COMPOSITIO MATHEMATICA

Galois representations attached to moments of Kloosterman sums and conjectures of Evans

Zhiwei Yun and Christelle Vincent

FOUNDATION
COMPOSITIO
MATHEMATICA

LONDON
MATHEMATICAL
SOCIETY
150 YEARS

Galois representations attached to moments of Kloosterman sums and conjectures of Evans

Zhiwei Yun

With Appendix B by Christelle Vincent

Abstract

Kloosterman sums for a finite field $\mathbb{F}_p$ arise as Frobenius trace functions of certain local systems defined over $\mathbb{G}_{m,\mathbb{F}_p}$. The moments of Kloosterman sums calculate the Frobenius traces on the cohomology of tensor powers (or symmetric powers, exterior powers, etc.) of these local systems. We show that when p ranges over all primes, the moments of the corresponding Kloosterman sums for $\mathbb{F}_p$ arise as Frobenius traces on a continuous ℓ-adic representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ that comes from geometry. We also give bounds on the ramification of these Galois representations. All of this is done in the generality of Kloosterman sheaves attached to reductive groups introduced by Heinloth, Ngô and Yun [Ann. of Math. (2) **177** (2013), 241–310]. As an application, we give proofs of conjectures of Evans [Proc. Amer. Math. Soc. **138** (2010), 517–531; Israel J. Math. **175** (2010), 349–362] expressing the seventh and eighth symmetric power moments of the classical Kloosterman sum in terms of Fourier coefficients of explicit modular forms. The proof for the eighth symmetric power moment conjecture relies on the computation done in Appendix B by C. Vincent.

Contents

1. Introduction

1.1 Kloosterman sums and their moments

We first recall the definition of Kloosterman sums.

DEFINITION 1.1.1. Let p be a prime number. Fix a nontrivial additive character $\psi : \mathbb{F}_p \to \mathbb{C}^\times$. Let $n \geqslant 2$ be an integer. Then the n-variable Kloosterman sum over $\mathbb{F}_p$ is a complex valued function on $\mathbb{F}_p^\times$ whose value at $a \in \mathbb{F}_p^\times$ is

$$\mathrm{Kl}_n(p; a) = \sum_{x_1,\ldots,x_n \in \mathbb{F}_p^\times; \, x_1 x_2 \cdots x_n = a} \psi(x_1 + \cdots + x_n).$$

These exponential sums arise naturally in the study of automorphic forms for GL_n.

1.1.2 *Deligne's Kloosterman sheaf.* Deligne [Del77] has given a geometric interpretation of the Kloosterman sum. He considers the following diagram consisting of schemes over $\mathbb{F}_p$.

$$
\begin{array}{ccc}
 & \mathbb{G}_m^n & \\
 {\scriptstyle \pi}\swarrow & & \searrow {\scriptstyle F} \\
\mathbb{G}_m & & \mathbb{A}^1
\end{array}
$$
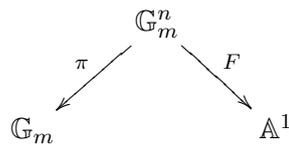
Here the morphism π is taking the product and F is the morphism of taking the sum. Deligne defines

$$\mathrm{Kl}_n := \mathbf{R}^{n-1} \pi_! F^* \mathrm{AS}_\psi,$$

where AS_ψ is the rank-one local system on $\mathbb{A}^1_{\mathbb{F}_p}$ corresponding to the character $\psi : \mathbb{F}_p \xrightarrow{\sim} \mu_p(\mathbb{C}) \cong \mu_p(\overline{\mathbb{Q}}_\ell)$ (we make a choice of the latter isomorphism). He then shows that Kl_n is a $\mathbb{Q}_\ell(\mu_p)$-local system on $\mathbb{G}_{m,\mathbb{F}_p}$ of rank n. The relationship between the local system Kl_n and the Kloosterman

sum $\mathrm{Kl}_n(p; a)$ is explained by the following identity

$$\mathrm{Kl}_n(p; a) = (-1)^{n-1}\mathrm{Tr}(\mathrm{Frob}_a, (\mathrm{Kl}_n)_a).$$

Here Frob_a is the geometric Frobenius operator acting on the geometric stalk $(\mathrm{Kl}_n)_a$ of the local system Kl_n at $a \in \mathbb{G}_m(\mathbb{F}_p) = \mathbb{F}_p^\times$. The Kloosterman sheaf has been studied in depth by Deligne [Del77] and Katz [Kat88].

For each representation V of GL_n, we may form the corresponding local system Kl_n^V. For example, if $V = \mathrm{Sym}^d$, the dth symmetric power of the standard representation of GL_n, then $\mathrm{Kl}_n^V = \mathrm{Sym}^d(\mathrm{Kl}_n)$. We define

$$\mathrm{Kl}_n^V(p; a) := \mathrm{Tr}(\mathrm{Frob}_a, (\mathrm{Kl}_n^V)_a).$$

Note that when V is the standard representation, $\mathrm{Kl}_n^V(p; a)$ differs from the Kloosterman sum $\mathrm{Kl}_n(p; a)$ by the sign $(-1)^{n-1}$.

When $n = 2$, Deligne's interpretation gives an expression $\mathrm{Kl}_2(p; a) = -(\alpha_a + \beta_a)$ where $\alpha_a, \beta_a \in \overline{\mathbb{Q}}_\ell$ are the eigenvalues of the operator Frob_a on $(\mathrm{Kl}_n)_a$. Taking $V = \mathrm{Sym}^d$, we have

$$\mathrm{Kl}_n^{\mathrm{Sym}^d}(p; a) = \sum_{i=0}^{d} \alpha_a^i \beta_a^{d-i}.$$

DEFINITION 1.1.3. Let V be a representation of GL_n. The V-*moment* of Kl_n is the sum

$$m_n^V(p) := \sum_{a \in \mathbb{F}_p^\times} \mathrm{Kl}_n^V(p; a).$$

One may ask whether there is a closed formula for $m_n^V(p)$ in terms of more familiar arithmetic functions of the prime p. A related question is whether the behavior of the sequence of numbers $\{m_n^V(p)\}$ exhibits certain 'motivic' nature. A typical example of a motivic sequence is the Fourier coefficients $\{a_f(p)\}$ of a holomorphic Hecke eigenform f. Another typical example is $\{\#X(\mathbb{F}_p)\}$ where X is an algebraic variety defined over $\mathbb{Z}$. In this direction, Fu and Wan [FW08] showed that for $V = \mathrm{Sym}^d, \wedge^d$ or $\otimes^d$, the moments $m_n^V(p)$ are the Frobenius traces on a *virtual* motivic Galois representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

Ron Evans (see [Eva10a, p. 523] and [Eva10b, p. 350]) has made very precise conjectures about the moments $m_2^V(p)$ when $V = \mathrm{Sym}^d$ and $d = 5, 6, 7$ or 8, the first two being proved and the last two open. We denote $m_2^{\mathrm{Sym}^d}(p)$ by $m_2^d(p)$.

(i) When $d = 5$, $(-m_2^5(p) - 1)/p^2$ is the pth Fourier coefficient of a holomorphic cuspdial Hecke eigenform of weight 3, level $\Gamma_0(15)$ and quadratic nebentypus $(\cdot/15)$. This is proved by Livné [Liv95] and Peters *et al.* [PTV92].

(ii) When $d = 6$, $(-m_2^6(p) - 1)/p^2$ is the pth Fourier coefficient of a holomorphic Hecke eigenform of weight 4 and level $\Gamma_0(6)$. This is proved by Hulek *et al.* [HSVV01].

We now state the open conjectures of Evans about $m_2^7(p)$ and $m_2^8(p)$.

CONJECTURE 1.1.4 (Evans). (i) (See [Eva10b, Conjecture 1.1].) There exists a holomorphic cuspdial Hecke eigenform f of weight 3, level $\Gamma_0(525)$ and nebentypus $\epsilon_f = (\cdot/21)\epsilon_5$, where $(\cdot/21)$ is the quadratic character with conductor 21 and ϵ_5 is a quartic character with conductor 5, such that for all primes $p > 7$, we have

$$\left(\frac{p}{105}\right)(-m_2^7(p) - 1)/p^2 = a_f(p)^2 \epsilon_f(p)^{-1} - p^2, \tag{1.1}$$

where $a_f(p)$ is the pth normalized Fourier coefficient of f.

(ii) (See [Eva10a, p. 523].) There exists a holomorphic cuspidal Hecke eigenform f of weight 6, level $\Gamma_0(6)$, such that for all odd primes p, we have

$$-m_2^8(p) - 1 - p^4 = p^2 a_f(p),$$

where $a_f(p)$ is the pth normalized Fourier coefficient f.

The main goal of this paper is to confirm these two conjectures.

THEOREM 1.1.5. *Conjecture 1.1.4 is true.*

Part (i) of Conjecture 1.1.4 will be proved in §4.7.7. Part (ii) of Conjecture 1.1.4 will be reduced to a finite calculation in Theorem 4.6.1, and will be verified by computational softwares in Appendix B by Christelle Vincent.

The proof of these conjectures are based on two key ingredients. One is Serre's modularity conjecture [Ser87] proved by Khare and Wintenberger [KW09a, KW09b]; the other is the following existence result on the Galois representations (not just a virtual one) underlying the moments $m_2^d(p)$, which is a very special case of our main result (to be stated later).

THEOREM 1.1.6 (Proved in §4.4). (i) *Let $d \geqslant 3$ be an odd integer. Then for each prime ℓ, there exists an orthogonal $\mathbb{Q}_\ell$-vector space M_ℓ of dimension $(d-1)/2$ and a continuous Galois representation*

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{O}(M_\ell)$$

with the following properties.

(a) *The determinant of ρ_ℓ is the quadratic character given by the Jacobi symbol $(\cdot/d!!)$, where $d!! = d(d-2)(d-4)\cdots 1$.*

(b) *The representation ρ_ℓ comes from geometry. More precisely, there is a smooth projective algebraic variety X of dimension $d-1$ over $\mathbb{Q}$ (independent of ℓ) such that the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module M_ℓ appears as a subquotient of $\mathrm{H}^{d-1}(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)((d-1)/2)$.*

(c) *For all primes $p \neq \ell$ satisfying $p > d$ or $p = 2$, ρ_ℓ is unramified at p, and we have*

$$-m_2^d(p) - 1 = p^{(d+1)/2}\mathrm{Tr}(\mathrm{Frob}_p, M_\ell). \tag{1.2}$$

(ii) *Let $d \geqslant 4$ be an even integer. Then for each prime ℓ, there exists a symplectic $\mathbb{Q}_\ell$-vector space M_ℓ of dimension $2[(d+2)/4] - 2$ and a continuous Galois representation*

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GSp}(M_\ell)$$

with the following properties.

(a) *The similitude character of ρ_ℓ is the $(-d-1)$th power of the ℓ-adic cyclotomic character.*

(b) *The representation ρ_ℓ comes from geometry. More precisely, there is a smooth projective algebraic variety X of dimension $d-1$ over $\mathbb{Q}$ (independent of ℓ) such that the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module M_ℓ appears as a subquotient of $\mathrm{H}^{d-1}(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)(-1)$.*

(c) *For a prime $p \neq 2, p \neq \ell$, M_ℓ as a $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$-module can be decomposed into a direct sum of symplectic $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$-modules*

$$M_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)} \cong J_2(-d/2)^{[d/2p]} \oplus U_p. \tag{1.3}$$

Here the action of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ on U_p is unramified, and J_2 is the unique two-dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ which is an extension of $\mathbb{Q}_\ell(-1)$ by $\mathbb{Q}_\ell$ and on which the

action of the inertia group $\mathcal{I}_p$ is unipotent but nontrivial (i.e. J_2 corresponds to the Steinberg representation under the local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$). Moreover, we have

$$-m_2^d(p) - 1 = \mathrm{Tr}(\mathrm{Frob}_p, M_\ell^{\mathcal{I}_p}) + \begin{cases} p^{d/2}, & d \equiv 0 \bmod 4, \\ 0, & d \equiv 2 \bmod 4. \end{cases} \tag{1.4}$$

In particular, ρ_ℓ is unramified at primes $p > d/2, p \neq \ell$.

We mention that when $d = 7$, Katz (private correspondence) proposed that there should exist a Galois representation into O_3 as in the above theorem that underlies the modular form predicted by Evans.

1.2 General case of the main result

In the main body of the paper, we work with more general Kloosterman sheaves than Deligne's. Motivated by the work of Frenkel and Gross [FG09], Heinloth *et al.* [HNY13] construct a Kloosterman sheaf $\mathrm{Kl}_{\widehat{G}}$ for each almost simple split algebraic group $\widehat{G}$ over $\mathbb{Q}_\ell$. It has the following properties.

(i) $\mathrm{Kl}_{\widehat{G}}$ is a $\widehat{G}^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p))$-local system over $\mathbb{G}_{m,\mathbb{F}_p} = \mathbb{P}^1_{\mathbb{F}_p} - \{0, \infty\}$. Here $\widehat{G}^{\mathrm{Tt}}$ is a slight modification of $\widehat{G}$ (after Deligne) in order to avoid half Tate twists, see § 2.1.4. In other words, for every representation V of $\widehat{G}^{\mathrm{Tt}}$ there is a local system $\mathrm{Kl}_{\widehat{G}}^V$ over $\mathbb{G}_{m,\mathbb{F}_p}$ of rank $\dim V$, such that the assignment $V \mapsto \mathrm{Kl}_{\widehat{G}}^V$ is compatible with the formation of tensor products.

(ii) The local geometric monodromy of $\mathrm{Kl}_{\widehat{G}}$ at the puncture $0 \in \mathbb{P}^1$ is a regular unipotent element in $\widehat{G}$.

(iii) The local geometric monodromy of $\mathrm{Kl}_{\widehat{G}}$ at the puncture $\infty \in \mathbb{P}^1$ is a 'simple wild parameter' à la Gross and Reeder [GR10]. This roughly means that the Swan conductor at ∞ of the adjoint local system $\mathrm{Kl}_{\widehat{G}}^{\mathrm{Ad}}$ takes the smallest possible nonzero value.

(iv) When $\widehat{G} = \mathrm{SL}_n$, $\mathrm{Kl}_{\widehat{G}}$ is the Kloosterman sheaf Kl_n defined by Deligne.

Let V be a representation of $\widehat{G}^{\mathrm{Tt}}$. We define a 'balanced' version of the V-moment for $\mathrm{Kl}_{\widehat{G}}$ to be the Frobenius module

$$M_{!*,\mathbb{F}_p}^V = \mathrm{Im}(\mathrm{H}_c^1(\mathbb{G}_{m,\overline{\mathbb{F}}_p}, \mathrm{Kl}_{\widehat{G}}^V) \to \mathrm{H}^1(\mathbb{G}_{m,\overline{\mathbb{F}}_p}, \mathrm{Kl}_{\widehat{G}}^V)). \tag{1.5}$$

When $\widehat{G} = \mathrm{SL}_n$, we have defined the V-moment of Kl_n in Definition 1.1.3, which is equal to the alternating trace of Frobenius on the cohomology $\mathrm{H}_c^*(\mathbb{G}_m, \mathrm{Kl}_n^V)$ according to the Lefschetz trace formula. Therefore, $\mathrm{Tr}(\mathrm{Frob}, M_{!*,\mathbb{F}_p}^V)$ does *not* directly give the moment $m_n^V(p)$. However, the difference between $\mathrm{Tr}(\mathrm{Frob}, M_{!*,\mathbb{F}_p}^V)$ and the moment $m_n^V(p)$ is explicitly computable in many examples. It turns out that $M_{!*,\mathbb{F}_p}^V$ is a better object to work with.

Our main result is a generalization of Theorem 1.1.6 to all moments of all Kloosterman sheaves $\mathrm{Kl}_{\widehat{G}}$.

THEOREM 1.2.1 (Proved in §§ 3.4.2 and 3.3.4). *Let V be the irreducible representation of $\widehat{G}$ with highest weight λ and we form the V-moment $M_{!*,\mathbb{F}_p}^V$ of $\mathrm{Kl}_{\widehat{G}}$ as in (1.5).[1] Suppose $V^{\widehat{G}^{\mathrm{geom}}} = 0$*

---

[1] We will see in § 2.1.5 that V can be viewed as a $\widehat{G}^{\mathrm{Tt}}$-representation in a natural way, which is pure of weight $w = \langle 2\rho, \lambda \rangle$. Therefore $\mathrm{Kl}_{\widehat{G}}^V$ and $M_{!*,\mathbb{F}_p}^V$ are defined.

(here $\widehat{G}^{\mathrm{geom}} \subset \widehat{G}$ is the Zariski closure of the geometric monodromy of $\mathrm{Kl}_{\widehat{G}}$ for large p, to be recalled in Theorem 2.3.2). Let $w = \langle 2\rho, \lambda \rangle$ where 2ρ denotes the sum of positive coroots for $\widehat{G}$. Then for each prime ℓ there exists a finite-dimensional $\mathbb{Q}_\ell$-vector space M_ℓ with a continuous $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action and a $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-equivariant $(-1)^{w+1}$-symmetric perfect pairing

$$M_\ell \otimes M_\ell \to \mathbb{Q}_\ell(-w-1)$$

so that the Galois representation

$$\rho_\ell^V : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GO}(M_\ell) \text{ when } w \text{ is odd}$$
$$\text{or } \rho_\ell^V : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GSp}(M_\ell) \text{ when } w \text{ is even}$$

has the following properties.

(i) *It comes from geometry.* More precisely, there exists a smooth projective algebraic variety X over $\mathbb{Q}$ (independent of ℓ), of dimension $w-1$, such that ρ_ℓ^V is a subquotient of the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module $\mathrm{H}^{w-1}(X, \mathbb{Q}_\ell)(-1)$.

(ii) Let $p \neq \ell$ be a prime such that $\mathrm{Kl}_{\widehat{G}}^V$ does not have geometrically trivial subquotients. Then we have a canonical inclusion of Frob_p-modules

$$M_{!*,\mathbb{F}_p}^V \hookrightarrow \mathbb{Q}_\ell(\mu_p) \otimes \mathrm{Im}(M_\ell^{\mathcal{I}_p} \to (M_\ell)_{\mathcal{I}_p}). \tag{1.6}$$

The proof of the above theorem again has two key ingredients.

First is the interpretation of the moments $M_{!*,\mathbb{F}_p}^V$ in terms of *homogeneous Fourier transform* (Proposition 2.6.6). Unlike Fourier–Deligne transform which relies on working with a single characteristic p, homogeneous Fourier transform (introduced by Laumon) is a 'characteristic-free' version of Fourier–Deligne transform, and it serves as the bridge between moments in characteristic p and motives over $\mathbb{Q}$. We will give a quick review of Laumon's homogeneous Fourier transform in Appendix A. The algebraic variety X in Theorem 1.2.1(i) is closely related to affine Schubert varieties in the affine Grassmannian.

Another ingredient is that in order to study the ramification behavior of the Galois representations ρ_ℓ^V we need to work with $\mathbb{Z}_p$-models of various geometric objects such as affine Grassmannian and perverse sheaves on them. Here the results of Pappas and Zhu [PZ13] play a crucial role.

1.3 Towards more conjectures of Evans type

In view of Theorem 1.2.1 and the Langlands correspondence, Evans's conjectures in §1.1.4 should be the first few examples of a large list of modularity problems relating moments of generalized Kloosterman sums and automorphic forms. For example, we may start with a simply connected almost simple group $\widehat{G}$ and consider V-moments of the Kloosterman sheaf $\mathrm{Kl}_{\widehat{G}}$, where V is the irreducible representation of $\widehat{G}$ of highest weight λ. We then compute the dimension $d_{!*}^V$ of $M_{!*,\mathbb{F}_p}^V$ for large p.

(a) If $d_{!*}^V = 2$, and $\langle 2\rho, \lambda \rangle$ is odd, then we get Galois representations into GO_2. The V-moment of $\mathrm{Kl}_{\widehat{G}}$ should be expressible in terms of Fourier coefficients of a CM modular form.

Examples: $(\widehat{G}, V) = (\mathrm{SL}_2, \mathrm{Sym}^5)$, $(\mathrm{SL}_4 \text{ or } \mathrm{Sp}_4, \mathrm{Sym}^3)$.

(b) If $d_{!*}^V = 2$, and $\langle 2\rho, \lambda \rangle$ is even, then we get two-dimensional Galois representations, and the V-moment of $\mathrm{Kl}_{\widehat{G}}$ should be expressible in terms of Fourier coefficients of a modular form.

Examples:

$$(\widehat{G}, V) = (\mathrm{SL}_2, \mathrm{Sym}^6), (\mathrm{SL}_2, \mathrm{Sym}^8),$$
$$(\mathrm{SL}_3, \mathrm{Sym}^4), (\mathrm{SL}_3, V_{2\omega_1+\omega_2}), (\mathrm{SL}_3, V_{2\omega_1+2\omega_2}),$$
$$(\mathrm{SL}_4 \text{ or } \mathrm{Sp}_4, \mathrm{Sym}^4) \text{ and its dual}, (\mathrm{Sp}_4, V_{3\omega_2}).$$

Here $\{\omega_1, \omega_2\}$ denote the fundamental weights of SL_3 or Sp_4, with ω_1 corresponding to the standard representation.

(c) If $d_{!*}^V = 3$ (in this case $\langle 2\rho, \lambda \rangle$ is necessarily odd), then we get Galois representations into $\mathrm{O}_3 \cong \mathrm{PGL}_2 \times \{\pm 1\}$. The PGL_2-part of these representations should be liftable to GL_2, and we should again expect that the V-moment of $\mathrm{Kl}_{\widehat{G}}$ be expressible in terms of a quadratic Dirichlet character and the Fourier coefficients of a modular form.

Example: $(\widehat{G}, V) = (\mathrm{SL}_2, \mathrm{Sym}^7)$.

(d) If $d_{!*}^V = 4$ and $\langle 2\rho, \lambda \rangle$ is odd, then we get Galois representations into O_4. If, moreover, the determinant character corresponds to a *real* quadratic field $F/\mathbb{Q}$ and if $\mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathrm{SO}_4$ lifts to $\mathrm{GSpin}_4 \hookrightarrow \mathrm{GL}_2 \times \mathrm{GL}_2$, then the V-moment of $\mathrm{Kl}_{\widehat{G}}$ should be expressible in terms of Fourier coefficients of two Hilbert modular forms for GL_2 over F.

Example: $(\widehat{G}, V) = (\mathrm{SL}_2, \mathrm{Sym}^9)$.

(e) If $d_{!*}^V = 4$ and $\langle 2\rho, \lambda \rangle$ is even, then we get Galois representations into GSp_4. We should then expect that the V-moment of $\mathrm{Kl}_{\widehat{G}}$ be expressible in terms of a Siegel modular form for GSp_4 over $\mathbb{Q}$.

Examples:

$$(\widehat{G}, V) = (\mathrm{SL}_2, \mathrm{Sym}^{10}), (\mathrm{SL}_2, \mathrm{Sym}^{12}),$$
$$(\mathrm{SL}_3, \mathrm{Sym}^5), (\mathrm{SL}_3, V_{3\omega_1+\omega_2}), (\mathrm{SL}_3, \mathrm{Sym}^6) \text{ and their duals},$$
$$(G_2, V_{\omega_1+\omega_2}), (G_2, V_{2\omega_2}), (G_2, V_{3\omega_1}).$$

Here ω_1 (respectively ω_2) are the dominant short (respectively long) root of G_2.

And the list continues. We plan to investigate some of these examples in detail in the future.

1.4 Convention

Throughout the paper, G will be a connected, split, simple (hence *adjoint*) group over a field k. We fix a maximal torus $T \subset G$ and a Borel subgroup B containing T. This way we get a based root system and the Weyl group W.

We also fix a prime ℓ which is different from $\mathrm{char}(k)$. All sheaves will be $\overline{\mathbb{Q}}_\ell$-sheaves in the étale topology of schemes or stacks.

Caution: all sheaf-theoretic functors are derived functors unless otherwise stated.

When k is a finite field, Frob_k will denote the *geometric* Frobenius element in $\mathrm{Gal}(\overline{k}/k)$. Similarly, for a prime p, Frob_p denotes the *geometric* Frobenius element in $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$.

2. Geometric setup for the moments

In this section, we give geometric interpretation of moments of Kloosterman sums. We will work with the more general Kloosterman sheaves constructed in our earlier work [HNY13]. After some preparation on the geometric Satake equivalence in §2.1, we review the construction of generalized Kloosterman sheaves and their monodromy properties in §§2.2–2.4. Finally in §§2.5–2.7 we define moments of Kloosterman sheaves and study them via homogeneous Fourier transform.

2.1 Deligne's modification of the Langlands dual group and the Satake–Tate category

In this subsection, k is either a finite field or a number field. Deligne has introduced a modification of the Langlands dual group in the setup of the global Langlands correspondence, in order to avoid making half Tate twists. We reformulate his construction from the viewpoint of the geometric Satake equivalence. For a quick algebraic account, see [FG09, § 2].

2.1.1 *Tate sheaves.* The Tate sheaf $\mathbb{Q}_\ell(1)$ is defined as a $\mathbb{Q}_\ell$-sheaf on $\mathrm{Spec}\, k$, using the inverse system $\varprojlim_n \mu_{\ell^n}$ of ℓ-power roots of unity in $\overline{k}$. Let δ^{Tt} be the full subcategory of $\mathbb{Q}_\ell$-sheaves on $\mathrm{Spec}\, k$ consisting of finite direct sums of $\mathbb{Q}_\ell(n)$ for $n \in \mathbb{Z}$. Then δ^{Tt} admits a structure of a tensor category using tensor product of sheaves. Fixing an algebraic closure $\overline{k}$ of k, the functor of taking the $\overline{k}$-stalk of a sheaf gives a fiber functor $\omega : \delta^{\mathrm{Tt}} \to \mathrm{Vect}_{\mathbb{Q}_\ell}$, and realizes a tensor equivalence

$$\delta^{\mathrm{Tt}} \cong \mathrm{Rep}(\mathbb{G}_m^{\mathrm{Tt}}).$$

Here $\mathbb{G}_m^{\mathrm{Tt}}$ is simply the multiplicative group $\mathbb{G}_m$. Under this equivalence, the tautological one-dimensional representation of $\mathbb{G}_m^{\mathrm{Tt}}$ corresponds to $\mathbb{Q}_\ell(1)$.

2.1.2 *The geometric Satake equivalence.* Let LG be the loop group of G: this is an ind-scheme representing the functor $R \mapsto G(R[[t]])$. Let L^+G the positive loops of G: this is a scheme of infinite type representing the functor $R \mapsto G(R[[t]])$. The fppf quotient $\mathrm{Gr} = LG/L^+G$ is called the *affine Grassmannian* of G. Then L^+G acts on Gr via left translation. The L^+G-orbits on Gr are indexed by dominant coweights $\lambda \in \mathbb{X}_*(T)^+$. The orbit containing the element $t^\lambda \in T(k((t)))$ is denoted by Gr_λ and its closure is denoted $\mathrm{Gr}_{\leqslant\lambda}$. We have $\dim \mathrm{Gr}_\lambda = \langle 2\rho, \lambda \rangle$, where 2ρ denotes the sum of positive roots in G. The reduced scheme structure on $\mathrm{Gr}_{\leqslant\lambda}$ is a projective variety called the affine Schubert variety attached to λ. We denote the intersection complex of $\mathrm{Gr}_{\leqslant\lambda}$ by IC_λ: this is the middle extension of the shifted constant sheaf $\mathbb{Q}_\ell[\langle 2\rho, \lambda\rangle]$ on Gr_λ.

The *Satake category* $\mathrm{Sat} = \mathrm{Perv}_{L^+G}(\mathrm{Gr})$ is the category of L^+G-equivariant perverse étale sheaves on Gr (over k) whose support is of finite type. Similarly, we define $\mathrm{Sat}^{\mathrm{geom}}$ by considering the base change of the situation to $\overline{k}$. Then we have a pullback functor $\mathrm{Sat} \to \mathrm{Sat}^{\mathrm{geom}}$.

In [Lus83, Gin95] and [MV07], it was shown that $\mathrm{Sat}^{\mathrm{geom}}$ carries a natural tensor structure (which is also defined for Sat), such that the global cohomology functor $h = H^*(\mathrm{Gr}, -) : \mathrm{Sat}^{\mathrm{geom}} \to \mathrm{Vect}$ is a fiber functor. It is also shown that the Tannakian group of the tensor category $\mathrm{Sat}^{\mathrm{geom}}$ is a connected reductive split group over $\mathbb{Q}_\ell$ whose root system is dual to that of G. We denote this group by $\widehat{G}$. The Tannakian formalism gives the *geometric Satake equivalence* of tensor categories

$$\mathrm{Sat}^{\mathrm{geom}} \cong \mathrm{Rep}(\widehat{G}, \mathbb{Q}_\ell).$$

2.1.3 *The Satake–Tate category.* Let $\mathrm{Sat}^{\mathrm{Tt}}$ be the full subcategory of Sat whose objects are finite direct sums of the form $\bigoplus_{\lambda_i, n_i} \mathrm{IC}_{\lambda_i}(n_i)$, where $\lambda_i \in \mathbb{X}_*(T)^+$ and $n_i \in \mathbb{Z}$. By [AB09, § 3.5], the category $\mathrm{Sat}^{\mathrm{Tt}}$ is closed under the tensor product in Sat, and $\mathrm{Sat}^{\mathrm{Tt}}$ is itself a rigid tensor category. Alternatively, one can define $\mathrm{Sat}^{\mathrm{Tt}}$ as the smallest tensor subcategory of Sat containing the objects IC_λ which is also stable under Tate twists. We have an embedding of tensor categories $\delta^{\mathrm{Tt}} \subset \mathrm{Sat}^{\mathrm{Tt}}$ sending $\mathbb{Q}_\ell(n)$ to $\mathrm{IC}_0(n)$, here IC_0 is the skyscraper sheaf supported at the singleton orbit Gr_0.

We have tensor functors

$$
\begin{array}{ccc}
\mathrm{Sat}^{w=0} & & \delta^{\mathrm{Tt}} \\
& \searrow \quad \swarrow & \\
& \mathrm{Sat}^{\mathrm{Tt}} & \\
& \swarrow \quad \searrow \mathrm{H}^{\mathrm{Wt}} & \\
\mathrm{Sat}^{\mathrm{geom}} & & \mathrm{Vect}^{\mathrm{gr}}
\end{array}
\tag{2.1}
$$

We explain the notation. The functor H^{Wt} is the global section functor H^* with the grading by weights. If we write an object $\mathcal{K} \in \mathrm{Sat}^{\mathrm{Tt}}$ as $\mathcal{K} = \bigoplus_i \mathcal{K}_i$, where $\mathcal{K}_i$ is pure of weight i, then $\mathrm{H}^{\mathrm{Wt}}(\mathcal{K})$ is the graded vector space $\bigoplus_i h(\mathcal{K}_i)$. The category $\mathrm{Sat}^{w=0}$ is the full subcategory of $\mathrm{Sat}^{\mathrm{Tt}}$ consisting of perverse sheaves which are pure of weight zero.

2.1.4 *Modified Langlands dual group.* The global section functor $h^{\mathrm{Tt}} = \mathrm{H}^*(\mathrm{Gr}, -) : \mathrm{Sat}^{\mathrm{Tt}} \to \mathrm{Vect}_{\mathbb{Q}_\ell}$ is a fiber functor, which gives the Tannakian group $\widehat{G}^{\mathrm{Tt}} := \mathrm{Aut}^\otimes(h^{\mathrm{Tt}})$ so that we have a tensor equivalence

$$
\mathrm{Sat}^{\mathrm{Tt}} \cong \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}}, \mathbb{Q}_\ell).
$$

For $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, we shall denote the corresponding object in $\mathrm{Sat}^{\mathrm{Tt}}$ by IC_V.

By Tannakian duality, the diagram (2.1) gives a diagram of reductive groups over $\mathbb{Q}_\ell$

$$
\begin{array}{ccc}
\widehat{G}^{\mathrm{ev}} & & \mathbb{G}_m^{\mathrm{Tt}} \\
\uparrow \nwarrow & & \nearrow \uparrow \\
\pi \quad & \searrow^{\mathrm{Tt}} \widehat{G}^{\mathrm{Tt}} \swarrow^{[-2]} & \quad [-2] \\
\downarrow \swarrow^\iota & \nwarrow_{\mathrm{Wt}} & \downarrow \\
\widehat{G} & & \mathbb{G}_m^{\mathrm{Wt}}
\end{array}
$$

where the two diagonal sequences are short exact. The reductive group $\widehat{G}^{\mathrm{ev}}$ is defined as the Tannakian dual of $\mathrm{Sat}^{w=0}$, and $\mathbb{G}_m^{\mathrm{Wt}}$ is a one-dimensional torus.

2.1.5 *Irreducible objects.* Let λ be a dominant coweight of T, then IC_λ is a complex pure of weight $\langle 2\rho, \lambda \rangle$. The corresponding object in $\widehat{G}^{\mathrm{Tt}}$ is an irreducible representation V_λ^{Tt} of $\widehat{G}^{\mathrm{Tt}}$. What is this representation? Its restriction to $\widehat{G}$ is the irreducible representation of highest weight λ. Since IC_λ is pure of weight $\langle 2\rho, \lambda \rangle$, $\mathbb{G}_m^{\mathrm{Wt}}$ acts on V_λ^{Tt} by scalar multiplication via the $\langle 2\rho, \lambda \rangle$th power of the tautological character. If we look at the action of $\widehat{G} \times \mathbb{G}_m^{\mathrm{Wt}}$ on V_λ^{Tt}, then the element $((-1)^{2\rho}, -1) \in \widehat{G} \times \mathbb{G}_m^{\mathrm{Wt}}$ acts trivially. Here 2ρ is viewed as a coweight of $\widehat{G}$. Note that $(-1)^{2\rho}$ always lies in the center of $\widehat{G}$.

We use $\mathbb{Q}_\ell(n)$ to denote the one-dimensional representation of $\widehat{G}^{\mathrm{Tt}}$ given by the nth power of the character $\mathrm{Tt} : \widehat{G}^{\mathrm{Tt}} \to \mathbb{G}_m^{\mathrm{Tt}}$. We also use $V(n)$ to mean $V \otimes \mathbb{Q}_\ell(n)$, as we do for sheaves.

Since every irreducible object in $\mathrm{Sat}^{\mathrm{Tt}}$ is of the form $\mathrm{IC}_\lambda(n)$ for some dominant coweight λ and $n \in \mathbb{Z}$, every irreducible representation of $\widehat{G}^{\mathrm{Tt}}$ is of the form $V_\lambda^{\mathrm{Tt}}(n)$. The action of $\widehat{G}$ on $V_\lambda^{\mathrm{Tt}}(n)$ is still the same as its action on V_λ, and the action of $\mathbb{G}_m^{\mathrm{Wt}}$ on it is through the $(\langle 2\rho, \lambda \rangle - 2n)$th power of the tautological character.

LEMMA 2.1.6. *The map* $\mathrm{Wt} : \mathbb{G}_m^{\mathrm{Wt}} \to \widehat{G}^{\mathrm{Tt}}$ *is central. The homomorphism* $\iota \times \mathrm{Wt} : \widehat{G} \times \mathbb{G}_m^{\mathrm{Wt}} \to \widehat{G}^{\mathrm{Tt}}$ *is an isogeny whose kernel is isomorphic to* μ_2 *generated by* $((-1)^{2\rho}, -1)$ *(which is central).*

Proof. Every irreducible object in $\mathrm{Sat}^{\mathrm{Tt}}$ is pure of some weight, hence $\mathbb{G}_m^{\mathrm{Wt}}$ acts on the corresponding representation by scalars. Therefore $\mathbb{G}_m^{\mathrm{Wt}}$ is central.

Let $\widehat{G}_1 := \widehat{G} \times \mathbb{G}_m^{\mathrm{Wt}}/\mu_2$ where the μ_2 is generated by $((-1)^{2\rho}, -1)$. We have seen from the discussion in § 2.1.5 that the action of $\widehat{G} \times \mathbb{G}_m^{\mathrm{Wt}}$ on every irreducible representations of $\widehat{G}^{\mathrm{Tt}}$ factors through $\widehat{G}_1$. Therefore, $\iota \times \mathrm{Wt}$ factors through a homomorphism $\phi : \widehat{G}_1 \to \widehat{G}^{\mathrm{Tt}}$. We then have a commutative diagram

$$
\begin{array}{ccccccccc}
1 & \longrightarrow & \widehat{G} & \longrightarrow & \widehat{G}_1 & \longrightarrow & \mathbb{G}_m^{\mathrm{Wt}}/\mu_2 & \longrightarrow & 1 \\
& & \| & & \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\iota} & & \\
1 & \longrightarrow & \widehat{G} & \longrightarrow & \widehat{G}^{\mathrm{Tt}} & \xrightarrow{\ \mathrm{Tt}\ } & \mathbb{G}_m^{\mathrm{Tt}} & \longrightarrow & 1
\end{array}
$$

where the right-most map is the isomorphism induced from $[-2] : \mathbb{G}_m^{\mathrm{Wt}} \to \mathbb{G}_m^{\mathrm{Tt}}$. Both the upper and lower sequences are exact, we conclude that ϕ is an isomorphism. $\qquad\square$

DEFINITION 2.1.7. An object $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$ is *pure of weight w*, if the action of the weight torus $\mathbb{G}_m^{\mathrm{Wt}}$ on V (via $\mathbb{G}_m^{\mathrm{Wt}} \xrightarrow{\mathrm{Wt}} \widehat{G}^{\mathrm{Tt}} \to \mathrm{GL}(V)$) is by the wth power (as scalar multiplication).

If $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$ is pure of weight w, the corresponding perverse sheaf IC_V is pure of the same weight w.

The following two examples will be important for our study of classical Kloosterman sums.

Example 2.1.8. Let $n \geqslant 2$ be an integer and $G = \mathrm{PGL}_n$. We have $\widehat{G} = \mathrm{SL}_n$. In this case, $(-1)^{2\rho} = (-1)^{n-1}I_n \in \widehat{G}$. We have

$$
\widehat{G}^{\mathrm{Tt}} \cong \begin{cases} \mathrm{SL}_n \times (\mathbb{G}_m^{\mathrm{Wt}}/\mu_2) & \text{if } n \text{ is odd,} \\ (\mathrm{SL}_n \times \mathbb{G}_m^{\mathrm{Wt}})/\mu_2 & \text{if } n \text{ is even (μ_2 is the diagonally embedded central subgroup).} \end{cases}
$$

We define a 'standard representation' of dimension n and weight $n-1$:

$$
\mathrm{St} : \widehat{G}^{\mathrm{Tt}} \to \mathrm{GL}_n
$$
$$
(g, t) \in \mathrm{SL}_n \times \mathbb{G}_m^{\mathrm{Wt}} \mapsto t^{n-1}g \in \mathrm{GL}_n.
$$

The formula we gave above is a homomorphism from $\mathrm{SL}_n \times \mathbb{G}_m^{\mathrm{Wt}}$ to GL_n, and it is easy to check that it factors through $\widehat{G}^{\mathrm{Tt}}$.

Example 2.1.9. Let $n \geqslant 2$ be an even integer and $G = \mathrm{SO}_{n+1}$. We have $\widehat{G} = \mathrm{Sp}_n$. In this case, $(-1)^{2\rho} = -I_n \in \widehat{G}$ and $\widehat{G}^{\mathrm{Tt}} \cong (\mathrm{Sp}_n \times \mathbb{G}_m^{\mathrm{Wt}})/\mu_2^{\Delta} \cong \mathrm{GSp}_n$. We also define a 'standard representation' of dimension n and weight $n-1$

$$
\mathrm{St} : \widehat{G}^{\mathrm{Tt}} \to \mathrm{GSp}_n \subset \mathrm{GL}_n
$$
$$
(g, t) \in \mathrm{Sp}_n \times \mathbb{G}_m^{\mathrm{Wt}} \mapsto t^{n-1}g \in \mathrm{GSp}_n.
$$

2.2 Construction of Kloosterman sheaves

In this subsection, the base field k is a finite field of characteristic $p > 0$. We fix a nontrivial additive character $\psi : \mathbb{F}_p \to \overline{\mathbb{Q}}_\ell(\mu_p)^\times$, which gives rise to an Artin–Schreier sheaf AS_ψ on $\mathbb{A}_k^1$.

We recall the construction of the $\widehat{G}$-Kloosterman local system on $\mathbb{G}_m$ in [HNY13]. Here we will actually give an enhancement of it to a $\widehat{G}^{\mathrm{Tt}}$-local system.

Let $\pi : \mathrm{GR} \to \mathbb{G}_m$ be the Beilinson–Drinfeld affine Grassmannian over the curve $\mathbb{G}_m$ (see [HNY13, Remark 2.8(1)]). In our case, there is a one-dimensional torus $\mathbb{G}_m^{\mathrm{rot}}$ acting on $\mathbb{G}_m$ by multiplication. This action induces an action of $\mathbb{G}_m^{\mathrm{rot}}$ on GR such that π is $\mathbb{G}_m^{\mathrm{rot}}$-equivariant.

Using the $\mathbb{G}_m^{\mathrm{rot}}$-action, the fibration π can be trivialized, and we may identify GR with $\mathrm{Gr} \times \mathbb{G}_m$. Let $\pi_{\mathrm{Gr}} : \mathrm{GR} \to \mathrm{Gr}$ be the projection. For $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, let $\mathrm{IC}_{V,\mathrm{GR}} = \pi_{\mathrm{Gr}}^* \mathrm{IC}_V$. Note that $\mathrm{IC}_{V,\mathrm{GR}}[1]$ is a perverse sheaf.

In [HNY13, §5.2] we defined an open sub-ind-scheme $\mathrm{GR}^\circ \subset \mathrm{GR}$ with a morphism $f : \mathrm{GR}^\circ \to \mathbb{G}_a^{r+1}$, where r is the semisimple rank of G:

$$
\begin{array}{c}
\mathrm{GR}^\circ \\
\pi^\circ \swarrow \quad f \searrow \quad \searrow F \\
\mathbb{G}_m \qquad\qquad \mathbb{G}_a^{r+1} \xrightarrow{\sigma} \mathbb{A}^1
\end{array}
\tag{2.2}
$$

Let $\mathrm{IC}_{V,\mathrm{GR}^\circ}$ be the restriction of $\mathrm{IC}_{V,\mathrm{GR}}$ to GR°.

One of the main results of [HNY13] can be stated as follows.

THEOREM 2.2.1 (Heinloth *et al.* [HNY13, Theorem 1(1)]). *For* $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, *we have the following isomorphism in* $D^b(\mathbb{G}_m, \mathbb{Q}_\ell(\mu_p))$

$$
\pi_!^\circ(\mathrm{IC}_{V,\mathrm{GR}^\circ} \otimes F^* \mathrm{AS}_\psi) \cong \pi_*^\circ(\mathrm{IC}_{V,\mathrm{GR}^\circ} \otimes F^* \mathrm{AS}_\psi)
\tag{2.3}
$$

given by the natural transformation $\pi_!^\circ \to \pi_*^\circ$. *(The tensor product above is taken over* $\mathbb{Q}_\ell$.*) Let* Kl_G^V *be any one of the two complexes above. Then* Kl_G^V *is a local system on* $\mathbb{G}_m$. *The assignment* $V \mapsto \mathrm{Kl}_G^V$ *gives a tensor functor*

$$
\mathrm{Kl}_{\widehat{G}} : \mathrm{Rep}_{\mathbb{Q}_\ell(\mu_p)}(\widehat{G}^{\mathrm{Tt}}) \to \mathrm{Loc}_{\mathbb{Q}_\ell(\mu_p)}(\mathbb{G}_m).
$$

In other words, we get a $\widehat{G}^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p))$-*local system* $\mathrm{Kl}_{\widehat{G}}$ *on* $\mathbb{G}_m$.

Note that the only difference between the above theorem and [HNY13, Theorem 1] is that we have replaced the usual Langlands dual group $\widehat{G}$ by Deligne's modified Langlands dual group $\widehat{G}^{\mathrm{Tt}}$. The proof in [HNY13] works without change. When $\widehat{G}$ is clear from the context, we simply write Kl^V for $\mathrm{Kl}_{\widehat{G}}^V$.

Remark 2.2.2. The Kloosterman sheaf $\mathrm{Kl}_{\widehat{G}}$ gives rise to a monodromy representation, well-defined up to conjugacy

$$
\rho_{\widehat{G}}^{\mathrm{Tt}} : \pi_1(\mathbb{G}_m/k) \to \widehat{G}^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p)).
$$

The monodromy representation $\rho_{\widehat{G}}^{\mathrm{Tt}}$ is compatible with the Tate torus in the following sense. The composition

$$
\pi_1(\mathbb{G}_m/k) \xrightarrow{\rho_{\widehat{G}}^{\mathrm{Tt}}} \widehat{G}^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p)) \xrightarrow{\mathrm{Tt}} \mathbb{G}_m^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p)) = \mathbb{Q}_\ell(\mu_p)^\times
$$

factors through $\pi_1(\mathbb{G}_m/k) \to \mathrm{Gal}(\overline{k}/k)$ and is given by the ℓ-adic cyclotomic character $\chi_{\mathrm{cyc}} : \mathrm{Gal}(\overline{k}/k) \to \mathbb{Z}_\ell^\times = \mathrm{Aut}(\mathbb{Z}_\ell(1))$, where $\mathbb{Z}_\ell(1) = \varprojlim_n \mu_{\ell^n}(\overline{k})$.

On the other hand, if an object $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$ is pure of weight w, the corresponding local system Kl^V is also pure of weight w.

2.3 Geometric monodromy of Kloosterman sheaves

2.3.1 *Global monodromy.* We also consider the restriction of $\rho_{\widehat{G}}^{\mathrm{Tt}}$ to the geometric fundamental group

$$
\rho_{\widehat{G}}^{\mathrm{geom}} : \pi_1(\mathbb{G}_m/\overline{k}) \to \widehat{G}^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p)).
$$

Let $\widehat{G}^{\mathrm{geom}}$ be the Zariski closure of the image of $\rho_{\widehat{G}}^{\mathrm{geom}}$. Since the composition $\pi_1(\mathbb{G}_m/\overline{k}) \to \widehat{G}^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p)) \to \mathbb{G}_m^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p))$ factors through $\mathrm{Gal}(\overline{k}/k)$ (Remark 2.2.2), the image of $\rho_{\widehat{G}}^{\mathrm{geom}}$ lies in $\widehat{G} = \ker(\widehat{G}^{\mathrm{Tt}} \to \mathbb{G}_m^{\mathrm{Tt}})$, and hence $\widehat{G}^{\mathrm{geom}} \subset \widehat{G}$.

THEOREM 2.3.2 (Heinloth *et al.* [HNY13, Theorem 3]). *The subgroup $\widehat{G}^{\mathrm{geom}}$ of $\widehat{G}$ is a connected and simply-connected group of types given by the following table (with restrictions on p stated in the third column)*

$\widehat{G}$	$\widehat{G}^{\mathrm{geom}}$	condition
A_{2n}	A_{2n}	$p > 2$
A_{2n-1}, C_n	C_n	$p > 2$
B_n, D_{n+1} $(n \geqslant 4)$	B_n	$p > 2$
E_7	E_7	$p > 2$
E_8	E_8	$p > 2$
E_6, F_4	F_4	$p > 2$
B_3, D_4, G_2	G_2	$p > 3$

We have $\widehat{G}^{\mathrm{geom}} = \widehat{G}^{\mathrm{Out}(\widehat{G}),\circ}$ ($\mathrm{Out}(\widehat{G})$ acting on $\widehat{G}$ as pinned automorphisms) unless $\widehat{G}$ is of type A_{2n} or B_3.

Let $\mathcal{I}_0$ and $\mathcal{I}_\infty$ be the inertia groups at 0 and ∞. These are subgroups of the geometric fundamental group $\pi_1(\mathbb{G}_m/\overline{k})$, well-defined up to conjugacy. We shall describe the image of $\mathcal{I}_0$ and $\mathcal{I}_\infty$ under the monodromy representation $\rho_{\widehat{G}}^{\mathrm{geom}} : \pi_1(\mathbb{G}_m/\overline{k}) \to \widehat{G}(\mathbb{Q}_\ell(\mu_p))$.

2.3.3 *Local monodromy at 0.* By [HNY13, Theorem 1(2)], we know that $\rho_{\widehat{G}}^{\mathrm{geom}}$ is trivial on the wild inertia $\mathcal{I}_0^w$ (i.e. $\mathrm{Kl}_{\widehat{G}}$ is tame at 0) and sends a topological generator of the tame inertia $\mathcal{I}_0^t$ to a regular unipotent element in $\widehat{G}$.

2.3.4 *Local monodromy at ∞ for large p.* The description of the local monodromy of $\mathrm{Kl}_{\widehat{G}}$ at ∞ for large p is obtained in [HNY13] based on the work of Gross and Reeder on simple wild parameters [GR10]. We review the results here. We assume p does not divide $\#W$ (W is also identified with the Weyl group of $\widehat{G}$).

We have the lower numbering filtration for the image of $\mathcal{I}_\infty$ under $\rho_{\widehat{G}}^{\mathrm{geom}}$ (see [Ser68, ch. IV, Proposition 1])

$$D_0 = \rho_{\widehat{G}}^{\mathrm{geom}}(\mathcal{I}_\infty) \rhd D_1 = \rho_{\widehat{G}}^{\mathrm{geom}}(\mathcal{I}_\infty^w) \rhd D_2 \rhd D_3 \rhd \cdots \tag{2.4}$$

where $\mathcal{I}_\infty^w$ is the wild inertia at ∞. Let $\widehat{G}^{\mathrm{ad}}$ be the adjoint form of $\widehat{G}$, which is also the adjoint form of $\widehat{G}^{\mathrm{Tt}}$. Let $\overline{D}_0 = \rho_{\widehat{G}^{\mathrm{ad}}}^{\mathrm{geom}}(\mathcal{I}_\infty) \rhd \overline{D}_1 = \rho_{\widehat{G}^{\mathrm{ad}}}^{\mathrm{geom}}(\mathcal{I}_\infty^w) \rhd \overline{D}_2 \rhd \cdots$ be the lower numbering filtration for the image of $\mathcal{I}_\infty$ in $\widehat{G}^{\mathrm{ad}}$. We first describe $\overline{D}_0$ and $\overline{D}_1$ by combining [HNY13, Theorem 2, Corollary 2.15] and [GR10, Proposition 5.6].

There is a unique maximal torus $\widehat{T}^{\mathrm{ad}}$ of $\widehat{G}^{\mathrm{ad}}$ such that $\overline{D}_1 \subset \widehat{T}^{\mathrm{ad}}[p]$. The group $\overline{D}_0$ lies in the normalizer $N(\widehat{T}^{\mathrm{ad}})$ of $\widehat{T}^{\mathrm{ad}}$. The quotient $\overline{D}_0/\overline{D}_1$ is cyclic of order h (the Coxeter number of G), and any generator of $\overline{D}_0/\overline{D}_1$ maps to a Coxeter element Cox in W under the map $\overline{D}_0/\overline{D}_1 \to N(\widehat{T}^{\mathrm{ad}})/\widehat{T}^{\mathrm{ad}} = W$. Moreover, $\overline{D}_1 \subset \widehat{T}^{\mathrm{ad}}[p]$ is isomorphic to the additive group of some finite field $\mathbb{F}_q = \mathbb{F}_p[\zeta]$ where ζ is a primitive hth root of unity in $\overline{\mathbb{F}}_p$. The action of a generator of $\overline{D}_0/\overline{D}_1$ on $\overline{D}_1 \cong \mathbb{F}_p[\zeta]$ is by multiplication by ζ.

We then describe D_0 and D_1 (which is a p-group mapping onto $\overline{D}_1$). Since p is prime to $\#Z\widehat{G}$, $D_1 \subset \widehat{G}$ is the unique p-group mapping onto $\overline{D}_1$, and D_1 lies in $\widehat{T}[p]$ (where $\widehat{T}$ is the preimage of $\widehat{T}^{\mathrm{ad}}$ in $\widehat{G}$). Again $D_0 \subset N(\widehat{T})$ because $N(\widehat{T})$ is the preimage of $N(\widehat{T}^{\mathrm{ad}})$ in $\widehat{G}$. Since D_0/D_1 is cyclic, D_0 is of the form $D_1 \rtimes \langle \widetilde{\mathrm{Cox}} \rangle$ where $\widetilde{\mathrm{Cox}} \in N(\widehat{T})$ is a lifting of a Coxeter element $\mathrm{Cox} \in W$. Any two liftings of Cox to $N(\widehat{T})$ are conjugate under $\widehat{T}$. Note that when $\widehat{G}$ is not adjoint, the order of $\widetilde{\mathrm{Cox}}$ is a multiple eh of h, but may not be h.

Example 2.3.5. Let $G = \mathrm{PGL}_n$, $\widehat{G} = \mathrm{SL}_n$. The Coxeter elements in the Weyl group $W = S_n$ are cyclic permutations of length n. When n is odd, any lifting of such an element to $N_{\widehat{G}}(\widehat{T})$ still has order n. However, when n is even, any lifting of a Coxeter element in S_n to $N_{\widehat{G}}(\widehat{T})$ has order $2n$.

When n is even and $G = \mathrm{SO}_{n+1}$, $\widehat{G} = \mathrm{Sp}_n$. Then we are in the same situation as $\widehat{G} = \mathrm{SL}_n$: the Coxeter element has order n but any lifting of it to $N_{\widehat{G}}(\widehat{T})$ has order $2n$.

Recall the Swan conductor of the representation $\rho_{\widehat{G}}^{\mathrm{geom}}|_{\mathcal{I}_\infty} : \mathcal{I}_\infty \to \widehat{G} \to \mathrm{GL}(V)$ is defined as

$$\mathrm{Swan}_\infty(V) = \sum_{j \geqslant 1} \frac{\dim V - \dim V^{D_j}}{[D_0 : D_j]}$$

where D_j are the lower numbering filtration subgroups of D_0 in (2.4) (see [Ser68, ch. VI, Proposition 2] for the closely related Artin conductor; see also [GR10, §2]).

LEMMA 2.3.6. *Let $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$. Assume $p \nmid \#W$. Then*

$$\mathrm{Swan}_\infty(V) = \frac{\dim V - \dim V^{D_1}}{h}.$$

Proof. Let eh be the order of $\widetilde{\mathrm{Cox}}$, then $D_0/D_1 \cong \mathbb{Z}/eh\mathbb{Z}$. The lower numbering filtration of D_0 takes the form

$$D_0 \rhd D_1 = D_2 = \cdots = D_e \rhd D_{e+1} = \{1\}. \tag{2.5}$$

In fact, let $K = \overline{k}((t^{-1}))$ and let L/K be Galois extension with Galois group isomorphic to $D_0 = \rho_{\widehat{G}}^{\mathrm{geom}}(\mathcal{I}_\infty)$. The quotient $\overline{D}_0$ corresponds to a subextension $L' \subset L$. Let K_t (respectively K_t') be the maximal tamely ramified extension of K in L (respectively L'). Since $D_1 = \overline{D}_1$, L is the compositum $L'K_t$ with $L' \cap K_t = K_t'$. Pick a generating element $x \in \mathcal{O}_{L'}$ over $\mathcal{O}_{K_t'}$. Then it is also a generating element of $\mathcal{O}_L$ over $\mathcal{O}_{K_t}$. For any $g \in D_1$, we then have $\mathrm{val}_L(g(x) - x) = e\,\mathrm{val}_{L'}(g(x) - x)$. Therefore, the jumps of the lower number filtration $\{D_i\}$ are exactly e times the jumps of $\{\overline{D}_i\}$. Since we have $\overline{D}_2 = \{1\}$, the sequence $\{D_i\}$ must look like (2.5). We then have

$$\mathrm{Swan}_\infty(V) = \sum_{j \geqslant 1} \frac{\dim V - \dim V^{D_j}}{[D_0 : D_j]} = \sum_{j=1}^{e} \frac{\dim V - \dim V^{D_1}}{eh} = \frac{\dim V - \dim V^{D_1}}{h}. \qquad \square$$

2.4 The classical Kloosterman sheaf

When $G = \mathrm{PGL}_n$, let us consider the following diagram.

$$\begin{array}{ccc} & \mathbb{G}_m^n & \\ {\scriptstyle \pi}\swarrow & \downarrow{\scriptstyle f} \searrow{\scriptstyle F} & \\ \mathbb{G}_m & \mathbb{G}_a^n \xrightarrow{\ \sigma\ } \mathbb{A}^1 & \end{array} \tag{2.6}$$

Here the morphism π is given by taking the product $\pi(x_1, \ldots, x_n) = x_1 \cdots x_n$; f is the natural inclusion; σ is given by taking the sum $\sigma(x_1, \ldots, x_n) = x_1 + \cdots + x_n$; finally $F = \sigma \circ f$.

Deligne defined

$$\mathrm{Kl}_n := m_! F^* \mathrm{AS}_\psi[n-1].$$

It is known [Del77, Théorème 7.8] that Kl_n is concentrated in degree zero, and is a local system on $\mathbb{G}_{m,\mathbb{F}_p}$ of rank n. Moreover, Kl_n is pure of weight $n-1$.

2.4.1 *Determinant.* For the one-dimensional representation det of GL_n, $\mathrm{Kl}_n^{\det} = \det(\mathrm{Kl}_n)$. Deligne [Del77, (7.15.2)] has determined that

$$\det \mathrm{Kl}_n = \mathbb{Q}_\ell \left(-\frac{n(n-1)}{2} \right)$$

as Frob-modules.

2.4.2 *Symplectic pairing.* When n is even, there is a symplectic pairing (see [Del77, § 7.5])

$$\mathrm{Kl}_n \otimes \mathrm{Kl}_n \to \mathbb{Q}_\ell(-n+1).$$

The diagram (2.6) is very similar to (2.2). In fact, when we work with $G = \mathrm{PGL}_n$ and V the standard representation of $\widehat{G}^{\mathrm{Tt}}$ (see Example 2.1.8), the sheaf IC_V is the shifted constant sheaf $\mathbb{Q}_\ell[n-1]$ supported on a Schubert variety in Gr which isomorphic to $\mathbb{P}^{n-1}$. Therefore, $\mathrm{IC}_{V,\mathrm{GR}}$ is the shifted constant sheaf $\mathbb{Q}_\ell[n-1]$ on $\mathbb{G}_m \times \mathbb{P}^{n-1} \subset \mathrm{GR}$. The intersection of the open subset GR° with $\mathbb{G}_m \times \mathbb{P}^{n-1}$ is isomorphic to $\mathbb{G}_m^n$. Thus, we recover (2.6) as part of (2.2), and Kl_n is the same as $\mathrm{Kl}_{\mathrm{SL}_n}^{\mathrm{St}}$.

The next proposition gives the precise sense in which $\mathrm{Kl}_{\widehat{G}}$ generalizes Deligne's Kloosterman sheaves Kl_n.

PROPOSITION 2.4.3. (i) *For $n \geqslant 2$, we have an isomorphism of local systems on $\mathbb{G}_m$*

$$\mathrm{Kl}_{\mathrm{SL}_n}^{\mathrm{St}} \cong \mathrm{Kl}_n.$$

Here St is the standard representation of $\mathrm{SL}_n^{\mathrm{Tt}}$ given in Example 2.1.8.

(ii) *When n is even, there exists $x \in \mathbb{G}_m(\mathbb{F}_p)$ and a character $\epsilon : \mathrm{Gal}(\overline{k}/k) \to \{\pm 1\}$ such that*

$$\mathrm{Kl}_{\mathrm{Sp}_n}^{\mathrm{St}} \cong m_x^* \mathrm{Kl}_n \otimes \mathbb{Q}_\ell(\epsilon). \tag{2.7}$$

Here $m_x : \mathbb{G}_m \to \mathbb{G}_m$ is the multiplication by x, $\mathbb{Q}_\ell(\epsilon)$ is the rank-one local system on $\mathrm{Spec}\, k$ obtained from the character ϵ, and St is the standard representation of $\mathrm{Sp}_n^{\mathrm{Tt}}$ given in Example 2.1.9.

Proof. (1) is proved in [HNY13, Proposition 3.4]. When n is even, according to the description of local monodromy of $\mathrm{Kl}_{\mathrm{Sp}_n}$ in §§ 2.3.3 and 2.3.4, the local monodromy of $\mathrm{Kl}_{\mathrm{Sp}_n}^{\mathrm{St}}$ is unipotent with a single Jordan block; at ∞, $\mathrm{Kl}_{\widehat{G}_n}^{\mathrm{St}}$ is totally wild with $\mathrm{Swan}_\infty(\mathrm{Kl}_{\widehat{G}_n}^{\mathrm{St}}) = 1$. Applying Theorem [Kat88, 8.7.1] we conclude that there exists $x \in \mathbb{G}_m(\mathbb{F}_p)$ and a continuous character $\epsilon : \mathrm{Gal}(\overline{k}/k) \to \overline{\mathbb{Q}}_\ell^\times$ such that (2.7) holds. Since both $\mathrm{Kl}_{\widehat{G}_n}^{\mathrm{St}}$ and Kl_n admit a symplectic autoduality with values in $\mathbb{Q}_\ell(-n+1)$, ϵ must be quadratic. □

The monodromy representation of Kl_n reads

$$\rho_n : \pi_1(\mathbb{G}_m/\mathbb{F}_p) \to \mathrm{SL}_n^{\mathrm{Tt}}(\mathbb{Q}_\ell(\mu_p)),$$

which is well-defined up to conjugacy. Let ρ_n^{geom} be the restriction of ρ_n to the geometric fundamental group $\pi_1(\mathbb{G}_m/\overline{\mathbb{F}}_p)$, and let $\widehat{G}_n^{\mathrm{geom}}$ be the Zariski closure of the image of ρ_n^{geom}.

Katz gave a complete description of $\widehat{G}_n^{\mathrm{geom}}$.

THEOREM 2.4.4 (Katz [Kat88, Main Theorem 11.1]). *We have*

$$\widehat{G}_n^{\mathrm{geom}} = \begin{cases} \mathrm{Sp}_n, & n \text{ even}, \\ \mathrm{SL}_n, & n \text{ odd}, p > 2, \\ \mathrm{SO}_n, & n \neq 7 \text{ odd}, p = 2, \\ G_2, & n = 7, p = 2. \end{cases}$$

2.5 Moments

In this subsection we define moments of Kloosterman sheaves and study their basic properties. Let k be a finite field.

DEFINITION 2.5.1. For the Kloosterman sheaf $\mathrm{Kl}_{\widehat{G}}$ and a representation $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, we define three versions of *the V-moment* of $\mathrm{Kl}_{\widehat{G}}$:

$$M_!^V := \mathbf{R}\Gamma_c(\mathbb{G}_m, \mathrm{Kl}^V)[1];$$
$$M_*^V := \mathbf{R}\Gamma(\mathbb{G}_m, \mathrm{Kl}^V)[1];$$
$$M_{!*}^V := \mathrm{Im}(\mathrm{H}^0 M_!^V \to \mathrm{H}^0 M_*^V) = \mathrm{Im}(\mathrm{H}_c^1(\mathbb{G}_m, \mathrm{Kl}^V) \to \mathrm{H}^1(\mathbb{G}_m, \mathrm{Kl}^V)).$$

Each of these are objects in $D_c^b(\mathrm{Spec}\, k)$. When we need to emphasize their dependence on k, we denote them by $M_{!,k}^V, M_{*,k}^V$ and $M_{!*,k}^V$, respectively.

Remark 2.5.2. The complex $M_!^V$ lies in cohomological degrees 0 and 1, and $\mathrm{H}^1 M_!^V = V_{\pi_1(\mathbb{G}_m/\overline{k})}(-1) = V_{\widehat{G}^{\mathrm{geom}}}(-1)$. Therefore, if $V^{\widehat{G}^{\mathrm{geom}}} = 0$ (which is equivalently to $V_{\widehat{G}^{\mathrm{geom}}} = 0$ since $\widehat{G}^{\mathrm{geom}}$ is reductive) then $M_!^V$ is concentrated in degree zero (i.e. is a plain $\mathrm{Gal}(\overline{k}/k)$-module). Dually, M_*^V lies in cohomological degrees -1 and 0, and is concentrated in degree zero if $V^{\widehat{G}^{\mathrm{geom}}} = 0$.

2.5.3 *Basic properties of moments.* Let $j : \mathbb{G}_m \hookrightarrow \mathbb{P}^1$ be the open inclusion, and let i_0 and i_∞ be the inclusions of the point 0 and ∞ into $\mathbb{P}^1$. The distinguished triangle

$$j_! \mathrm{Kl}^V \to j_* \mathrm{Kl}^V \to i_{0,*} i_0^* j_* \mathrm{Kl}^V \oplus i_{\infty,*} i_\infty^* j_* \mathrm{Kl}^V \to j_! \mathrm{Kl}^V[1]$$

gives a long exact sequence

$$0 \to \mathrm{H}^{-1} M_*^V \to \mathrm{H}^0 i_0^* j_* \mathrm{Kl}^V \oplus \mathrm{H}^0 i_\infty^* j_* \mathrm{Kl}^V \to \mathrm{H}^0 M_!^V \to \mathrm{H}^0 M_*^V$$
$$\to \mathrm{H}^1 i_0^* j_* \mathrm{Kl}^V \oplus \mathrm{H}^1 i_\infty^* j_* \mathrm{Kl}^V \to \mathrm{H}^1 M_!^V \to 0. \tag{2.8}$$

We can rewrite the exact sequence (2.8) as

$$0 \to V^{\widehat{G}^{\mathrm{geom}}} \to V^{\mathcal{I}_0} \oplus V^{\mathcal{I}_\infty} \to \mathrm{H}^0 M_!^V \to \mathrm{H}^0 M_*^V$$
$$\to V_{\mathcal{I}_0}(-1) \oplus V_{\mathcal{I}_\infty}(-1) \to V_{\widehat{G}^{\mathrm{geom}}}(-1) \to 0. \tag{2.9}$$

LEMMA 2.5.4. *Let $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$ be a pure object of weight w. As $\mathrm{Gal}(\overline{k}/k)$-modules, $\mathrm{H}^0 M_!^V$ is of weights at most $w+1$ and $\mathrm{H}^0 M_*^V$ is of weights at least $w+1$. Moreover,*

$$M_{!*}^V = \mathrm{H}^0(\mathbb{P}^1, j_{!*}\mathrm{Kl}^V[1]) = \mathrm{Gr}_{w+1}^W \mathrm{H}^0 M_!^V = \mathrm{Gr}_{w+1}^W \mathrm{H}^0 M_*^V$$

which is pure of weight $w+1$. Here $j_{!*}\mathrm{Kl}^V[1]$ is the middle extension of the perverse sheaf $\mathrm{Kl}^V[1]$ from $\mathbb{G}_m$ to $\mathbb{P}^1$.

Proof. The local system Kl^V is pure of weight w by Remark 2.2.2. By [Del80, Théorème 3.3.1], $\mathrm{H}^0 M_!^V = \mathrm{H}_c^1(\mathbb{G}_m, \mathrm{Kl}^V)$ has weights at most $w+1$. A dual argument shows that $\mathrm{H}^0 M_*^V$ is of weights at least $w+1$. Since $M_{!*}^V = \mathrm{Im}(\mathrm{H}^0 M_!^V \to \mathrm{H}^0 M_*^V)$, it is pure of weight $w+1$.

By (2.9), the kernel of the natural map $\mathrm{H}^0 M_!^V \to \mathrm{H}^0 M_*^V$ is a quotient of $\mathrm{H}^0 i_0^* j_* \mathrm{Kl}^V \oplus \mathrm{H}^0 i_\infty^* j_* \mathrm{Kl}^V$, which has weight at most w because $\mathrm{H}^0 i_0^* j_* \mathrm{Kl}^V$ and $\mathrm{H}^0 i_\infty^* j_* \mathrm{Kl}^V$ are stalks of the perverse sheaf $j_{!*} \mathrm{Kl}^V$ which is pure of weight w. Therefore, $M_{!*}^V = \mathrm{Gr}_{w+1}^W \mathrm{H}^0 M_!^V$. A dual argument shows that $M_{!*}^V = \mathrm{Gr}_{w+1}^W \mathrm{H}^0 M_*^V$. $\qquad\square$

Using the description of local monodromy of the Kloosterman sheaves, we can give a formula for the dimension of $M_!^V, M_*^V$ and $M_{!*}^V$.

LEMMA 2.5.5. *Let $N \in \mathrm{Lie}\,\widehat{G}$ be a regular nilpotent element. Then we have*

$$\dim M_!^V = \dim M_*^V = \mathrm{Swan}_\infty(V), \tag{2.10}$$

$$\dim M_{!*}^V = \mathrm{Swan}_\infty(V) - \dim V^N - \dim V^{\mathcal{I}_\infty} + \dim V^{\widehat{G}^{\mathrm{geom}}}. \tag{2.11}$$

When $p \nmid \#W$, we have

$$\dim M_{!*}^V = \frac{\dim V - \dim V^{D_1}}{h} - \dim V^N - \dim V^{D_1, \widetilde{\mathrm{Cox}}} + \dim V^{\widehat{G}^{\mathrm{geom}}}.$$

Proof. Equation (2.10) follows from the Grothendieck–Ogg–Shafarevich formula. Equation (2.11) follows from (2.10) and the sequence (2.9). Note that $V^{\mathcal{I}_0} = V^N$ by the result stated in § 2.3.3. Finally (2.5.5) follows from (2.11) and Lemma 2.3.6. $\qquad\square$

2.5.6 *Independence of dimension of moments for large p.* We are going to show that the dimensions of $M_{?,\mathbb{F}_p}^V$ ($? = !, *$ and $!*$) are independent of p when p is sufficiently large.

Let $\mathbb{X}_*(\widehat{T})_{\mathbb{Q},\mathrm{prim}} \subset \mathbb{X}_*(\widehat{T})_\mathbb{Q}$ be the subspace where Cox acts via a primitive hth root of unity. Then $\mathbb{X}_*(\widehat{T})_{\mathbb{Q},\mathrm{prim}}$ is a simple $\mathbb{Q}[\mathrm{Cox}]$-module isomorphic to $\mathbb{Q}[x]/(\phi_h(x))$, the cyclotomic field generated by hth roots of unity. Let $\mathbb{X}_*(\widehat{T})_{\mathrm{prim}} = \mathbb{X}_*(\widehat{T})_{\mathbb{Q},\mathrm{prim}} \cap \mathbb{X}_*(\widehat{T})$. Let $\mathbb{X}^*(\widehat{T})_{\mathrm{prim}}^\perp \subset \mathbb{X}^*(\widehat{T})$ be the annihilator of $\mathbb{X}_*(\widehat{T})_{\mathrm{prim}}$.

LEMMA 2.5.7. *Assume $p \nmid \#W$ so that we have $D_1 \subset \widehat{T}[p]$ for some maximal torus $\widehat{T} \subset \widehat{G}$ (see § 2.3.4). Let $V \in \mathrm{Rep}(\widehat{G})$ and let $V(\lambda)$ be the λ-weight space of V for $\lambda \in \mathbb{X}^*(\widehat{T})$. Then for sufficiently large p (depending on V), we have*

$$V^{D_1} = \bigoplus_{\lambda \in \mathbb{X}^*(\widehat{T})_{\mathrm{prim}}^\perp} V(\lambda).$$

In particular, according to Lemma 2.5.5, $\dim M_{!,\mathbb{F}_p}^V, \dim M_{*,\mathbb{F}_p}^V$ and $\dim M_{!*,\mathbb{F}_p}^V$ are constant for p sufficiently large.

Proof. For each weight $\lambda \in \mathbb{X}_*(\widehat{T})$, D_1 acts on $V(\lambda)$ via the character $\lambda|_{D_1} : D_1 \to \mathbb{Q}_\ell(\mu_p)^\times$. Therefore, we need to show that $\lambda|_{D_1}$ is trivial for sufficiently large p if and only if $\lambda \in \mathbb{X}_*(\widehat{T})_{\mathrm{prim}}^\perp$.

We define $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}} \subset \mathbb{X}_*(\widehat{T})_{\mathbb{F}_p} = \mathbb{X}_*(\widehat{T}) \otimes_\mathbb{Z} \mathbb{F}_p$ to be the $\mathbb{F}_p[\mathrm{Cox}]$-submodules generated by those on which Cox acts via a primitive hth root of unity (so $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}}$ may not be a simple $\mathbb{F}_p[\mathrm{Cox}]$-module). For sufficiently large p, $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}}$ is the reduction mod p of $\mathbb{X}_*(\widehat{T})_{\mathrm{prim}}$, i.e. $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}} = \mathbb{X}_*(\widehat{T})_{\mathrm{prim},\mathbb{F}_p}$.

We identify $\widehat{T}[p]$ with $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p}$ hence view D_1 as a $\mathbb{F}_p[\mathrm{Cox}]$-submodule of $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p}$. By the description of D_1 in §2.3.4, we have $D_1 \subset \mathbb{X}_*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}}$. By the discussion above, for p large, we may view D_1 as a $\mathbb{F}_p[\mathrm{Cox}]$-submodule of $\mathbb{X}_*(\widehat{T})_{\mathrm{prim},\mathbb{F}_p}$.

If $\lambda \perp \mathbb{X}_*(\widehat{T})_{\mathrm{prim}}$, then λ restricts trivially to $\mathbb{X}_*(\widehat{T})_{\mathrm{prim},\mathbb{F}_p}$. Hence, for sufficiently large p, λ mod p is trivial on $\mathbb{X}_*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}}$, hence on D_1.

On the other hand, if λ is nonzero on $\mathbb{X}_*(\widehat{T})_{\mathrm{prim}}$, we claim that $\lambda|_{D_1}$ is also nontrivial for sufficiently large p. Let $\overline{\lambda}$ be the image of $\overline{\lambda}$ in $\mathbb{X}^*(\widehat{T})_{\mathrm{prim}} = \mathbb{X}_*(\widehat{T})^{\vee}_{\mathrm{prim}}$. The vectors $\overline{\lambda}, \mathrm{Cox}(\overline{\lambda}), \ldots,$ $\mathrm{Cox}^{\phi(h)-1}(\overline{\lambda})$ span the $\mathbb{Q}$-vector space $\mathbb{X}^*(\widehat{T})_{\mathbb{Q},\mathrm{prim}}$. Therefore, for sufficiently large p, $\overline{\lambda}$ mod p generate $\mathbb{X}^*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}}$ as a $\mathbb{F}_p[\mathrm{Cox}]$-module. If $\overline{\lambda}|_{D_1} = 1$, then $\overline{\lambda}$ mod p could only generate a proper $\mathbb{F}_p[\mathrm{Cox}]$-submodule of $\mathbb{X}^*(\widehat{T})_{\mathbb{F}_p,\mathrm{prim}}$ (contained in the annihilator of D_1). Therefore, $\overline{\lambda}|_{D_1}$ is nontrivial. $\qquad\square$

2.6 Moments via homogeneous Fourier transform

We would like to express the moments $M_!^V$ using homogeneous Fourier transform, which has the advantage that the new formula makes sense over any base field (i.e. the Artin–Schreier sheaf disappears in the new formula). Basic definitions and properties of the homogeneous Fourier transform are recalled in Appendix A, following the original article of Laumon [Lau03].

2.6.1 *An alternative description of* GR°. Here we give an alternative description of the open part GR° of the Beilinson–Drinfeld Grassmannian defined in [HNY13, §5.2]. We would like to show that GR° is a group ind-scheme over $\mathbb{G}_m$ and the morphism $f : \mathrm{GR}^\circ \to \mathbb{G}_a^{r+1}$ is a group homomorphism.

Fix a pair of opposite Borel subgroups $B, B^{\mathrm{opp}} \subset G$ with intersection $T = B \cap B^{\mathrm{opp}}$ a maximal torus of G. Let LG_0 and LG_∞ be the loop group of G, with the formal parameter chosen to be the local coordinate t around $0 \in \mathbb{P}^1$ and $s = t^{-1}$ around $\infty \in \mathbb{P}^1$. We use B^{opp} to define an Iwahori subgroup $\mathbf{I}_0^{\mathrm{opp}} \subset LG_0$, i.e. $\mathbf{I}_0$ is the preimage of B^{opp} under the evaluation map $L^+G_0 \to G$. Similarly, we use B to define an Iwahori subgroup $\mathbf{I}_\infty \subset LG_\infty$. We also consider the following Moy–Prasad subgroups of $\mathbf{I}_\infty$ (see [HNY13, §1.2]):

$$\mathbf{I}_\infty = \mathbf{I}_\infty(0) \rhd \mathbf{I}_\infty(1) \rhd \mathbf{I}_\infty(2).$$

Here $\mathbf{I}_\infty(1)$ is the preimage of N (radical of B) under the evaluation map $L^+G_\infty \to G$, so that $\mathbf{I}_\infty(0)/\mathbf{I}_\infty(1) = T$. On the other hand $\mathbf{I}_\infty(1)/\mathbf{I}_\infty(2)$ can be identified with the $(r+1)$th product of the additive group $\mathbb{G}_a$, where r is the rank of G; each copy of $\mathbb{G}_a$ is the root subgroup of LG_∞ corresponding to a simple affine root.

Recall that in [HNY13, §1.4] we introduced moduli stacks $\mathrm{Bun}_{G(0,i)}$ for $i = 0, 1, 2$. It classifies G-torsors over $\mathbb{P}^1$ with an $\mathbf{I}_0^{\mathrm{opp}}$-level structure at 0 and an $\mathbf{I}_\infty(i)$-level structure at ∞. There is an open point $\star \in \mathrm{Bun}_{G(0,1)}$ (with no automorphisms), which corresponds to the trivial G-torsor over $\mathbb{P}^1$ with the tautological $\mathbf{I}_0^{\mathrm{opp}}$ and $\mathbf{I}_\infty(1)$-level structures. We denote this G-torsor with level structures by $\mathcal{E}_{0,1}^\star$. Its image in $\mathrm{Bun}_{G(0,0)}$ will be denoted $\mathcal{E}_{0,0}^\star$, with automorphism group T.

The connected components of $\mathrm{Bun}_{G(0,i)}$ are indexed by $\Omega = \mathbb{X}_*(T)/\mathbb{Z}\Phi^\vee$ where $\mathbb{Z}\Phi^\vee$ is the coroot lattice. For $\omega \in \Omega$, we let $\mathrm{Bun}_{G(0,i)}^\omega$ be the corresponding component. We may identify Ω with the quotient $N_{LG_0}(\mathbf{I}_0^{\mathrm{opp}})/\mathbf{I}_0^{\mathrm{opp}}$, hence it acts on $\mathrm{Bun}_{G(0,i)}$ by changing the G-torsors in the formal neighborhood of 0. This action permutes the components of $\mathrm{Bun}_{G(0,i)}$ simply transitively. Using Ω-translates of $\mathcal{E}_{0,i}^\star$ we get an open point $\omega(\star) \in \mathrm{Bun}_{G(0,i)}^\omega$, which corresponds to a G-torsor $\mathcal{E}_{0,i}^{\omega(\star)}$ with level structures at 0 and ∞. The Ω-action only changes a G-torsor in the formal

neighborhood of 0, hence there are canonical isomorphisms $\mathcal{E}_{0,i}^{\omega(\star)}|_{\mathbb{P}^1-\{0\}} \overset{\sim}{\to} \mathcal{E}_{0,i}^{\star}|_{\mathbb{P}^1-\{0\}}$ for all $\omega \in \Omega$.

For $x \in \mathbb{G}_m(k)$, and $\omega \in \Omega$, define an ind-scheme

$$\mathfrak{G}_x^\omega = \mathrm{Isom}_{\mathbb{P}^1-\{x\}}(\mathcal{E}_{0,1}^{\omega(\star)}, \mathcal{E}_{0,1}^{\star}).$$

When $\omega = 1$, $\mathfrak{G}_x^1$ is the automorphism group (ind-scheme) of $\mathcal{E}_{0,1}^{\star}|_{\mathbb{P}^1-\{x\}}$, which acts on each $\mathfrak{G}_x^\omega$. When x varies over $\mathbb{G}_m$, we get an ind-scheme $\mathfrak{G}^\omega$ over $\mathbb{G}_m$, and we define

$$\mathfrak{G}_x := \bigsqcup_{\omega \in \Omega} \mathfrak{G}_x^\omega, \quad \mathfrak{G} := \bigsqcup_{\omega \in \Omega} \mathfrak{G}^\omega.$$

We define a group structure on $\mathfrak{G}$ as follows. For $\varphi \in \mathfrak{G}^\omega$ and $\varphi' \in \mathfrak{G}^{\omega'}$, we define $\varphi \circ \varphi'$ to be the composition $\mathcal{E}_{0,1}^{\omega\omega'(\star)} \xrightarrow{\omega(\varphi')} \mathcal{E}_{0,1}^{\omega(\star)} \xrightarrow{\varphi} \mathcal{E}_{0,1}^{\star}$. It is easy to check that this defines a group ind-scheme structure on $\mathfrak{G}$ over $\mathbb{G}_m$.

For each $\varphi \in \mathfrak{G}_x^\omega$, which is an isomorphism $\mathcal{E}_{0,1}^{\omega(\star)}|_{\mathbb{P}^1-\{x\}} \to \mathcal{E}_{0,1}^{\star}|_{\mathbb{P}^1-\{x\}}$, we may consider its restriction φ_∞ to the formal neighborhood of $\infty \in \mathbb{P}^1$. Since $\mathcal{E}_{0,1}^{\omega(\star)}$ is the same as $\mathcal{E}_{0,1}$ away from 0, φ_∞ gives an element in L^+G_∞. Since φ_∞ preserves the $\mathbf{I}_\infty(1)$-level structures, $\varphi_\infty \in \mathbf{I}_\infty(1)$. This defines a group homomorphism

$$\mathrm{ev}_\infty : \mathfrak{G} \to \mathbf{I}_\infty(1),$$

and hence a group homomorphism by composition

$$f_{\mathfrak{G}} : \mathfrak{G} \to \mathbf{I}_\infty(1) \to \mathbf{I}_\infty(1)/\mathbf{I}_\infty(2) \cong \mathbb{G}_a^{r+1}.$$

LEMMA 2.6.2. *There is a canonical isomorphism of ind-schemes over $\mathbb{G}_m$*

$$\iota : \mathfrak{G} \overset{\sim}{\to} \mathrm{GR}^\circ$$

such that $f_{\mathfrak{G}} = f \circ \iota$.

Proof. We first recall the definition of GR° in [HNY13]. We defined the Hecke correspondence (see [HNY13, §2.3])

In [HNY13, §5.2], the Beilinson–Drinfeld Grassmannian is defined to be the fiber $\mathrm{pr}_2^{-1}(\star)$ of the open point $\star \in \mathrm{Bun}_{G(0,1)}$ in $\mathrm{Hk}_{G(0,1)}$. The open subscheme $\mathrm{GR}^\circ \subset \mathrm{GR}$ is defined as the preimage of the union of the open points $\bigsqcup_{\omega \in \Omega} \omega(\star) \subset \mathrm{Bun}_{G(0,1)}$ under the first projection $\mathrm{pr}_1 : \mathrm{GR} \to \mathrm{Bun}_{G(0,1)}$. Therefore, GR° is the union of the preimages of $(\omega(\star), \star)$ under the projections $\mathrm{pr}_1 \times \mathrm{pr}_2 : \mathrm{Hk}_{G(0,1)} \to \mathrm{Bun}_{G(0,1)} \times \mathrm{Bun}_{G(0,1)}$. By the moduli interpretation of $\mathrm{Hk}_{G(0,1)}$, this preimage is exactly $\mathfrak{G}$. The rest of the lemma is clear. $\square$

Remark 2.6.3. Let us give the morphism $\mathfrak{G}_x \hookrightarrow \mathrm{GR}_x$ explicitly. The G-torsors $\mathcal{E}^{\star}_{0,1}$ and $\mathcal{E}^{\omega(\star)}_{0,1}$ are canonically trivialized at $x \in \mathbb{G}_m$ by definition. Therefore, restricting an isomorphism $\varphi \in \mathfrak{G}_x$ to the formal punctured neighborhood of x gives a homomorphism of group ind-schemes

$$\mathrm{ev}_x : \mathfrak{G}_x \to LG_x \tag{2.12}$$

where LG_x is the loop group of G at x. Now the morphism $\mathfrak{G}_x \to \mathrm{GR}_x$ is given by composing (2.12) with the projection $LG_x \to \mathrm{GR}_x = LG_x/L^+G_x$.

In §4.1 we will give a more explicit description of $\mathfrak{G}$ in the case $G = \mathrm{PGL}_n$.

2.6.4 *Torus action on* $\mathfrak{G}$. We may define a variant of $\mathfrak{G}$ as follows. Instead of considering isomorphisms between $\mathcal{E}^{\omega(\star)}_{0,1}$ we consider $\mathcal{E}^{\omega(\star)}_{0,0}$. This way we get an ind-scheme $\mathfrak{G}'^{\omega} \to \mathbb{G}_m$ whose fiber over $x \in \mathbb{G}_m$ is $\mathrm{Isom}_{\mathbb{P}^1 - \{x\}}(\mathcal{E}^{\omega(\star)}_{0,0}, \mathcal{E}^{\star}_{0,0})$. Let $\mathfrak{G}' = \bigsqcup_{\omega \in \Omega} \mathfrak{G}'^{\omega}$. Similarly, we have the evaluation map $\mathrm{ev}'_\infty : \mathfrak{G}' \to \mathbf{I}_\infty$ and hence a morphism

$$\tau : \mathfrak{G}' \to \mathbf{I}_\infty \to \mathbf{I}_\infty/\mathbf{I}_\infty(1) = T.$$

By construction, $\mathfrak{G} = \tau^{-1}(1)$. Note also that $T = \mathrm{Aut}_{\mathbb{P}^1}(\mathcal{E}^{\omega(\star)}_{0,0}) = \mathrm{Aut}_{\mathbb{P}^1}(\mathcal{E}^{\star}_{0,0})$ acts on each $\mathfrak{G}'^{\omega}_x$ both from the right and from the left by pre- and post-composition. Combining these two actions gives an adjoint action of T on $\mathfrak{G}'$. The morphism ev'_∞ is equivariant under the adjoint T-actions: T also acts on $\mathbf{I}_\infty$ by conjugation. Since the homomorphism τ is invariant under the adjoint T-action, the T-action preserves $\mathfrak{G}$. Therefore we obtain an action of T on $\mathfrak{G}$, such that the morphism $f_{\mathfrak{G}} : \mathfrak{G} \to \mathbf{I}_\infty(1)/\mathbf{I}_\infty(2)$ is T-equivariant.

Recall $\mathbb{G}^{\mathrm{rot}}_m$ is the one-dimensional torus acting on $\mathbb{P}^1$ by scaling the natural coordinate: $s \in \mathbb{G}^{\mathrm{rot}}_m$ as $t \mapsto s^{-1}t$. This action induces a $\mathbb{G}^{\mathrm{rot}}_m$-action on $\mathfrak{G} \cong \mathrm{GR}^\circ$ which commutes with the adjoint T-action. This way we conclude.

LEMMA 2.6.5. *There is an action of $T \times \mathbb{G}^{\mathrm{rot}}_m$ on $\mathfrak{G}$ making the morphism $f_{\mathfrak{G}} : \mathfrak{G} \to \mathbf{I}_\infty(1)/\mathbf{I}_\infty(2) \cong \mathbb{G}^{r+1}_a$ equivariant under $T \times \mathbb{G}^{\mathrm{rot}}_m$. Here, for $i = 0, 1, \ldots, r$, $T \times \mathbb{G}^{\mathrm{rot}}_m$ acts on the ith coordinate of $\mathbb{G}^{r+1}_a$ through the simple affine root α_i of the loop group LG_∞. Identifying $\mathfrak{G}_x$ with GR°_x, the T-action on $\mathrm{GR}^\circ_x \subset LG_x/L^+G_x$ is via left multiplication.*

Define a one-dimensional subtorus $\mathbb{G}^{(\rho^\vee, h)}_m \subset T \times \mathbb{G}^{\mathrm{rot}}_m$ by

$$\mathbb{G}^{(\rho^\vee, h)}_m \ni s \mapsto (s^{\rho^\vee}, s^h) \in T \times \mathbb{G}^{\mathrm{rot}}_m.$$

where h is the Coxeter number of G. For example, when $G = \mathrm{PGL}_n$, $\mathbb{G}^{(\rho^\vee, h)}_m \subset T \times \mathbb{G}^{\mathrm{rot}}_m$ is given by $(\mathrm{diag}(s^{n-1}, s^{n-2}, \ldots, s^1, 1), s^n)$. This subtorus $\mathbb{G}^{(\rho^\vee, h)}_m$ has the property that it acts on each simple affine root space by dilation. Therefore, the map $F = \sigma \circ f_{\mathfrak{G}} : \mathfrak{G} \cong \mathrm{GR}^\circ \to \mathbb{A}^1$ is $\mathbb{G}^{(\rho^\vee, h)}_m$-equivariant, where $\mathbb{G}^{(\rho^\vee, h)}_m$ acts on $\mathbb{A}^1$ still by dilation.

Using the $\mathbb{G}^{\mathrm{rot}}_m$-action we may trivialize the group ind-scheme $\mathfrak{G} \to \mathbb{G}_m$ and write $\mathfrak{G} = \mathfrak{G}_1 \times \mathbb{G}_m$. We have a canonical identification $\mathfrak{G}_1 \cong \mathrm{GR}^\circ_1$. We may also identify GR_1 with the affine Grassmannian $\mathrm{Gr} = LG/L^+G$ (by using the standard local coordinate $t - 1$ at $1 \in \mathbb{G}_m$). We write Gr° for GR°_1 under this identification. Now T acts only on the $\mathfrak{G}_1$-factor (by conjugation) and $\mathbb{G}^{\mathrm{rot}}_m$ only acts on the $\mathbb{G}_m$-factor (by translation). Therefore,

$$[\mathrm{GR}^\circ/\mathbb{G}^{(\rho^\vee, h)}_m] \cong [\mathrm{Gr}^\circ/\mu^{\rho^\vee}_h] \cong [\mathfrak{G}_1/\mathrm{Ad}(\mu^{\rho^\vee}_h)].$$

The notation emphasizes that μ_h acts on Gr° via the cocharacter $\rho^\vee : \mathbb{G}_m \to T$ and the adjoint T-action on $\mathfrak{G}_1 \cong \mathrm{Gr}^\circ$.

Because F is equivariant under $\mathbb{G}_m^{(\rho^\vee,h)}$, it descends to a morphism of stacks $[\mathrm{GR}^\circ/\mathbb{G}_m^{(\rho^\vee,h)}] \to [\mathbb{A}^1/\mathbb{G}_m^{(\rho^\vee,h)}]$. Therefore, we may rewrite F as a morphism

$$\overline{F} : [\mathrm{Gr}^\circ/\mu_h^{\rho^\vee}] \cong [\mathfrak{G}_1/\mathrm{Ad}(\mu_h^{\rho^\vee})] \to [\mathbb{A}^1/\mathbb{G}_m]. \tag{2.13}$$

Every object in $\mathrm{Sat}^{\mathrm{Tt}}$ is also T-equivariant, and in particular μ_h-equivariant. We use the same notation IC_V to mean a μ_h-equivariant perverse sheaf on Gr, whose restriction to Gr° is still denoted by IC_V°. The complexes $F_!\mathrm{IC}_{V,\mathrm{GR}^\circ}$ and $F_*\mathrm{IC}_{V,\mathrm{GR}^\circ}$ belong to $D^b_{\mathbb{G}_m}(\mathbb{A}^1, \mathbb{Q}_\ell)$, and can be identified with $\overline{F}_!\mathrm{IC}_V^\circ$ and $\overline{F}_*\mathrm{IC}_V^\circ$.

Let S be any base scheme. For an object $\mathcal{F} \in D^b_{\mathbb{G}_{m,S}}(\mathbb{A}^1_S, \mathbb{Q}_\ell)$, we use $\mathcal{F}_\eta$ to denote its restriction to the 'open point' $S = [\mathbb{G}_{m,S}/\mathbb{G}_{m,S}]$. The following proposition gives a formula for $M_!^V$, which has the potential of making sense over a general base scheme S other than $\mathrm{Spec}\, k$.

PROPOSITION 2.6.6. *There are canonical isomorphisms in $D^b(\mathrm{Spec}\, k, \mathbb{Q}_\ell(\mu_p))$*

$$M_!^V \cong \mathrm{HFour}(\overline{F}_!\mathrm{IC}_V^\circ)_\eta \otimes \mathbb{Q}_\ell(\mu_p). \tag{2.14}$$

Proof. By definition of Kl^V and $M_!^V$, and proper base change, we have

$$M_!^V = \mathbf{R}\Gamma_c(\mathbb{G}_m, \mathrm{Kl}^V)[1] = \mathbf{R}\Gamma_c(\mathrm{GR}^\circ, \mathrm{IC}_{V,\mathrm{GR}^\circ} \otimes F^*\mathrm{AS}_\psi)[1] = \mathbf{R}\Gamma_c(\mathbb{A}^1, F_!\mathrm{IC}_{V,\mathrm{GR}^\circ} \otimes \mathrm{AS}_\psi)[1].$$

The last object is the stalk at $1 \in \mathbb{A}^1(k)$ of the Fourier–Deligne transform $\mathrm{Four}_\psi(F_!\mathrm{IC}_{V,\mathrm{GR}^\circ})$, see § A.1. By Theorem A.2.1(ii), we conclude that $M_!^V$ is the same as the stalk at 1 of $\mathrm{HFour}(\overline{F}_!\mathrm{IC}_V^\circ)$, which is the same as $\mathrm{HFour}(\overline{F}_!\mathrm{IC}_V^\circ)_\eta \otimes \mathbb{Q}_\ell(\mu_p)$. □

2.7 Autoduality of moments

For $V, W \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, the isomorphism $\mathrm{Kl}^V \otimes \mathrm{Kl}^W \xrightarrow{\sim} \mathrm{Kl}^{V\otimes W}$ induces a cup product

$$\mathrm{H}^*_c(\mathbb{G}_m, \mathrm{Kl}^V) \otimes \mathrm{H}^*_c(\mathbb{G}_m, \mathrm{Kl}^W) \to \mathrm{H}^*_c(\mathbb{G}_m, \mathrm{Kl}^{V\otimes W}),$$

which is equivalent to a map

$$M_!^V \otimes M_!^W \to M_!^{V\otimes W}[1]. \tag{2.15}$$

When $W = V^\vee$, $\mathrm{Kl}^{V^\vee}$ is the dual local system of Kl^V, therefore (2.15) composed with the evaluation map $V \otimes V^\vee \to \mathbb{Q}_\ell$ (the trivial representation) gives

$$M_!^V \otimes M_!^{V^\vee} \to M_!^{V\otimes V^\vee}[1] \cong \mathrm{H}^*_c(\mathbb{G}_m, \mathbb{Q}_\ell(\mu_p))[2] \to \mathbb{Q}_\ell(\mu_p)(-1) \tag{2.16}$$

where the last map is given by projecting to $\mathrm{H}^2_c(\mathbb{G}_m, \mathbb{Q}_\ell(\mu_p))$.

2.7.1 *The opposition σ.* Recall that in [HNY13, § 6.1] we studied the behavior of $\mathrm{Kl}_{\widehat{G}}$ under pinned automorphisms of G. Let σ be the opposition on G: this is the pinned automorphism of G which acts as $-w_0$ on $\mathbb{X}_*(T)$. Then σ induces an involution (which we still denote by σ) on $\widehat{G}^{\mathrm{Tt}}$ as well which is the identity on $\mathbb{G}_m^{\mathrm{Wt}}$. Then [HNY13, Corollary 6.5 and Table 2] says that when G is not of type A_{2n}, the monodromy of $\mathrm{Kl}_{\widehat{G}}$ lies in $\widehat{G}^{\mathrm{Tt},\sigma}$, i.e. we have a tensor functor

$$\mathcal{K} : \mathrm{Rep}(\widehat{G}^{\mathrm{Tt},\sigma}) \to \mathrm{Loc}(\mathbb{G}_m)$$

(we suppress the coefficient field $\mathbb{Q}_\ell(\mu_p)$ in the notation), denoted $V \mapsto \mathcal{K}^V$, such that the following diagram is made commutative by a canonical natural isomorphism.

$$\mathrm{Rep}(\widehat{G}^{\mathrm{Tt}}) \xrightarrow{\mathrm{Res}} \mathrm{Rep}(\widehat{G}^{\mathrm{Tt},\sigma}) \xrightarrow{\mathcal{K}} \mathrm{Loc}(\mathbb{G}_m)$$
$$\underbrace{\qquad\qquad\qquad}_{\mathrm{Kl}_{\widehat{G}}}$$

When G is of type A_{2n}, then [HNY13, Corollary 6.5 and Table 2] says that there is a $\widehat{G} \rtimes \langle \sigma \rangle$-local system $\mathcal{K}$ on $\mathbb{G}_m$ such that the following diagram is made commutative by a canonical natural isomorphism.

$$
\begin{array}{ccc}
\mathrm{Rep}(\widehat{G}^{\mathrm{Tt}} \rtimes \langle \sigma \rangle) & \xrightarrow{\mathcal{K}} & \mathrm{Loc}(\mathbb{G}_m) \\
\downarrow{\scriptstyle \mathrm{Res}} & & \downarrow{\scriptstyle [2]^*} \\
\mathrm{Rep}(\widehat{G}^{\mathrm{Tt}}) & \xrightarrow{\mathrm{Kl}_{\widehat{G}}} & \mathrm{Loc}(\mathbb{G}_m)
\end{array}
$$

Here $[2]^*$ is the pullback functor along the square morphism $[2] : \mathbb{G}_m \to \mathbb{G}_m$.

For $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, let V^σ be the same vector space with the action of $\widehat{G}^{\mathrm{Tt}}$ given by the composition $\widehat{G}^{\mathrm{Tt}} \xrightarrow{\sigma} \widehat{G}^{\mathrm{Tt}} \to \mathrm{GL}(V)$.

LEMMA 2.7.2. *Let $V \in \mathrm{Rep}(\widehat{G})$. When G is not of type A_{2n} there is a canonical isomorphism*

$$
\mathrm{Kl}_{\widehat{G}}^V \xrightarrow{\sim} \mathrm{Kl}_{\widehat{G}}^{V^\sigma}. \tag{2.17}
$$

When G is of type A_{2n}, there is a canonical isomorphism

$$
[2]_* \mathrm{Kl}_{\widehat{G}}^V \xrightarrow{\sim} [2]_* \mathrm{Kl}_{\widehat{G}}^{V^\sigma}. \tag{2.18}
$$

In particular, for all G, we always have a canonical isomorphism

$$
M_!^V \xrightarrow{\sim} M_!^{V^\sigma}. \tag{2.19}
$$

Proof. We shall use the local system $\mathcal{K}$ introduced in §2.7.1. When G is not of type A_{2n}, both sides of (2.17) can be identified with $\mathcal{K}^V$ (since $V = V^\sigma$ as $\widehat{G}^{\mathrm{Tt},\sigma}$-representations). When G is of type A_{2n}, both sides of (2.18) can be identified with $\mathcal{K}^W$ where $W = \mathrm{Ind}_{\widehat{G}^{\mathrm{Tt}}}^{\widehat{G}^{\mathrm{Tt}} \rtimes \langle \sigma \rangle} V = \mathrm{Ind}_{\widehat{G}^{\mathrm{Tt}}}^{\widehat{G}^{\mathrm{Tt}} \rtimes \langle \sigma \rangle} V^\sigma \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}} \rtimes \langle \sigma \rangle)$. $\qquad \square$

PROPOSITION 2.7.3. *Let V be an irreducible object in $\mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$ of weight w. Then there is a canonical isomorphism*

$$
M_!^V \cong M_!^{V^\vee}(-w) \tag{2.20}
$$

such that the pairing

$$
M_!^V \otimes M_!^V \cong M_!^V \otimes M_!^{V^\vee}(-w) \xrightarrow{(2.16)} \mathbb{Q}_\ell(\mu_p)(-w-1) \tag{2.21}
$$

is $(-1)^{w+1}$-symmetric.

Proof. Since V is irreducible of weight w, we have $V^\vee(-w) \cong V^\sigma$. The isomorphism (2.20) comes from this isomorphism and (2.19).

We may rewrite the isomorphism $V^\vee(-w) \cong V^\sigma$ as a perfect pairing

$$
(\cdot, \cdot) : V \otimes V \to \mathbb{Q}_\ell(-w) \tag{2.22}
$$

such that $(gu, \sigma(g)v) = (u, v)$ for all $g \in \widehat{G}$. When G is not of type A_{2n}, let $\widehat{H} = \widehat{G}^{\mathrm{Tt},\sigma}$ and let $W = V$ viewed as an $\widehat{H}$ representation. Then we have $\mathcal{K}^W = \mathrm{Kl}_{\widehat{G}}^V$. When G is of type A_{2n}, let $\widehat{H} = \widehat{G}^{\mathrm{Tt}} \rtimes \langle \sigma \rangle$ and let $W = \mathrm{Ind}_{\widehat{G}^{\mathrm{Tt}}}^{\widehat{H}} V$ as an $\widehat{H}$ representation, then $\mathcal{K}^W$ is also defined. From the

proof of Lemma 2.7.2, the pairing (2.21) is induced from the pairing of local systems on $\mathbb{G}_m$:

$$\mathcal{K}^W \otimes \mathcal{K}^W \to \mathbb{Q}_\ell(\mu_p)(-w). \tag{2.23}$$

The above pairing comes from a pairing $W \otimes W \to \mathbb{Q}_\ell(-w)$ of $\widehat{H}$-modules, which in turn comes from the pairing (2.22) (via an induction process when G is of type A_{2n}). Hence, the pairing (2.23) has the same symmetry type as (2.22).

We now claim that (2.22) is $(-1)^w$-symmetric. In fact, since V is an irreducible representation of $\widehat{G}$, pairings on V satisfying $(gu, \sigma(g)v) = (u, v)$ (for all $g \in \widehat{G}$) are unique up to a scalar. In particular, the transposed pairing $(u, v)' := (v, u)$ also satisfies the same condition, hence $(v, u) = (u, v)' = c(u, v)$ for some constant c, and clearly $c = \pm 1$. Let $\varphi : \mathrm{SL}_2 \to \widehat{G}^\sigma \subset \widehat{G}$ be the principal SL_2 whose restriction to the diagonal torus of SL_2 is given by the cocharacter 2ρ of $\widehat{G}$. Then (2.22) is $\varphi(\mathrm{SL}_2)$-invariant. Decomposing V as a direct sum of irreducible SL_2-modules, the irreducible representation Sym^n with n largest appears with multiplicity one. In fact, the highest weight space under the SL_2-action is the same as the highest weight space for $\widehat{G}$ (which is one-dimensional), and we have $n = \langle 2\rho, \lambda \rangle$ if λ is the highest weight of V. The symmetry type c of V is then the same as the symmetric type when restricted to Sym^n. From the discussion in §2.1.5, the weight w of V is also $\langle 2\rho, \lambda \rangle$. Since any SL_2-invariant autoduality on Sym^n is $(-1)^n$-symmetric, the constant c must be $(-1)^n = (-1)^w$.

Now we have shown that the pairing (2.23) is $(-1)^w$-symmetric. Since taking H_c^1 changes the sign of the pairing, the pairing (2.21) is $(-1)^{w+1}$-symmetric. $\qquad\square$

Remark 2.7.4. Since $M_{!*}^V = \mathrm{H}^1(\mathbb{P}^1, j_{!*}\mathrm{Kl}^V)$ by Lemma 2.5.4, the pairing (2.16) factors through a perfect pairing

$$M_{!*}^V \otimes M_{!*}^{V^\vee} \to \mathbb{Q}_\ell(\mu_p)(-1).$$

Therefore, the pairing (2.21) also factors through a $(-1)^{w+1}$-symmetric perfect pairing

$$M_{!*}^V \otimes M_{!*}^V \to \mathbb{Q}_\ell(\mu_p)(-w - 1).$$

3. The Galois representations attached to moments

The goal of this section is to prove the main Theorem 1.2.1, which attaches a Galois representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ to each pair $(\widehat{G}, V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}}))$, such that the local behavior at p of this Galois representation is closely related to the moment $M_{!*,\mathbb{F}_p}^V$ defined earlier in Definition 2.5.1. The proof starts by 'gluing' the moments defined over $\mathbb{F}_p$ into a moment defined over $\mathrm{Spec}\,\mathbb{Z}$.

3.1 Moments over $\mathbb{Z}$

In this subsection, we consider schemes over $S = \mathrm{Spec}\,\mathbb{Z}[\ell^{-1}]$. We use underlined letters to denote schemes over S. We use the following notation for inclusions of closed points and the generic point of S:

$$\mathrm{Spec}\,\mathbb{F}_p \xrightarrow{i_p} S \xleftarrow{j} \mathrm{Spec}\,\mathbb{Q}.$$

Let $\underline{G}$ be the Chevalley group over S with the same root datum as the simple group G we fixed in the beginning. The affine Grassmannian GR, Gr, as well as $\mathrm{GR}^\circ \cong \mathfrak{G}$, $\mathrm{Gr}^\circ \cong \mathfrak{G}_1$ have natural models $\underline{\mathrm{GR}}, \underline{\mathrm{Gr}}$, etc. over S. For example, the affine Grassmannian $\underline{\mathrm{Gr}}$ over S can be defined as the fppf quotient $L\underline{G}/L^+\underline{G}$. See [PZ13, §5.b.1, Proposition 5.3]. Its generic fiber $\mathrm{Gr}_\mathbb{Q}$ and special fibers $\mathrm{Gr}_{\mathbb{F}_p}$ are the old affine Grassmannians defined over a field, see [PZ13, Corollary 5.6].

LEMMA 3.1.1. *Let $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$. Let $j : \mathrm{Gr}_{\mathbb{Q}} \hookrightarrow \underline{\mathrm{Gr}}$ be the inclusion of the generic fiber and $i_p : \mathrm{Gr}_{\mathbb{F}_p} \hookrightarrow \underline{\mathrm{Gr}}$ be the inclusion of the mod p special fiber. Let $\underline{\mathrm{IC}}_V = j_{!*}\mathrm{IC}_V$. Then we have*

$$i^* \underline{\mathrm{IC}}_V \cong \mathrm{IC}_{V,\mathbb{F}_p}.$$

Here $\mathrm{IC}_{V,\mathbb{F}_p} \in \mathrm{Sat}_{\mathbb{F}_p}^{\mathrm{Tt}}$ is the perverse sheaf on $\mathrm{Gr}_{\mathbb{F}_p}$ corresponding to V.

Proof. Since we are dealing with a single prime p at a time, we may restrict the situation to Spec $\mathbb{Z}_p$, and retain the same notation. By [PZ13, Proposition 9.15], the nearby cycles sheaf $\Psi(\mathrm{IC}_V)$ (a perverse sheaf on $\mathrm{Gr}_{\mathbb{F}_p}$) is isomorphic to $\mathrm{IC}_{V,\mathbb{F}_p}$. Also, by [PZ13, Proposition 9.12], the monodromy action on $\Psi(\mathrm{IC}_V)$ is trivial (because we are in the situation of a split G and hyperspecial parahoric level structure, in the context of [PZ13]). Then, by [PZ13, Lemma 9.19], $i^* j_{!*}\mathrm{IC}_V$ is isomorphic to $\Psi(\mathrm{IC}_V)$, which is in turn isomorphic to $\mathrm{IC}_{V,\mathbb{F}_p}$. □

Remark 3.1.2. We may similarly define the convolution product of $\underline{\mathrm{IC}}_V * \underline{\mathrm{IC}}_W$ on $\underline{\mathrm{Gr}}$, using the integral model of the convolution diagram [MV07, §4]. We claim that there is a natural isomorphism

$$\underline{\mathrm{IC}}_{V \otimes W} \cong \underline{\mathrm{IC}}_V * \underline{\mathrm{IC}}_W.$$

In fact, both sides above are isomorphic when restricted to $\mathrm{Gr}_{\mathbb{Q}}$, and they both have trivial monodromy on the nearby cycles towards the special fiber $\mathrm{Gr}_{\mathbb{F}_p}$ (in the case of $\underline{\mathrm{IC}}_V * \underline{\mathrm{IC}}_W$, use fact that the nearby cycles commute with proper push-forward). Therefore, by [PZ13, Lemma 9.19] again, both sides above are middle extensions of their restriction to $\mathrm{Gr}_{\mathbb{Q}}$, hence they are canonically isomorphic to each other.

Although we do not have Kloosterman sheaves defined over $\mathbb{G}_{m,S}$, the moments of Kloosterman sheaves do have natural analogs over S. In fact, from the description of $\mathfrak{G} \cong \mathrm{GR}^\circ$ in §2.6.1, we see that the morphism $\overline{F}$ in (2.13) can be extended to a morphism over S:

$$\overline{F} : [\underline{\mathrm{Gr}}^\circ / \underline{\mu}_h^{\rho^\vee}] \cong [\underline{\mathfrak{G}}_1 / \mathrm{Ad}(\underline{\mu}_h^{\rho^\vee})] \to [\mathbb{A}^1 / \underline{\mathbb{G}_m}]. \tag{3.1}$$

From Proposition 2.6.6, we arrive at the following natural definition.

DEFINITION 3.1.3. For $V \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, we define

$$\underline{M}_!^V = \mathrm{HFour}(\overline{F}_! \underline{\mathrm{IC}}_V^\circ)_\eta$$

as an object in $D_c^b(S, \mathbb{Q}_\ell)$. Here, as usual, $\underline{\mathrm{IC}}_V^\circ$ denotes the restriction of $\underline{\mathrm{IC}}_V$ to $\underline{\mathrm{Gr}}^\circ$.

LEMMA 3.1.4. *We have an isomorphism in $D^b(\mathrm{Spec}\, \mathbb{F}_p, \mathbb{Q}_\ell(\mu_p))$*

$$i_p^* \underline{M}_!^V \otimes \mathbb{Q}_\ell(\mu_p) \cong M_{!,\mathbb{F}_p}^V. \tag{3.2}$$

Here $M_{!,\mathbb{F}_p}^V$ is defined in Definition 2.5.1 for the base field $k = \mathbb{F}_p$.

Proof. For arbitrary base change $i : S' \to S$, the formation of HFour commutes with i^*: this follows from proper base change. Therefore, we have

$$i_p^* \underline{M}_!^V \cong i_p^* \mathrm{HFour}(\overline{F}_! \underline{\mathrm{IC}}_V^\circ)_\eta \cong \mathrm{HFour}(i_p^* \overline{F}_! \underline{\mathrm{IC}}_V^\circ)_\eta \cong \mathrm{HFour}(\overline{F}_! i_p^* \underline{\mathrm{IC}}_V^\circ)_\eta.$$

By Lemma 3.1.1, we have $i_p^* \underline{\mathrm{IC}}_V^\circ \cong \mathrm{IC}_{V,\mathbb{F}_p}^\circ$. Therefore,

$$i_p^* \underline{M}_!^V \cong \mathrm{HFour}(\overline{F}_! \mathrm{IC}_{V,\mathbb{F}_p}^\circ)_\eta.$$

Using Proposition 2.6.6, we get (3.2). □

The next proposition will be used in the next subsection to establish a pairing on $\underline{M}_!^V$ analogous to the pairing on $M_{!,\mathbb{F}_p}^V$ defined in Proposition 2.7.3. What we need there is a definition of the cup product (2.15) without appealing to Kloosterman sheaves.

PROPOSITION 3.1.5. *For $V, W \in \mathrm{Rep}(\widehat{G}^{\mathrm{Tt}})$, there is a bifunctorial morphism*

$$\underline{M}_!^V \otimes \underline{M}_!^W \to \underline{M}_!^{V \otimes W}[1] \tag{3.3}$$

such that after taking i_p^* for any prime $p \nmid \ell$ and using the isomorphism (3.2), the above morphism becomes the cup product (2.15) for the base field $\mathbb{F}_p$.

Proof. The moduli stack $\mathrm{Bun}_{G(0,1)}$ defined in § 2.6.1 has a natural integral model $\underline{\mathrm{Bun}}_{G(0,1)}$ over S, which contains an open substack $\underline{\mathrm{Bun}}_{G(0,1)}^\circ = \bigsqcup_{\omega \in \Omega} \omega(\star)$ consisting of the open S-points of $\underline{\mathrm{Bun}}_{G(0,1)}$. Consider the moduli functor $\mathfrak{G}^{(2)}$ classifying the data $(x, \mathcal{E}', \mathcal{E}, \alpha, \beta)$ where $x \in \underline{\mathbb{G}}_m$, $\mathcal{E}' \in \underline{\mathrm{Bun}}_{G(0,1)}, \mathcal{E} \in \underline{\mathrm{Bun}}_{G(0,1)}^\circ$, $\alpha : \mathcal{E}|_{\mathbb{P}^1 - \{x\}} \xrightarrow{\sim} \mathcal{E}'|_{\mathbb{P}^1 - \{x\}}$ and $\beta : \mathcal{E}'|_{\mathbb{P}^1 - \{x\}} \xrightarrow{\sim} \mathcal{E}_{0,1}^\star|_{\mathbb{P}^1 - \{x\}}$.

We have morphisms

where $p_1(x, \mathcal{E}', \mathcal{E}, \alpha, \beta) = (x, \mathcal{E}', \beta); p_2(x, \mathcal{E}', \mathcal{E}, \alpha, \beta) = (x, \mathcal{E}, \beta \circ \alpha)$ and $p_3(x, \mathcal{E}', \mathcal{E}, \alpha, \beta) = (x, \mathcal{E}, \mathcal{E}', \alpha)$. The fibers of p_1 are isomorphic to the affine Grassmannian $\underline{\mathrm{Gr}}$, hence $\mathfrak{G}^{(2)}$ is also represented by an ind-scheme.

We have

$$\underline{\mathrm{IC}}_{V \otimes W}^\circ \cong (\underline{\mathrm{IC}}_V * \underline{\mathrm{IC}}_W)|_{\underline{\mathrm{GR}}^\circ} \cong p_{2,!}(p_1^* \underline{\mathrm{IC}}_V \otimes p_3^* \underline{\mathrm{IC}}_W). \tag{3.4}$$

Here the first equality follows from Remark 3.1.2. In the last term, we use $\underline{\mathrm{IC}}_W$ to denote also the complex on $\underline{\mathrm{Hk}}_{G(0,1)}$ whose restriction to each fiber of the projection $\underline{\mathrm{Hk}}_{G(0,1)} \to \underline{\mathbb{G}}_m \times \underline{\mathrm{Bun}}_{G(0,1)}$ (the fibers are identified with the affine Grassmannian) is isomorphic to $\underline{\mathrm{IC}}_W$. The second equality follows from the definition of the convolution product for sheaves on the affine Grassmannian in [MV07, § 4].

We consider the open ind-scheme $j : \mathfrak{G}^{(2),\circ} \hookrightarrow \mathfrak{G}^{(2)}$ classifying the data $(x, \mathcal{E}', \mathcal{E}, \alpha, \beta)$ with the open condition that $\mathcal{E}' \in \underline{\mathrm{Bun}}_{G(0,1)}^\circ$. Then there is an isomorphism

$$\mathfrak{G} \times_{\underline{\mathbb{G}}_m} \mathfrak{G} \xrightarrow{\sim} \mathfrak{G}^{(2),\circ}$$

mapping $(x, \varphi' : \mathcal{E}_{0,1}^{\omega'(\star)}|_{\mathbb{P}^1 - \{x\}} \to \mathcal{E}_{0,1}^\star|_{\mathbb{P}^1 - \{x\}}, \varphi : \mathcal{E}_{0,1}^{\omega(\star)}|_{\mathbb{P}^1 - \{x\}} \to \mathcal{E}_{0,1}^\star|_{\mathbb{P}^1 - \{x\}})$ to $(x, \mathcal{E}_{0,1}^{\omega'\omega(\star)}, \mathcal{E}_{0,1}^{\omega'(\star)}, \omega'(\varphi), \varphi')$. Under this isomorphism, the morphism $p_2^\circ = p_2|_{\mathfrak{G}^{(2),\circ}} : \mathfrak{G}^{(2),\circ} \to \underline{\mathrm{GR}}^\circ \cong \mathfrak{G}$ becomes the multiplication map under the group structure of $\mathfrak{G}$

$$m : \mathfrak{G} \times_{\underline{\mathbb{G}}_m} \mathfrak{G} \to \mathfrak{G}.$$

Moreover,

$$j^*(p_1^* \underline{\mathrm{IC}}_V \otimes p_3^* \underline{\mathrm{IC}}_W) \cong \underline{\mathrm{IC}}_V^\circ \boxtimes_{\underline{\mathbb{G}}_m} \underline{\mathrm{IC}}_W^\circ.$$

Therefore, by adjunction, we get a map

$$m_!(\underline{\mathrm{IC}}_V^\circ \boxtimes_{\underline{\mathbb{G}}_m} \underline{\mathrm{IC}}_W^\circ) \cong p_{2,!} j_! j^*(p_1^* \underline{\mathrm{IC}}_V \otimes p_3^* \underline{\mathrm{IC}}_W) \to p_{2,!}(p_1^* \underline{\mathrm{IC}}_V \otimes p_3^* \underline{\mathrm{IC}}_W) \cong \underline{\mathrm{IC}}_{V \otimes W}^\circ. \tag{3.5}$$

Since $\underline{F} : \underline{\mathfrak{G}} \to \underline{\mathbb{A}}^1$ is a group homomorphism, we have an isomorphism in $D_{\underline{\mathbb{G}_m}}^b(\underline{\mathbb{A}}^1)$:

$$\underline{F}_! m_!(\mathrm{IC}_V^\circ \boxtimes_{\underline{\mathbb{G}_m}} \mathrm{IC}_W^\circ) \cong \underline{F}_! \mathrm{IC}_V^\circ *_+ \underline{F}_! \mathrm{IC}_W^\circ.$$

Here $\mathcal{F}_1 *_+ \mathcal{F}_2 = a_!(\mathcal{F}_1 \boxtimes \mathcal{F}_2)$ means the convolution on $\underline{\mathbb{A}}^1$ using the addition map $a : \underline{\mathbb{A}}^1 \times \underline{\mathbb{A}}^1 \to \underline{\mathbb{A}}^1$. Descending to the stack $[\underline{\mathbb{A}}^1/\underline{\mathbb{G}_m}]$ and combining with (3.5), we get a morphism in $D_c^b([\underline{\mathbb{A}}^1/\underline{\mathbb{G}_m}])$, bifunctorial in V and W

$$\overline{F}_! \mathrm{IC}_V^\circ *_+ \overline{F}_! \mathrm{IC}_W^\circ \to \overline{F}_! \mathrm{IC}_{V \otimes W}^\circ.$$

Applying HFour and using the compatibility between convolution $*_+$ and the tensor product (see Theorem A.2.1(v)), we get

$$\mathrm{HFour}(\overline{F}_! \mathrm{IC}_V^\circ)_\eta \otimes \mathrm{HFour}(\overline{F}_! \mathrm{IC}_W^\circ)_\eta \cong \mathrm{HFour}(\overline{F}_! \mathrm{IC}_V^\circ *_+ \overline{F}_! \mathrm{IC}_W^\circ)_\eta[1] \to \mathrm{HFour}(\overline{F}_! \mathrm{IC}_{V \otimes W}^\circ)_\eta[1],$$

which gives (3.3).

Next we try to identify the mod p stalk of (3.3) with (2.15). Now we work over $\mathbb{F}_p$. We have the following commutative diagram.

$$
\begin{array}{ccc}
& \mathfrak{G} \times_{\mathbb{G}_m} \mathfrak{G} & \xrightarrow{(F,F)} \mathbb{A}^1 \times \mathbb{A}^1 \\
\scriptstyle(\pi,\pi) \swarrow & \downarrow{\scriptstyle m} & \downarrow{\scriptstyle a} \\
\mathbb{G}_m \xleftarrow{\quad \pi \quad} & \mathfrak{G} \xrightarrow{\quad F \quad} & \mathbb{A}^1
\end{array}
$$

We get a commutative diagram of cohomology groups.

$$\mathrm{H}_c^*(\mathfrak{G}, \mathrm{IC}_V^\circ \otimes F^* \mathrm{AS}_\psi) \otimes \mathrm{H}_c^*(\mathfrak{G}, \mathrm{IC}_W^\circ \otimes F^* \mathrm{AS}_\psi) \xrightarrow{\sim} \mathrm{H}_c^*(\mathbb{G}_m, \mathrm{Kl}^V) \otimes \mathrm{H}_c^*(\mathbb{G}_m, \mathrm{Kl}^W)$$

$$\mathrm{H}_c^*(\mathfrak{G} \times_{\mathbb{G}_m} \mathfrak{G}, \mathrm{IC}_V^\circ \boxtimes_{\mathbb{G}_m} \mathrm{IC}_W^\circ \otimes (F,F)^*(\mathrm{AS}_\psi \boxtimes \mathrm{AS}_\psi)) \xrightarrow{\sim} \mathrm{H}_c^*(\mathbb{G}_m, \mathrm{Kl}^V \otimes \mathrm{Kl}^W)$$

$$\downarrow{\scriptstyle \alpha} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \downarrow{\scriptstyle \wr}$$

$$\mathrm{H}_c^*(\mathfrak{G}, m_!(\mathrm{IC}_V^\circ \boxtimes_{\mathbb{G}_m} \mathrm{IC}_W^\circ) \otimes F^* \mathrm{AS}_\psi)$$

$$\downarrow{\scriptstyle (3.5)}$$

$$\mathrm{H}_c^*(\mathfrak{G}, \mathrm{IC}_{V \otimes W}^\circ \otimes F^* \mathrm{AS}_\psi) \xrightarrow{\qquad\qquad\qquad\qquad\qquad} \mathrm{H}_c^*(\mathbb{G}_m, \mathrm{Kl}^{V \otimes W})$$

The isomorphism α in the middle left column uses the composition of

$$(F,F)^*(\mathrm{AS}_\psi \boxtimes \mathrm{AS}_\psi) \cong (F,F)^* a^* \mathrm{AS}_\psi \cong m^* F^* \mathrm{AS}_\psi$$

and the projection formula

$$m_!((\mathrm{IC}_V^\circ \boxtimes_{\mathbb{G}_m} \mathrm{IC}_W^\circ) \otimes m^* F^* \mathrm{AS}_\psi) \cong m_!(\mathrm{IC}_V^\circ \boxtimes_{\mathbb{G}_m} \mathrm{IC}_W^\circ) \otimes F^* \mathrm{AS}_\psi.$$

It is easy to see that the composition of the left column is the same as i_p^* applied to the map (3.3); the right column is the cup product giving (2.15). Therefore, the result of i_p^* on (3.3) is the same as (2.15). □

3.2 Construction of the Galois representation

In this subsection we fix an irreducible representation V of $\widehat{G}^{\mathrm{Tt}}$, pure of weight w such that $V^{\widehat{G}^{\mathrm{geom}}} = 0$. Here $\widehat{G}^{\mathrm{geom}}$ is the Zariski closure of the geometric monodromy of $\mathrm{Kl}_{\widehat{G}}$ for large p as tabulated in Theorem 2.3.2. We note that this condition is rather weak: when $\widehat{G}^{\mathrm{geom}} = \widehat{G}$ this simply means that V is not the trivial representation.

3.2.1 *The open subset S'.* Let S_{bad} be the set of primes p such that $V^{\widehat{G}_p^{\mathrm{geom}}} \neq 0$, where $\widehat{G}_p^{\mathrm{geom}}$ is the Zariski closure of the geometric monodromy of $\mathrm{Kl}_{\widehat{G}}$ over $\mathbb{G}_{m,\mathbb{F}_p}$. According to the table in Theorem 2.3.2, $S_{\mathrm{bad}} \subset \{2, 3\}$. Let $S' = S - S_{\mathrm{bad}}$. This is a Zariski open subset of S.

When $\widehat{G} = \mathrm{SL}_{2n}$ or Sp_{2n}, $\mathrm{Kl}_{\widehat{G}}^{\mathrm{St}}$ is essentially the same as Kl_{2n} according to Proposition 2.4.3. Then by Katz's Theorem 2.4.4, $\widehat{G}_p^{\mathrm{geom}} = \mathrm{Sp}_{2n}$ for all primes p, hence $S' = S$ in these cases.

LEMMA 3.2.2. *The complex $\underline{M}_!^V[1]$ is a perverse sheaf on S, and $\underline{M}_!^V|_{S'}$ is a sheaf.*

Proof. We first show that $\underline{M}_!^V[1] \in {}^{\mathrm{p}}D^{\geqslant 0}(S, \mathbb{Q}_\ell)$. In fact, $\underline{\mathrm{IC}}_V[1]$ is perverse (because the base scheme S now has dimension one) and $\overline{F}$ when restricted to $\underline{\mathrm{Gr}}_{\leqslant\lambda}^\circ$ is affine (because $\underline{\mathrm{Gr}}_{\leqslant\lambda}^\circ$ is a closed subscheme of finite type of the ind algebraic group $\underline{\mathfrak{G}}_1$). Therefore, $\overline{F}_!\underline{\mathrm{IC}}_V[1] \in {}^{\mathrm{p}}D^{\geqslant 0}([\mathbb{A}^1/\underline{\mathbb{G}_m}], \mathbb{Q}_\ell)$. Since the functor HFour is t-exact with respect to the perverse t-structure (Theorem A.2.1(iv)), $\mathrm{HFour}(\overline{F}_!\underline{\mathrm{IC}}_V^\circ)[1] \in {}^{\mathrm{p}}D^{\geqslant 0}([\mathbb{A}^1/\underline{\mathbb{G}_m}], \mathbb{Q}_\ell)$ and hence $\underline{M}_!^V[1] \in {}^{\mathrm{p}}D^{\geqslant 0}(S, \mathbb{Q}_\ell)$.

We then show that the stalks of $\underline{M}_!^V$ lie in degree at most 1, which, combined with the previous paragraph will prove that $\underline{M}_!^V[1]$ is perverse. Since $\underline{M}_!^V$ is constructible, it suffices to show that $i_p^*\underline{M}_!^V$ lies in degree at most 1 for every prime $p \nmid \ell$, and in degree at most 0 for almost all p. In fact, by (3.2), the mod p stalk (up to tensoring with $\mathbb{Q}_\ell(\mu_p)$) is $M_{!,\mathbb{F}_p}^V \cong \mathrm{H}_c^*(\mathbb{G}_m, \mathrm{Kl}^V)[1]$, which lies in degree 0 and 1. For p appearing in S', the Zariski closure of the geometric monodromy of $\mathrm{Kl}_{\widehat{G}}$ is the same as $\widehat{G}^{\mathrm{geom}}$, hence $\mathrm{H}_c^2(\mathbb{G}_m, \mathrm{Kl}^V) = V_{\widehat{G}^{\mathrm{geom}}}(-1) = 0$ by our assumption on V, which implies that $i_p^*\underline{M}_!^V$ lies in degree 0. This last statement also shows that $\underline{M}_!^V|_{S'}$ is a sheaf. This finishes the proof. $\qquad\square$

PROPOSITION 3.2.3. *Recall w is the weight of V. There is a canonical pairing (of complexes on S):*

$$\underline{M}_!^V \otimes \underline{M}_!^V \to \mathbb{Q}_\ell(-w-1) \tag{3.6}$$

which is $(-1)^{w+1}$-symmetric when restricted to S', i.e. it factors through $\mathrm{Sym}^2(\underline{M}_!^V|_{S'})$ if w is odd and through $\wedge^2(\underline{M}_!^V|_{S'})$ if w is even.

Proof. For the trivial representation $\mathbf{1}$, $\mathrm{IC}_{\mathbf{1}}$ is the constant sheaf supported at the unit section $S \hookrightarrow \underline{\mathrm{Gr}}$. The zero section $z : S \to [\mathbb{A}^1/\underline{\mathbb{G}_m}]$ can be factored as $S \xrightarrow{g} \mathbb{B}\underline{\mathbb{G}_m} \xrightarrow{\alpha} [\mathbb{A}^1/\underline{\mathbb{G}_m}]$. Direct calculation shows that

$$\mathrm{HFour}(z_!\mathbb{Q}_\ell)_\eta \cong g^*g_!\mathbb{Q}_\ell[1].$$

Hence,

$$\underline{M}_!^{\mathbf{1}} = \mathrm{HFour}(\overline{F}_!\mathrm{IC}_{\mathbf{1}}^\circ)_\eta = \mathrm{HFour}(z_!\mathbb{Q}_\ell)_\eta \cong g^*g_!\mathbb{Q}_\ell[1].$$

Applying Proposition 3.1.5 to $W = V^\vee$ (dual representation) we get a canonical map

$$\underline{M}_!^V \otimes \underline{M}_!^{V^\vee} \to \underline{M}_!^{V \otimes V^\vee}[1] \xrightarrow{\epsilon} \underline{M}_!^{\mathbf{1}}[1] \cong g^*g_!\mathbb{Q}_\ell[2] \to \mathbb{Q}_\ell(-1). \tag{3.7}$$

Here the map ϵ is induced by the evaluation map $V \otimes V^\vee \to \mathbf{1}$; the last map is induced by the truncation $g_!\mathbb{Q}_\ell \to \mathbf{R}^2g_!\mathbb{Q}_\ell[-2] \cong \mathbb{Q}_\ell[-2](-1)$.

Next, we would like to produce an isomorphism

$$\underline{M}_!^V \cong \underline{M}_!^{V^\sigma} \tag{3.8}$$

whose mod p fiber is the same as the isomorphism (2.19) in Lemma 2.7.2. Now the argument of Lemma 2.7.2 no longer works because we do not have Kloosterman sheaves over S. What we do instead is to unravel the argument of [HNY13, §6.1] for the σ-invariance of $\mathrm{Kl}_{\widehat{G}}$ and give a construction of the isomorphism (3.8) which only uses the morphism $\overline{F}$ in (3.1). Let $\epsilon = -1$ if G is of type A_{2n} and $\epsilon = 1$ otherwise. We view ϵ as a point of $\mathbb{G}_m^{\mathrm{rot}}$. Consider the involution (σ, ϵ) acting on $\underline{\mathfrak{G}} \cong \underline{\mathrm{GR}}^\circ$, where σ is the pinned involution of G which induces $-w_0$ on $\mathbb{X}_*(T)$ and ϵ acts via $\mathbb{G}_m^{\mathrm{rot}}$. This action commutes with the $\underline{\mathbb{G}_m}^{(\rho^\vee, h)}$-action on $\underline{\mathfrak{G}}$ and descends to an involution on the stack $[\underline{\mathfrak{G}}/\underline{\mathbb{G}_m}^{(\rho^\vee, h)}]$. It is easy to check that $\overline{F}$ in (3.1) is invariant under this involution. Hence, we get an isomorphism

$$\underline{M}_!^V = \mathrm{HFour}(\overline{F}_! \underline{\mathrm{IC}}_V^\circ)_\eta \cong \mathrm{HFour}(\overline{F}_! (\sigma, \epsilon)^* \underline{\mathrm{IC}}_V^\circ)_\eta = \mathrm{HFour}(\overline{F}_! \underline{\mathrm{IC}}_{V^\sigma}^\circ)_\eta = \underline{M}_!^{V^\sigma}.$$

Since $V^\sigma \cong V^\vee(-w)$, we can apply (3.8) to the pairing (3.7) as we did in Proposition 2.7.3, and get the desired pairing (3.6). The construction guarantees that after taking the mod p stalk, (3.6) becomes the pairing (2.21).

Finally we determine the parity of the pairing (3.6). We write $\underline{M}_!^V|_{S'} \otimes \underline{M}_!^V|_{S'}$ as a direct sum $\mathrm{Sym}^2(\underline{M}_!^V|_{S'}) \oplus \wedge^2(\underline{M}_!^V|_{S'})$. Suppose first that w is even. Let

$$\alpha : \underline{\mathcal{S}} := \mathrm{Sym}^2(\underline{M}_!^V|_{S'}) \to \mathbb{Q}_\ell(-w-1)$$

be the restriction of the pairing (3.6) to the symmetric part of $\underline{M}_!^V|_{S'} \otimes \underline{M}_!^V|_{S'}$. By Proposition 2.7.3, each mod p stalk of the pairing (3.6) is skew symmetric, therefore $i_p^* \alpha = 0$. Since $\underline{\mathcal{S}}$ is a local system over an open subset $U \subset S'$ and $i_p^* \alpha = 0$ for those p appearing in U, $\alpha|_U$ must be zero. Hence, α factors through $\underline{\mathcal{S}} \to i_* i^* \underline{\mathcal{S}} \to \mathbb{Q}_\ell(-w-1)$ where $i : S' - U \hookrightarrow S'$ is the inclusion. The latter map becomes $\beta : i^* \underline{\mathcal{S}} \to i^! \mathbb{Q}_\ell(-w-1)$ by adjunction. But now $i^* \underline{\mathcal{S}}$ lies in degrees 0 by Lemma 3.2.2 while $i^! \mathbb{Q}_\ell$ lies in degree 2, β must also be zero. This shows that $\alpha = 0$ and the pairing (3.6) factors through $\wedge^2(\underline{M}_!^V|_{S'})$. The case where w is odd is proved in the same way. □

The pairing (3.6) induces a morphism

$$\underline{M}_!^V[1] \to (\mathbb{D}(\underline{M}_!^V[1]))(-w-2). \tag{3.9}$$

from the left copy of $\underline{M}_!^V$ to the dual of the right copy of $\underline{M}_!^V$. Here we use the convention that the dualizing complex on S is $\mathbb{Q}_\ell[2](1)$, so that the Verdier duality is $\mathbb{D}(-) = \mathbf{R}\underline{\mathrm{Hom}}_S(-, \mathbb{Q}_\ell[2](1))$. By Lemma 3.2.2, both the source and the target of the above morphism are perverse sheaves.

DEFINITION 3.2.4. We define an object $\underline{M}_{!*}^V \in D_c^b(S, \mathbb{Q}_\ell)$ by requiring that $\underline{M}_{!*}^V[1]$ be the image of morphism (3.9) in the abelian category of perverse sheaves on S.

By definition, we have a morphism $\underline{M}_!^V \to \underline{M}_{!*}^V$ which is a surjection of perverse sheaves after applying a shift. We denote the generic stalk of $\underline{M}_{!*}^V$ by $M_{!*,\mathbb{Q}}^V$, which is viewed as a continuous $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module (concentrated in degree 0).

By construction, $\ker(\underline{M}_!^V \to \underline{M}_{!*}^V)$ lies in the radical of the pairing (3.6). Therefore, (3.6) factors through a $(-1)^{w+1}$-symmetric pairing on $\underline{M}_{!*}^V|_{S'}$ such that following diagram is

commutative.

$$\begin{array}{ccc}
\underline{M}_!^V|_{S'} \otimes \underline{M}_!^V|_{S'} & \xrightarrow{(3.6)} & \mathbb{Q}_\ell(-w-1) \\
\downarrow \quad\quad \downarrow & & \| \\
\underline{M}_{!*}^V|_{S'} \otimes \underline{M}_{!*}^V|_{S'} & \longrightarrow & \mathbb{Q}_\ell(-w-1)
\end{array} \tag{3.10}$$

Taking the generic stalk $M_{!*,\mathbb{Q}}^V$ of $\underline{M}_{!*}^V$, we obtain a $(-1)^{w+1}$-symmetric Galois equivariant pairing on $M_{!*,\mathbb{Q}}^V$:

$$\langle\,,\,\rangle^V : M_{!*,\mathbb{Q}}^V \otimes M_{!*,\mathbb{Q}}^V \to \mathbb{Q}_\ell(-w-1). \tag{3.11}$$

This pairing is perfect because by construction, $\ker(M_{!,\mathbb{Q}}^V \to M_{!*,\mathbb{Q}}^V)$ is exactly the radical of the generic stalk of the pairing (3.6).

The Galois representation $M_{!*,\mathbb{Q}}^V$ is the Galois representation M_ℓ mentioned in Theorem 1.2.1. The rest of the theorem will be proved in the subsequent subsections.

3.3 Proof of Theorem 1.2.1(ii)

We keep the same notation from the previous subsection. We shall need the following notion. Let $U \subset S$ be a nonempty Zariski open subset.

DEFINITION 3.3.1. A $\mathbb{Q}_\ell$-sheaf $\underline{P}$ on U is called *punctual* if it is a direct sum of sheaves supported at closed points of U. A $\mathbb{Q}_\ell$-sheaf $\underline{M}$ on U is said to be a *middle extension sheaf*, if $\underline{M}[1]$ is a middle extension perverse sheaf, i.e. $\underline{M}[1] = u_{!*}u^*\underline{M}[1]$ for some non-empty open subset $u : U' \hookrightarrow U$.

We collect some elementary properties about middle extension $\mathbb{Q}_\ell$-sheaves, which are standard exercises in perverse sheaf theory and we omit the proof.

LEMMA 3.3.2. *Let* $\mathrm{ME}(U, \mathbb{Q}_\ell)$ *be the category of middle extension $\mathbb{Q}_\ell$-sheaves on U.*

(i) *The category* $\mathrm{ME}(U, \mathbb{Q}_\ell)$ *is a Serre subcategory of $\mathbb{Q}_\ell$-sheaves on U.*

(ii) *Let* $\underline{M} \in \mathrm{ME}(U, \mathbb{Q}_\ell)$, *then its geometric generic stalk M affords a continuous Galois representation*

$$\rho_{\underline{M}} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(M).$$

The functor $\underline{M} \mapsto (M, \rho_{\underline{M}})$ *gives an equivalence of categories*

$$\mathrm{ME}(U, \mathbb{Q}_\ell) \xrightarrow{\sim} \mathrm{Rep}_{\mathrm{cont}}^{\mathrm{aeur}}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathbb{Q}_\ell)$$

where $\mathrm{Rep}_{\mathrm{cont}}^{\mathrm{aeur}}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \mathbb{Q}_\ell)$ *is the abelian category of continuous representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into finite-dimensional $\mathbb{Q}_\ell$-vector spaces which are unramified at almost all primes p.*

(iii) *Let* $\mathrm{Spec}\,\mathbb{F}_p$ *be a closed point of U and let M be the geometric generic stalk of $\underline{M}$ which affords the* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action. *Fix an embedding of* $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ *and use it to define the decomposition and the inertia groups at p, i.e.* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) > \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) > \mathcal{I}_p$. *Then we have canonical isomorphisms of* Frob_p-modules

$$i_p^*\underline{M} \cong M^{\mathcal{I}_p}.$$

PROPOSITION 3.3.3. *The complex* $\underline{M}_{!*}^V|_{S'}$ *is a middle extension sheaf on S'.*

Proof. Let $\underline{M} = \underline{M}_!^V|_{S'}$. Then by Lemma 3.2.2, $\underline{M}[1]$ is a perverse sheaf satisfying the following stronger condition: the mod p stalks of $\underline{M}$ all lie in degree -1. We make the following claims.

(i) There is no nonzero map from $\underline{M}[1]$ to punctual sheaves.

(ii) We claim that $\underline{M}[1]$ fits into an exact sequence of perverse sheaves

$$0 \to \underline{P} \to \underline{M}[1] \to \underline{N}[1] \to 0$$

where $\underline{P}$ is a punctual sheaf and $\underline{N}$ is a middle extension sheaf.

For property (1), suppose $i_{p,*}P$ is a punctual sheaf supported at some closed point $\operatorname{Spec} \mathbb{F}_p \hookrightarrow S'$, with P in degree zero, then a map $\underline{M}[1] \to i_{p,*}P$ is adjoint to $i_p^*\underline{M}[1] \to P$. Since the stalk $i_p^*\underline{M}[1]$ lies in degree -1, this map must be zero.

For property (2), let $u : S' \hookrightarrow S'$ be the open locus where $\underline{M}$ is lisse, and consider the natural map of perverse sheaves $\alpha : u_!u^*\underline{M}[1] \to \underline{M}[1]$. Since the cokernel of α is punctual, and by claim (1) $\underline{M}[1]$ does not map to punctual sheaves, α must be surjective. Since the kernel of α is also punctual, it is contained in the maximal punctual subsheaf $\ker(u_!u^*\underline{M}[1] \to u_{!*}u^*\underline{M}[1])$. Therefore, we can write $\underline{M}[1]$ as an extension of the middle extension perverse sheaf $\underline{N}[1] = u_{!*}u^*\underline{M}[1]$ and a punctual sheaf as in property (2).

Now let $\underline{M}' = \mathbb{D}(\underline{M}_!^V)|_{S'}[-2](-w - 2)$. By Verdier duality, the perverse sheaf $\underline{M}'[1]$ satisfies the dual of the two properties (1) and (2): $\underline{M}'[1]$ does not admit any nonzero map from punctual sheaves and it fits into an exact sequence

$$0 \to \underline{N}'[1] \to \underline{M}'[1] \to \underline{P}' \to 0$$

where $\underline{N}' \in \operatorname{ME}(S', \mathbb{Q}_\ell)$ and $\underline{P}'$ is punctual.

Consider the morphism $\underline{M}[1] \to \underline{M}'[1]$ given by (3.9). The properties (1) and (2) of $\underline{M}[1]$ and the dual properties of $\underline{M}'$ imply that $\underline{M}[1] \to \underline{M}'[1]$ necessarily factors as

$$\underline{M}[1] \twoheadrightarrow \underline{N}[1] \xrightarrow{\gamma} \underline{N}'[1] \hookrightarrow \underline{M}'[1].$$

Therefore, $\underline{M}_{!*}^V \cong \operatorname{Im}(\underline{N} \xrightarrow{\gamma[-1]} \underline{N}')$ is a middle extension sheaf, since both $\underline{N}$ and $\underline{N}'$ are. $\quad\square$

3.3.4 *Proof of Theorem 1.2.1(ii).* Let $p \neq \ell$. The condition that Kl_G^V over $\mathbb{G}_{m,\overline{\mathbb{F}}_p}$ does not contain a trivial sub-local system is equivalent to $V^{\widehat{G}_p^{\operatorname{geom}}} = 0$. For such a prime p, $\operatorname{Spec} \mathbb{F}_p$ is contained in S'. Consider the perfect pairing $\langle\,,\rangle^V$ (3.11) of the $\operatorname{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$-module $M := M_{!*,\mathbb{Q}}^V$. Let $\mathcal{I}_p < \operatorname{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ be the inertia group. A general fact says that the restriction of the pairing $\langle\,,\rangle^V$ to $\mathcal{I}_p$-invariants may not be perfect, but it factors through a *perfect* pairing between $\overline{M} := \operatorname{Im}(M^{\mathcal{I}_p} \to M_{\mathcal{I}_p})$:

$$\begin{array}{ccc}
M^{\mathcal{I}_p} \otimes M^{\mathcal{I}_p} & \longrightarrow & \mathbb{Q}_\ell(-w-1) \\
\downarrow \quad \downarrow & & \| \\
\overline{M} \otimes \overline{M} & \xrightarrow{\text{perfect}} & \mathbb{Q}_\ell(-w-1)
\end{array} \tag{3.12}$$

Here the vertical maps are the quotient map $M^{\mathcal{I}_p} \twoheadrightarrow \overline{M}$.

By Proposition 3.3.3, $\underline{M}_{!*}^V|_{S'}$ is a middle extension sheaf, hence $i_p^*\underline{M}_{!*}^V \cong M^{\mathcal{I}_p}$ by Lemma 3.3.2(iii). On the other hand, taking the mod p stalks of the diagram (3.10) and using Lemma 3.1.4, we get another diagram.

$$\begin{array}{ccc}
M_{!,\mathbb{F}_p}^V \otimes M_{!,\mathbb{F}_p}^V & \longrightarrow & \mathbb{Q}_\ell(\mu_p)(-w-1) \\
\downarrow \quad \downarrow & & \| \\
(M^{\mathcal{I}_p} \otimes \mathbb{Q}_\ell(\mu_p)) \otimes_{\mathbb{Q}_\ell(\mu_p)} (M^{\mathcal{I}_p} \otimes \mathbb{Q}_\ell(\mu_p)) & \longrightarrow & \mathbb{Q}_\ell(\mu_p)(-w-1)
\end{array} \tag{3.13}$$

Combining (3.12) and (3.13), we get a commutative diagram of pairings.

$$
\begin{array}{ccc}
M^V_{!,\mathbb{F}_p} \otimes M^V_{!,\mathbb{F}_p} & \longrightarrow & \mathbb{Q}_\ell(\mu_p)(-w-1) \\
\downarrow \quad \downarrow & & \Big\| \\
(\overline{M} \otimes \mathbb{Q}_\ell(\mu_p)) \otimes_{\mathbb{Q}_\ell(\mu_p)} (\overline{M} \otimes \mathbb{Q}_\ell(\mu_p)) & \xrightarrow{\text{perfect}} & \mathbb{Q}_\ell(\mu_p)(-w-1)
\end{array}
$$

In other words, the natural map $M^V_{!,\mathbb{F}_p} \to \overline{M} \otimes \mathbb{Q}_\ell(\mu_p)$ kills the radical R of the pairing on $M^V_{!,\mathbb{F}_p}$. On the other hand, we also know from Remark 2.7.4 that the same pairing on $M^V_{!,\mathbb{F}_p}$ factors through a *perfect* pairing on the quotient $M^V_{!*,\mathbb{F}_p}$, identifying $M^V_{!*,\mathbb{F}_p}$ as $M^V_{!,\mathbb{F}_p}/R$. Therefore, we get the desired inclusion (1.6).

We have the following immediate corollary of Theorem 1.2.1(ii). Recall from Lemma 2.5.7 that $\dim M^V_{!*,\mathbb{F}_p}$ is a constant for p large; we denote this dimension by $d^V_{!*}$.

COROLLARY 3.3.5. *Assumptions as in Theorem 1.2.1. Let B be the set of primes p such that either $V^{\widehat{G}_p^{\mathrm{geom}}} \neq 0$ or $\dim_{\mathbb{Q}_\ell(\mu_p)} M^V_{!*,\mathbb{F}_p} \neq d^V_{!*}$. Then the system of Galois representations $\{\rho^V_\ell\}$ is Frobenius-compatible (same as the notion of 'strictly compatible' in [Ser89, § 2.3]) with exceptional set B. More precisely, when $p \notin B \cup \{\ell\}$, the Galois representation ρ^V_ℓ is unramified at p, and $M^V_{!*,\mathbb{F}_p}$ is isomorphic to $M^V_{!*,\mathbb{Q}} \otimes \mathbb{Q}_\ell(\mu_p)$ as Frob_p-modules. The characteristic polynomial of $\rho^V_\ell(\mathrm{Frob}_p)$ has $\mathbb{Z}$-coefficients which are independent of ℓ, and all of its roots are p-Weil numbers of weight $w+1$.*

3.4 Proof of Theorem 1.2.1(i)

3.4.1 *Weight filtration on Galois representations.* By [Del80, § 6.1.1], we may talk about mixed sheaves over varieties $X_\mathbb{Q}$ defined over $\mathbb{Q}$. A mixed $\mathbb{Q}_\ell$-sheaf $\mathcal{F}$ over $X_\mathbb{Q}$ is a sheaf that extends to a sheaf $\underline{\mathcal{F}}$ on an integral model $\underline{X}$ of $X_\mathbb{Q}$ over $\mathbb{Z}[1/N]$ (for some N) such that $\underline{\mathcal{F}}$ admits a filtration with pure subquotients on each special fiber of $\underline{X}$. In particular we may talk about mixed $\mathbb{Q}_\ell$-sheaves over Spec $\mathbb{Q}$, i.e. continuous $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules on $\mathbb{Q}_\ell$-vector spaces with weight filtrations. In [Del80, § 6.1], Deligne proved that the category of mixed sheaves on varieties over $\mathbb{Q}$ are stable under the usual sheaf-theoretic operations. In particular, for a mixed sheaf $\mathcal{F}$ over $X_\mathbb{Q}$, its cohomology $\mathrm{H}^i_c(X_{\overline{\mathbb{Q}}}, \mathcal{F})$ and $\mathrm{H}^i(X_{\overline{\mathbb{Q}}}, \mathcal{F})$ are $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules with canonical weight filtrations.

Notation: in what follows, when we write $\mathrm{H}^i(X, \mathbb{Q}_\ell)$ or $\mathrm{H}^i_c(X, \mathbb{Q}_\ell)$ for an algebraic variety X over $\mathbb{Q}$, we always mean $\mathrm{H}^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$ or $\mathrm{H}^i_c(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$, as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules.

3.4.2 *Proof of Theorem 1.2.1(i).* Up to Tate twist we may assume $\mathrm{IC}_V = \mathrm{IC}_\lambda$, which is pure of weight $w = \langle 2\rho, \lambda \rangle$. To simplify notation, we denote $\mathrm{IC}_{\lambda,\mathrm{GR}}$ by IC_λ, and denote $\mathrm{IC}_{\lambda,\mathrm{GR}^\circ}$ by $\mathrm{IC}^\circ_\lambda$.

Applying Theorem A.2.1(i) to the object $\overline{F}_!\mathrm{IC}^\circ_\lambda$, we get a long exact sequence

$$\cdots \to \mathrm{H}^0_c(F^{-1}(1), \mathrm{IC}^\circ_\lambda) \to \mathrm{H}^0 M^V_{!,\mathbb{Q}} \to \mathrm{H}^1_c(F^{-1}(0), \mathrm{IC}^\circ_\lambda) \to \cdots . \tag{3.14}$$

Every object in this sequence carries a weight filtration. Since $\mathrm{IC}^\circ_\lambda$ is pure of weight w, $F_!\mathrm{IC}^\circ_\lambda$ is of weight at most w. This implies that $\mathrm{H}^1(F^{-1}(0), \mathrm{IC}^\circ_\lambda)$ has weight at most $w+1$ and $\mathrm{H}^0(F^{-1}(1), \mathrm{IC}^\circ_\lambda)$ has weight at most w. Since $M^V_{!*,\mathbb{Q}}$ is pure of weight $w+1$ (Corollary 3.3.5), the exact sequence (3.14) implies that $M^V_{!*,\mathbb{Q}}$ (as a quotient of $\mathrm{H}^0 M^V_{!,\mathbb{Q}}$) can be identified with a subspace of $\mathrm{Gr}^W_{w+1} \mathrm{H}^1_c(F^{-1}(0), \mathrm{IC}^\circ_\lambda)$.

The complex $\mathrm{IC}_\lambda^\circ$ is the intersection cohomology complex of $\mathrm{GR}_{\leqslant\lambda}^\circ$, which is an open subset of $\mathrm{GR}_{\leqslant\lambda}$. We have a $T \times \mathbb{G}_m^{\mathrm{rot}}$-equivariant resolution of singularities $\nu : Y = Y_1 \times \mathbb{G}_m \to \mathrm{Gr}_{\leqslant\lambda} \times \mathbb{G}_m = \mathrm{GR}_{\leqslant\lambda}$ given by the Bott–Samelson resolution of $\mathrm{Gr}_{\leqslant\lambda}$. Let $\nu^\circ : Y^\circ \to \mathrm{GR}_{\leqslant\lambda}^\circ$ be the restriction of ν to $\mathrm{GR}_{\leqslant\lambda}^\circ$, which is again $T \times \mathbb{G}_m^{\mathrm{rot}}$-equivariant. By the decomposition theorem [BBD82, Théorème 6.2.5], $\mathrm{IC}_\lambda^\circ$ is geometrically (i.e. over $\overline{\mathbb{Q}}$) a direct summand of $^{\mathrm{p}}\mathrm{H}^0 \nu_*^\circ \mathbb{Q}_\ell[w]$. It is necessarily invariant under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ because it is the only direct summand of $^{\mathrm{p}}\mathrm{H}^0\nu_*^\circ\mathbb{Q}_\ell[w]$ with full support. Therefore, $\mathrm{H}_c^1(F^{-1}(0), \mathrm{IC}_\lambda^\circ)$ is a direct summand of $\mathrm{H}_c^1(F^{-1}(0), {}^{\mathrm{p}}\mathrm{H}^0\nu_*^\circ\mathbb{Q}_\ell[w])$. On the other hand, the Leray spectral sequence attached to the perverse filtration on $\nu_*^\circ\mathbb{Q}_\ell[w]$ degenerates at the E_1 page by the decomposition theorem again, therefore $\mathrm{H}_c^1(F^{-1}(0), {}^{\mathrm{p}}\mathrm{H}^0\nu_*^\circ\mathbb{Q}_\ell[w])$ is in turn a subquotient of $\mathrm{H}_c^1(F^{-1}(0), \nu_*^\circ\mathbb{Q}_\ell[w])$. Let $Z = (F \circ \nu^\circ)^{-1}(0)$, then $\mathrm{H}_c^1(F^{-1}(0), \nu_*^\circ\mathbb{Q}_\ell[w]) = \mathrm{H}_c^{w+1}(Z, \mathbb{Q}_\ell)$. At this point we have shown that $M_{!*,\mathbb{Q}}^V$ is a subquotient of $\mathrm{Gr}_{w+1}^W \mathrm{H}_c^{w+1}(Z, \mathbb{Q}_\ell)$ as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules.

Since $F \circ \nu^\circ : Y^\circ \to \mathbb{A}^1$ is dominant and Y° is smooth irreducible of dimension $w+1$, we have $\dim Z = w$. Since $F \circ \nu^\circ$ is $\mathbb{G}_m^{(\rho^\vee,h)}$-equivariant, the torus $\mathbb{G}_m^{(\rho^\vee,h)}$ still acts on $Z = (F \circ \nu^\circ)^{-1}(0)$. Moreover, we have a projection $\pi_Z : Z \subset Y^\circ \subset Y_1 \times \mathbb{G}_m \to \mathbb{G}_m$ which is equivariant under the homomorphism $\mathbb{G}_m^{(\rho^\vee,h)} \to \mathbb{G}_m^{\mathrm{rot}}$. Let $Z_1 = \pi_Z^{-1}(1)$, then $\dim Z_1 = w - 1$ and $Z = Z_1 \overset{\mu_h}{\times} \mathbb{G}_m^{(\rho^\vee,h)}$. By Künneth formula, we then have

$$\mathrm{Gr}_{w+1}^W \mathrm{H}_c^{w+1}(Z, \mathbb{Q}_\ell) \cong \mathrm{Gr}_{w+1}^W \mathrm{H}_c^{w+1}(Z_1 \times \mathbb{G}_m^{(\rho^\vee,h)}, \mathbb{Q}_\ell)^{\mu_h}$$
$$\cong \mathrm{Gr}_{w+1}^W \mathrm{H}_c^w(Z_1, \mathbb{Q}_\ell)^{\mu_h} \otimes \mathrm{Gr}_0^W \mathrm{H}_c^1(\mathbb{G}_m)$$
$$\oplus \mathrm{Gr}_{w-1}^W \mathrm{H}_c^{w-1}(Z_1, \mathbb{Q}_\ell)^{\mu_h} \otimes \mathrm{Gr}_2^W \mathrm{H}_c^2(\mathbb{G}_m). \qquad (3.15)$$

Since $\mathrm{H}_c^*(Z_1, \mathbb{Q}_\ell)$ is the stalk at 0 of the complex $(F \circ \nu_1^\circ)_! \mathbb{Q}_\ell$ (where $\nu_1^\circ : Y_1^\circ \to \mathrm{Gr}_{\leqslant\lambda}^\circ$ is the resolution), which has weight at most 0, therefore $\mathrm{H}_c^w(Z_1, \mathbb{Q}_\ell)$ has weight at most w. Therefore, the first summand in (3.15) is zero, and we have

$$\mathrm{Gr}_{w+1}^W \mathrm{H}_c^{w+1}(Z, \mathbb{Q}_\ell) \cong \mathrm{Gr}_{w-1}^W \mathrm{H}_c^{w-1}(Z_1, \mathbb{Q}_\ell)^{\mu_h}(-1).$$

At this point, we have shown that $M_{!*,\mathbb{Q}}^V$ is a subquotient of $\mathrm{Gr}_{w-1}^W \mathrm{H}_c^{w-1}(Z_1, \mathbb{Q}_\ell)(-1)$ as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules, where Z_1 is an algebraic variety over $\mathbb{Q}$ of dimension $w - 1$.

Let $\overline{Z}_1$ be the closure of Z_1 in Y_1, which is projective and defined over $\mathbb{Q}$. We then have $\mathrm{Gr}_{w-1}^W \mathrm{H}_c^{w-1}(Z_1, \mathbb{Q}_\ell) \hookrightarrow \mathrm{Gr}_{w-1}^W \mathrm{H}^{w-1}(\overline{Z}_1, \mathbb{Q}_\ell)$. Finally let $X \to \overline{Z}_1$ be a resolution of singularities over $\mathbb{Q}$, so that X is smooth and projective of dimension $w - 1$, then $\mathrm{Gr}_{w-1}^W \mathrm{H}^{w-1}(\overline{Z}_1, \mathbb{Q}_\ell) \hookrightarrow \mathrm{H}^{w-1}(X, \mathbb{Q}_\ell)$ (see [Del74, Proposition 8.2.5]). Combining the previous steps we have shown that $M_{!*,\mathbb{Q}}^V$ is a subquotient of $\mathrm{H}_c^{w-1}(X, \mathbb{Q}_\ell)(-1)$ as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules, where X is smooth and projective of dimension $w - 1$ over $\mathbb{Q}$. This finishes the proof of Theorem 1.2.1(i).

4. Symmetric power moments of Kl_2

In this section we shall prove Evans's conjectures. We start by giving more precise information on the motives underlying the symmetric power moments of the classical Kloosterman sheaves in § 4.1. In §§ 4.2 and 4.3, we make explicit calculations on moments of Kl_2, which are necessary for the proof of Theorem 1.1.6 in § 4.4. Evans's conjectures are proved in §§ 4.5–4.7.

4.1 The motives underlying the symmetric power moments
In this subsection, we explicitly describe the algebraic varieties over $\mathbb{Z}$ whose ℓ-adic cohomology accommodates the Galois representations attached to the symmetric power moments of Kl_n.

4.1.1 *Bifiltered vector bundles.* We shall work in the slightly different context where $G = \mathrm{GL}_n$ instead of PGL_n. We first describe the moduli stacks $\mathrm{Bun}_n(0, i) := \mathrm{Bun}_{\mathrm{GL}_n(0,i)}$ (as recalled in § 2.6.1) more explicitly in this case.

The stack $\mathrm{Bun}_n(0, 0)$ classifies triples $(\mathcal{E}, F^*\mathcal{E}, F_*\mathcal{E})$ where:

(i) $\mathcal{E}$ is a vector bundle of rank n on $\mathbb{P}^1$;
(ii) a decreasing filtration $F^*\mathcal{E}$ giving a complete flag of the fiber of $\mathcal{E}$ at ∞,

$$\mathcal{E} = F^0\mathcal{E} \supset F^1\mathcal{E} \supset \cdots \supset F^n\mathcal{E} = \mathcal{E}(-\{\infty\});$$

(iii) an increasing filtration $F_*\mathcal{E}$ giving a complete flag of the fiber of $\mathcal{E}$ at 0,

$$\mathcal{E}(-\{0\}) = F_0\mathcal{E} \subset F_1\mathcal{E} \subset \cdots \subset F_n\mathcal{E} = \mathcal{E};$$

Remark 4.1.2. Given a triple $(\mathcal{E}, F^*\mathcal{E}, F_*\mathcal{E})$ as above, we can extend the filtration $F_*\mathcal{E}$ to all integer indices such that $F_{i+n}\mathcal{E} = F_i\mathcal{E} \otimes \mathcal{O}(\{0\})$. Similarly we can extend the filtration $F^*\mathcal{E}$ to all integer indices such that $F^{i+n}\mathcal{E} = F^i\mathcal{E} \otimes \mathcal{O}(-\{\infty\})$.

The moduli stack $\mathrm{Bun}_n(0, 1)$ classifies the data $(\mathcal{E}, F^*\mathcal{E}, F_*\mathcal{E}, \{\overline{v}_i\}_{i=1}^n)$ where $(\mathcal{E}, F^*\mathcal{E}, F_*\mathcal{E})$ is as above and $\overline{v}_i$ is a basis of $F^{i-1}\mathcal{E}/F^i\mathcal{E}$ for $i = 1, 2, \ldots, n$.

The moduli stack $\mathrm{Bun}_n(0, 2)$ classifies the data $(\mathcal{E}, F^*\mathcal{E}, F_*\mathcal{E}, \{\widetilde{v}_i\}_{i=1}^n)$ where $(\mathcal{E}, F^*\mathcal{E}, F_*\mathcal{E})$ is as above and $\widetilde{v}_i \in F^{i-1}\mathcal{E}/F^{i+1}\mathcal{E}$ is a vector whose projection to $F^{i-1}\mathcal{E}/F^i\mathcal{E}$ is nonzero ($i = 1, 2, \ldots, n$).

4.1.3 *Open points.* The stack $\mathrm{Bun}_n(0, i)$ is decomposed into connected components $\mathrm{Bun}_n^d(0, i)$ according to the integer $d = \deg(\mathcal{E})$. For each component of $\mathrm{Bun}_n^d(0, 1)$, there is a unique open point $(\mathcal{E}_d, F^*, F_*, \{\overline{e}_i\})$ which we recall now (see also [HNY13, § 3.1]).

When $d = 0$, we take $\mathcal{E}_0 = \mathcal{O}e_1 \oplus \cdots \oplus \mathcal{O}e_n$. The symbols $e_1, \ldots, e_n$ are incorporated only to emphasize the ordering of the various factors. Define $F^i\mathcal{E}_0 = \mathcal{O}(-\{\infty\})e_1 \oplus \cdots \oplus \mathcal{O}(-\{\infty\})e_i \oplus \mathcal{O}e_{i+1} \oplus \cdots \oplus \mathcal{O}e_n$ for $i = 0, \ldots, n-1$, and extend this definition to $F^i\mathcal{E}_0$ for all $i \in \mathbb{Z}$ as in Remark 4.1.2. Define $F_i\mathcal{E}_0 = \mathcal{O}e_1 \oplus \cdots \oplus \mathcal{O}e_i \oplus \mathcal{O}(-\{0\})e_{i+1} \oplus \cdots \oplus \mathcal{O}(-\{0\})e_n$ for $i = 0, \ldots, n-1$ and extend this definition to $F_i\mathcal{E}_0$ for all $i \in \mathbb{Z}$ as in Remark 4.1.2. This finishes the definition of $(\mathcal{E}_0, F^*, F_*)$.

For arbitrary integer d, we define $\mathcal{E}_d = F_{n+d}\mathcal{E}_0$ and $F_i\mathcal{E}_d = F_{i+d}\mathcal{E}_0$. Therefore, we have a canonical isomorphism between $\mathcal{E}_d$ and $\mathcal{E}_0$ over $\mathbb{P}^1 - \{0\}$. We define $F^i\mathcal{E}_d|_{\mathbb{P}^1 - \{0\}}$ to be the image of $F^i\mathcal{E}_0$ under this isomorphism, and let $F^i\mathcal{E}_d$ be the vector bundle obtained by gluing $F^i\mathcal{E}_d|_{\mathbb{P}^1 - \{0\}}$ and $\mathcal{E}_d|_{\mathbb{P}^1 - \{\infty\}}$ over $\mathbb{G}_m$.

The basis vectors $\{e_i\}$ give a canonical basis $\overline{e}_i \in F^{i-1}\mathcal{E}_0/F^i\mathcal{E}_0 = F^{i-1}\mathcal{E}_d/F^i\mathcal{E}_d$. It also gives a vector $\widetilde{e}_i \in F^{i-1}\mathcal{E}_0/F^{i+1}\mathcal{E}_0 = F^{i-1}\mathcal{E}_d/F^{i+1}\mathcal{E}_d$.

For each $d \geqslant 0$, let $\mathfrak{G}_d$ be the space of homomorphisms $\phi : \mathcal{E}_0 \to \mathcal{E}_d$ of coherent sheaves such that:

(a) $\phi(F^i\mathcal{E}_0) \subset F^i\mathcal{E}_d$ and $\phi(F_i\mathcal{E}_0) \subset F_i\mathcal{E}_d$ for all $i \in \mathbb{Z}$;
(b) $\phi(\overline{e}_i) = \overline{e}_i$, under the natural identification $F^{i-1}\mathcal{E}_0/F^i\mathcal{E}_0 = F^{i-1}\mathcal{E}_d/F^i\mathcal{E}_d$;
(c) the zeros of ϕ are concentrated at a single point in $\mathbb{G}_m$.

4.1.4 *Torus action.* Let $T_n = \mathbb{G}_m^n$ be the diagonal torus of GL_n. The object $(\mathcal{E}_d, F^*, F_*)$ admits an action of T_n by $(t_1, \ldots, t_n) \cdot e_i = t_i e_i$. Let s be the local coordinate at $\infty \in \mathbb{P}^1$. The torus $\mathbb{G}_m^{\mathrm{rot}}$ acts on $\mathbb{P}^1$ by scaling s. We normalize the equivariant structure of $\mathcal{E}_0$ by making $\mathbb{G}_m^{\mathrm{rot}}$

act trivially on e_i, which then induces a canonical $\mathbb{G}_m^{\mathrm{rot}}$-equivariant structure on all $\mathcal{E}_d$ and their filtrations.

The torus T_n acts on $\mathfrak{S}^d$ by $t \cdot \phi = t \circ \phi \circ t^{-1}$, where the t (respectively t^{-1}) refers to the action of T_n on $\mathcal{E}_0$ (respectively $\mathcal{E}_d$). This action factors through the adjoint torus $T_n^{\mathrm{ad}} \subset \mathrm{PGL}_n$. The $\mathbb{G}_m^{\mathrm{rot}}$-equivariant structures on both $(\mathcal{E}_0, F^*, F_*)$ and $(\mathcal{E}_d, F^*, F_*)$ induce an $\mathbb{G}_m^{\mathrm{rot}}$-action on $\mathfrak{S}^d$, which commutes with the action of T_n.

Let $\phi : \mathcal{E}_0 \to \mathcal{E}_d$ be a point in $\mathfrak{S}^d$. For $i = 1, \ldots, n-1$ we can write $\phi(\widetilde{e}_i) = \widetilde{e}_i + a_i \overline{e}_{i+1}$ mod $F^{i+1}\mathcal{E}_d$ for some $a_i \in \mathbb{A}^1$. For $i = n$, $\phi(\widetilde{e}_n) = \widetilde{e}_n + a_n s \overline{e}_1$ mod $F^{i+1}\mathcal{E}_d$ for some $a_n \in \mathbb{A}^1$. The assignment $\phi \mapsto (a_1, \ldots, a_n)$ defines a morphism $f = f_{n,d} : \mathfrak{S}^d \to \mathbb{A}^n$. A direct calculation shows that the map $f_{n,d}$ is $T_n \times \mathbb{G}_m^{\mathrm{rot}}$-equivariant.

We thus get the following diagram which is analogous to (2.2) and (2.6).

$$
\begin{array}{ccc}
 & \mathfrak{S}^d & \\
{\scriptstyle \pi}\swarrow & {\scriptstyle f_n^d}\downarrow & \searrow{\scriptstyle F_n^d} \\
\mathbb{G}_m & & \mathbb{A}^n \xrightarrow{\ \sigma\ } \mathbb{A}^1
\end{array}
$$

Here π takes $\phi \in \mathfrak{S}^d$ to the support of the zeros of ϕ, which is a unique point in $\mathbb{G}_m$ by assumption on ϕ. The schemes in the above diagram are defined over $\mathbb{Q}$.

To emphasize the dependence on n and on the prime ℓ, we introduce the notation

$$M_{n,\ell}^d := M_{!*,\mathbb{Q}}^{\mathrm{Sym}^d} \tag{4.1}$$

where the right-hand side is the generic stalk of the object defined in Definition 3.2.4 for the Sym^d moment of the usual Kloosterman sheaf Kl_n.

PROPOSITION 4.1.5. *Let $Z_0 = F_n^{d,-1}(0)$ and $Z_1 = F_n^{d,-1}(1)$. These are affine varieties of dimension $d(n-1)$ over $\mathbb{Q}$.*

(i) *The $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module $M_{n,\ell}^d$ fits into a short exact sequence*

$$0 \to M_{n,\ell}^d \to \mathrm{Gr}_{d(n-1)+1}^W \mathrm{H}_c^{d(n-1)+1}(Z_{0,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \xrightarrow{\mathrm{Sp}} \mathrm{Gr}_{d(n-1)+1}^W \mathrm{H}_c^{d(n-1)+1}(Z_{1,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \to 0. \tag{4.2}$$

where the map Sp is induced from the specialization map $\mathrm{Sp} : \mathrm{H}_c^*(Z_0, \mathbb{Q}_\ell) \to \mathrm{H}_c^*(Z_1, \mathbb{Q}_\ell)$.

(ii) *Define a $\mathbb{Q}$-Hodge structure $M_{n,\mathrm{Hod}}^d$ of weight $d(n-1)+1$ to be the kernel of the specialization map*

$$\mathrm{Sp}_{\mathrm{Hod}} : \mathrm{Gr}_{d(n-1)+1}^W \mathrm{H}_c^{d(n-1)+1}(Z_0(\mathbb{C}), \mathbb{Q}) \to \mathrm{Gr}_{d(n-1)+1}^W \mathrm{H}_c^{d(n-1)+1}(Z_1(\mathbb{C}), \mathbb{Q}).$$

Then the multiset of Hodge–Tate weights of $M_{n,\ell}^d$ (viewed as a $\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$-module) is the same as the multiset where $i \in \mathbb{Z}$ appears $\dim \mathrm{Gr}_F^{-i} M_{n,\mathrm{Hod}}^d$ (where F is the Hodge filtration on $M_{n,\mathrm{Hod}}^d \otimes \mathbb{C}$). In particular, the multiset of Hodge–Tate weights of $M_{n,\ell}^d$ is independent of ℓ.

Proof. (1) The proof is a more refined version of the first part of the argument in §3.4.2. By Lemma 3.1.4, Corollary 3.3.5 and Lemma 2.5.4, for large p we may identify the mod p stalk of $\underline{M}_{!*}^{\mathrm{Sym}^d}$ with the weight $d(n-1)+1$-quotient of the mod p stalk of $\underline{M}_!^{\mathrm{Sym}^d}$. Therefore, we have

$$M_{n,\ell}^d = M_{!*,\mathbb{Q}}^{\mathrm{Sym}^d} = \mathrm{Gr}_{d(n-1)+1}^W M_{!,\mathbb{Q}}^{\mathrm{Sym}^d}.$$

The morphism F_n^d is equivariant under the action of $\mathbb{G}_m^{(\rho^\vee, n)} \subset T_n \times \mathbb{G}_m^{\text{rot}}$ on $\mathfrak{S}^d$ and the dilation action on $\mathbb{A}^1$, hence it descends to a morphism of stacks

$$\overline{F}_n^d : [\mathfrak{S}^d / \mathbb{G}_m^{(\rho^\vee, n)}] \to [\mathbb{A}^1 / \mathbb{G}_m].$$

The scheme $\mathfrak{S}^d$ is naturally a subscheme of $\mathfrak{S}$ in the context of § 2.6.1 for the group $G = \text{PGL}_n$. When $V = \text{Sym}^d$, the intersection complex $\text{IC}_{V, \text{GR}^\circ}$ is the shifted constant sheaf $\mathbb{Q}_\ell[d(n-1)]$ on $\mathfrak{S}^d$. Therefore by Definition 3.1.3, we have

$$M_{!,\mathbb{Q}}^{\text{Sym}^d} \cong \text{HFour}(\overline{F}_{n,!}^d \mathbb{Q}_\ell)_\eta [d(n-1)].$$

By Theorem A.2.1(i), we get an exact sequence

$$\text{H}_c^{d(n-1)}(Z_{1,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \to M_{!,\mathbb{Q}}^{\text{Sym}^d} \to \text{H}_c^{d(n-1)+1}(Z_{0,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \to \text{H}_c^{d(n-1)+1}(Z_{1,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \to 0. \qquad (4.3)$$

Here the maps $\text{H}_c^*(Z_{0,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \to \text{H}_c^*(Z_{0,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$ are the maps called 'Sp' in Theorem A.2.1(i). Since $\mathbb{Q}_\ell[d(n-1)]$ is the same as the intersection complex on $\mathfrak{S}^d$, $\text{H}_c^i(Z_{\alpha,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$ is of weight at most i for all i and $\alpha = 0$ or 1. Taking the weight $d(n-1)+1$ part of the sequence (4.3), we get the short exact sequence (4.2).

(2) Fix an embedding $\mathbb{Q}_\ell \hookrightarrow \mathbb{C}$. We introduce an abelian category $\mathcal{H}$ consisting of triples $(H_\ell, H_{\text{Hod}}, c)$ where H_ℓ is a de Rham representation of $\text{Gal}(\overline{\mathbb{Q}}_\ell / \mathbb{Q}_\ell)$ over $\mathbb{Q}_\ell$, H_{Hod} is a $\mathbb{Q}$-Hodge structure and c is an isomorphism compatible with Hodge filtrations on both sides

$$c : H_{\text{Hod}} \otimes_{\mathbb{Q}} \mathbb{C} \cong (H_\ell \otimes_{\mathbb{Q}_\ell} B_{\text{dR}})^{\text{Gal}(\overline{\mathbb{Q}}_\ell / \mathbb{Q}_\ell)} \otimes_{\mathbb{Q}_\ell} \mathbb{C}.$$

The morphisms in $\mathcal{H}$ are required to respect the isomorphisms c.

Since both specialization maps Sp and Sp_{Hod} are surjective, to show (2), we only need to show that, for $\alpha = 0$ and 1, we can complete the pair $(\text{Gr}_{d(n-1)+1}^W \text{H}_c^{d(n-1)+1}(Z_{\alpha,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell),$ $\text{Gr}_{d(n-1)+1}^W \text{H}_c^{d(n-1)+1}(Z_\alpha(\mathbb{C}), \mathbb{Q}))$ into an object in $\mathcal{H}$, i.e. providing a comparison between the pure weight pieces of the étale and singular cohomologies (after tensoring with period rings) which preserve Hodge filtrations. In fact, we will show a stronger statement. For any variety Z over $\mathbb{Q}$, and any $m, w \in \mathbb{Z}$, we know that the $\text{Gal}(\overline{\mathbb{Q}}_\ell / \mathbb{Q}_\ell)$-module $\text{Gr}_w^W \text{H}_c^m(Z_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$ (in fact the whole $\text{H}_c^m(Z_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$) is de Rham by a theorem of Kisin [Kis02, Theorem 3.2]. We will show that there is an isomorphism of $\mathbb{C}$-vector spaces equipped with Hodge filtrations

$$(\text{Gr}_w^W \text{H}_c^m(Z(\mathbb{C}), \mathbb{Q})) \otimes_{\mathbb{Q}} \mathbb{C} \cong (\text{Gr}_w^W \text{H}_c^m(Z_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \otimes_{\mathbb{Q}_\ell} B_{\text{dR}})^{\text{Gal}(\overline{\mathbb{Q}}_\ell / \mathbb{Q}_\ell)} \otimes_{\mathbb{Q}_\ell} \mathbb{C}. \qquad (4.4)$$

In other words, we will complete the pair $(\text{Gr}_w^W \text{H}_c^m(Z_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)), \text{Gr}_w^W \text{H}_c^m(Z(\mathbb{C}), \mathbb{Q})$ into an object $\text{Gr}_w^W h_c^m(Z) \in \mathcal{H}$.

For Z proper smooth over $\mathbb{Q}$, the isomorphism (4.4) follows from Faltings's theorem [Fal02] and classical Hodge theory: both sides of (4.4) can be identified with the algebraic de Rham cohomology of Z (with scalars extended to $\mathbb{C}$) equipped with the Hodge filtration. We now have the object $h^m(Z) \in \mathcal{H}$.

In general, when Z is not necessarily smooth, one can find a proper hypercovering $\{Z_n\}$ of Z with each Z_n a smooth variety over $\mathbb{Q}$, and a simplicial compactification $\{Z_n\} \hookrightarrow \{X_n\}$ with normal crossing divisors $\{D_n = X_n - Z_n\}$, as in [Del74, § 6.2.8]. For each $i \geqslant 0$, let $\widetilde{D}_{n,i}$ be the normalization of closed subscheme of X_n where at least i local components of D_n meet.

Then $\widetilde{D}_{n,i}$ is proper smooth over $\mathbb{Q}$ and $\widetilde{D}_{n,0} = X_n$. According to the H_c^* variant of [Del74, Proposition 8.1.20(2)], $\mathrm{Gr}_w^W \mathrm{H}_c^m(Z(\mathbb{C}), \mathbb{Q})$ is the $(m-w)$th cohomology of the simple complex associated with the double complex $C_{\mathrm{Hod}}^{a,b} = \mathrm{H}^w(D_{a,b}(\mathbb{C}), \mathbb{Q})$ (the differentials are given by pullback maps). The same argument shows that $\mathrm{Gr}_w^W \mathrm{H}_c^m(Z_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$ is the $(m-w)$th cohomology of the simple complex associated with the double complex $C_\ell^{a,b} = \mathrm{H}^w(D_{a,b,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$. Since $D_{a,b}$ is proper smooth over $\mathbb{Q}$, the pair $(C_\ell^{a,b}, C_{\mathrm{Hod}}^{a,b})$ can be completed into the object $h^w(D_{a,b}) \in \mathcal{H}$. This way we get a double complex $C_{\mathcal{H}}^{*,*}$ with $C_{\mathcal{H}}^{a,b} = h^w(D_{a,b})$ in the category $\mathcal{H}$. The $(m-w)$th cohomology of the simple complex associated with $C_{\mathcal{H}}^{*,*}$ is also an object in $\mathcal{H}$, giving the desired object $\mathrm{Gr}_w^W h_c^m(Z) \in \mathcal{H}$. $\qquad\square$

4.1.6 *The case* $n = 2$. When $d = 2k+1$, $k \geqslant 0$, $\mathfrak{G}^d$ classifies maps $\phi : \mathcal{O}e_1 \oplus \mathcal{O}e_2 \to \mathcal{O}(k)e_1 \oplus \mathcal{O}(k+1)e_2$ such that $\phi_0(e_1) \in \mathrm{Span}\{e_2\}$, and $\phi_\infty(e_1) = e_1 + \mathrm{Span}\{e_2\}, \phi_\infty(e_2) = e_2$. Using the local coordinate s at ∞, every point of $\mathfrak{G}^d$ is uniquely represented by a matrix

$$\phi = \begin{pmatrix} 1 + x_1 s + \cdots + x_k s^k & y_1 s + \cdots + y_{k+1} s^{k+1} \\ z_0 + z_1 s + \cdots + z_k s^k & 1 + w_1 s + \cdots + w_k s^k \end{pmatrix}$$

subject to the equation

$$(1+\cdots+x_k s^k)(1+\cdots+w_k s^k) - (y_1 s + \cdots + y_{k+1} s^{k+1})(z_0 + \cdots + z_k s^k) = \left(1 + \frac{x_1 + w_1 - y_1 z_0}{2k+1} s\right)^{2k+1}$$

and an open condition

$$x_1 + w_1 - y_1 z_0 \neq 0. \tag{4.5}$$

When $d = 2k$, $k \geqslant 0$, we can similarly write every point $\phi \in \mathfrak{G}^d$ as a matrix

$$\phi = \begin{pmatrix} 1 + x_1 s + \cdots + x_k s^k & y_1 s + \cdots + y_k s^k \\ z_0 + z_1 s + \cdots + z_{k-1} s^{k-1} & 1 + w_1 s + \cdots + w_k s^k \end{pmatrix}$$

subject to the equation

$$(1+\cdots+x_k s^k)(1+\cdots+w_k s^k) - (y_1 s + \cdots + y_k s^k)(z_0 + \cdots + z_{k-1} s^{k-1}) = \left(1 + \frac{x_1 + w_1 - y_1 z_0}{2k} s\right)^{2k}$$

and the same open condition as (4.5).

In both cases, $\mathfrak{G}^d$ is a $d+1$-dimensional affine subvariety of $\mathbb{A}^{2d}$ defined by imposing $d-1$ equations and removing one divisor (4.5). The action of $t \in T_2^{\mathrm{ad}}$ on $\mathfrak{G}^d$ sends $y_i \mapsto t^{-1} y_i$ and $z_i \mapsto t z_i$. The action of $t \in \mathbb{G}_m^{\mathrm{rot}}$ on $\mathfrak{G}^d$ sends $(?)_i \mapsto t^i(?)_i$, where $(?) = x, y, z$ or w. The morphism $f_{n,d}$ is given by

$$\phi \mapsto (z_0, y_1).$$

4.2 Dimension of moments: $p > 2$

From now on, until the end of the paper, we shall restrict our consideration to Kl_2. We introduce the notation

$$M_{2,?,\mathbb{F}_p}^d := M_{?,\mathbb{F}_p}^{\mathrm{Sym}^d} \quad \text{for } ? = !, * \text{ or } !*.$$

Here the right-hand side is understood in the context $\widehat{G} = \mathrm{SL}_2$ and Definition 2.5.1 for the base field $k = \mathbb{F}_p$.

We would like to give explicit formula for the dimension of $M_{2,?}^d$ for $? =!, *$ and $!*$. Let $V = \text{Sym}^d$. According to Theorem 2.4.4, we have $V^{\widehat{G}_p^{\text{geom}}} = 0$ for *all* primes p. Therefore, $M_{2,!}^V$ and $M_{2,*}^V$ are both concentrated in degree zero, and the exact sequence (2.9) simplifies to

$$0 \to V^{\mathcal{I}_0} \oplus V^{\mathcal{I}_\infty} \to M_{2,!,\mathbb{F}_p}^d \to M_{2,*,\mathbb{F}_p}^d \to V_{\mathcal{I}_0}(-1) \oplus V_{\mathcal{I}_\infty}(-1) \to 0. \tag{4.6}$$

We first calculate $V^{\mathcal{I}_0}$ and $V^{\mathcal{I}_\infty}$ as Frob_p-modules.

LEMMA 4.2.1 (Special case of [Kat88, Theorem 7.3.2(3)]). *For all primes p, we have $V^{\mathcal{I}_0} \cong \mathbb{Q}_\ell$ as Frob_p-modules.*

LEMMA 4.2.2. *Let $p > 2$ be a prime. As Frob_p-modules*

$$V^{\mathcal{I}_\infty} \cong \begin{cases} \mathbb{Q}_\ell(\mu_p)(-d/2)^{[d/2p]+1}, & d \equiv 0 \bmod 4, \\ \mathbb{Q}_\ell(\mu_p)(-d/2)^{[d/2p]}, & d \equiv 2 \bmod 4, \\ 0, & d \text{ odd.} \end{cases}$$

Proof. By the discussion in §2.3.4, we may assume that the image of the wild inertia $\mathcal{I}_\infty^w$ under the monodromy representation of Kl_2 is contained in the diagonal torus $\widehat{T}_1 \subset \text{SL}_2$. Then the wild inertia $\mathcal{I}_\infty^w$ surjects onto $\widehat{T}_1[p]$. The Coxeter element acts as $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \in \text{SL}_2$. Let $\phi \in \text{GL}_2 = \widehat{G}^{\text{Tt}}$ be the image of a lifting of Frob_∞. Since ϕ normalizes the image of $\mathcal{I}_\infty^w$, we must have $\phi \in N_{\text{GL}_2}(\widehat{T}_1) = N_{\text{GL}_2}(\widehat{T})$ where $\widehat{T}$ is the diagonal torus in GL_2. Up to multiplying ϕ by the Coxeter element (changing the lifting of the Frobenius) we may assume that $\phi \in \widehat{T}$. From §2.4.1 we know that $\det \phi = p$, therefore $\phi = \left(\begin{smallmatrix} \sqrt{p}\alpha & 0 \\ 0 & \sqrt{p}\alpha^{-1} \end{smallmatrix}\right)$ for some $\alpha \in \overline{\mathbb{Q}}_\ell^\times$. The commutation relation between Frob_∞ and $\mathcal{I}_\infty^t$ forces that

$$\begin{pmatrix} \sqrt{p}^{-1}\alpha^{-1} & 0 \\ 0 & \sqrt{p}^{-1}\alpha \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{p}\alpha & 0 \\ 0 & \sqrt{p}\alpha^{-1} \end{pmatrix} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^p \mod \widehat{T}_1[p].$$

This implies that $\alpha \in \pm\mu_p$.

Under the action of $\widehat{T}_1 \cong \mathbb{G}_m$, we have a weight decomposition

$$V = \bigoplus_{-d \leqslant i \leqslant d, i \equiv d \bmod 2} V(i).$$

We have

$$V^{D_1} = \bigoplus_{-d \leqslant jp \leqslant d, j \equiv d \bmod 2} V(jp).$$

When d is odd, Cox^2 acts as -1 on V^{D_1} and hence $V^{\mathcal{I}_\infty} = 0$.

When d is even, the action of $\phi = \sqrt{p}\text{diag}(\alpha, \alpha^{-1})$ on $V(jp)$ is via $p^{d/2}\alpha^{jp}$. Since $\alpha^{2p} = 1$ and $2p|jp$, we conclude that ϕ acts on V^{D_1} by the scalar $p^{d/2}$.

It remains to calculate the dimension of $V^{\mathcal{I}_\infty} = V^{D_1,\text{Cox}}$ when d is even. The Coxeter element acts on V as an involution, and it permutes the factors $V(j)$ and $V(-j)$ for $j \neq 0$, and acts on $V(0)$ by $(-1)^{d/2}$. Therefore

$$\dim V^{D_1,\text{Cox}} = \begin{cases} \left[\dfrac{d}{2p}\right] + 1, & d \equiv 0 \bmod 4, \\[2mm] \left[\dfrac{d}{2p}\right], & d \equiv 2 \bmod 4, \\[2mm] 0, & d \text{ odd.} \end{cases} \qquad \square$$

Applying Lemmas 2.5.5 and 4.2.2 we get the following corollary.

COROLLARY 4.2.3 (Fu and Wan [FW05]). *For $p > 2$, we have*

$$\dim M^d_{2,!,\mathbb{F}_p} = \dim M^d_{2,*,\mathbb{F}_p} = \begin{cases} \dfrac{d}{2} - \left[\dfrac{d}{2p}\right], & d \text{ even}, \\ \dfrac{d+1}{2} - \left[\dfrac{d}{2p} + \dfrac{1}{2}\right], & d \text{ odd}; \end{cases}$$

$$\dim M^d_{2,!*,\mathbb{F}_p} = \begin{cases} \dfrac{d}{2} - 2\left[\dfrac{d}{2p}\right] - 2, & d \equiv 0 \bmod 4, \\ \dfrac{d}{2} - 2\left[\dfrac{d}{2p}\right] - 1, & d \equiv 2 \bmod 4, \\ \dfrac{d-1}{2} - \left[\dfrac{d}{2p} + \dfrac{1}{2}\right], & d \text{ odd}. \end{cases}$$

We shall also use the following formula for the determinant of the Frob-action on $M^d_{2,!*,\mathbb{F}_p}$, proved by Fu–Wan [FW10, Theorem 0.1].

THEOREM 4.2.4 (Fu–Wan[FW10]). *We have*

$$p^{-(d+1)/2 \dim M^d_{2,!*,\mathbb{F}_p}} \det(\mathrm{Frob}, M^d_{2,!*,\mathbb{F}_p})$$

$$= \begin{cases} 1, & d \text{ even}, \\ \left(\dfrac{-2}{p}\right)^{[(d/2p)+1/2]} \displaystyle\prod_{0 \leqslant j \leqslant (d-1)/2, p \nmid 2j+1} \left(\dfrac{(-1)^j(2j+1)}{p}\right), & d \text{ odd}. \end{cases}$$

4.3 Dimension of moments: $p = 2$

4.3.1 *Inertia image at ∞ when $p = 2$.* Let ℓ be an odd prime, and $\rho : \mathcal{I}_\infty \to \mathrm{SL}_2(\mathbb{Q}_\ell)$ be the restriction of the monodromy representation of Kl_2 over $\mathbb{G}_{m,\mathbb{F}_2}$ to the inertia at ∞. Let $D_0 = \rho(\mathcal{I}_\infty) < \mathrm{SL}_2(\mathbb{Q}_\ell)$. We first determine the image $\overline{D}_0$ of $\rho(\mathcal{I}_\infty)$ in $\mathrm{PGL}_2(\mathbb{Q}_\ell)$. Note that $\overline{D}_0$ is a finite subgroup of PGL_2 which has a normal Sylow 2-subgroup. According to the classification of Platonic groups, we have the following possibilities:

(1) $\overline{D}_0$ is cyclic;
(2) $\overline{D}_0$ is dihedral of order 2^k, $k > 1$;
(3) $\overline{D}_0 \cong A_4$.

Suppose that $\overline{D}_0$ is cyclic, so is D_0. Therefore, ρ decomposes into a sum of two characters which are inverse to each other. In this case $\mathrm{Swan}_\infty(\mathrm{Kl}_2)$ must be an even number and cannot be one. Thus, we can eliminated the first possibility.

Suppose $\overline{D}_0$ is dihedral of order 2^k for some $k > 1$, then up to conjugacy $D_0 = \mu_{2^k} \cdot \langle w_0 \rangle$ where μ_{2^k} is embedded into the diagonal torus of SL_2 and $w_0 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Then $\mathrm{Sym}^2(\rho)$ can be decomposed as $V_2 \oplus V_1$ with $\dim V_2 = 2$, $\mathrm{Swan}_\infty(V_2) > 0$, $\dim V_1 = 1$ on which D_0 acts nontrivially through its quotient $D_0/\mu_{2^k} \cong \langle w_0 \rangle$. This implies that $\mathrm{Swan}_\infty(\mathrm{Sym}^2(\mathrm{Kl}_2)) \geqslant 2$. However, Katz proved [Kat88, Proposition 10.4.1] that $\mathrm{Swan}_\infty(\mathrm{Kl}_2 \otimes \mathrm{Kl}_2) = 1$. This is a contradiction. Therefore, we can eliminate the second possibility as well.

In conclusion, we must have $\overline{D}_0 \cong A_4$. This also implies that $D_0 \cong \widetilde{A}_4$, the preimage of $A_4 < \mathrm{PGL}_2(\mathbb{Q}_\ell)$ in $\mathrm{SL}_2(\mathbb{Q}_\ell)$. Let $K_4 < A_4$ be the Klein group of order 4. The preimage

of K_4 in $\widetilde{A}_4$ is the quarternion group $Q_8 = \{\pm 1, \pm I, \pm J, \pm K\}$. The lower numbering filtration of $D_0 = \rho(\mathcal{I}_\infty) \cong \widetilde{A}_4$ is a decreasing filtration

$$D_0 \rhd D_1 \rhd D_2 \rhd \cdots \rhd D_m = \{1\}.$$

Here $D_1 = \rho(\mathcal{I}_\infty^w) = Q_8$, each quotient D_i/D_{i+1} is an elementary abelian 2-group, and m is the smallest integer such that D_m is trivial. Then the Swan conductor of ρ (or of Kl_2 at ∞) is

$$\mathrm{Swan}_\infty(\mathrm{Kl}_2) = \sum_{j \geqslant 1} \frac{2 - \dim \rho^{D_j}}{[D_0 : D_j]} = \frac{2}{3}\left(1 + \sum_{j=2}^{m-1} \frac{1}{[D_1 : D_j]}\right).$$

Here we use the fact that $\rho^{D_{m-1}} = 0$ because ρ is irreducible. Since $\mathrm{Swan}_\infty(\mathrm{Kl}_2) = 1$, $D_1 = Q_8$, we have $\sum_{j=2}^{m-1} \#D_j = 4$. Note each $\#D_j = 2$ or 4. There are only two cases:

(i) $\#D_2 = \#D_3 = 2, D_4 = \{1\}$;

(ii) $\#D_2 = 4, D_3 = \{1\}$.

The second case is impossible because the order-4 subgroups of Q_8 are not elementary abelian 2-groups. We are left with the first possibility, and the lower numbering filtration looks like

$$D_0 = \widetilde{A}_4 \rhd D_1 = Q_8 \rhd D_2 = D_3 = \{\pm 1\} \rhd D_4 = \{1\}.$$

4.3.2 *Image of Frobenius.* Finally we determine the image of Frobenius at ∞. Let $\phi \in \mathrm{GL}_2$ be the image of a lifting of the geometric Frobenius Frob_∞. By §2.4.1 we may write $\phi = \sqrt{2}\phi_1$ where $\phi_1 \in \mathrm{SL}_2$. Since ϕ_1 normalizes $\widetilde{A}_4$, it lies in $N_{\mathrm{SL}_2}(\widetilde{A}_4) = \widetilde{S}_4$, the preimage of $S_4 = N_{\mathrm{PGL}_2}(A_4)$ in SL_2. The tame quotient of the local Galois group then maps to $\widetilde{S}_4/Q_8 \cong S_3$. The image of a generator of the tame inertia is a cyclic permutation $C \in S_3$. The relation $\phi^{-1}C\phi = C^2$ forces the image of ϕ in S_3 to be a transposition. In particular, $\phi_1 \in \widetilde{S}_4 - \widetilde{A}_4$.

With the preparations above, we may now perform calculations.

LEMMA 4.3.3. *When $p = 2$, the Swan conductor of the action of $\mathcal{I}_\infty$ on $V = \mathrm{Sym}^d$ is given by*

$$\mathrm{Swan}_\infty(\mathrm{Sym}^d) = \begin{cases} \dfrac{d+1}{2}, & d \text{ odd}, \\[2mm] \left[\dfrac{d+2}{4}\right], & d \text{ even}. \end{cases} \tag{4.7}$$

Proof. Use the standard basis $\{x^r y^{d-r}\}_{r=0,\dots,d}$ for Sym^d. We extend scalars from $\mathbb{Q}_\ell$ to $\mathbb{Q}_\ell(i)$ where $i = \sqrt{-1}$. Recall the quarternion group $Q_8 = \{\pm 1, \pm I, \pm J, \pm K\}$ acts on the standard two-dimensional representation $\mathrm{Span}\{x, y\}$ as

$$I : (x, y) \mapsto (ix, -iy); \quad J : (x, y) \mapsto (-y, x).$$

When d is odd, the action of $D_3 = \{\pm 1\}$ on Sym^d is by -1. Therefore, D_1, D_2, D_3 do not have invariants on Sym^d. We get

$$\mathrm{Swan}_\infty(\mathrm{Sym}^d) = \sum_{j=1}^{3} \frac{d+1}{[D_0 : D_j]} = \frac{d+1}{2}. \tag{4.8}$$

When d is even, the action of $\mathcal{I}_\infty$ on Sym^d factors through $A_4 < \mathrm{PGL}_2(\mathbb{Q}_\ell)$. Therefore, D_2 acts trivially on Sym^d and D_1 acts via its quotient K_4. In terms of a standard basis $x^i y^{d-i}$ for Sym^d, $I : x^i y^{d-i} \mapsto (-1)^{d/2+i} x^i y^{d-i}$ and $J : x^i y^{d-i} \mapsto (-1)^i x^{d-i} y^i$. Therefore, $\dim(\mathrm{Sym}^d)^{K_4} = d/4 + 1$

if $4|d$ and $\dim(\mathrm{Sym}^d)^{K_4} = [d/4]$ if $d \equiv 2 \bmod 4$. The Swan conductor is

$$\mathrm{Swan}_\infty(\mathrm{Sym}^d) = \frac{1}{3}(d+1 - \dim(\mathrm{Sym}^d)^{K_4}) = \left\lceil \frac{d+2}{4} \right\rceil. \tag{4.9}$$

$\square$

LEMMA 4.3.4. *When $p = 2$ and $V = \mathrm{Sym}^d$, then we have an isomorphism of Frob-modules*

$$V^{\mathcal{I}_\infty} \cong \begin{cases} 0, & d \text{ odd,} \\ \mathbb{Q}_\ell(d/2)^{[d/24]+1} \oplus \mathrm{sgn}(d/2)^{[d/24]}, & d \equiv 0 \bmod 8, \\ \mathbb{Q}_\ell(d/2)^{[d/24]} \oplus \mathrm{sgn}(d/2)^{[d/24]+1}, & d \equiv 6 \bmod 8, \\ \mathbb{Q}_\ell(d/2)^{[d/24]} \oplus \mathrm{sgn}(d/2)^{[d/24]}, & d \equiv 2, 4, 10 \bmod 24, \\ \mathbb{Q}_\ell(d/2)^{[d/24]+1} \oplus \mathrm{sgn}(d/2)^{[d/24]+1}, & d \equiv 12, 18, 20 \bmod 24. \end{cases} \tag{4.10}$$

Here sgn *stands for the one-dimensional Frob-module on which Frob acts as -1.*

Proof. We have $V^{\mathcal{I}_\infty} = (\mathrm{Sym}^d)^{\widetilde{A}_4}$, and we would like to calculate $\dim(\mathrm{Sym}^d)^{\widetilde{A}_4}$ and $\mathrm{Tr}(\phi, (\mathrm{Sym}^d)^{\widetilde{A}_4})$ (where ϕ is the image of a lift of Frob_∞ as we discussed in §4.3.2).

We first make a general remark about the calculation. If G is a finite group with a normal subgroup $N \lhd G$, let V be a representation of G, and let $\bar{a} \in G/N$; then $\bar{a}$ acts on V^N and its trace is given by

$$\mathrm{Tr}(\bar{a}, V^N) = \frac{1}{\#N} \sum_{x \in \bar{a}N} \mathrm{Tr}(x, V).$$

Applying this remark to $N = \widetilde{A}_4 \lhd G = \widetilde{S}_4$, taking $\bar{a} = 1$ we get

$$\sum_{d \geqslant 0} \dim(\mathrm{Sym}^d)^{\widetilde{A}_4} t^d = \frac{1}{24} \sum_{g \in \widetilde{A}_4} \frac{1}{1 - \mathrm{Tr}(g)t + t^2}.$$

Here we write $\mathrm{Tr}(g)$ for the trace of $g \in \widetilde{A}_4$ under the standard two-dimensional representation. Similarly, taking $\bar{a} = \phi_1$ (note $\phi = \sqrt{2}\phi_1$ where $\phi_1 \in \widetilde{S}_4$), we get

$$\sum_{d \geqslant 0} \mathrm{Tr}(\phi_1, (\mathrm{Sym}^d)^{\widetilde{A}_4}) t^d = \frac{1}{24} \sum_{g \in \phi\widetilde{A}_4} \frac{1}{1 - \mathrm{Tr}(g)t + t^2}.$$

We may identify $\widetilde{S}_4$ with the quarternions

$$\pm 1, \pm I, \pm J, \pm K, \frac{1}{2}(\pm 1 \pm I \pm J \pm K), \tag{4.11}$$

$$\frac{1}{\sqrt{2}}(\pm 1 \pm I), \frac{1}{\sqrt{2}}(\pm 1 \pm J), \frac{1}{\sqrt{2}}(\pm 1 \pm K), \frac{1}{\sqrt{2}}(\pm I \pm J), \frac{1}{\sqrt{2}}(\pm J \pm K), \frac{1}{\sqrt{2}}(\pm K \pm I). \tag{4.12}$$

The line (4.11) consists of elements in $\widetilde{A}_4$. Since $\mathrm{Tr}(\pm 1) = \pm 2$, $\mathrm{Tr}(\pm I) = \mathrm{Tr}(\pm J) = \mathrm{Tr}(\pm K) = 0$, we see that one element in $\widetilde{A}_4$ has trace 2, one has trace -2, 8 have trace 1, 8 have trace -1 and the remaining 6 have trace 0. Therefore,

$$\sum_{d \geqslant 0} \dim(\mathrm{Sym}^d)^{\widetilde{A}_4} t^d = \frac{1}{24}\left(\frac{1}{(1-t)^2} + \frac{1}{(1+t)^2} + \frac{6}{1+t^2} + \frac{8}{1-t+t^2} + \frac{8}{1+t+t^2} \right)$$

$$= \frac{1}{12}\sum_{j \geqslant 0}(2j+1)t^{2j} + \frac{1}{4}\sum_{j \geqslant 0}(-1)^j t^{2j} + \frac{2}{3}\sum_{j \geqslant 0}(t^{6j} - t^{6j+4}).$$

Computing coefficients, we get

$$
\dim V^{\mathcal{I}_\infty} = \begin{cases} 0, & d \text{ odd}, \\ \left[\dfrac{d}{12}\right] + 1, & d \equiv 0, 6, 8 \bmod 12, \\ \left[\dfrac{d}{12}\right], & d \equiv 2, 4, 10 \bmod 12. \end{cases} \tag{4.13}
$$

Since $\phi_1 \notin \widetilde{A}_4$, the coset $\phi_1 \widetilde{A}_4$ consists of the second line (4.12) of the above list. We see that 6 of these elements have trace $\sqrt{2}$, 6 of them have trace $-\sqrt{2}$ and the rest have trace zero. Therefore,

$$
\sum_{d \geqslant 0} \mathrm{Tr}(\phi_1, \dim(\mathrm{Sym}^d)^{\widetilde{A}_4}) t^d = \frac{1}{24}\left(\frac{6}{1 - \sqrt{2}t + t^2} + \frac{6}{1 + \sqrt{2}t + t^2} + \frac{12}{1 + t^2}\right)
$$
$$
= \frac{1}{2}\left(\sum_{j \geqslant 0}(-1)^j t^{4j} + \sum_{j \geqslant 0}(-1)^j t^{4j+2} + \sum_{j \geqslant 0}(-1)^j t^{2j}\right).
$$

Therefore,

$$
\mathrm{Tr}(\phi_1, \dim(\mathrm{Sym}^d)^{\widetilde{A}_4}) = \begin{cases} 1, & d \equiv 0 \bmod 8, \\ -1, & d \equiv 6 \bmod 8, \\ 0, & \text{otherwise}. \end{cases}
$$

Since $\phi = \sqrt{2}\phi_1$, we get

$$
\mathrm{Tr}(\mathrm{Frob}_\infty, V^{\mathcal{I}_\infty}) = \begin{cases} 2^{d/2}, & d \equiv 0 \bmod 8, \\ -2^{d/2}, & d \equiv 6 \bmod 8, \\ 0, & \text{otherwise}. \end{cases} \tag{4.14}
$$

Finally, since ϕ_1^2 lies in the image of $\mathcal{I}_\infty$, it acts trivially on $V^{\mathcal{I}_\infty}$. Therefore, the eigenvalues of $\mathrm{Frob} = \sqrt{2}\phi_1$ on $V^{\mathcal{I}_\infty}$ are $\pm 2^{d/2}$. Combining information from (4.13) and from (4.14) we get the multiplicities of the eigenvalues $\pm 2^{d/2}$, and hence the formula (4.10). $\qquad\square$

Combining the previous lemmas we get the following result.

COROLLARY 4.3.5. *When $p = 2$, $d > 1$, we have*

$$
\dim M_{2,!,\mathbb{F}_2}^d = \dim M_{2,*,\mathbb{F}_2}^d = \begin{cases} \dfrac{d+1}{2}, & d \text{ odd}, \\ \left[\dfrac{d+2}{4}\right], & d \text{ even}. \end{cases}
$$

$$
\dim M_{2,!*,\mathbb{F}_2}^d = \begin{cases} \dfrac{d-1}{2}, & d \text{ odd}, \\ 2\left[\dfrac{d+2}{12}\right], & d \equiv 2, 4, 6, 8, 10 \bmod 12, \\ \dfrac{d}{6} - 2, & d \equiv 0 \bmod 12. \end{cases}
$$

In the following table, we summarize the dimension of $M_{2,!*,\mathbb{F}_p}^d$ for the first few primes p.

d	1	2	3	4	5	6	7	8	9	10	11	12	13
good p	0	0	1	0	2	2	3	2	4	4	5	4	6
duality			+		+	-	+	-	+	-	+	-	+
$p=2$			good		good	0	good	0	good	2	good	0	good
$p=3$			0		1	0	2	0	2	2	3	0	4
$p=5$			good		1	good	2	good	3	2	4	2	5
$p=7$			good		good	good	2	good	3	good	4	good	5
$p=11$			good		good	good	good	good	good	good	4	good	5
$p=13$			good		good	good	good	good	good	good	good	good	5

4.4 Proof of Theorem 1.1.6

We recall the notation $M_{2,\ell}^d$ from (4.1), which is equipped with the continuous $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action

$$\rho_{2,\ell}^d : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(M_{2,\ell}^d).$$

From Corollary 4.2.3, we see that the dimension of $M_{2,!*,\mathbb{F}_p}^d$ for large p is

$$\dim M_{2,!*,\mathbb{F}_p}^d = \begin{cases} 2\left[\dfrac{d+2}{4}\right] - 2, & d \text{ even}, p > d/2, \\ \dfrac{d-1}{2}, & d \text{ odd}, p > d \text{ or } p = 2. \end{cases} \tag{4.15}$$

This gives the dimension of the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module $M_{2,\ell}^d$.

4.4.1 *When d is odd.* Let $M_\ell := M_{2,\ell}^d((d+1)/2)$. It is equipped with the orthogonal pairing

$$M_\ell \otimes M_\ell \to \mathbb{Q}_\ell$$

induced from the orthogonal (since $(-1)^{d+1} = 1$) pairing $M_{2,\ell}^d \otimes M_{2,\ell}^d \to \mathbb{Q}_\ell(-d-1)$ from Theorem 1.2.1. We then get an orthogonal Galois representation

$$\rho_\ell = \rho_{2,\ell}^d \chi_{\mathrm{cyc},\ell}^{(d+1)/2} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{O}(M_\ell) \cong \mathrm{O}_{(d-1)/2}(\mathbb{Q}_\ell).$$

We can compute $\det(\rho_\ell)$ using Theorem 4.2.4. For $p > d$, we have

$$\det(\rho_\ell(\mathrm{Frob}_p)) = p^{-((d+1)/2)((d-1)/2)} \det(\mathrm{Frob}_p, M_{2,!*,\mathbb{F}_p}^d) = \left(\frac{(-3)5(-7)\cdots(\pm d)}{p}\right) = \left(\frac{p}{d!!}\right).$$

Therefore, $\det(\rho_\ell)$ is the quadratic Dirichlet character $(\cdot/d!!)$.

Finally, by the Lefschetz trace formula and the sequence (4.6) we have

$$-m_2^d(p) = \mathrm{Tr}(\mathrm{Frob}, M_{2,!,\mathbb{F}_p}^d) = \mathrm{Tr}(\mathrm{Frob}_0, V^{\mathcal{I}_0}) + \mathrm{Tr}(\mathrm{Frob}_\infty, V^{\mathcal{I}_\infty}) + \mathrm{Tr}(\mathrm{Frob}, M_{2,!*,\mathbb{F}_p}^d).$$

For all primes p we have $\mathrm{Tr}(\mathrm{Frob}_0, V^{\mathcal{I}_0}) = 1$ by Lemma 4.2.1, and $V^{\mathcal{I}_\infty} = 0$ by Lemmas 4.2.2 and 4.3.4. Therefore,

$$-m_2^d(p) = 1 + \mathrm{Tr}(\mathrm{Frob}, M_{2,!*,\mathbb{F}_p}^d).$$

For $p > d, p \neq \ell$, Corollary 3.3.5 together with (4.15) implies that ρ_ℓ is unramified, and that $M_{2,!*,\mathbb{F}_p}^d$ is isomorphic to $M_{2,\ell}^d$ as Frob_p modules. Therefore, $\mathrm{Tr}(\mathrm{Frob}, M_{2,!*,\mathbb{F}_p}^d)$ is equal to $p^{(d+1)/2}\mathrm{Tr}(\mathrm{Frob}_p, M_\ell)$. Combining these facts we get the formulae (1.2).

4.4.2 *When d is even.* Let $M_\ell := M_{2,\ell}^d$. It is equipped with the symplectic (since $(-1)^{d+1} = -1$) pairing

$$M_\ell \otimes M_\ell \to \mathbb{Q}_\ell(-d-1)$$

from Theorem 1.2.1. We then get a symplectic Galois representation

$$\rho_\ell = \rho_{2,\ell}^d : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GSp}(M_\ell) \cong \mathrm{GSp}_{2[(d+2)/4]-2}(\mathbb{Q}_\ell)$$

with the similitude character equal to $\chi_{\mathrm{cyc}}^{-d-1}$. Corollary 3.3.5 together with (4.15) implies that ρ_ℓ is unramified at primes $p > d/2$, $p \nmid \ell$.

4.4.3 *Proof of (1.3) and (1.4).* The argument for (1.2 works as well for d even. In particular, we know that ρ_ℓ is unramified for $p > d/2$, $p \neq \ell$, in which case $M_{2,!*,\mathbb{F}_p}^d$ is isomorphic to $M_{2,\ell}^d = M_\ell$ as Frob_p modules.

Let us deal with the general case $p > 2$ ($p \neq \ell$ is always assumed). Let $\underline{M}_! = \underline{M}_!^{\mathrm{Sym}^d}$ be the sheaf on $S = \mathrm{Spec}\,\mathbb{Z}[\ell^{-1}]$ defined in Definition 3.1.3 for the Sym^d-moment of Kl_2. In the proof of Proposition 3.3.3 we have shown that $\underline{M}_!$ fits into an exact sequence

$$0 \to \underline{P}[-1] \to \underline{M}_! \to \underline{N} \to 0$$

where $\underline{P} = \bigoplus_p i_{p,*}\delta_p$ is a punctual sheaf (the sum is finite) and $\underline{N}$ is a middle extension sheaf on S. By definition, the middle extension sheaf $\underline{M}_{!*}$ is a quotient of $\underline{N}$. Let $\underline{K} = \ker(\underline{N} \to \underline{M}_{!*})$, which is also a middle extension sheaf. Taking the stalk at a prime $p \neq \ell$, we get an exact sequence

$$0 \to i_p^*\underline{M}_! \to i_p^*\underline{N} \to \delta_p \to 0.$$

By Lemma 3.1.4, we have $i_p^*\underline{M}_! \cong M_{!,\mathbb{F}_p}^d$. Recall the notation: for a sheaf $\underline{N}$ over S we use $N_\mathbb{Q}$ to denote its generic stalk, which is a continuous $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module. Since $\underline{N}$ is a middle extension sheaf, we have $i_p^*\underline{N} = N_\mathbb{Q}^{\mathcal{I}_p}$, which is an extension of $M_\mathbb{Q}^{\mathcal{I}_p}$ by $K_\mathbb{Q}^{\mathcal{I}_p}$. For large enough p, $\delta_p = 0$ and i_p^*K is equal to $\ker(i_p^*\underline{M}_! \to i_p^*\underline{M}_{!*}) = \ker(M_{2,!,\mathbb{F}_p}^d \to M_{2,!*,\mathbb{F}_p}^d) = V_p^{\mathcal{I}_0} \oplus V_p^{\mathcal{I}_\infty}$. (Here we write $V_p^{\mathcal{I}_\infty}$ etc. to emphasize its dependence on p.) By Lemma 4.2.1 and Lemma 4.2.2, we see that for p large,

$$\mathrm{Tr}(\mathrm{Frob}_p, K_\mathbb{Q}) = 1 + \begin{cases} 0, & d \equiv 2 \bmod 4, \\ p^{d/2}, & d \equiv 0 \bmod 4. \end{cases} \tag{4.16}$$

Therefore, up to semisimplification, we have $K_\mathbb{Q} \cong \mathbb{Q}_\ell \oplus \mathbb{Q}_\ell(-d/2)$ as Galois representations, which is unramified over S. In particular, (4.16) holds for *all* primes $p \neq \ell$.

For every $p \neq \ell$, let L_p be the kernel of the map $N_\mathbb{Q}^{\mathcal{I}_p} \twoheadrightarrow M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p}$. Then L_p is an extension of $\ker(M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p})$ by $K_\mathbb{Q}$. The inclusion $i_p^*\underline{M}_! \otimes \mathbb{Q}_\ell(\mu_p) \cong M_{!,\mathbb{F}_p}^d \hookrightarrow i_p^*\underline{N} \otimes \mathbb{Q}_\ell(\mu_p) = N_\mathbb{Q}^{\mathcal{I}_p} \otimes \mathbb{Q}_\ell(\mu_p)$ fits into an exact sequence of maps

$$\begin{array}{ccccccccc} 0 & \longrightarrow & V_p^{\mathcal{I}_0} \oplus V_p^{\mathcal{I}_\infty} & \longrightarrow & M_{!,\mathbb{F}_p}^d & \longrightarrow & M_{!*,\mathbb{F}_p}^d & \longrightarrow & 0 \\ & & \cup \downarrow & & \cup \downarrow & & \cup \downarrow & & \\ 0 & \longrightarrow & L_p \otimes \mathbb{Q}_\ell(\mu_p) & \longrightarrow & N_\mathbb{Q}^{\mathcal{I}_p} \otimes \mathbb{Q}_\ell(\mu_p) & \longrightarrow & \mathrm{Im}(M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p}) \otimes \mathbb{Q}_\ell(\mu_p) & \longrightarrow & 0 \end{array} \tag{4.17}$$

We claim that the vertical maps are all injective: the argument is already contained in the proof of Theorem 1.2.1(ii) in §3.3.4, namely for both rows the first term is the radical of the natural

pairing on the middle term. When $p > 2$, comparing the formula for $V_p^{\mathcal{I}_\infty}$ in Lemma 4.2.2 and the formula for $K_{\mathbb{Q}}$ in (4.16), we see that $\ker(M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p}) = L_p/K_{\mathbb{Q}}$ should contain at least $\mathbb{Q}_\ell(-d/2)^{[d/2p]}$ as a Frob_p submodule. In other words, M_ℓ contains at least $[d/2p]$ unipotent Jordan blocks of size $\geqslant 2$ under the action of $\mathcal{I}_p$. We then have

$$\dim M_\ell \geqslant 2\left[\frac{d}{2p}\right] + \dim \mathrm{Im}(M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p}) \geqslant 2\left[\frac{d}{2p}\right] + \dim M_{2,!*,\mathbb{F}_p}^d.$$

However, by Lemma 4.2.3, the above inequalities have to be equalities. Equivalently, all vertical maps in (4.17) have to be isomorphisms. This means that for $p > 2$ and $p \neq \ell$, $\mathcal{I}_p$ acts unipotently on M_ℓ with $[d/2p]$ Jordan blocks of size 2 and trivial Jordan blocks elsewhere. Moreover, the action of Frob_p on $M_\ell^{\mathcal{I}_p}$ is $\mathbb{Q}_\ell(-d/2)^{[d/2p]} \oplus M_{2,!*,\mathbb{F}_p}^d$. This proves the decomposition (1.3) with $U_p \cong M_{2,!*,\mathbb{F}_p}^d$ as Frob_p modules.

Finally, by Lefschetz trace formula and using the fact that $M_{2,!,\mathbb{F}_p}^d = N_{\mathbb{Q}}^{\mathcal{I}_p} \otimes \mathbb{Q}_\ell(\mu_p)$, we have

$$-m_2^d(p) = \mathrm{Tr}(\mathrm{Frob}, M_{2,!,\mathbb{F}_p}^d) = \mathrm{Tr}(\mathrm{Frob}_p, N_{\mathbb{Q}}^{\mathcal{I}_p}) = \mathrm{Tr}(\mathrm{Frob}_p, K_{\mathbb{Q}}) + \mathrm{Tr}(\mathrm{Frob}_p, M_\ell^{\mathcal{I}_p}).$$

Combined with (4.16), we get (1.4).

In the following subsections, we shall study the cases $d = 5, 6, 7$ and 8 in more details.

4.5 Sym^5 of Kl_2

For $d = 5$, Theorem 1.1.6 gives a Frobenius-compatible system of continuous Galois representations

$$\rho_\ell = \rho_{2,\ell}^5 \chi_{\mathrm{cyc}}^3 : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{O}_2(\mathbb{Q}_\ell)$$

with determinant $(\cdot/15)$ and ramified at $3, 5$ and ℓ.

At $p = 3$ or 5, according to Theorem 1.2.1(ii), we have $\dim \mathrm{Im}(M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p}) \geqslant \dim M_{2,!*,\mathbb{F}_p}^5 = 1$. Since ρ_ℓ does ramify at p, the only possibility is that $\mathcal{I}_p$ acts through its tame quotient and a generator of $\mathcal{I}_p^t$ maps to a matrix conjugate to $\mathrm{diag}(1, -1)$.

The Hodge–Tate weights of $\rho_{2,\ell}^5$ lie in $\{-1, -2, \ldots, -5\}$ by Theorem 1.2.1(i), hence the Hodge–Tate weights of ρ_ℓ are $(k, -k)$ where $k \in \{0, 1, 2\}$ (we see from Proposition 4.1.5 that k is independent of ℓ).

Let $K = \mathbb{Q}(\sqrt{-15})$. The restriction $\rho_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}$ is an abelian representation into $\mathrm{SO}_2(\mathbb{Q}_\ell)$ which is ramified only at ℓ. If $k = 0$, then $\rho_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}$ form a Frobenius-compatible family of Galois representations with finite image. We may choose ℓ such that $\rho_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}$ is everywhere unramified, which then factors through the class group $\mathrm{Cl}(\mathcal{O}_K) = \mathbb{Z}/2\mathbb{Z}$. But this possibility can be excluded by numerical calculation of $m_2^5(p)$ for the first few p.

Therefore, $k = 1$ or 2, and the Frobenius-compatible family $\rho_\ell \chi_{\mathrm{cyc}}^{-k} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Q}_\ell)$ then comes from a CM form of weight $2k + 1$, level $\Gamma_0(15)$ and nebentypus $(\cdot/15)$ (using the converse theorem since the L-function of ρ_ℓ is the same as the L-function of the Galois character $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}$). The value of k can again be checked by numerics.

4.6 Sym^6 and Sym^8 of Kl_2

When $d = 6$ or $d = 8$, Theorem 1.1.6(ii) gives a Frobenius-compatible system of continuous Galois representations

$$\rho_\ell = \rho_{2,\ell}^d : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Q}_\ell),$$

with determinant $\det(\rho_\ell) = \chi_{\mathrm{cyc}}^{-d-1}$, which are at most ramified at $2, 3$ and ℓ. Moreover, at $p = 3$, the action of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ is via $J_2(-d/2)$. By Theorem 1.2.1(i), the Hodge–Tate weights of $\rho_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)}$ are $(-d/2 - 1 + k, -d/2 - k)$ for $1 \leqslant k \leqslant d/2$.

The next theorem reduces the proof of Conjecture 1.1.4(ii) to a finite calculation.

THEOREM 4.6.1. *There exist integers* $1 \leqslant k \leqslant 4$, $0 \leqslant e \leqslant 8$ *and a holomorphic cuspdial Hecke eigenform* f *of weight* $2k$, *level* $\Gamma_0(2^e \cdot 3)$ *and rational Fourier coefficients such that for all primes* $p \geqslant 3$, *we have*

$$-m_2^8(p) - 1 - p^4 = p^{5-k} a_f(p) \tag{4.18}$$

where $a_f(p)$ *is the* pth *normalized Fourier coefficient of* f.

Proof. Set $d = 8$, then $1 \leqslant k \leqslant 4$ for all ℓ. Let $\rho'_\ell = \rho_\ell \chi_{\mathrm{cyc},\ell}^{5-k}$. Then the Hodge–Tate weights of ρ'_ℓ are $(0, 1 - 2k)$.

We shall use the Serre conjecture, proved by Khare and Wintenberger [KW09a, KW09b], to deduce the modularity of ρ'_ℓ, following an argument originally due to Serre in [Ser87, §4.8]. One can find a similar argument from Kisin's survey article [Kis07, Theorem 1.4.3]. Let $\overline{\rho}'_\ell$ be the reduction mod ℓ of ρ'_ℓ, well-defined up to semisimplification. Let L be the set of primes ℓ such that:

(a) $\ell > 7$;

(b) the variety X in Theorem 1.2.1(i) has good reduction at ℓ, so that ρ_ℓ is crystalline at ℓ; by Fontaine–Laffaille theory (note that $\ell > \dim X = 7$, see [FL82], see also [BLZ04, Corollary 4.3]), the semisimplification of $\overline{\rho}'_\ell|_{\mathcal{I}_\ell}$ is either of the form $1 \oplus \omega^{1-2k}$ or $\omega'^{1-2k} \oplus \omega''^{1-2k}$; here ω is the mod ℓ cyclotomic character and $\omega', \omega'' : \mathcal{I}_\ell \to \mathbb{F}_{\ell^2}^\times$ are the level two fundamental characters; by the definition of Serre weights [Ser87, §2], the Serre weight of $\overline{\rho}'_\ell$ is equal to $2k$;

(c) $\ell \not\equiv \pm 1 \bmod 8$ (see [Ser87, (4.8.7)]).

This set L is infinite. Strong Serre conjecture then implies that for each $\ell \in L$, $\overline{\rho}'_\ell$ is modular of weight $2k$, level of the form $2^e 3$ for some $0 \leqslant e \leqslant 8$ (see the estimate of Artin conductors in [Ser87, Corollary to Proposition 9] and [Ser87, Last paragraph of p. 216]; this is why we need the third condition when defining L). We have 3 in the level because $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ acts under ρ_ℓ by $J_2(-d/2)$. The finiteness of such modular forms then implies the modularity of the compatible system $\{\rho'_\ell\}$ (for details, we refer to [Ser87, §4.8]). Therefore, there is a Hecke eigenform f of weight $2k$, level $N_f = 2^e 3$ for some $0 \leqslant e \leqslant 8$ (and trivial nebentypus) such that the attached Galois representation $\rho_{f,\ell}$ is isomorphic to $\rho_\ell \chi_{\mathrm{cyc},\ell}^{5-k}$ for every prime $\ell \in L$. Since $\rho_\ell \chi_{\mathrm{cyc},\ell}^{5-k}$ form a Frobenius-compatible family, $\rho_{f,\ell} \cong \rho_\ell \chi_{\mathrm{cyc},\ell}^{5-k}$ for all ℓ. The theorem now follows from Theorem 1.1.6(ii). $\qquad\square$

Remark 4.6.2. The same proof works for $m_2^6(p)$, with the weight of the modular form now $2k$ with $1 \leqslant k \leqslant 3$. We actually get more precise information at $p = 2$ thanks to the fact $(\mathrm{Sym}^6)^{\mathcal{I}_\infty} \cong \mathrm{sgn}(-3)$ proved in Lemma 4.3.4. Using this fact, the same argument in §4.4.3 allows us to show that $\rho_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}_2/\mathbb{Q}_2)} \cong J_2 \otimes \mathrm{sgn}(-3)$, and in particular the modular form f in question should have $\Gamma_0(6)$. Finally (4.18) should change to

$$-m_2^6(p) - 1 = p^{4-k} a_f(p)$$

which holds for *all* primes p.

4.7 Sym^7 of Kl$_2$

Theorem 1.1.6(i) gives a Frobenius-compatible system of Galois representations

$$\rho_\ell = \rho_{2,\ell}^7 \chi_{\mathrm{cyc},\ell}^4 : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{O}_3(\mathbb{Q}_\ell)$$

with determinant $(\cdot/105)$ and ramified only at $3, 5, 7$ and ℓ. By Theorem 1.2.1(i), the Hodge–Tate weights of $\rho_{2,\ell}^7|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)}$ are in $\{-1, \ldots, -7\}$, hence the Hodge–Tate weights for $\rho_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)}$ are $(-k, 0, k)$ for $0 \leqslant k \leqslant 3$ independent of ℓ (see Proposition 4.1.5(2)).

We then define

$$\rho_\ell' = \rho_\ell\left(\frac{\cdot}{105}\right) : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{SO}_3(\mathbb{Q}_\ell) \cong \mathrm{PGL}_2(\mathbb{Q}_\ell).$$

For $p = 3, 5$ and 7, Theorem 1.2.1(ii) implies that $\dim \mathrm{Im}(M_\ell^{\mathcal{I}_p} \to M_{\ell,\mathcal{I}_p}) \geqslant \dim M_{2,!*,\mathbb{F}_p}^7 = 2$. Since ρ_ℓ does ramify at p, this inequality must be an equality, the action of $\mathcal{I}_p$ factors through its tame quotient, and a generator of $\mathcal{I}_p^t$ maps under ρ_ℓ to a matrix conjugate to $\mathrm{diag}(1, 1, -1) \in \mathrm{O}_3(\overline{\mathbb{Q}}_\ell)$. Equivalently, a generator of $\mathcal{I}_p^t$ maps under ρ_ℓ' to a matrix conjugate to $\zeta = \mathrm{diag}(-1, 1) \in \mathrm{PGL}_2(\overline{\mathbb{Q}}_\ell)$. We fix a lifting of Frob_p to $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ and let $\phi_p \in \mathrm{PGL}_2(\overline{\mathbb{Q}}_\ell)$ be its image under ρ_ℓ'. Then the relation $\phi_p^{-1}\zeta\phi_p = \zeta^p = \zeta$ forces ϕ_p to lie in $N_{\mathrm{PGL}_2}(\zeta) = T \cup Tw$, where T is the diagonal torus in $\mathrm{PGL}_2(\overline{\mathbb{Q}}_\ell)$ and $w = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.

We say p is of *abelian type* if $\phi_p \in T$; we say p is of *dihedral type* if $\phi_p \in Tw$.

LEMMA 4.7.1. *Here $p = 3$ and 7 are of abelian type, and $p = 5$ is dihedral type.*

Proof. Whether $\phi_p \in T$ or $\phi_p \in Tw$ can be detected by computing the determinant of Frob_p on $M_{2,!*,\mathbb{F}_p}^7$. In fact, under $\rho_{2,\ell}^7$, Frob_p acts on the two-dimensional orthogonal space $M_\ell^{\mathcal{I}_p}$ with similitude p^4. If $\det(\mathrm{Frob}_p, M_{2,!*,\mathbb{F}_p}^7) = p^8$, then $\phi \in T$; if $\det(\mathrm{Frob}_p, M_{2,!*,\mathbb{F}_p}^7) = -p^8$, then $\phi \in Tw$. We then use Theorem 4.2.4 to see that $\det(\mathrm{Frob}_3, M_{2,!*,\mathbb{F}_3}^7) = \det(\mathrm{Frob}_7, M_{2,!*,\mathbb{F}_7}^7) = p^8$ while $\det(\mathrm{Frob}_5, M_{2,!*,\mathbb{F}_5}^7) = -p^8$. $\qquad\square$

LEMMA 4.7.2. *For any prime $\ell > 7$, ρ_ℓ' is absolutely irreducible as a three-dimensional orthogonal Galois representation.*

Proof. If not, there should be a continuous character $\chi : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \overline{\mathbb{Q}}_\ell^\times$ such that ρ_ℓ' is isomorphic to $\chi \oplus \chi^{-1} \oplus 1$ up to semisimplification. By the Hodge–Tate property, we have $\chi = \chi_{\mathrm{cyc}}^k \chi_0$ where χ_0 has finite order. Therefore, for each prime $p \neq 3, 5, 7$ or ℓ, we have

$$\left(\frac{p}{105}\right)\frac{-m_2^7(p) - 1}{p^4} = \mathrm{Tr}(\rho_\ell'(\mathrm{Frob}_p)) = p^k \chi_0(p) + p^{-k}\chi_0(p)^{-1} + 1. \tag{4.19}$$

In particular, we have

$$|m_2^7(p) + 1| \geqslant p^4(p^k - p^{-k} - 1) \geqslant p^5 - p^4 - p^3.$$

However, from [Eva10b, Table 2.1], we see that neither $p = 11$ nor $p = 13$ (at least one of them is not equal to ℓ) satisfies the above inequality. $\qquad\square$

LEMMA 4.7.3. *For $\ell > 7$, the Galois representation ρ_ℓ' does not have finite image.*

Proof. We need to eliminate the possibility that $\mathrm{Im}(\rho'_\ell)$ is cyclic, dihedral, A_4, S_4 or A_5. Since ρ'_ℓ fits into a Frobenius-compatible system, it is unramified at ℓ.

By Lemma 4.7.2, the image of ρ'_ℓ cannot be cyclic.

Suppose $\mathrm{Im}(\rho'_\ell) = D_{2n}$, then we have a surjection $\epsilon : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \twoheadrightarrow D_{2n} \twoheadrightarrow \{\pm 1\}$ which is unramified away from $3, 5, 7$. Therefore, for some $N|105$ we have $\epsilon = (\cdot/N)$. For those prime p such that $(p/N) = -1$, Frob_p lies in the non-neutral component of $\mathrm{O}_2 \subset \mathrm{SO}_3 \cong \mathrm{PGL}_2$, hence has trace -1. Therefore, whenever $(p/N) = -1$, we must have

$$\left(\frac{p}{105}\right)\frac{-m_2^7(p) - 1}{p^4} = -1.$$

Checking all possible N and numerics of $m_2^7(p)$ from [Eva10b, Table 2.1] we can eliminate this possibility.

Suppose $\mathrm{Im}(\rho'_\ell) = A_4, S_4$ or A_5, then $\mathrm{Tr}(\rho'_\ell(\mathrm{Frob}_p))$ can only take at most five different values. However, numerics from [Eva10b, Table 2.1] show that $(p/105)(-m_2^7(p) - 1)/p^4$ takes more than five values for varying p. $\square$

COROLLARY 4.7.4. *The Hodge–Tate weights of ρ'_ℓ are not all zero.*

Proof. Since ρ'_ℓ is a subquotient of the cohomology of some fixed smooth projective variety X over $\mathbb{Q}$ (independent of ℓ) by Theorem 1.2.1(i), we may take ℓ such that ρ'_ℓ is crystalline. If the Hodge–Tate weights of ρ'_ℓ are all zero, it should be unramified at ℓ. Then we know that ρ'_ℓ is only ramified at $3, 5$ and 7 with inertial image of order 2. Hence, for each m the image of the reduction $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_2(\mathbb{Z}/\ell^m)$ is a finite Galois group $\mathrm{Gal}(K_m/\mathbb{Q})$ for K_m of bounded discriminant independent of m. The inductive system $\varinjlim_m K_m$ should then stabilize and ρ'_ℓ must have finite image. But this is impossible by Lemma 4.7.3. $\square$

COROLLARY 4.7.5. *Let c be a complex conjugation in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, then $\rho'_\ell(c) \in \mathrm{PGL}_2(\overline{\mathbb{Q}}_\ell)$ is conjugate to $\mathrm{diag}(1, -1)$.*

Proof. We only need to show that $\rho'_\ell(c) = \rho_\ell(c) \neq 1$.

Consider the $\mathbb{Q}$-Hodge structure $M_{2,\mathrm{Hod}}^7$ defined in Proposition 4.1.5(2). By the comparison of singular and étale cohomology, there is an isomorphism $M_{2,\mathrm{Hod}}^7 \otimes \mathbb{Q}_\ell \xrightarrow{\sim} M_{2,\ell}^7$ which intertwines the action of the complex conjugation on $M_{2,\mathrm{Hod}}^7$ (coming from the complex conjugation acting on the $\mathbb{C}$-points of varieties defined over $\mathbb{R}$), and the action of c on $M_{2,\ell}^7$. By Proposition 4.1.5(2), the Hodge–Tate weights of $M_{2,\mathrm{Hod}}^7$ are also $(4 - k, 4, 4 + k)$, i.e. the Hodge decomposition for $M_{2,\mathrm{Hod}}^7 \otimes_\mathbb{Q} \mathbb{C}$ reads $H^{4+k,4-k} \oplus H^{4,4} \oplus H^{4-k,4+k}$, with each summand one-dimensional. We know from Corollary 4.7.4 that $k > 0$.

Since the Galois group $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ acts on $M_{2,\mathrm{Hod}}^7 \otimes_\mathbb{Q} \mathbb{C}$ (obtained by extending scalars from its action on $M_{2,\mathrm{Hod}}^7$) by permuting $H^{4+k,4-k}$ and $H^{4-k,4+k}$, we conclude that $\rho_\ell(c) \neq 1$. $\square$

LEMMA 4.7.6. *For sufficiently large prime ℓ, there exists a continuous Galois representation*

$$\widetilde{\rho}_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell).$$

lifting the Galois representation $\rho'_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_2(\mathbb{Q}_\ell) \hookrightarrow \mathrm{PGL}_2(\overline{\mathbb{Q}}_\ell)$. Moreover, the lifting $\widetilde{\rho}_\ell$ may be chosen to satisfy the following properties.

(i) *It is unramified away from* $3, 5, 7$ *and* ℓ.

(ii) *When* $p = 3$ *or* 7, $\widetilde{\rho}_\ell|_{\mathcal{I}_p}$ *is tame, and maps a generator of* $\mathcal{I}_p^t$ *to a matrix conjugate to* $\mathrm{diag}(1, -1)$ *in* $\mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$.

(iii) *When* $p = 5$, $\widetilde{\rho}_\ell|_{\mathcal{I}_p}$ *is tame, and maps a generator of* $\mathcal{I}_p^t$ *to a matrix conjugate to* $\mathrm{diag}(x, -x)$ *in* $\mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$, *where* x *is a primitive* 8th *root of unity*;

(iv) *It is crystalline at* ℓ *with Hodge–Tate weights* $(0, -k)$.

Proof. By a result of Patrikis [Pat12, Corollary 13.0.15], there exists a lifting $\widetilde{\rho}^\dagger$ which is 'geometric' in the sense of Fontaine–Mazur, namely it is almost everywhere unramified and de Rham at ℓ. Below we will define continuous characters $c_p : \mathbb{Z}_p^\times \to \overline{\mathbb{Q}}_\ell^\times$, one for each prime p, and trivial for almost all p. Ultimately the product of $\{c_p\}$ will give a character c of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\widetilde{\rho}^\dagger \otimes c^{-1}$ satisfies all of the requirements in the lemma.

For $p \neq 3, 5, 7$ or ℓ, the restriction $\widetilde{\rho}^\dagger|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is abelian because $\widetilde{\rho}^\dagger|_{\mathcal{I}_p}$ maps to scalar matrices in $\mathrm{GL}_2(E_\lambda)$. Therefore, $\widetilde{\rho}^\dagger|_{\mathcal{I}_p}$ corresponds to a character $c_p : \mathbb{Z}_p^\times \to \overline{\mathbb{Q}}_\ell^\times$ by local class field theory. Since $\widetilde{\rho}^\dagger$ is ramified only at finitely many places, c_p is trivial for almost all p.

For $p = 3$ or 7, $\rho_\ell'|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is of abelian type (i.e. its image lies in a torus), hence $\widetilde{\rho}^\dagger(\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p))$ is also abelian and can be conjugated to $\mathrm{diag}(c_p, c_p')$. Here $c_p, c_p' : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \overline{\mathbb{Q}}_\ell^\times$ are characters such that c_p/c_p' restricted to $\mathcal{I}_p$ has order two.

For $p = 5$, $\rho_\ell'|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is of dihedral type, the restriction $\widetilde{\rho}^\dagger|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_{p^2})}$ is still abelian, and can be conjugated to $\mathrm{diag}(d_1, d_2)$ where $d_1, d_2 : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_{p^2}) \to \overline{\mathbb{Q}}_\ell^\times$ are characters. One easily sees that $d_2 = d_1^\sigma$ where $\sigma \in \mathrm{Gal}(\mathbb{Q}_{p^2}/\mathbb{Q}_p)$ is the nontrivial element. Also $d_1/d_2 = d_1/d_1^\sigma$ when restricted to $\mathcal{I}_p$ has order exactly two. Using class field theory we may write $d_1|_{\mathcal{I}_p} : \mathbb{Z}_{p^2}^\times \to \overline{\mathbb{Q}}_\ell^\times$. Now d_1/d_1^σ is trivial on $(\mathbb{Z}_{p^2}^\times)^2$, i.e. $d_1|_{(\mathbb{Z}_{p^2}^\times)^2}$ is invariant under $\mathrm{Gal}(\mathbb{Q}_{p^2}/\mathbb{Q}_p)$. Identify $\mathbb{Z}_{p^2}^\times$ with $(1 + p\mathbb{Z}_{p^2}) \times \mathbb{F}_{p^2}^\times$, we conclude that d_1 factors through

$$\mathbb{Z}_{p^2}^\times \cong (1 + p\mathbb{Z}_{p^2}) \times \mathbb{F}_{p^2}^\times \xrightarrow{\mathrm{Nm}, (-)^{(p+1)/2}} (1 + p\mathbb{Z}_p) \times \mu_{2(p-1)}(\overline{\mathbb{F}}_p).$$

We denote the resulting character $(1 + p\mathbb{Z}_p) \times \mu_{2(p-1)}(\overline{\mathbb{F}}_p) \to \overline{\mathbb{Q}}_\ell^\times$ by $\overline{d}$. Since d_1/d_1^σ is nontrivial, d_1 does not factor further through $(1 + p\mathbb{Z}_p) \times \mathbb{F}_p^\times = \mathbb{Z}_p^\times$. Then there exists a character $c_p : \mathbb{Z}_p^\times = (1 + p\mathbb{Z}_p) \times \mathbb{F}_p^\times \to \overline{\mathbb{Q}}_\ell^\times$ which is equal to $\overline{d}$ on $(1 + p\mathbb{Z}_p)$. Then $d_1 c_p^{-1}$ (which is now tame) maps a generator of $\mathcal{I}_p^t$ to a primitive $2(p-1) = 8$th root of unity.

We suppose ℓ is sufficiently large so that ρ_ℓ' is crystalline. At ℓ, the Hodge–Tate weights are $(a, a-k)$ for some $a \in \mathbb{Z}$. Since $\rho_\ell'|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)}$ admits a de Rham lifting, it also admits a crystalline lifting $\phi : \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$, by a result of Conrad [Con, Corollary 6.7]. By making a Tate twist we may assume ϕ has Hodge–Tate weights $(0, -k)$. We then have $\widetilde{\rho}_\ell^\dagger|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)} = \phi \cdot d_\ell$ for some continuous character $d_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)^{ab} \to \overline{\mathbb{Q}}_\ell^\times$. The restriction of d_ℓ to the inertia group then gives a character $c_\ell : \mathbb{Z}_\ell^\times \to \overline{\mathbb{Q}}_\ell^\times$.

Let $c : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})^{ab} \cong \prod_p \mathbb{Z}_p^\times \to \overline{\mathbb{Q}}_\ell^\times$ be the product of the c_p defined above for all primes p. This makes sense because c_p is trivial for almost all p. Also c is continuous since each c_p is. Then $\widetilde{\rho}_\ell := \widetilde{\rho}^\dagger \otimes c^{-1}$ satisfies all of the requirements. $\square$

4.7.7 *Proof of Conjecture 1.1.4(i).* Fix a lifting $\widetilde{\rho}_\ell$ as in Lemma 4.7.6 (ℓ is sufficiently large).

The Artin conductor of $\widetilde{\rho}_\ell$ at $3, 5$ and 7 are $1, 2$ and 1, respectively, according to the description of local behavior of $\widetilde{\rho}_\ell$ in Lemma 4.7.6.

From the description on the behavior of $\widetilde{\rho}_\ell$ at $p = 3, 5$ and 7 again, we see that under $\widetilde{\rho}_\ell$ a generator of $\mathcal{I}_3$ or $\mathcal{I}_7$ gets mapped to $\mathrm{diag}(1, -1)$, with determinant -1. A generator of $\mathcal{I}_5$ gets mapped to $\mathrm{diag}(x, -x)$, where x is a primitive 8th root of unity, with determinant a primitive 4th root of unity. Therefore,

$$\det(\widetilde{\rho}_\ell) \cong \left(\frac{\cdot}{21}\right)\epsilon_5\epsilon_\ell\chi_{\mathrm{cyc},\ell}^{-k}.$$

Here $\epsilon_5 : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \twoheadrightarrow (\mathbb{Z}/5\mathbb{Z})^\times \xrightarrow{\sim} \mu_4(\overline{\mathbb{Q}}_\ell)$ is a quartic Dirichlet character of conductor 5; ϵ_ℓ is a Dirichlet character of conductor a power of ℓ.

Let $\overline{\widetilde{\rho}}_\ell$ be the reduction mod ℓ of $\widetilde{\rho}_\ell$, i.e. a continuous representation $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$. Since $\widetilde{\rho}_\ell$ is crystalline at ℓ with Hodge–Tate weights $(0, -k)$, and ℓ is assumed to be sufficiently large, Fontaine–Laffaille theory implies that the Serre weight of $\overline{\widetilde{\rho}}_\ell$ is $k + 1$ (as in the proof of Theorem 4.6.1). Therefore the ℓ-part of $\det(\overline{\widetilde{\rho}}_\ell)$ should be ω_ℓ^{-k} here ω_ℓ is $\chi_{\mathrm{cyc},\ell}$ mod ℓ. Hence, ϵ_ℓ is trivial modulo ℓ, and therefore admitting a square root $\epsilon_\ell^{1/2}$. Changing $\widetilde{\rho}_\ell$ to $\widetilde{\rho}_\ell\epsilon_\ell^{-1/2}$, our new lifting, which we still denote by $\widetilde{\rho}_\ell$ (and still satisfying the properties of Lemma 4.7.6), satisfies

$$\det(\widetilde{\rho}_\ell) \cong \left(\frac{\cdot}{21}\right)\epsilon_5\chi_{\mathrm{cyc},\ell}^{-k}. \tag{4.20}$$

By Corollary 4.7.5, a complex conjugation c gets mapped to $\mathrm{diag}(1, -1) \in \mathrm{PGL}_2(\overline{\mathbb{Q}}_\ell)$ under ρ_ℓ', therefore $\widetilde{\rho}_\ell$ is odd. Since $(\cdot/21)\epsilon_5$ takes value -1 on c as well, we conclude from (4.20) that k is even. Since $1 \leqslant k \leqslant 3$, we get $k = 2$.

Let us summarize our knowledge about $\overline{\widetilde{\rho}}_\ell$ for large ℓ at this point. It is absolutely irreducible (by Lemma 4.7.2), odd, with Artin conductor $525 = 3 \cdot 5^2 \cdot 7$ (away from ℓ part), Serre weight equal to $k + 1 = 3$ and determinant given by (4.20) modulo ℓ.

We now appeal to Serre's argument for modularity [Ser87, § 4.8] again. Although we do not know whether we can choose the liftings $\{\widetilde{\rho}_\ell\}$ to form a Frobenius-compatible family, we have the following weaker property. The lifting $\widetilde{\rho}_\ell$ satisfying the condition (4.20) is unique up to twisting by a quadratic Dirichlet character. Hence, for $p \neq 3, 5$ or 7, $\mathrm{Tr}(\widetilde{\rho}_\ell(\mathrm{Frob}_p))^2$ is independent of ℓ. This property will play the role of Frobenius-compatibility in the argument below. Therefore, by Serre's conjecture, for sufficiently large ℓ, $\overline{\widetilde{\rho}}_\ell$ comes from the mod ℓ Galois representation of a Hecke eigenform of weight 3, level 525 and nebentypus $\epsilon_f = (\cdot/21)\epsilon_5$. Since there are only finitely many possibilities for such eigenforms, we conclude that there is such a newform f such that $\overline{\rho}_{f,\ell} \cong \overline{\widetilde{\rho}}_\ell$ for infinitely many ℓ. Therefore, for each prime $p \neq 3, 5$ or 7, $\mathrm{Tr}(\widetilde{\rho}_\ell(\mathrm{Frob}_p))^2$ is congruent mod ℓ to the Fourier coefficient $a_f(p)^2$ for infinitely many ℓ. Hence, for each prime $p \neq 3, 5$ or 7, we have

$$\mathrm{Tr}(\widetilde{\rho}_\ell(\mathrm{Frob}_p))^2 = a_f(p)^2.$$

This means that

$$\mathrm{Tr}(\rho_\ell'(\mathrm{Frob}_p)) = \frac{\mathrm{Tr}(\widetilde{\rho}_\ell(\mathrm{Frob}_p))^2}{\det(\widetilde{\rho}_\ell(\mathrm{Frob}_p))} - 1 = a_f(p)^2 p^{-2}\epsilon_f(p)^{-1} - 1. \tag{4.21}$$

By Theorem 1.1.6(i), we see that

$$-m_2^7(p) = 1 + p^4\mathrm{Tr}(\rho_\ell(\mathrm{Frob}_p)) = 1 + p^4\left(\frac{p}{105}\right)\mathrm{Tr}(\rho_\ell'(\mathrm{Frob}_p)).$$

Plugging in (4.21), we get (1.1).

Appendix A. Fourier transform

In this appendix, we collect some facts about homogeneous Fourier transform that we need in the main body of the paper. We follow Laumon's original paper [Lau03].

A.1 Fourier–Deligne transform

Let S be a scheme of finite type over k. Let $\mathbb{A}$ and $\mathbb{A}^\vee$ both be the affine line $\mathbb{A}^1_S$ over S. We recall the Fourier–Deligne transform

$$\mathrm{Four}_\psi : D^b_c(\mathbb{A}, \mathbb{Q}_\ell(\mu_p)) \to D^b_c(\mathbb{A}^\vee, \mathbb{Q}_\ell(\mu_p)).$$

It is defined as follows. Consider the following diagram

where $\mathrm{pr}, \mathrm{pr}^\vee$ are projections and m is the multiplication map. For $\mathcal{F} \in D^b_c(\mathbb{A}, \mathbb{Q}_\ell(\mu_p))$, its Fourier transform is defined as

$$\mathrm{Four}_\psi(\mathcal{F}) = \mathrm{pr}^\vee_!(\mathrm{pr}^*\mathcal{F} \otimes m^*\mathrm{AS}_\psi)[1].$$

A.2 Laumon's homogeneous Fourier transform

We still let $\mathbb{A}$ and $\mathbb{A}^\vee$ be affine lines over S, except now S is allowed to be any scheme over finite type over $\mathbb{Z}[\ell^{-1}]$. The notation $\mathbb{G}_m$ now means the torus $\mathbb{G}_{m,S}$ over S, which acts on $\mathbb{A}$ and $\mathbb{A}^\vee$. We shall consider the equivariant derived categories $D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell)$ and $D^b_{\mathbb{G}_m}(\mathbb{A}^\vee, \mathbb{Q}_\ell)$. Laumon defines a homogeneous Fourier transform

$$\mathrm{HFour} : D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell) \to D^b_{\mathbb{G}_m}(\mathbb{A}^\vee, \mathbb{Q}_\ell).$$

Its definition uses a similar diagram of stacks over S.

Let $\beta : S = [\mathbb{G}_m/\mathbb{G}_m] \hookrightarrow [\mathbb{A}^1_S/\mathbb{G}_m]$ be the open inclusion, and let $\Psi = \beta_*\mathbb{Q}_\ell \in D^b_{\mathbb{G}_m}(\mathbb{A}^1_S, \mathbb{Q}_\ell)$. For $\mathcal{F} \in D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell)$, its homogeneous Fourier transform is defined as

$$\mathrm{HFour}(\mathcal{F}) = \overline{\mathrm{pr}}^\vee_!(\overline{\mathrm{pr}}^*\mathcal{F} \otimes \overline{m}^*\Psi).$$

We summarize the major properties of the homogeneous Fourier transform.

THEOREM A.2.1. *Let S be a scheme of finite type over $\mathbb{Z}[\ell^{-1}]$. Let $\mathcal{F} \in D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell)$.*

(i) *Let $\mathcal{F}_0$ and $\mathcal{F}_1$ be restriction of $\mathcal{F}$ to the sections 0_S and 1_S of the projection $\mathbb{A} \to S$ (recall $\mathbb{A} = \mathbb{A}^1_S$). Then there is a specialization morphism*

$$\mathrm{Sp}(\mathcal{F}) : \mathcal{F}_0 \to \mathcal{F}_1$$

functorial in $\mathcal{F}$, such that the generic stalk $\mathrm{HFour}(\mathcal{F})_\eta$ of the homogeneous Fourier transform $\mathrm{HFour}(\mathcal{F})$ (i.e. the restriction of $\mathrm{HFour}(\mathcal{F})$ to $[\mathbb{G}_m/\mathbb{G}_m] = S$) fits into a distinguished triangle

$$\mathcal{F}_0 \xrightarrow{\mathrm{Sp}(\mathcal{F})} \mathcal{F}_1 \to \mathrm{HFour}(\mathcal{F})_\eta \to \mathcal{F}_0[1]. \tag{A.1}$$

(ii) *Suppose S is over $\mathbb{F}_p$, then we have a canonical isomorphism in $D^b_c(\mathbb{A}^\vee, \mathbb{Q}_\ell(\mu_p))$*

$$\mathrm{Four}_\psi(\mathcal{F} \otimes \mathbb{Q}_\ell(\mu_p)) \cong \mathrm{HFour}(\mathcal{F}) \otimes \mathbb{Q}_\ell(\mu_p).$$

(iii) *There is an isomorphism of functors:*

$$\mathrm{HFour} \circ \mathbb{D} \cong (\mathbb{D} \circ \mathrm{HFour})(-1).$$

(iv) *The functor HFour is t-exact with respect to the perverse t-structure on $D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell)$.*

(v) *Let $*_+$ denote the convolution on $D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell)$ using the additive group structure. Then for $\mathcal{F}_1, \mathcal{F}_2 \in D^b_{\mathbb{G}_m}(\mathbb{A}, \mathbb{Q}_\ell)$, we have a canonical isomorphism*

$$\mathrm{HFour}(\mathcal{F}_1 *_+ \mathcal{F}_2)[1] \cong \mathrm{HFour}(\mathcal{F}_1) \otimes \mathrm{HFour}(\mathcal{F}_2).$$

Proof. (1) Let $h : \mathbb{A} \to S$ be the structure morphism. By definition and proper base change, we have

$$\mathrm{HFour}(\mathcal{F})_\eta = h_!(\beta_* \otimes \mathcal{F}). \tag{A.2}$$

Let $\alpha : \mathbb{B}\mathbb{G}_m \hookrightarrow [\mathbb{A}/\mathbb{G}_m]$ be the closed embedding of the origin. Consider the distinguished triangle in $D^b_{\mathbb{G}_m}(\mathbb{A})$:

$$\beta_!\mathbb{Q}_\ell \to \beta_*\mathbb{Q}_\ell \to \alpha_*\alpha^*\beta_*\mathbb{Q}_\ell \to \beta_!\mathbb{Q}_\ell[1]. \tag{A.3}$$

Also let $g : S \to \mathbb{B}\mathbb{G}_m$ be the quotient morphism, then by [Lau03, Lemme 1.4(ii)] we have an isomorphism in $D^b_{\mathbb{G}_m}(S, \mathbb{Q}_\ell)$

$$\alpha^*\beta_*\mathbb{Q}_\ell \cong g_!\mathbb{Q}_\ell[1].$$

Therefore (A.3), after rotating the triangle, becomes

$$\alpha_*g_!\mathbb{Q}_\ell \to \beta_!\mathbb{Q}_\ell \to \beta_*\mathbb{Q}_\ell \to \alpha_*g_!\mathbb{Q}_\ell[1].$$

Tensoring with $\mathcal{F}$ and applying the functor $h_!$, we get

$$h_!(\alpha_*g_!\mathbb{Q}_\ell \otimes \mathcal{F}) \to h_!(\beta_!\mathbb{Q}_\ell \otimes \mathcal{F}) \to h_!(\beta_* \otimes \mathcal{F}) \to . \tag{A.4}$$

By projection formula

$$h_!(\alpha_*g_!\mathbb{Q}_\ell \otimes \mathcal{F}) = h_!\alpha_!g_!(g^*\alpha^*\mathcal{F}) = \mathcal{F}_0, \tag{A.5}$$
$$h_!(\beta_!\mathbb{Q}_\ell \otimes \mathcal{F}) = h_!\beta_!(\beta^*\mathcal{F}) = \mathcal{F}_1. \tag{A.6}$$

Plugging (A.5), (A.6) and (A.2) into (A.4), we get the desired triangle (A.1), which also includes the definition of the functorial map $\mathrm{Sp}(\mathcal{F}) : \mathcal{F}_0 \to \mathcal{F}_1$ as part of the triangle.

For part (ii), see [Lau03, Théorème 2.2].

For part (iii), see [Lau03, Théorème 4.1].

For part (iv), see [Lau03, Théorème 4.2].

Part (v) is the special case of [Lau03, Lemme 1.7] in the case of the linear map $\mathbb{A}^2 \to \mathbb{A}^1$ given by the addition of two coordinates. $\qquad\square$

Appendix B. Proof of Conjecture 1.1.4(ii): computational

Christelle Vincent

In this appendix we give a brief account of the computational part of the proof of Conjecture 1.1.4(ii). This computation was carried out using the open source software Sage (http://www.sagemath.org), and part of the computation was carried out on the Sage Cloud (http://cloud.sagemath.com). All Sage code written for this project is available online (http://math.stanford.edu/~cvincent/research.html).

Fix a prime p. We first compute the numbers $m_2^8(p)$. In the notation of § 1, for each $a \in \mathbb{F}_p^\times$, we have

$$\mathrm{Kl}_2(p; a) = -(\alpha_a + \beta_a),$$

for α_a and β_a the eigenvalues of Frob_a acting on $(\mathrm{Kl}_2)_a$, and

$$m_2^8(p) = \sum_{a \in \mathbb{F}_p^\times} \sum_{i=0}^{8} \alpha_a^i \beta_a^{8-i}.$$

Sage already has a function to compute twisted Kloosterman sums. Since in our case the Dirichlet character is trivial, we modified the code to remove the dependence on the Dirichlet character. Even though it is possible to return the exact value of the Kloosterman sum, we compute it as an approximate complex number to speed up computations.

Using these values, we may compute, for $n \geqslant 1$, the sums

$$S_n(p) = \sum_{a \in \mathbb{F}_p^\times} (\mathrm{Kl}_2(p; a))^n.$$

Since we only have an approximation for $\mathrm{Kl}_2(p; a)$, to recover the exact value of $S_n(p)$ we use the congruence

$$S_n(p) \equiv (-1)^{n-1}(n-1)p \pmod{p^2}$$

proved in [CE07].

Using $S_n(p)$ and the fact that $\alpha_a \beta_a = p$, we have

$$m_2^8(p) = S_8(p) - 7pS_6(p) + 15p^2 S_4(p) - 10p^3 S_2(p) + p^5 - p^4 + 10p^3 - 15p^2 + 7p - 1.$$

From there we can obtain the Fourier coefficients of the form f which we are looking for:

$$a_f(p) = -(m_2^8(p) + 1 + p^4)/p^{5-k},$$

where k is half of the weight of the form f.

We begin by noticing that $m_2^8(5) = 1024$. If $2k = 2$ or $2k = 4$, we have that the predicted Hecke eigenvalue is not an integer. Since Hecke eigenvalues are integral we do not need to check these cases. We are left with the possibility of a form f of weight 6 or of weight 8.

For each

$$N \in \{3, 6, 12, 24, 48, 96, 192, 384\},$$

and $2k = 68$ and each

$$N \in \{3, 6, 12, 24, 48, 96, 192\},$$

and $2k = 8$, we compute the space of cuspidal new modular symbols of weight k and level $\Gamma_0(N)$. We then decompose it into invariant subspaces under the action of the Hecke operators of index

coprime to the level. Each of these spaces is simple, and we compute the eigenvalue of the Hecke operator T_5 on each space.

When $2k = 6$, if this eigenvalue is not equal to $a_f(5) = -66$, we can immediately discard the space. In only two cases is the eigenvalue equal to -66: when $N = 6$, and in one of the subspaces obtained when $N = 48$. In the second case, we check the eigenvalue of the Hecke operator T_7, and it is -176, which is not equal to $a_f(7) = 176$. We can therefore discard this space too.

When $2k = 8$, the eigenvalue is never equal to $a_f(5) = -330$.

When $N = 768$ and $2k = 6$, $N = 384$ and $2k = 8$ or $N = 768$ and $2k = 8$ the spaces become too large and the computation becomes prohibitively long. So we instead compute the space of cuspidal new modular symbols over the field $\mathbb{F}_{13}$. In each case we then compute the characteristic polynomial of the Hecke operator T_5 and check whether the number -66 when $2k = 6$ or -330 when $2k = 8$ is a root of this polynomial. It is not, and therefore neither of these subspaces contains an eigenform with eigenvalue for T_5 congruent to -66, respectively -330, modulo 13. We can therefore discard those spaces as well.

This leaves us with the unique new cusp form of weight 6 and level $\Gamma_0(6)$. For good measure, we check that its Hecke eigenvalues agree with our list of $a_f(p)$, and they do for all primes p such that $3 \leqslant p < 1000$.

References

AB09   S. Arkhipov and R. Bezrukavnikov, *Perverse sheaves on affine flags and Langlands dual group (with an appendix by R. Bezrukavnikov and I. Mirković)*, Israel J. Math. **170** (2009), 135–183.

BBD82   A. Beilinson, J. Bernstein and P. Deligne, *Faisceaux pervers*, in *Analysis and topology on singular spaces, I*, Astérisque, vol. 100 (Société Mathématique de France, Paris, 1982), 5–171.

BLZ04   L. Berger, H. Li and H. Zhu, *Construction of some families of 2-dimensional crystalline representations*, Math. Ann. **329** (2004), 365–377.

CE07   H. T. Choi and R. Evans, *Congruences for sums of powers of Kloosterman sums*, Int. J. Number Theory **3** (2007), 105–117.

Con   B. Conrad, Lifting global representations with local properties, Preprint available at http://math.stanford.edu/~conrad/.

Del74   P. Deligne, *Théorie de Hodge. III*, Publ. Math. Inst. Hautes Études Sci. (1974), 5–77.

Del77   P. Deligne, *Applications de la formule des traces aux sommes trigonométriques*, in *Cohomologie étale. SGA* 412, Lecture Notes in Mathematics, vol. 569 (Springer, New York, 1977).

Del80   P. Deligne, *La conjecture de Weil, II*, Publ. Math. Inst. Hautes Études Sci. (1980), 137–252.

Eva10a R. Evans, *Hypergeometric $_3F_2(1/4)$ evaluations over finite fields and Hecke eigenforms*, Proc. Amer. Math. Soc. **138** (2010), 517–531.

Eva10b R. Evans, *Seventh power moments of Kloosterman sums*, Israel J. Math. **175** (2010), 349–362.

Fal02   G. Faltings, *Almost étale extensions*, in *Cohomologies p-adiques et applications arithmétiques, II*, Astérisque, vol. 279 (Société Mathématique de France, Paris, 2002), 185–270.

FL82   J.-M. Fontaine and G. Laffaille, *Construction de représentations p-adiques*, Ann. Sci. Éc. Norm. Supér. (4) **15** (1982), 547–608.

FG09   E. Frenkel and B. Gross, *A rigid irregular connection on the projective line*, Ann. of Math. (2) **170** (2009), 1469–1512.

FW05   L. Fu and D. Wan, *L-functions for symmetric products of Kloosterman sums*, J. Reine Angew. Math. **589** (2005), 79–103.

FW08   L. Fu and D. Wan, *L-functions of symmetric products of the Kloosterman sheaf over $\mathbb{Z}$*, Math. Ann. **342** (2008), 387–404.

FW10   L. Fu and D. Wan, *Functional equations of L-functions for symmetric products of the Kloosterman sheaf*, Trans. Amer. Math. Soc. **362** (2010), 5947–5965.

Gin95   V. Ginzburg, *Perverse sheaves on a Loop group and Langlands duality*, Preprint (1995), arXiv:math/9511007.

GR10   B. Gross and M. Reeder, *Arithmetic invariants of discrete Langlands parameters*, Duke Math. J. **154** (2010), 431–508.

HNY13   J. Heinloth, B.-C. Ngô and Z. Yun, *Kloosterman sheaves for reductive groups*, Ann. of Math. (2) **177** (2013), 241–310.

HSVV01   K. Hulek, J. Spandaw, B. van Geemen and D. van Straten, *The modularity of the Barth–Nieto quintic and its relatives*, Adv. Geom. **1** (2001), 263–289.

Kat88   N. Katz, *Gauss sums, Kloosterman sums and monodromy groups*, Annals of Mathematics Studies, vol. 116 (Princeton University Press, Princeton, NJ, 1988).

KW09a   C. Khare and J.-P. Wintenberger, *Serres modularity conjecture (I)*, Invent. Math. **178** (2009), 485–504.

KW09b   C. Khare and J.-P. Wintenberger, *Serres modularity conjecture (II)*, Invent. Math. **178** (2009), 505–586.

Kis02   M. Kisin, *Potential semi-stability of p-adic étale cohomology*, Israel J. Math. **129** (2002), 157–173.

Kis07   M. Kisin, *Modularity of 2-dimensional Galois representations*, in *Current Developments in Mathematics* (International Press, Somerville, MA, 2007), 191–230.

Lau03   G. Laumon, *Transformation de Fourier homogène*, Bull. Soc. Math. France **131** (2003), 527–551.

Liv95   R. Livné, *Motivic orthogonal two-dimensional representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Israel J. Math. **92** (1995), 149–156.

Lus83   G. Lusztig, *Singularities, character formulas, and a q-analog of weight multiplicities*, in *Analysis and topology on singular spaces, II, III*, Astérisque, vol. 101–102 (Société Mathématique de France, Paris, 1983), 208–229.

MV07   I. Mirković and K. Vilonen, *Geometric Langlands duality and representations of algebraic groups over commutative rings*, Ann. of Math. (2) **166** (2007), 95–143.

PZ13   G. Pappas and X. Zhu, *Local models of Shimura varieties and a conjecture of Kottwitz*, Invent. Math. **194** (2013), 147–254.

Pat12   S. Patrikis, *Variations on a theorem of Tate*, Preprint (2012), arXiv:1207.6724.

PTV92   C. Peters, J. Top and M. van der Vlugt, *The Hasse zeta function of a K3 surface related to the number of words of weight 5 in the Melas codes*, J. Reine Angew. Math. **432** (1992), 151–176.

Ser68   J.-P. Serre, *Corps Locaux*, Publications de l'Université de Nancago, vol. VIII, second edition (Hermann, Paris, 1968).

Ser87   J.-P. Serre, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Duke Math. J. **54** (1987), 179–230.

Ser89   J.-P. Serre, *Abelian l-adic representations and elliptic curves*, Advanced book classics (Addison-Wesley, Redwood City, CA, 1989).

Zhiwei Yun   zwyun@stanford.edu
Department of Mathematics, Stanford University, 450 Serra Mall, Building 380, Stanford, CA 94305, USA

Christelle Vincent   cvincent@stanford.edu
Department of Mathematics, Stanford University, 450 Serra Mall, Building 380, Stanford, CA 94305, USA