepsilon\right)\sqrt{\log\log X} + O(1), \end{aligned}$$

so that by Halasz's theorem, for every $|t| \le T$,

$$F(1+it) \ll (\log X)^{-1/16} + \frac{1}{1+|t-t_1|}.$$

In the region $|t-t_1| \ge (\log X)^{1/16}$, the above implies the following in exactly the same way as [Matomäki and Radziwiłł 2015, Lemma 3].

Lemma A.4. *Let $X \ge Q \ge P \ge 2$. Let t_1 be as above and let*

$$G(s) = \sum_{X \le n \le 2X} \frac{f(n)}{n^s} \cdot \frac{1}{\#\{p \in [P,Q] : p|n\} + 1}.$$

Then, for any $t \in \mathcal{T}_2$,

$$|G(1+it)| \ll \frac{\log Q}{(\log X)^{1/16}\log P} + \log X \cdot \exp\left(-\frac{\log X}{3\log Q}\log\frac{\log X}{\log Q}\right).$$

This was the only part in the proof [Matomäki and Radziwiłł 2015, Proposition 1] that needed f to be real-valued, and thus we get

$$\int_{\mathcal{T}_2} |F(1+it)|^2\,dt \ll \left(\frac{T}{X/Q_1} + 1\right)\left(\frac{(\log Q_1)^{1/3}}{P_1^{1/6-\eta}} + \frac{1}{(\log X)^{1/50}}\right).$$

Using the estimate $F(1+it) \ll 1/|t-t_1|$ for $t \in \mathcal{T}_1$ and, from Halasz's theorem, the estimate $F(1+it) \ll \exp(-M(f;X))M(f;X)$ for $t \in \mathcal{T}_0$, we obtain

$$\int_{\mathcal{T}_0 \cup \mathcal{T}_1} |F(1+it)|^2\,dt \ll \exp(-M(f;X))M(f;X),$$

and the claim follows. □

Proof of Theorem A.1. Let $\eta = 1/12$, $P_1 = (\log h)^{480}$, $Q_1 = h$, let P_j and Q_j for $j \geq 2$ be as in Definition 2.1, and let $\mathcal{S}$ be as above. Then

$$\frac{1}{X}\int_X^{2X}\left|\frac{1}{h}\sum_{x\leq n\leq x+h} f(n)\right|^2 dx \leq \frac{1}{X}\int_X^{2X}\left|\frac{1}{h}\sum_{\substack{x\leq n\leq x+h\\ n\in\mathcal{S}}} f(n)\right|^2 dx + \frac{1}{X}\int_X^{2X}\left|\frac{1}{h}\sum_{\substack{x\leq n\leq x+h\\ n\notin\mathcal{S}}} 1\right|^2 dx.$$

The contribution from the first integral is acceptable by Theorem A.2. We rewrite the second integrand as

$$\begin{aligned}\left|\frac{1}{h}\sum_{\substack{x\leq n\leq x+h\\ n\notin\mathcal{S}}} 1\right| &= \left|1 + O(1/h) - \frac{1}{h}\sum_{\substack{x\leq n\leq x+h\\ n\in\mathcal{S}}} 1\right| \\ &\leq \left|\frac{1}{X}\sum_{\substack{X\leq n\leq 2X\\ n\in\mathcal{S}}} 1 - \frac{1}{h}\sum_{\substack{x\leq n\leq x+h\\ n\in\mathcal{S}}} 1\right| + \left|\frac{1}{X}\sum_{\substack{X\leq n\leq 2X\\ n\notin\mathcal{S}}} 1\right| + O(1/h),\end{aligned}$$

and the claim follows from [Matomäki and Radziwiłł 2015, Theorem 3 with $f = 1$] and Lemma 2.2. □

Appendix B: Counterexample to the uncorrected Elliott conjecture

In this appendix we present a counterexample to Conjecture 1.5. More precisely:

Theorem B.1 (counterexample). *There exists a* 1*-bounded multiplicative function* $g : \mathbb{N} \to \mathbb{C}$ *such that*

$$\sum_p \frac{1 - \operatorname{Re}(g(p)\overline{\chi(p)}p^{-it})}{p} = \infty \tag{B-1}$$

for all Dirichlet characters χ *and* $t \in \mathbb{R}$ (*i.e., one has* $M(g; \infty, \infty) = \infty$), *but such that*

$$\left|\sum_{n\leq t_m} g(n)\overline{g(n+1)}\right| \gg t_m \tag{B-2}$$

for all sufficiently large m *and some sequence* t_m *going to infinity.*

Proof. For each prime p, we choose $g(p)$ from the unit circle $S^1 := \{z : |z| = 1\}$ by the following iterative procedure involving a sequence $t_1 < t_2 < t_3 < \cdots$:

(1) Initialize $t_1 := 100$ and $m := 1$, and set $g(p) := 1$ for all $p \leq t_1$.

(2) Now suppose recursively that $g(p)$ has been chosen for all $p \leq t_m$. As the quantities $\log p$ are linearly independent over the integers, the (continuous) sequence $t \mapsto (t \log p \bmod 1)_{p\leq t_m}$ is equidistributed in the torus $\prod_{p\leq t_m}\mathbb{T}$ and, equivalently, the sequence $t \mapsto (p^{it})_{p\leq t_m}$ is equidistributed in the torus $\prod_{p\leq t_m} S^1$.

Thus one can find a quantity $s_{m+1} > \exp(t_m)$ such that, for all $p \le t_m$,

$$p^{is_{m+1}} = g(p)\left(1 + O\left(\frac{1}{t_m^2}\right)\right). \tag{B-3}$$

(3) Set $t_{m+1} := s_{m+1}^2$, and then set

$$g(p) := p^{is_{m+1}} \tag{B-4}$$

for all $t_m < p \le t_{m+1}$. Now increment m to $m+1$ and return to step (2).

Clearly the t_m go to infinity, so $g(p)$ is defined for all primes p. We then define

$$g(n) := \mu(n)^2 \prod_{p|n} g(p), \tag{B-5}$$

which is clearly a 1-bounded multiplicative function.

Suppose that $n \le t_{m+1}$ is squarefree. Then n is the product of distinct primes less than or equal to t_{m+1}, including at most t_m primes less than or equal to t_m. From (B-5) we then have

$$g(n) = n^{is_{m+1}}\left(1 + O\left(\frac{1}{t_m^2}\right)\right)^{O(t_m)} = n^{is_{m+1}} + O\left(\frac{1}{t_m}\right).$$

If n is not squarefree, then $g(n)$ of course vanishes. Thus, for $t_{m+1}^{3/4} \le n \le t_{m+1} - 1$, we have

$$\begin{aligned} g(n)\overline{g(n+1)} &= \mu^2(n)\mu^2(n+1)\left(\frac{n+1}{n}\right)^{is_{m+1}} + O\left(\frac{1}{t_m}\right) \\ &= \mu^2(n)\mu^2(n+1) + O\left(\frac{s_{m+1}}{t_{m+1}^{3/4}}\right) + O\left(\frac{1}{t_m}\right) \\ &= \mu^2(n)\mu^2(n+1) + O\left(\frac{1}{t_m}\right), \end{aligned}$$

and the claim (B-2) then easily follows since the sequence $\mu^2(n)\mu^2(n+1)$ has positive mean value.

Now we prove (B-1). From (B-4), we have

$$\begin{aligned} \sum_p \frac{1 - \mathrm{Re}(g(p)\overline{\chi(p)}p^{-it})}{p} &\ge \sum_{t_m < p \le t_{m+1}} \frac{1 - \mathrm{Re}(\overline{\chi(p)}p^{i(s_{m+1}-t)})}{p} \\ &\ge \sum_{\exp((\log t_{m+1})^{5/6}) < p \le t_{m+1}} \frac{1 - \mathrm{Re}(\overline{\chi(p)}p^{i(s_{m+1}-t)})}{p} \end{aligned}$$

since $\exp((\log t_{m+1})^{5/6}) \ge \exp((2t_m)^{5/6}) \ge t_m$. We see as in (1-12) that the right-hand side goes to infinity as $m \to \infty$ for any fixed χ, t, and the claim follows. $\square$

It is easy to see that the function g constructed in the above counterexample violates (1-9), and so is not a counterexample to the corrected form of Conjecture 1.5. It is also not difficult to modify the above counterexample so that the function g is completely multiplicative instead of multiplicative, using the fact that most numbers up to t_{m+1} have fewer than t_m prime factors less than t_m (counting multiplicity); we leave the details to the interested reader.

Appendix C: An argument of Granville and Soundararajan

In this appendix we show the equivalence of the hypotheses (1-7) and (1-9) for Elliott's conjecture in the case that the multiplicative function g_{j_0} is real. The key lemma is the following estimate, essentially due to Granville and Soundararajan.

Lemma C.1. *Let $f:\mathbb{N}\to[-1,1]$ be a multiplicative function, let $x\geq 100$, and let χ be a fixed Dirichlet character. For $1\leq|\alpha|\leq x$, one has*

$$\mathbb{D}(f, n\mapsto \chi(n)n^{i\alpha}; x)\geq \frac{1}{4}\sqrt{\log\log x}+O_\chi(1). \tag{C-1}$$

When χ^2 is nonprincipal, this holds for all $|\alpha|\leq x$.

If χ^2 is principal (i.e., χ is a quadratic character), then, for $|\alpha|\leq 1$, one has

$$\mathbb{D}(f, n\mapsto \chi(n)n^{i\alpha}; x)\geq \frac{1}{3}\mathbb{D}(f,\chi;x)+O(1). \tag{C-2}$$

Proof. To establish (C-1), we notice that, by conjugation symmetry and the triangle inequality,

$$\begin{aligned}\mathbb{D}(f, n\mapsto \chi(n)n^{i\alpha}; x) &= \frac{1}{2}(\mathbb{D}(f, n\mapsto \chi(n)n^{i\alpha}; x)+\mathbb{D}(f, n\mapsto \overline{\chi(n)}n^{-i\alpha}; x))\\ &\geq \frac{1}{2}\mathbb{D}(n\mapsto \overline{\chi(n)}n^{-i\alpha}, n\mapsto \chi(n)n^{i\alpha}; x)\\ &= \frac{1}{2}\biggl(\sum_{p\leq x}\frac{1-\operatorname{Re}\chi^2(p)p^{2i\alpha}}{p}\biggr)^{1/2}\end{aligned}$$

which implies the claim for $|\alpha|\geq 1$ or for nonprincipal χ^2 by the zero-free (and pole-free) region for Dirichlet L-functions (see (1-12) for a related argument).

To establish (C-2), notice first that since χ^2 is principal, χ is real-valued which together with the triangle inequality implies

$$\mathbb{D}(f, n\mapsto \chi(n)n^{i\alpha}; x)=\mathbb{D}(f\chi, n\mapsto n^{i\alpha}; x)\geq \mathbb{D}(1, f\chi; x)-\mathbb{D}(1, n\mapsto n^{i\alpha}; x).$$

Now $\mathbb{D}(1, n\mapsto n^{i\alpha}; x)=\mathbb{D}(1, n\mapsto n^{2i\alpha}; x)+O(1)$ for $|\alpha|\leq 1$ since, from the prime number theorem, $\mathbb{D}(1, n\mapsto n^{i\alpha}; x)^2=\log(1+|\alpha|\log x)+O(1)$, so that the claim follows unless $\mathbb{D}(1, n \mapsto n^{2i\alpha}; x) \geq (2/3)\mathbb{D}(1, f\chi; x)$. But in the latter case, the triangle inequality gives

$$\begin{aligned}\frac{2}{3}\mathbb{D}(f, \chi; x) &= \frac{2}{3}\mathbb{D}(1, f\chi; x)\\ &\leq \mathbb{D}(1, n \mapsto n^{2i\alpha}; x)\\ &= \mathbb{D}(n \mapsto n^{-i\alpha}, n \mapsto n^{i\alpha}; x)\\ &\leq \mathbb{D}(f\chi, n \mapsto n^{-i\alpha}; x) + \mathbb{D}(f\chi; n \mapsto n^{i\alpha}; x)\\ &= 2\mathbb{D}(f, n \mapsto \chi(n)n^{i\alpha}; x),\end{aligned}$$

and the claim (C-2) follows. □

From this lemma, we see that if g_{j_0} is a real 1-bounded multiplicative function, then, for given Q, the condition (1-9) is equivalent to

$$\mathbb{D}(g_{j_0}, \chi; X) \to \infty$$

when $X \to \infty$ for all quadratic characters χ of modulus at most Q. But this follows from (1-7). The converse implication is trivial.

Acknowledgments

Tao was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and NSF grant DMS-1266164. The authors thank Andrew Granville and the anonymous referee for useful comments and corrections.

References

[Balog 1990] A. Balog, "The prime k-tuplets conjecture on average", pp. 47–75 in *Analytic number theory* (Allerton Park, IL, 1989), edited by B. C. Berndt et al., Progr. Math. **85**, Birkhäuser, Boston, 1990. MR 92e:11105 Zbl 0719.11066

[Bourgain et al. 2013] J. Bourgain, P. Sarnak, and T. Ziegler, "Disjointness of Moebius from horocycle flows", pp. 67–83 in *From Fourier analysis and number theory to radon transforms and geometry*, edited by H. M. Farkas et al., Dev. Math. **28**, Springer, New York, 2013. MR 2986954 Zbl 06150358

[Chowla 1965] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*, Mathematics and Its Applications **4**, Gordon and Breach Science Publishers, New York-London-Paris, 1965. MR 31 #2201 Zbl 0133.30003

[Daboussi and Delange 1982] H. Daboussi and H. Delange, "On multiplicative arithmetical functions whose modulus does not exceed one", *J. London Math. Soc.* (2) **26**:2 (1982), 245–264. MR 83k:10082 Zbl 0499.10052

[Davenport 1937] H. Davenport, "On some infinite series involving arithmetical functions, II", *Quart. J. Math. Oxf.* **8** (1937), 313–320. Zbl 0017.39101

[Elliott 1992] P. D. T. A. Elliott, "On the correlation of multiplicative functions", *Notas Soc. Mat. Chile* **11**:1 (1992), 1–11. MR 95j:11076 Zbl 0810.11055

[Frantzikinakis and Host 2015] N. Frantzikinakis and B. Host, "Asymptotics for multilinear averages of multiplicative functions", preprint, 2015. arXiv 1502.02646

[Friedlander and Iwaniec 2010] J. Friedlander and H. Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications **57**, American Mathematical Society, Providence, RI, 2010. MR 2011d:11227 Zbl 1226.11099

[Granville and Soundararajan 2007] A. Granville and K. Soundararajan, "Large character sums: pretentious characters and the Pólya–Vinogradov theorem", *J. Amer. Math. Soc.* **20**:2 (2007), 357–384. MR 2007k:11128 Zbl 1210.11090

[Green and Tao 2010] B. Green and T. Tao, "Linear equations in primes", *Ann. of Math.* (2) **171**:3 (2010), 1753–1850. MR 2011j:11177 Zbl 1242.11071

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications **53**, American Mathematical Society, Providence, RI, 2004. Zbl 1059.11001

[Kátai 1986] I. Kátai, "A remark on a theorem of H. Daboussi", *Acta Math. Hungar.* **47**:1–2 (1986), 223–225. MR 87j:11072 Zbl 0607.10034

[Kawada 1993] K. Kawada, "The prime k-tuplets in arithmetic progressions", *Tsukuba J. Math.* **17**:1 (1993), 43–57. MR 94i:11068 Zbl 0797.11076

[Kawada 1995] K. Kawada, "A Montgomery–Hooley type theorem for prime k-tuplets", *Acta Math. Hungar.* **66**:3 (1995), 177–200. MR 96h:11096 Zbl 0869.11069

[Matomäki and Radziwiłł 2015] K. Matomäki and M. Radziwiłł, "Multiplicative functions in short intervals", 2015. To appear in *Ann. of Math.* arXiv 1501.04585

[Mikawa 1992] H. Mikawa, "On prime twins in arithmetic progressions", *Tsukuba J. Math.* **16**:2 (1992), 377–387. MR 94e:11101 Zbl 0778.11053

[Montgomery 1994] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics **84**, American Mathematical Society, Providence, RI, 1994. MR 96i:11002 Zbl 0814.11001

[Montgomery and Vaughan 1977] H. L. Montgomery and R. C. Vaughan, "Exponential sums with multiplicative coefficients", *Invent. Math.* **43**:1 (1977), 69–82. MR 56 #15579 Zbl 0362.10036

[Montgomery and Vaughan 2007] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory, I: Classical theory*, Cambridge Studies in Advanced Mathematics **97**, Cambridge Univ. Press, 2007. MR 2009b:11001 Zbl 1142.11001

Communicated by Andrew Granville
Received 2015-04-17 Revised 2015-08-28 Accepted 2015-10-06

ksmato@utu.fi — *Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland*

maksym.radziwill@gmail.com — *Department of Mathematics, Rutgers University, Hill Center for the Mathematical Sciences, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, United States*

tao@math.ucla.edu — *Department of Mathematics, University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1555, 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 9 No. 9 2015