Cambridge Journal of Mathematics
Volume 9, Number 3, 577–691, 2021

Hypergeometric sheaves and finite symplectic and unitary groups*

Nicholas M. Katz and Pham Huu Tiep†

We construct hypergeometric sheaves whose geometric monodromy groups are the finite symplectic groups $\mathrm{Sp}_{2n}(q)$ for any **odd** $n \geq 3$, for q any power of an odd prime p. We construct other hypergeometric sheaves whose geometric monodromy groups are the finite unitary groups $\mathrm{GU}_n(q)$, for any **even** $n \geq 2$, for q any power of **any** prime p. Suitable Kummer pullbacks of these sheaves yield local systems on $\mathbb{A}^1$, whose geometric monodromy groups are $\mathrm{Sp}_{2n}(q)$, respectively $\mathrm{SU}_n(q)$, in their total Weil representation of degree q^n, and whose trace functions are simple-to-remember one-parameter families of two-variable exponential sums. The main novelty of this paper is three-fold. First, it treats unitary groups $\mathrm{GU}_n(q)$ with n **even** via hypergeometric sheaves for the first time. Second, in both the symplectic and the unitary cases, it uses a maximal torus which is a product of two sub-tori to furnish a generator of local monodromy at 0. Third, this is the first natural occurrence of families of two-variable exponential sums in the context of finite classical groups.

AMS 2000 subject classifications: Primary 11T23, 14D05, 20C33; secondary 20C15, 20D06, 20G40.
Keywords and phrases: Local systems, Hypergeometric sheaves, Monodromy groups, Finite simple groups, Weil representations.

[*] The authors are grateful to the referee for careful reading and helpful comments on the paper.

[†] The second author gratefully acknowledges the support of the NSF (grant DMS-1840702), the Joshua Barlaz Chair in Mathematics, and the Charles Simonyi Endowment at the Institute for Advanced Study (Princeton).

1. Introduction

Throughout this paper, p is a prime, and q is a power of p. In our previous paper [KT3], we considered the problem of realizing the finite symplectic groups $\mathrm{Sp}_{2n}(q)$ when p is odd as monodromy groups of "simple to remember" families of exponential sums on the affine line $\mathbb{A}^1$ in characteristic p, with the proviso that these families themselves be closely related to hypergeometric sheaves, and the analogous problem for the finite unitary groups $\mathrm{SU}_n(q)$. In the $\mathrm{Sp}_{2n}(q)$ case, we succeeded for **even** n with $p > 2$ (and $\mathrm{Sp}_2(q)$ was treated in [KT1]). In the $\mathrm{SU}_n(q)$ case, we succeeded for **odd** $n \geq 3$, again when p was odd. For a long time, we did not believe that $\mathrm{SU}_n(q)$ for $n \geq 2$ **even** could be obtained from hypergeometric sheaves.

In this paper, we make use of a new approach, which allows us to treat the $\mathrm{Sp}_{2n}(q)$ case for $n \geq 3$ **odd**, still with p odd, and the $\mathrm{SU}_n(q)$ case for

$n \geq 2$ **even** and any p. This approach is based on the novel idea of constructing hypergeometric sheaves whose local monodromy at 0 uses a generator of a maximal torus in the group which is a product of two sub-tori. Previously, known hypergeometric sheaves for finite classical groups all have local monodromy at 0 that utilizes only cyclic maximal tori of the classical group in question. This novel approach allows us to treat unitary groups $\mathrm{GU}_n(q)$ with n **even** via hypergeometric sheaves for the first time. Another principal novelty of this paper is the use of the operation **Cancel** on hypergeometric sheaves [K2, 9.3.1] as a way to obtain the explicit trace functions of our candidates for total Weil representations of symplectic and unitary groups. Note that, even though these hypergeometric sheaves have a fairly explicit shape predicted by local monodromy considerations, they individually do not have nice trace functions. The use of **Cancel** allows us to show that suitable direct sums of them do have nice trace functions. These trace functions are then used to prove that their monodromy groups are finite, but these trace functions alone give us no clue what the finite monodromy groups are, and in what representations they are occurring. We then prove group-theoretic recognition results that identify these geometric monodromy groups as finite symplectic and unitary groups acting in their total Weil representations. Further group-theoretic results are then established to identify the occurring arithmetic monodromy groups.

This paper may also be viewed as a companion piece to [KT7], which determines which almost quasisimple groups can possibly occur as monodromy groups of hypergeometric sheaves. The main results, Theorems 6.6 and 7.3 of [KT7], show that if a finite classical group G in characteristic r can be realized as the geometric monodromy group of a hypergeometric sheaf $\mathcal{H}$ on $\mathbb{G}_m/\overline{\mathbb{F}_p}$, then, aside from a small and explicit list of exceptions, we necessarily have that $r = p$ and that G is the image of a general linear group $\mathrm{GL}_n(q)$, a general unitary group $\mathrm{GU}_n(q)$, or a symplectic group $\mathrm{Sp}_{2n}(q)$ with q a power of p, and moreover the resulting representation of G is an irreducible Weil representation. The converse problem of showing that such a finite classical group G acting in a Weil representation does indeed occur as the geometric monodromy group of a hypergeometric sheaf $\mathcal{H}$ is the subject of the current paper and [KT6]. As mentioned above, a major difference between the local systems considered in this paper and the ones in [KT6] is the novel consideration of a product of two sub-tori as local monodromy at 0 in this paper, which necessitates the development of new algebro-geometric and group-theoretic tools.

Let us briefly discuss the "numerology" of our approach here. We are given a prime p, a strictly positive power q of p, and two positive integers

$$a, b.$$

We then define

$$M := \gcd(q^a + 1, q^b + 1),$$
$$A := (q^a + 1)/M, B := (q^b + 1)/M.$$

Thus

$$\gcd(A, B) = 1.$$

Grosso modo, when $M = 2$, we find that we are dealing with $\mathrm{Sp}_{2(a+b)}(q)$, and that when $M = q + 1$, we are dealing with $\mathrm{SU}_{a+b}(q)$. A moment's reflection shows that $M = 2$ is only possible if at least one of a, b is even; a further technical constraint requires that the other be odd and this is why we can only attain $\mathrm{Sp}_{2n}(q)$ for $n \geq 3$ **odd**. Similarly, $M = q + 1$ is only possible if both a, b are odd; this is why we can only attain $\mathrm{SU}_n(q)$ for n **even**. Because M is a divisor of each of q^a+1 and q^b+1, $M = 2$ is only possible if q is odd, whereas $M = q + 1$ imposes no parity restriction on q. This is what allows us to treat the $\mathrm{SU}_n(q)$ case, $n \geq 2$ even, in any characteristic.

It then turns out that one-parameter families of **two-variable** exponential sums, of the shape

$$t \in E \mapsto \sum_{x,w \in E} \psi_E(txw + x^{q^b+1} + w^{q^a+1})$$

are what provide the sought after total Weil representations. This is in sharp contrast to the case of $\mathrm{SU}_n(q)$ with n **odd** or any $\mathrm{Sp}_{2n}(q)$, where the total Weil representations are incarnated by families of **one-variable** exponential sums. Moreover, with $n = a + b$, it comes as a pleasant surprise that the local systems with these trace functions realize total Weil representations of $\mathrm{Sp}_{2n}(q)$ and of $\mathrm{SU}_n(q)$. In the "overlap" case of $\mathrm{Sp}_{2n}(q)$ with n **odd**, where we have both this two-variable exponential sum approach and the already developed one-variable exponential sum approach to total Weil representations, it would be interesting to understand the relation between the two approaches.

The main result for symplectic groups is Theorem 15.7, which explicitly constructs hypergeometric sheaves whose arithmetic and geometric monodromy groups realize $\mathrm{Sp}_{2n}(q)$ in its irreducible Weil representations. Suitable Kummer pullbacks of these sheaves yield local systems over $\mathbb{A}^1/\mathbb{F}_q$ with the same monodromy groups and with trace functions being easy to remember one-parameter families of two-variable exponential sums. Similar results for unitary groups (in any characteristic) are established in Theorems 16.11, 16.12, and 17.5.

2. A variant approach to finite monodromy, especially of hypergeometric sheaves; the $[N]_\star$ method

We refer to [K2, 8.2.2, 8.4] or to [K5, 2.2] for the definitions and basic properties of hypergeometric sheaves. Let $\mathcal{H}yp_\psi(\chi_1,\dots,\chi_n;\rho_1,\dots,\rho_m)$ be a hypergeometric sheaf of type (n,m), defined over a finite field $\mathbb{F}_q$ in the strong sense that ψ is a nontrivial additive character of $\mathbb{F}_q$ and each χ_i and each ρ_j is a (possibly trivial) multiplicative character of $\mathbb{F}_q^\times$. We assume that no χ_i is any ρ_j.

If we pick an embedding of $\mathbb{Q}(\mu_{q-1})$ into $\overline{\mathbb{Q}_\ell}$, we can view the multiplicative characters χ_i and ρ_j as taking values in $\mathbb{Q}(\mu_{q-1})$. So viewed, it makes sense to ask if the set (with multiplicity) consisting of the upstairs characters χ_i, and the set (with multiplicity) consisting of the downstairs characters ρ_j, are each Galois stable (by the action of $\mathrm{Gal}(\mathbb{Q}(\mu_{q-1})/\mathbb{Q})$). [This notion does not depend on the choice of embedding of $\mathbb{Q}(\mu_{q-1})$ into $\overline{\mathbb{Q}_\ell}$, since any two embeddings differ by precomposition with an element of $\mathrm{Gal}(\mathbb{Q}(\mu_{q-1})/\mathbb{Q})$.] If both sets are Galois stable, we say that $\mathcal{H}yp_\psi(\chi_1,\dots,\chi_n;\rho_1,\dots,\rho_m)$ has Galois stable sets of characters.

Lemma 2.1. *Let $\mathcal{H}yp_\psi(\chi_1,\dots,\chi_n;\rho_1,\dots,\rho_m)$ be a hypergeometric of type (n,m) with Galois stable sets of characters. Then we have the following results.*

(i) *For any finite extension field $L/\mathbb{F}_q$, and any point $t\in L^\times$, the trace*

$$\mathrm{Trace}\big(Frob_{t,L}|\mathcal{H}yp_\psi(\chi_1,\dots,\chi_n;\rho_1,\dots,\rho_m)\big)$$

and the determinant

$$\det\big(Frob_{t,L}|\mathcal{H}yp_\psi(\chi_1,\dots,\chi_n;\rho_1,\dots,\rho_m)\big)$$

both lie in $\mathbb{Z}[\zeta_p]$.

(ii) *When p is odd, the "Gauss-twisted" sheaf*

$$\mathcal{H}yp_\psi(\chi_1,\dots,\chi_n;\rho_1,\dots,\rho_m)\big(-\mathsf{Gauss}(\psi,\chi_2)\big)^{-(n+m-1)\deg}$$

is pure of weight zero, for every embedding of $\overline{\mathbb{Q}_\ell}$ into $\mathbb{C}$, and has traces in $\mathbb{Z}[\zeta_p][1/p]$. When p is even and q is an even power of p, the same is true with $\mathsf{Gauss}(\psi,\chi_2)$ replaced by $\sqrt{q}$.

Proof. The two Galois extensions $\mathbb{Q}(\mu_p)/\mathbb{Q}$ and $\mathbb{Q}(\mu_{q-1})/\mathbb{Q}$ are linearly disjoint, so we may view $\mathrm{Gal}(\mathbb{Q}(\mu_{q-1})/\mathbb{Q})$ as $\mathrm{Gal}(\mathbb{Q}(\mu_{q-1},\mu_p)/\mathbb{Q}(\mu_p))$ and $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$ as $\mathrm{Gal}(\mathbb{Q}(\mu_{q-1},\mu_p)/\mathbb{Q}(\mu_{q-1}))$.

In the formulas below, we write ψ_L for $\psi\circ\mathrm{Trace}_{L/\mathbb{F}_q}$, and we write $\chi_{i,L}$ for $\chi_i\circ\mathrm{Norm}_{L/\mathbb{F}_q}$.

The assertion about the trace is obvious from the explicit formula [K2, 8.2.7] for this trace, namely

$$(-1)^{n+m-1}\sum_{\prod_i x_i=t\prod_j y_j}\psi_L\Big(\sum_i x_i-\sum_j y_j\Big)\prod_i\chi_{i,L}(x_i)\prod_j\overline{\rho_{j,L}}(y_j).$$

This formula makes clear that the trace is an algebraic integer, and that the effect of

$$\mathrm{Gal}\big(\mathbb{Q}(\mu_{q-1})/\mathbb{Q}\big)\cong\mathrm{Gal}\big(\mathbb{Q}(\mu_{q-1},\mu_p)/\mathbb{Q}(\mu_p)\big)$$

is simply to permute the x_i and to permute the y_j. Thus the trace is an algebraic integer in the field $\mathbb{Q}(\mu_p)$, so lies in $\mathbb{Z}[\zeta_p]$. This rationality, applied over finite extensions, gives the same rationality for the determinant (indeed for all the coefficients of the reversed characteristic polynomial

$$\det\big(1-TFrob_{t,L}|\mathcal{H}yp_\psi(\chi_1,\ldots,\chi_n;\rho_1,\ldots,\rho_m)\big)\big).$$

We know [K2, 8.4.13] that $\mathcal{H}yp_\psi(\chi_1,\ldots,\chi_n;\rho_1,\ldots,\rho_m)$ is pure of weight $n+m-1$, so the "Gauss-twisted" sheaf is pure of weight zero. It results from the first assertion that this variant of the Tate-twist has values in $\mathbb{Z}[\zeta_p][1/p]$, since the quadratic Gauss sum lies in $\mathbb{Z}[\zeta_p]$ and divides p. □

Proposition 2.2. *Let $\mathcal{F}$ be a lisse sheaf on $\mathbb{G}_m/\mathbb{F}_q$ which is pure of weight zero and which has all traces in $\mathbb{Z}[\zeta_p][1/p]$. Suppose that $\mathcal{F}$ is arithmetically semisimple. Fix an integer $N\geq 1$ prime to p. Then the following are equivalent.*

(a) *$G_{\mathrm{arith},\mathcal{F}}$ is finite.*
(b) *$G_{\mathrm{geom},\mathcal{F}}$ is finite.*
(c) *All traces of $\mathcal{F}$ are algebraic integers, i.e., lie in $\mathbb{Z}[\zeta_p]$.*
(d) *For any finite extension $L/\mathbb{F}_q$, for any chosen p-adic* ord *of $\mathbb{Q}(\mu_p,\mu_{|L^\times|})$ extending the usual p-adic* ord_p *on $\mathbb{Q}$ with $\mathrm{ord}_p(p^n)=n$, and for every*

multiplicative character χ *of* $L^\times$, *the sum*

$$\sum_{t\in L^\times}\chi(t)\mathrm{Trace}(Frob_{t,L}|\mathcal{F})$$

has $\mathrm{ord}_{p,L}\geq 0$.

(e) *For every finite extension* $L/\mathbb{F}_q$, *for any* p*-adic* ord *of* $\mathbb{Q}(\mu_p,\mu_{|L^\times|})$, *and for every multiplicative character* χ *of* $L^\times$, *the sum*

$$\sum_{t\in L^\times}\chi(t)^N\mathrm{Trace}(Frob_{t,L}|\mathcal{F})$$

has $\mathrm{ord}_{p,L}\geq 0$.

Proof. The equivalence of (c) and (d) results from the Mellin transform argument, cf. [KRLT1, 2.1, 2.2, 2.7], which also explains the equivalence of (a), (b), and (c). It is obvious that (d) implies (e). It remains to show that (e) implies (d). We use the identity

$$\begin{aligned}\sum_{t\in L^\times}\chi(t)^N\mathrm{Trace}(Frob_{t,L}|\mathcal{F}) &= \sum_{s\in L^\times}\chi(s)\sum_{t\in L^\times,t^N=s}\mathrm{Trace}(Frob_{t,L}|\mathcal{F})\\ &= \sum_{s\in L^\times}\chi(s)\mathrm{Trace}(Frob_{s,L}|[N]_\star\mathcal{F}).\end{aligned}$$

The Kummer direct image $[N]_\star\mathcal{F}$ remains pure of weight zero and arithmetically semisimple, with all traces in $\mathbb{Q}(\mu_p)$. Therefore by the equivalence of (a) through (b), applied to $[N]_\star\mathcal{F}$, we see that $[N]_\star\mathcal{F}$ has finite G_{arith}. Therefore its pullback $[N]^\star[N]_\star\mathcal{F}$ has finite G_{arith}. But $\mathcal{F}$ is a direct factor of this pullback, so $\mathcal{F}$ itself has finite G_{arith}. [When $\mathbb{F}_q$ contains the N^{th} roots of unity, this pullback is the direct sum of the multiplicative translates of $\mathcal{F}$ by the N^{th} roots of unity. If $\mathbb{F}_q$ does not contain the N^{th} roots of unity, we break up the sum of these translates into clumps according to the order of the N^{th} root of unity by which we translate. Each of these clumps lives over $\mathbb{F}_q$, and $\mathcal{F}$ is the clump for the trivial translate.] □

3. Overall set-up

Here p is a prime, q is a power of p,

$$a,b$$

are positive integers. We define

$$M := \gcd(q^a + 1, q^b + 1),$$
$$A := (q^a + 1)/M, B := (q^b + 1)/M.$$

Thus

$$\gcd(A, B) = 1.$$

We also fix integers α, β with

$$\alpha A - \beta B = 1.$$

We also fix a prime number $\ell \neq p$, so as to be able to use $\overline{\mathbb{Q}_\ell}$-cohomology, and we fix a choice of nontrivial additive character ψ of $\mathbb{F}_p$.

Given an integer $n \geq 1$ which is prime to p, we denote by $\mathsf{Char}(n)$ the group of multiplicative characters of order dividing n. Given a multiplicative character ρ of finite order, we denote by

$$\mathsf{Char}(n; \rho) := \{\chi | \chi^n = \rho\}.$$

Thus $\mathsf{Char}(n) = \mathsf{Char}(n; \mathbb{1})$.

In the above paragraph, the characters in question are the characters of finite order of the tame fundamental group of $\mathbb{G}_m/\overline{\mathbb{F}_p}$, which is the inverse limit of the multiplicative groups of the finite subfields of $\overline{\mathbb{F}_p}$, with transition maps the norm. Thus the group of characters in question is the direct limit of the groups $\mathrm{Hom}(k^\times, \overline{\mathbb{Q}_\ell}^\times)$ under the inclusion maps, whenever $k_2/k_1/\mathbb{F}_p$ are finite extensions, given by

$$\mathrm{Hom}(k_1^\times, \overline{\mathbb{Q}_\ell}^\times) \subset \mathrm{Hom}(k_2^\times, \overline{\mathbb{Q}_\ell}^\times), \chi \mapsto \chi \circ \mathrm{Norm}_{k_2/k_1}.$$

4. Kloosterman candidates

We refer to [K1, 4.1.1, with all b_i there taken to be 1] and to [K2, 8.4.3] for the definition and properties of Kloosterman sheaves. Given multiplicative characters χ and ρ, each of order dividing M, such that the two sets $\mathsf{Char}(A, \chi)$ and $\mathsf{Char}(B, \rho)$ are disjoint, we may speak of the Kloosterman sheaf

$$\mathcal{K}l_\psi\big(\mathsf{Char}(MAB) \smallsetminus (\mathsf{Char}(A, \chi) \sqcup \mathsf{Char}(B, \rho))\big).$$

Lemma 4.1. *The two sets* $\mathsf{Char}(A, \chi)$ *and* $\mathsf{Char}(B, \rho)$ *fail to be disjoint if and only if* $\rho^A = \chi^B$.

Proof. Suppose the two sets are not disjoint. Let Λ have $\Lambda^A = \chi, \Lambda^B = \rho$. Then $\Lambda^{AB} = \chi^B$ and $\Lambda^{AB} = \rho^A$. Conversely, if $\rho^A = \chi^B$, then using the fact that $\gcd(A, B) = 1$ we see that there is a (necessarily unique) Λ with $\Lambda^A = \chi, \Lambda^B = \rho$. Indeed, using the integers α, β with $\alpha A - \beta B = 1$, then we have $\Lambda = \chi^\alpha / \rho^\beta$. Notice that Λ itself has order dividing M. □

Lemma 4.2. *Choose integers k, l such that*

$$kA - lB = 1.$$

Consider the two injective group homomorphisms

$$\mathsf{Char}(M) \to \mathsf{Char}(M) \times \mathsf{Char}(M),$$

given by

$$\phi_{A,B} : \Lambda \mapsto (\Lambda^A, \Lambda^B), \ \ \phi_{l,k} : \sigma \mapsto (\sigma^l, \sigma^k),$$

with image groups $\mathrm{Im}_{A,B}$ *and* $\mathrm{Im}_{l,k}$. *Then the group* $\mathsf{Char}(M) \times \mathsf{Char}(M)$ *is the product*

$$\mathrm{Im}_{A,B} \times \mathrm{Im}_{l,k}.$$

Proof. As we have two subgroups, each of order M in an abelian group of order M^2, it suffices to show that the intersection $\mathrm{Im}_{A,B} \cap \mathrm{Im}_{l,k}$ consists of the single element $(\mathbb{1}, \mathbb{1})$. To see this, note that if (χ, ρ) lies in the intersection, then on the one hand $(\chi, \rho) = (\Lambda^A, \Lambda^B)$ for some Λ, hence $\chi^B/\rho^A = \mathbb{1}$. On the other hand, $(\chi, \rho) = (\sigma^l, \sigma^k)$ for some σ, so $\rho^A/\chi^B = \sigma$. Thus $\sigma = \mathbb{1}$, and hence $(\chi, \rho) = (\mathbb{1}, \mathbb{1})$. □

Given multiplicative characters χ and ρ, each of order dividing M, such that the two sets $\mathsf{Char}(A, \chi)$ and $\mathsf{Char}(B, \rho)$ are disjoint, we denote by

$$\mathcal{K}l(M, A, B, \chi, \rho)$$

the Kloosterman sheaf

$$\mathcal{K}l_\psi\big(\mathsf{Char}(MAB) \smallsetminus (\mathsf{Char}(A, \chi) \sqcup \mathsf{Char}(B, \rho))\big). \tag{4.2.1}$$

We have the following twisting formula:

Lemma 4.3. *For Λ a character of order dividing M, we have the twisting formula*

$$\mathcal{L}_\Lambda \otimes \mathcal{K}l(M, A, B, \chi, \rho) = \mathcal{K}l(M, A, B, \chi\Lambda^A, \rho\Lambda^B).$$

Remark 4.4. The rank of $\mathcal{K}l(M, A, B, \chi, \rho)$, namely $MAB - A - B$, is

$$(q^{a+b} - 1)/M.$$

Corollary 4.5. *The above $M(M-1)$ sheaves $\mathcal{K}l(M, A, B, \chi, \rho)$ with*

$$\chi^B \neq \rho^A$$

are precisely the $\mathcal{L}_\Lambda$ twists, with Λ of order dividing M, of the $M-1$ sheaves $\mathcal{K}l(A, B, \sigma^l, \sigma^k)$ with $\sigma \neq \mathbb{1}$ of order dividing M.

Remark 4.6. We will often make constant field twists of the sheaves under consideration to achieve weight zero. In odd characteristic, we will do this using the correct power of a choice of the quadratic Gauss sum. In characteristic $p = 2$, so long as our ground field k is an even degree extension of $\mathbb{F}_p$, we will use the correct power of

$$p^{\deg(k/\mathbb{F}_p)/2}.$$

In other words, we **define**

$$-\mathsf{Gauss}(\psi_k, \chi_2) := p^{\deg(k/\mathbb{F}_p)/2}$$

when $k/\mathbb{F}_p$ has even degree and $p = 2$, and proceed with the usual formalism of Gauss sums.

Theorem 4.7. *Each of the above $M(M-1)$ sheaves $\mathcal{K}l(M, A, B, \chi, \rho)$ with $\chi^B \neq \rho^A$ has finite geometric monodromy group G_{geom}. Over any finite field $k/\mathbb{F}_p$ containing the MAB roots of unity, the constant field twist*

$$\mathcal{K}l(M, A, B, \chi, \rho) \otimes (-1/\mathsf{Gauss}(\psi_k, \chi_2))^{\deg \times (\mathrm{rank}(\mathcal{K}l) - 1)}$$

is pure of weight zero and has finite arithmetic monodromy group G_{arith}.

Proof. The purity of weight zero is simply the fact that a Kloosterman sheaf is pure of weight one less than its rank. It suffices to show the finiteness of G_{arith}, since $G_{\mathrm{geom}} < G_{\mathrm{arith}}$. For this, we use the Kubert V-function, cf. [K4, §13] and cf. [KRL, Appendix]. The criterion is that for all $x \in (\mathbb{Q}/\mathbb{Z})_{\text{not } p}$, and for every pair of integers n, m mod M with $Bn \neq Am$ mod M, we have

$$V(MABx) - V(Ax + n/M) - V(Bx + m/M) + 1 \geq 0.$$

In fact, we will prove this for every pair of integers n, m mod M. We will make use of the $[N]_\star$ method, explained in §2, with $N := M$. This amounts to replacing the variable x in the above inequalities by $Nx = Mx$, exactly as in Proposition 2.2, which replaces χ by χ^N in passing from condition (d) to condition (e). The criterion becomes that for all $x \in (\mathbb{Q}/\mathbb{Z})_{\text{not } p}$, and for every pair of integers n, m mod M, we have

$$V((q^a+1)(q^b+1)x) - V((q^a+1)x + n/M) - V((q^b+1)x + m/M) + 1 \geq 0.$$

We rewrite this as

$$1 + V((q^a+1)(q^b+1)x) \geq V((q^a+1)x + n/M) + V((q^b+1)x + m/M).$$

Recall that the integrality of Jacobi sums gives

$$V(x) + V(y) \geq V(x+y),$$

which (replacing x, y by $-x, -y$) gives

$$1 + V(x+y) \geq V(x) + V(y).$$

Because both $q^a + 1$ and $q^b + 1$ are divisible by M, we have equalities in $(\mathbb{Q}/\mathbb{Z})_{\text{not } p}$

$$\begin{aligned}(q^a+1)(q^b+1)x &= (q^a+1)((q^b+1)x + m/M),\\ (q^a+1)(q^b+1)x &= (q^b+1)((q^a+1)x + n/M).\end{aligned}$$

Using the first of these, and the inequality $1 + V(x+y) \geq V(x) + V(y)$, we get

$$\begin{aligned}1 + V((q^a+1)(q^b+1)x) &= 1 + V((q^a+1)((q^b+1)x + m/M))\\ &\geq V(q^a((q^b+1)x+m/M) + V((q^b+1)x+m/M))\\ &= 2V((q^b+1)x + m/M).\end{aligned}$$

Using the second of these, we get

$$\begin{aligned}1 + V((q^a+1)(q^b+1)x) &= 1 + V((q^b+1)((q^a+1)x + n/M))\\ &\geq V(q^b((q^a+1)x + n/M)) + V((q^a+1)x + n/M)\\ &= 2V((q^a+1)x + n/M).\end{aligned}$$

Adding these last two inequalities, we get twice the asserted inequality. □

Lemma 4.8. *If qAB is odd, then of the $M(M-1)$ sheaves $\mathcal{K}l(M,A,B,\chi,\rho)$ with $\chi^B \neq \rho^A$, precisely two are geometrically self-dual. They are*

$$\mathcal{K}l(M,A,B,\mathbb{1},\chi_2) \text{ and } \mathcal{K}l(M,A,B,\chi_2,\mathbb{1}).$$

Each is symplectically self-dual. If qAB is even, none of these $M(M-1)$ sheaves is geometrically self-dual.

Proof. Suppose first that q is odd. Then M is even. If both A, B are odd, then these sheaves all have even rank (namely $MAB - A - B$). Because MAB is even, each has determinant $\chi_2/(\chi\rho)$. Their sets of characters are stable by complex conjugation precise when each of χ and ρ is its own complex conjugate, i.e., when each is either $\mathbb{1}$ or χ_2. They cannot both be $\mathbb{1}$ or both be χ_2, as these cases violate the disjointness condition. Both $\mathcal{K}l(M,A,B,\mathbb{1},\chi_2)$ and $\mathcal{K}l(M,A,B,\chi_2,\mathbb{1})$ are self-dual. Their determinants, being $\chi_2/(\chi\rho)$, are then both trivial, so the asserted symplectic autoduality follows from [K2, 8.8.1-2].

Suppose now that q is odd and precisely one of A, B is odd. [They cannot both be even because their gcd $= 1$.] Then each sheaf has odd rank $MAB - A - B$, and one knows [K2, 8.8.1] that in odd characteristic, no Kloosterman sheaf of odd rank is geometrically self-dual.

Suppose now that q is even. Then autoduality [K2, 8.8.1] would force each of χ and ρ to be its own complex conjugate, which in characteristic 2 forces them both to be trivial, and this violates disjointness. □

In the rest of this section, we prove the primitivity of the Kloosterman sheaves $\mathcal{K}l(M,A,B,\chi,\rho)$ with $\chi^B \neq \rho^A$ considered above (with one exception, see the statement of Corollary 4.12).

Lemma 4.9. *Let $N \geq 2$ be a prime to p integer, $A \geq 1$ and $B \geq 1$ two divisors of N with $\gcd(A,B) = 1$ and $A \neq B$. Let $A_1 \subset \mathsf{Char}(N)$ be a $\mathsf{Char}(A)$-coset, and let $B_1 \subset \mathsf{Char}(N)$ be a $\mathsf{Char}(B)$-coset. Suppose that A_1 and B_1 are disjoint. Then the Kloosterman sheaf*

$$\mathcal{K}l_\psi(\mathsf{Char}(N) \smallsetminus (A_1 \sqcup B_1))$$

is not Kummer induced (and hence not induced, by Pink's lemma [K3, Lemma 11]).

Proof. Suppose our Kloosterman sheaf $\mathcal{K}l$ is Kummer induced, say is $[d]_\star\mathcal{F}$ for some Kloosterman sheaf $\mathcal{F}$ and some prime to p integer d. Then every χ of order dividing d is a ratio of characters occurring in $\mathcal{K}l$, all of which have order dividing N. Thus d divides N. For χ a character of order d, we thus have

$$\mathcal{L}_\chi \otimes \mathcal{K}l \cong \mathcal{K}l,$$

which means precisely that we have an equality of sets

$$\chi\mathsf{Char}(N) \smallsetminus (\chi A_1 \sqcup \chi B_1) = \mathsf{Char}(N) \smallsetminus (A_1 \sqcup B_1).$$

Now $\chi\mathsf{Char}(N)$ is just $\mathsf{Char}(N)$, so inside $\mathsf{Char}(N)$ we have the equality of subsets

$$\chi A_1 \sqcup \chi B_1 = A_1 \sqcup B_1.$$

We will show that in fact $\chi A_1 = A_1$ and $\chi B_1 = B_1$. Once we know this, then χ lies in both $\mathsf{Char}(A)$ and in $\mathsf{Char}(B)$, so must be trivial (because $\gcd(A, B) = 1$). This results from the following elementary lemma.

Lemma 4.10. *Let $N \in \mathbb{Z}_{\geq 1}$ be prime to p and let A, B be divisors of N. Let $S \subset \mathsf{Char}(N)$ be a subset which is the disjoint union of a $\mathsf{Char}(A)$-coset A_1 and a $\mathsf{Char}(B)$-coset B_1. If $A \neq B$, then whenever $S = A_2 \sqcup B_2$ with A_2 a $\mathsf{Char}(A)$-coset and B_2 a $\mathsf{Char}(B)$-coset, we have $A_1 = A_2$ and $B_1 = B_2$.*

Proof. Suppose first that the intersection $A_1 \cap A_2$ is nonempty, say contains α. Then $A_1 = A_2 = \mathsf{Char}(A)\alpha$. From this, we have $S \smallsetminus A_1 = S \smallsetminus A_2$, which is to say $B_1 = B_2$.

Similarly, if $B_1 \cap B_2$ is nonempty, we again get the desired conclusion.

Suppose finally that both $A_1 \cap A_2$ and $B_1 \cap B_2$ are both empty. The $A_1 \subset S \smallsetminus A_2 = B_2$, and $B_2 \subset S \smallsetminus B_1 = A_1$. Thus $A_1 \subset B_2 \subset A_1$, hence $A_1 = B_2$. But this is impossible, because the two sets have different cardinalities A and B respectively. □

As explained above, once we have Lemma 4.10 we have proven Lemma 4.9. □

In Lemma 4.9, we omitted the case when $\gcd(A, B) = 1$ but $A = B$, i.e. the case when $A = B = 1$. Here the situation is as follows.

Lemma 4.11. *Let $N \geq 2$ be a prime to p integer, and $\chi \neq \rho$ two distinct characters in $\mathsf{Char}(N)$. Then the Kloosterman sheaf*

$$\mathcal{K}l_\psi(\mathsf{Char}(N) \smallsetminus \{\chi, \rho\})$$

is primitive, i.e., not Kummer induced, except in the case when N is even and $\chi = \chi_2\rho$, in which case it is the Kummer induction

$$[2]_\star\big(\mathcal{L}_{\rho^2} \otimes \mathcal{K}l_\psi(\mathsf{Char}(N/2) \smallsetminus \{\mathbb{1}\})\big).$$

Proof. Exactly as in the proof of Lemma 4.9, if our Kloosterman sheaf is $[d]_\star\mathcal{F}$ for some Kloosterman sheaf $\mathcal{F}$ and some prime to p integer $d > 1$, then $d|N$ and for any character σ of order d, we have an equality of sets

$$\{\sigma\chi, \sigma\rho\} = \{\chi, \rho\}.$$

As $\sigma\chi \neq \chi$, we must have $\sigma\chi = \rho$, and similarly $\sigma\rho = \chi$. Thus $\sigma^2 = \mathbb{1}$, $d = 2$, and $\chi = \chi_2\rho$. In this case, we indeed have the asserted Kummer induction. □

Corollary 4.12. *Given multiplicative characters χ and ρ, each of order dividing M, such that the two sets* $\mathsf{Char}(A,\chi)$ *and* $\mathsf{Char}(B,\rho)$ *are disjoint, the Kloosterman sheaf $\mathcal{K}l(M,A,B,\chi,\rho)$ satisfies condition* **(S+)** *of* [KT7, Definition 1.2]*,* **except** *in the situation*

$$M = q+1 \text{ is even},\ A = B = 1,\ \text{and } \chi = \chi_2\rho.$$

Proof. Immediate from the primitivity lemmas 4.9 and 4.11, applied with N taken to be MAB, thanks to [KT7, Theorem 1.7]. □

5. The hypergeometric candidate

In this section, we consider the hypergeometric sheaf

$$\mathcal{H}yp_\psi\big(\mathsf{Char}(MAB) \sqcup \{\mathbb{1}\} \smallsetminus (\mathsf{Char}(A) \sqcup \mathsf{Char}(B)); \mathbb{1}\big), \tag{5.0.1}$$

which we denote

$$\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1}).$$

Theorem 5.1. *The sheaf $\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})$ has finite geometric monodromy group $G_{\rm geom}$. Over any finite field $k/\mathbb{F}_p$ containing the $AB(q+1)$ roots of unity, the constant field twist*

$$\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1}) \otimes (-1/\mathsf{Gauss}(\psi_k,\chi_2))^{\deg \times \mathrm{rank}(\mathcal{H}yp)}$$

is pure of weight zero and has finite arithmetic monodromy group $G_{\rm arith}$.

Proof. The purity of weight zero is simply the fact that a hypergeometric sheaf of type (n,m) with disjoint upstairs and downstairs characters is pure of weight $n+m-1$ (in our case, $n = MAB - A - B + 1$, and $m = 1$).

It suffices to show the finiteness of G_{arith}. For this, we use the Kubert V function. The criterion is that for all $x \in (\mathbb{Q}/\mathbb{Z})_{\mathrm{not}\ p}$, we have

$$V(MABx) + V(x) - V(Ax) - V(Bx) + V(-x) \geq 0.$$

If $x = 0$, this trivially holds. If $x \neq 0$, then $V(x) + V(-x) = 1$, and the inequality becomes

$$1 + V(MABx) \geq V(Ax) + V(Bx).$$

This is the $n = m = 0$ case of what was proven in Theorem 4.7. □

Lemma 5.2. *If either qAB is odd or q is even, the sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ is, geometrically, orthogonally self-dual. Otherwise, it is not geometrically self-dual.*

Proof. This sheaf has rank $MAB + 1 - A - B$, and is of type

$$(MAB + 1 - A - B, 1).$$

Its sets of upstairs and downstairs characters are each stable by complex conjugation.

When q is odd, this sheaf is self-dual [K2, 8.8.1] precisely when its rank $MAB + 1 - A - B$ is odd. But when q is odd, M is even, so autoduality holds when $A + B$ is even. But as $\gcd(A, B) = 1$, $A + B$ is even precisely when both A, B are odd. In this case, the rank $MAB + 1 - A - B$ is odd, so the duality must be orthogonal.

When q is even, this sheaf is self-dual [K2, 8.8.1]. Each of M, A, B is odd in this q even case, so the rank n is even. By [K2, 8.8.1], no hypergeometric sheaf of type $(n, 1)$ with n even is symplectically self-dual. Therefore in this case as well, the sheaf is, geometrically, orthogonally self-dual (despite having even rank). □

Lemma 5.3. *We have the following results.*

(i) *If $M = 2$, the hypergeometric sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ is primitive.*
(ii) *If $M = q + 1$, then except in the case $q = 3, a = b = 1$, the hypergeometric sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ is primitive.*
(iii) *If $M = 1$, the hypergeometric sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ is primitive.*

Proof. One knows [KRLT2, Cor. 2.3] that any hypergeometric sheaf of type $(n, 1)$ whose rank n is not a power of p is primitive. If $M = 2$, the sheaf

$\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})$ has rank $(q^{a+b}+1)/2$, which is prime to p. If $M=q+1$ and $a+b>2$, then $\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})$ has rank

$$(q^{a+b}+q)/(q+1)=q(q^{a+b-1}+1)/(q+1),$$

which is q times a p-adic unit. Looking at the ord_p, we see that the rank can only be a power of p if the rank is q, and this happens only when $a+b=2$, i.e., when $a=b=1$.

To finish case (ii), we now treat the case when $M=q+1$ and $a=b=1$. Then $A=B=1$, and $\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})$ is $\mathcal{H}yp(\mathsf{Char}(q+1)\smallsetminus\{\mathbb{1}\};\mathbb{1})$, of type $(q,1)$. It cannot be Kummer induced. For it to be Belyi induced, there must exist positive integers A_0,B_0, both prime to p, with $A_0+B_0=q$, and a nontrivial (otherwise the two sets $\mathsf{Char}(A_0,\chi)$ and $\mathsf{Char}(B_0,\overline{\chi})$ will each contain $\mathbb{1}$) multiplicative character χ such that

$$\mathsf{Char}(q+1)\smallsetminus\{\mathbb{1}\}=\mathsf{Char}(A_0,\chi)\sqcup\mathsf{Char}(B_0,\overline{\chi}).$$

Pick a multiplicative character ρ of full order $q+1$. At the expense of interchanging A_0 and B_0, it suffices to treat the case when $\rho\in\mathsf{Char}(A_0,\chi)$, i.e. when $\rho^{A_0}=\chi$. Then χ, being a power of ρ, has order $d|(q+1)$, and so $\rho^{dA_0}=\mathbb{1}$. Thus $(q+1)|dA_0$. On the other hand, $\mathsf{Char}(A_0,\chi)$ contains a character Λ of full order dA_0. But any such character lies in $\mathsf{Char}(q+1)$, hence $dA_0|(q+1)$. Thus $dA_0=q+1$. Similarly, $\mathsf{Char}(B_0,\overline{\chi})$ contains a character of full order dB_0, so $dB_0|(q+1)$. But $q+1=dA_0$, so $dB_0|dA_0$, hence $B_0|A_0$. But $A_0+B_0=q$, and $p\nmid A_0B_0$, so in fact $\gcd(A_0,B_0)=1$. Therefore $B_0=1$, and hence $A=q-1$. But $dA_0=q+1$, so $d(q-1)=q+1$. This is only possible if $q=3$ and $d=2$. Indeed, in this case, we have $\chi=\chi_2$, $A_0=2,B_0=1$, and in fact we do have

$$\mathsf{Char}(4)\smallsetminus\{\mathbb{1}\}=\mathsf{Char}(2,\chi_2)\sqcup\{\chi_2\}.$$

We now turn to the case $M=1$. Then q is even, $A=q^a+1,B=q^b+1$, and $\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})$ is

$$\mathcal{H}yp(\mathsf{Char}(AB)\sqcup\{\mathbb{1}\}\smallsetminus(\mathsf{Char}(A)\sqcup\mathsf{Char}(B));\mathbb{1}),$$

of rank q^{a+b}. Just as above, this sheaf cannot be Kummer induced. If it is Belyi induced, there must exist positive integers A_0,B_0, both prime to 2, with $A_0+B_0=q^{a+b}$, and a nontrivial multiplicative character χ such that

$$\mathsf{Char}(AB)\sqcup\{\mathbb{1}\}\smallsetminus(\mathsf{Char}(A)\sqcup\mathsf{Char}(B))=\mathsf{Char}(A_0,\chi)\sqcup\mathsf{Char}(B_0,\overline{\chi}).$$

Pick a multiplicative character ρ of full order AB. At the expense of interchanging A_0 and B_0, it suffices to treat the case when $\rho \in \mathsf{Char}(A_0, \chi)$, i.e. when $\rho^{A_0} = \chi$. Then χ, being a power of ρ, has order $d|AB$, and so $\rho^{dA_0} = \mathbb{1}$. Thus $AB|dA_0$. On the other hand, $\mathsf{Char}(A_0, \chi)$ contains a character Λ of full order dA_0. But any such character lies in $\mathsf{Char}(AB)$, hence $dA_0|AB$. Thus $AB = dA_0$. Similarly, $\mathsf{Char}(B_0, \overline{\chi})$ contains a character of full order dB_0, so $dB_0|AB$. But $AB = dA_0$, so $dB_0|dA_0$, and hence $B_0|A_0$. As above, A_0, B_0 are both odd, but sum to a power ot 2, so $\gcd(A_0, B_0) = 1$. Therefore $B_0 = 1$, and hence $A_0 = q^{a+b} - 1$. Thus

$$d(q^{a+b} - 1) = AB = (q^a + 1)(q^b + 1).$$

Because χ is nontrivial, and of order prime to $p = 2$, we have $d \geq 3$. We cannot have a, b both odd, otherwise M is divisible by $q + 1$. The displayed equality is impossible, because $d \geq 3$, but

$$3(q^{a+b} - 1) > (q^a + 1)(q^b + 1).$$

Indeed, this is equivalent to

$$\begin{aligned} 2q^{a+b} - 3 > q^a + q^b + 1, \text{ i.e. } & q^{a+b} + q^{a+b} - q^a - q^b + 1 > 5, \\ \text{i.e. } & q^{a+b} + (q^a - 1)(q^b - 1) > 5. \end{aligned}$$

But a, b are not both odd, so $a + b \geq 3$, and already the q^{a+b} term forces the asserted inequality. □

Corollary 5.4. *We have the following results.*

(i) *If $M = 2$, the hypergeometric sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ satisfies* **(S+)**.

(ii) *If $M = q + 1$, then except in the case $a = b = 1$ and q is one of $\{2, 3, 4, 8, 9\}$ the hypergeometric sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ satisfies* **(S+)**.

(iii) *Suppose that $M = 1$. Then, except in the case $\{a, b\} = \{1, 2\}$ and $q = 2$, the hypergeometric sheaf $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ satisfies* **(S+)**.

Proof. If $M = 2$, then q must be odd, and one of a, b must be even (otherwise $(q + 1)|M$). Thus the rank of $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ is

$$(q^{a+b} + 1)/2 \geq (3^3 + 1)/2 = 14,$$

and we apply [KT7, Theorem 1.9].

If $M = q+1$, then the rank is $(q^{a+b}+q)/(q+1)$, and both of a, b are odd (simply because $q^{\text{even}}+1 \equiv 2 \bmod q+1$). If $a = b = 1$, the rank is q, so we must exclude $q = 2, 3, 4, 8, 9$. Otherwise $a+b \geq 4$, so either the rank is 6 (when $q = 2$ and $a+b = 4$) or it is ≥ 21, and we apply [KT7, Theorem 1.12].

If $M = 1$, then q must be even (otherwise $2|M$), and one of a, b must be even (otherwise $(q+1)|M$). The rank is q^{a+b}. So we exclude the case $q = 2, a+b = 3$ and apply [KT7, Theorem 1.9]. □

6. Candidate for the "total M- Weil representation"

Recall that p is a prime, q is a power of p, a and b are positive integers,

$$M := \gcd(q^a+1, q^b+1),\ A := (q^a+1)/M,\ B := (q^b+1)/M.$$

Thus $\gcd(A, B) = 1$. We also fix integers α, β with

$$\alpha A - \beta B = 1.$$

We wish to study the direct sum

$$\mathsf{Total}(M, A, B) := \mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1}) \bigoplus_{\sigma \in \mathsf{Char}(M), \sigma \neq \mathbb{1}} \mathcal{K}l(M, A, B, \sigma^{-\beta}, \sigma^{-\alpha}).$$

Theorem 6.1. *The local system* $\mathsf{Total}(M, A, B)$ *is geometrically isomorphic to the arithmetically semisimple local system on* $\mathbb{G}_m/\mathbb{F}_p$ *whose trace function at a point* $v \in E^\times = \mathbb{G}_m(E)$, *for* $E/\mathbb{F}_p$ *a finite extension, is given by*

$$v \mapsto (1/\#E) \sum_{x,w \in E} \psi_E(MABxw - v^{-\alpha}Ax^{q^b+1} - v^{\beta}Bw^{q^a+1}).$$

Subsequently, in §11, we will state and prove a more precise formulation, Theorem 11.4, of this theorem.

7. First steps toward the proof of Theorem 6.1: cancelling

Recall from [K2, 9.3.1], the operation **Cancel** on hypergeometric sheaves

$$\mathcal{H}yp_\psi(\chi_1, \ldots, \chi_n; \rho_1, \ldots, \rho_m),$$

defined whenever the sets, with multiplicity, of the upstairs and downstairs characters are not identical. Suppose that precisely r of the downstairs characters also occur upstairs. Renumber so that $\chi_i = \rho_i$ for $1 \le i \le r$. Then

$$\mathbf{Cancel}\mathcal{H}yp_\psi(\chi_1, \ldots, \chi_n; \rho_1, \ldots, \rho_m) := \mathcal{H}yp_\psi(\chi_{r+1}, \ldots, \chi_n; \rho_{r+1}, \ldots \rho_m),$$

a hypergeometric of type $(n-r, m-r)$ whose upstairs and downstairs characters are disjoint.

The key fact about cancelling is the following theorem, which is proven (but not stated (!)) in [K2, 8.4.7 and 8.4.13].

Theorem 7.1. *Suppose that $\mathcal{H}yp := \mathcal{H}yp_\psi(\chi_1, \ldots, \chi_n; \rho_1, \ldots, \rho_m)$ is a hypergeometric sheaf whose upstairs and downstairs characters are not identical, and which is defined over a finite field $k/\mathbb{F}_p$* (i.e., all the χ_i and ρ_j are of finite order dividing $(\#k) - 1$). *Suppose that precisely r of the downstairs characters also occur upstairs. Then $\mathcal{H}yp$ is lisse on $\mathbb{G}_m/k$, mixed of weight $\le n + m - 1$, and its highest weight quotient* [De, 3.4.1 (ii)] *is* $(\mathbf{Cancel}\mathcal{H}yp)(-r)$, *which is pure of weight $n + m - 1$. More precisely, we have a short exact sequence of lisse sheaves on $\mathbb{G}_m/k$,*

$$0 \to (\text{weight } \le n + m - 2) \to \mathcal{H}yp \to (\mathbf{Cancel}\mathcal{H}yp)(-r) \to 0.$$

The virtue of **Cancel** is that it gives a convenient expression for each of the summands of $\mathsf{Total}(M, A, B)$. We have the following two lemmas, which are immediate from the definitions.

Lemma 7.2. *Suppose that $\rho^A \ne \chi^B$, so that $\mathcal{K}l(M, A, B, \chi, \rho)$ exists. Consider the hypergeometric sheaf*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A, \chi) \sqcup \mathsf{Char}(B, \rho))$$

of type $(MAB, A + B)$. The $\mathcal{K}l(MAB, \chi, \rho)$ is its **Cancel**.

Lemma 7.3. *Consider the hypergeometric sheaf*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A) \sqcup \mathsf{Char}(B))$$

of type $(MAB, A + B)$. Then $\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})$ is its **Cancel**.

8. Computing traces

In this section, we take as ground field

$$E_1 := \mathbb{F}_p(\mu_{MAB}).$$

For each[1] divisor N of MAB, we define the following product of Gauss sums over E_1:

$$\mathsf{Gauss}(N) := \prod_{\chi \in \mathsf{Char}(N)} (-\mathsf{Gauss}(\psi_{E_1}, \chi)).$$

We then define the twisting factor $\mathsf{Gauss}(M, A, B)$ as

$$\mathsf{Gauss}(M, A, B) := \mathsf{Gauss}(MAB)\mathsf{Gauss}(A)\mathsf{Gauss}(B).$$

Theorem 8.1. *Let E/E_1 be a finite extension. The trace function of*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A, \chi) \sqcup \mathsf{Char}(B, \rho)) \otimes \mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)}$$

at a point $v \in E^\times$, is given by

$$\begin{aligned} v \mapsto & (-1)^{MAB-A-B-1} \times \\ & \times \sum_{x, w \in E^\times} \psi_E(MABx - v^{-\alpha} A x^{MB}/w^B - v^\beta B w^A)(\chi^B/\rho^A)(w)(\chi^\alpha/\rho^\beta)(v). \end{aligned}$$

Proof. The idea is to exploit the fact that

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A, \chi) \sqcup \mathsf{Char}(B, \rho))$$

is the multiple ! multiplicative convolution

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \emptyset) \star_{!,\times} \mathcal{H}yp_\psi(\emptyset; \mathsf{Char}(A, \chi)) \star_{!,\times} \mathcal{H}yp_\psi(\emptyset; \mathsf{Char}(B, \rho)).$$

We now make use of the direct image formula of [K1, 5.6.2, first line of proof] and the definition [K2, 8.2.1 (3)] to give simple formulas for the trace functions of each of the three factors.

The trace function of $\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \emptyset) \otimes \mathsf{Gauss}(MAB)^{-\deg(E/E_1)}$ is

$$s \in E^\times \mapsto \sum_{x \in E^\times, x^{MAB} = s} \psi_E(MABx).$$

The trace function of $\mathcal{H}yp_\psi(\emptyset; \mathsf{Char}(A, \chi)) \otimes \mathsf{Gauss}(A)^{-\deg(E/E_1)}$ is

$$t \in E^\times \mapsto \sum_{y \in E^\times, y^A = t} \psi_E(-A/y)\chi(y).$$

[1] If we are in characteristic 2, then both $q^a + 1$ and $q^b + 1$ are odd, hence their gcd $= M$ is also odd, and hence each of M, A, B is odd. So we will not need to "interpret" the quadratic Gauss sum here, because it will not arise.

The trace function of $\mathcal{H}yp_\psi\big(\emptyset; \mathsf{Char}(B,\rho)\big) \otimes \mathsf{Gauss}(B)^{-\deg(E/E_1)}$ is

$$u \in E^\times \mapsto \sum_{z \in E^\times, z^B = u} \psi_E(-B/z)\rho(z).$$

In general, the trace function of the ! multiplicative convolution of two hypergeometrics is **minus** the multiplicative convolution of their trace functions. So the trace function of a triple ! multiplicative convolution of two hypergeometrics is the multiplicative convolution of their trace functions, with no "extra" sign. In particular, the trace function of

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\chi) \sqcup \mathsf{Char}(B,\rho)) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}$$

is the multiplicative convolution of the above three trace functions. Thus it is

$$\begin{aligned} v \in E^\times \mapsto \sum_{\substack{s,t,u \in E^\times, \\ stu = v}} \; \sum_{\substack{x,y,z \in E^\times, \\ x^{MAB} = s, y^A = t, z^B = u}} & \psi_E(MABx - A/y - B/z)\chi(y)\rho(z) \\ = \sum_{x,y,z \in E^\times,\; x^{MAB} y^A z^B = v} & \psi_E(MABx - A/y - B/z)\chi(y)\rho(z). \end{aligned}$$

We now rewrite the range of summation as consisting of those triples $x, y, z \in E^\times$ satisfying

$$(x^{MB}y)^A z^B = v.$$

We then make use of $\alpha A - \beta B = 1$ to write $v = v^{\alpha A - \beta B}$, so the range of summation becomes those $x, y, z \in E^\times$ satisfying

$$(v^{-\alpha} x^{MB} y)^A = (1/(v^\beta z))^B.$$

Because $\gcd(A,B) = 1$, there exists a unique w such that

$$v^{-\alpha} x^{MB} y = w^B, \;\; 1/(v^\beta z) = w^A.$$

Using the first equation, we solve for y in terms of x, w,

$$1/y = v^{-\alpha} x^{MB}/w^B,$$

and using the second equation we solve for z in terms of w,

$$1/z = v^\beta w^A.$$

So the expression for the trace at time $v \in E^\times$ becomes

$$\sum_{x,w\in E^\times} \psi_E(MABx - Av^{-\alpha}x^{MB}/w^B - Bv^\beta w^A)\chi(v^\alpha w^B/x^{MB})\rho(v^{-\beta}/w^A)$$
$$= \sum_{x,w\in E^\times} \psi_E(MABx - Av^{-\alpha}x^{MB}/w^B - Bv^\beta w^A)(\chi^B\rho^{-A})(w)(\chi^\alpha\rho^{-\beta})(v),$$

the last equality because χ has order dividing M, thus $\chi(x^{MB}) = 1$. □

Corollary 8.2. *For $\sigma \in \mathsf{Char}(M)$, the trace function of*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\sigma^{-\beta}) \sqcup \mathsf{Char}(B,\sigma^{-\alpha})) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}$$

at a point $v \in E^\times$, is given by

$$v \mapsto \sum_{x,w\in E^\times} \psi_E(MABx - v^{-\alpha}Ax^{MB}/w^B - v^\beta Bw^A)\sigma(w).$$

Proof. In the case when $(\chi,\rho) = (\sigma^{-\beta},\sigma^{-\alpha})$, we have

$$\chi^B/\rho^A = \sigma^{-\beta B}/\sigma^{-\alpha A} = \sigma^{\alpha A-\beta B} = \sigma, \quad \chi^\alpha = \rho^\beta.$$ □

Lemma 8.3. *Suppose p is odd. Denote by $K/\mathbb{Q}$ the unique quadratic extension of $\mathbb{Q}$ inside $\mathbb{Q}(\zeta_p)$. If $M = 2$, then for each $\sigma \in \mathsf{Char}(M)$, the trace function of*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\chi) \sqcup \mathsf{Char}(B,\rho)) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)},$$

viewed on $\mathbb{G}_m/\mathbb{F}_p$, has values in K.

Proof. Take $\lambda \in \mathbb{F}_p^\times$, and make the substitution $(x,w) \mapsto (\lambda^2 x, \lambda^2 w)$. This does not change the sum

$$\sum_{x,w\in E^\times} \psi_E(MABx - v^{-\alpha}Ax^{MB}/w^B - v^\beta Bw^A)\sigma(w),$$

but it replaces ψ by its $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/K)$ conjugate $x \mapsto \psi(\lambda^2 x)$. Indeed, the term x is multiplied by λ^2, the monomial w^A is multiplied by the factor $\lambda^{2A} = \lambda^{q^a+1} = \lambda^2$, and the monomial x^{MB}/w^B is multiplied by $\lambda^{4B}/\lambda^{2B}$, which by the previous argument is equal to $\lambda^4/\lambda^2 = \lambda^2$. □

Lemma 8.4. *If $M = q+1$, or if $p = 2$, then for each $\sigma \in \mathsf{Char}(M)$, the trace function of*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\chi) \sqcup \mathsf{Char}(B,\rho)) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)},$$

viewed on $\mathbb{G}_m/\mathbb{F}_{q^2}$, has values in $\mathbb{Q}(\sigma)$.

Proof. If $p = 2$, this is obvious, because ψ_E takes values in ± 1.

Suppose now that $M = q+1$. Take $\lambda \in \mathbb{F}_p^\times$, and make the substitution $(x,w) \mapsto (\lambda x, \lambda w)$. This does not change the sum

$$\sum_{x,w \in E^\times} \psi_E(MABx - v^{-\alpha}Ax^{MB}/w^B - v^\beta Bw^A)\sigma(w),$$

but it replaces ψ by its $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/K)$ conjugate $x \mapsto \psi(\lambda x)$. Indeed, x is multiplied by λ, the monomial w^A is multiplied by $\lambda^A = \lambda^{(q^a+1)/(q+1)} = \lambda$ (because the exponent $(q^a+1)/(q+1) \equiv 1 \bmod (q-1)$), and the monomial x^{MB}/w^B is multiplied by $\lambda^{q^b+1}/\lambda^B$, which by the previous argument is equal to $\lambda^2/\lambda = \lambda$. Finally, the term $\sigma(w)$ is moved to $\sigma(\lambda w)$, but this is equal to $\sigma(w)$ because λ (or indeed any element of $\mathbb{F}_q^\times$), is the $q+1$ power of some element of $\mathbb{F}_{q^2}$ (surjectivity of the norm). □

Corollary 8.5. *The trace function of the direct sum $\bigoplus_{\sigma \in \mathsf{Char}(M)}$ of the sheaves*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\sigma^{-\beta}) \sqcup \mathsf{Char}(B,\sigma^{-\alpha})) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}$$

at a point $v \in E^\times$, is given by

$$v \mapsto \sum_{x,w \in E^\times} \psi_E(MABxw - v^{-\alpha}Ax^{q^b+1} - v^\beta Bw^{q^a+1}).$$

Proof. Indeed, the sum of the individual trace functions at a point $v \in E^\times$, is given by

$$\begin{aligned} v \mapsto & \sum_{x,w \in E^\times} \psi_E(MABx - v^{-\alpha}Ax^{MB}/w^B - v^\beta Bw^A) \sum_{\sigma \in \mathsf{Char}(M)} \sigma(w) \\ = & \sum_{x,w \in E^\times} \psi_E(MABx - v^{-\alpha}Ax^{MB}/w^{MB} - v^\beta Bw^{AM}). \end{aligned}$$

Now make the substitution $x \mapsto xw, w \mapsto w$. □

9. Descent results

A reformulation of Theorem 8.1, taking into account [K2, 8.4.6.2], is the following.

Theorem 9.1. *On $(\mathbb{G}_m)^3/E$, with coordinates (v,x,w), consider the lisse sheaf*

$$\mathcal{F}_{\chi,\rho} := \mathcal{L}_{\psi(MABx - v^{-\alpha}Ax^{MB}/w^B - v^{\beta}Bw^A)} \otimes \mathcal{L}_{(\chi^B/\rho^A)(w)} \otimes \mathcal{L}_{(\chi^\alpha/\rho^\beta)(v)}.$$

For the projection

$$\mathrm{pr}_1 : (\mathbb{G}_m)^3/E \mapsto \mathbb{G}_m/E, \quad (v,x,w) \mapsto v,$$

we have $R^i(\mathrm{pr}_1)_!(\mathcal{F}_{\chi,\rho}) = 0$ for $i \neq 2$, $R^2(\mathrm{pr}_1)_!(\mathcal{F}_{\chi,\rho})$ is lisse on $\mathbb{G}_m/E$, mixed of weight ≤ 2, and there is an arithmetic isomorphism between

$$R^2(\mathrm{pr}_1)_!(\mathcal{F}_{\chi,\rho})$$

and

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\chi) \sqcup \mathsf{Char}(B,\rho)) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}.$$

Proof. The situation is that we are given three Kloosterman sheaves $\mathcal{K}l_A$, $\mathcal{K}l_B$, and $\mathcal{K}l_C$ of ranks A, B, C with $A+B<C$, and we form the triple ! multiplicative convolution of $\mathcal{K}l_C$, $\mathrm{inv}^\star\mathcal{K}l_A$, and $\mathrm{inv}^\star\mathcal{K}l_B$. By definition, we first form their external tensor product

$$\mathcal{K}l_C \boxtimes \mathrm{inv}^\star\mathcal{K}l_A \boxtimes \mathrm{inv}^\star\mathcal{K}l_B$$

on $(\mathbb{G}_m)^3$, and then for the multiplication map

$$\mathrm{mult}_3 : (\mathbb{G}_m)^3 \to \mathbb{G}_m, \ (s,t,u) \mapsto stu$$

we form $R(\mathrm{mult}_3)_!(\mathcal{K}l_C \boxtimes \mathrm{inv}^\star\mathcal{K}l_A \boxtimes \mathrm{inv}^\star\mathcal{K}l_B)$. The key fact is that because $A+B<C$, we have $R^i(\mathrm{mult}_3)_! = 0$ for $i \neq 2$, and $R^2(\mathrm{mult}_3)_!$ is lisse, of rank C. To see this, we factor the multiplication map as

$$\mathrm{mult}_3 = \mathrm{mult} \circ (\mathrm{Id} \times \mathrm{mult}_{2,3}), (s,t,u) \mapsto (s,tu) \mapsto stu.$$

Then $R(\mathrm{Id} \times \mathrm{mult}_{2,3})_!(\mathcal{K}l_C \boxtimes \mathrm{inv}^\star\mathcal{K}l_A \boxtimes \mathrm{inv}^\star\mathcal{K}l_B)$ is the external tensor product on $\mathbb{G}_m \times \mathbb{G}_m$ of $\mathcal{K}l_C$ with $R(\mathrm{mult})_!(\mathrm{inv}^\star\mathcal{K}l_A \boxtimes \mathrm{inv}^\star\mathcal{K}l_B)$. The second factor has $R^i(\mathrm{mult})_! = 0$ for $i \neq 1$, and $R(\mathrm{mult})_!(\mathrm{inv}^\star \mathcal{K}l_A \boxtimes \mathrm{inv}^\star \mathcal{K}l_B)$ is $\mathrm{inv}^\star \mathcal{K}l_{A+B}$ for a Kloosterman sheaf of rank $A+B$, cf. [K1, 5.1]. Thus our triple convolution $R(\mathrm{mult}_3)_!$ is

$$R(\mathrm{mult})_!(\mathcal{K}_C \boxtimes \mathrm{inv}^\star \mathcal{K}l_{A+B}).$$

Fibre by fibre over $\mathbb{G}_m$, each stalk is the cohomology of the usual tensor product of $\mathcal{K}l_C$ with a multiplicative translate of $\mathcal{K}l_{A+B}$. The first factor is totally wild of rank C and has all ∞-slopes $1/C$, the second is totally wild of rank $A+B$ and has all ∞-slopes $1/(A+B) > 1/C$. Thus the tensor product has rank $A(B+C)$ with all ∞-slopes $1/(A+B)$. So each such tensor product has $H^i_c = 0$ for $i \neq 1$, and $h^1_c = \mathsf{Swan}_\infty = C$. Thus $R^i(\mathrm{mult})_!(\mathcal{K}_C \boxtimes \mathrm{inv}^\star \mathcal{K}l_{A+B}) = 0$ for $i \neq 1$, and the $R^1(\mathrm{mult})_!$ has constant rank C, hence is lisse because it is a sheaf of perverse origin. Combining these cohomological vanishings, we get the asserted vanishing of $R^i(\mathrm{mult}_3)_! = 0$ for $i \neq 2$, and the fact that $R^2(\mathrm{mult}_3)_!$ is lisse, of rank C.

In the case at hand, it is an exercise, using the explicit descriptions given in Theorem 8.1, of the particular sheaves $\mathcal{K}l_A$, $\mathcal{K}l_B$, and $\mathcal{K}l_C$ in play, to rewrite the $R(\mathrm{mult}_3)_!$ as the $R(\mathrm{pr}_1)$ of the statement of the theorem. □

Corollary 9.2. *Let $E_0 \subset E$ be any subfield over which χ and ρ are defined. Then $\mathcal{F}$ makes sense on $(\mathbb{G}_m)^3/E_0$, and $R^2(\mathrm{pr}_1)_!(\mathcal{F})$ on $\mathbb{G}_m/E_0$ is a lisse sheaf, mixed of weight ≤ 2, which, when pulled back to $\mathbb{G}_m/E_0$, is arithmetically isomorphic to*

$$\mathcal{H}yp_\psi(\mathsf{Char}(MAB); \mathsf{Char}(A,\chi) \sqcup \mathsf{Char}(B,\rho)) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}.$$

Its trace function is that given in Theorem 8.1, now valid on $\mathbb{G}_m/E_0$.

Proof. The trace formula results from the Lefschetz trace formula [Gr1]. □

Corollary 9.3. *In the situation of Corollary 8.2, we have the following results.*

(i) *For $\sigma \in \mathsf{Char}(M)$ nontrivial, taking χ, ρ in Theorem 9.1 to be $\sigma^{-\beta}, \sigma^{-\alpha}$, the weight two quotient* [De, 3.4.1 (ii)] $\mathrm{gr}_{\mathrm{wt}=2}(R^2(\mathrm{pr}_1)_!(\mathcal{F}_{\chi,\rho}))$ *on $\mathbb{G}_m/E_0$ is an arithmetic descent of*

$$\mathcal{K}l(M,A,B,\sigma^{-\beta},\sigma^{-\alpha})(-A-B) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}.$$

(ii) *Taking χ, ρ in Theorem 9.1 to be $\mathbb{1}, \mathbb{1}$, the weight two quotient* [De, 3.4.1 (ii)] $\mathrm{gr}_{\mathrm{wt}=2}(R^2(\mathrm{pr}_1)_!(\mathcal{F}_{\chi,\rho}))$ *on $\mathbb{G}_m/E_0$ is an arithmetic descent of*

$$\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})(-A-B+1) \otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}.$$

10. Interlude: rationality properties of highest weight quotients

In this section, we consider the following general situation. We are given $k/\mathbb{F}_p$ a finite extension, U/k smooth and geometrically connected of some dimension $d \geq 0$, and an integer w. Consider a lisse $\overline{\mathbb{Q}_\ell}$-sheaf $\mathcal{F}$ on U which is mixed of weight $\leq w$. We know [De, 3.4.9] that $\mathcal{F}$ admits a unique "filtration by the weight". In particular, $\mathcal{F}$ sits in a short exact sequence of lisse sheaves on U

$$0 \to \mathcal{F}_{\mathrm{wt}<w} \to \mathcal{F} \to \mathcal{F}_{\mathrm{wt}=w} \to 0,$$

in which $\mathcal{F}_{\mathrm{wt}<w}$ is mixed of weight $< w$ and $\mathcal{F}_{\mathrm{wt}=w}$ is pure of weight w.

Theorem 10.1. *On U/k suppose that $\mathcal{F}$ is a lisse sheaf, mixed of weight $\leq w$. Let $K/\mathbb{Q}$ be a finite extension. Suppose that $\mathcal{F}$ has all traces in K. Then $\mathcal{F}_{\mathrm{wt}=w}$ has all its traces in K.*

Proof. For each finite extension E/k, and each point $u \in U(E)$, the reversed characteristic polynomial

$$\det(1 - TFrob_{u,E}|\mathcal{F})$$

lies in $1 + TK[T]$. When we factor it over $\overline{\mathbb{Q}}$, say

$$\det(1 - TFrob_{u,E}|\mathcal{F}) = \prod_{i=1}^{\mathrm{rank}(\mathcal{F})} (1 - \alpha_i T).$$

After suitable renumbering, we have

$$\det(1 - TFrob_{u,E}|\mathcal{F}) = \left(\prod_{i=1}^{\mathrm{rank}(\mathcal{F}_{\mathrm{wt}=w})} (1 - \beta_i T)\right)\left(\prod_{j=1}^{\mathrm{rank}(\mathcal{F}_{\mathrm{w}<w})} (1 - \gamma_j T)\right),$$

in which each β_i, together will all its $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates, has complex absolute value $(\#E)^{w/2}$, while each γ_j together will all its $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates, has complex absolute value $(\#E)^{v_j/2}$ for some $v_j < w$.

We now exploit this Galois invariance. Because the entire polynomial $\det(1 - TFrob_{u,E}|\mathcal{F})$ is (coefficient-wise) fixed by $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$, it follows that each factor $\prod_{i=1}^{\mathrm{rank}(\mathcal{F}_{\mathrm{wt}=w})}(1 - \beta_i T)$ and $\prod_{j=1}^{\mathrm{rank}(\mathcal{F}_{\mathrm{w}<w})}(1 - \gamma_j T)$ separately has coefficients in K. The first of these factors is precisely

$$\det(1 - TFrob_{u,E}|\mathcal{F}_{\mathrm{wt}=w}).$$

This being true for every E/k and every $u \in U(E)$, $\mathcal{F}_{\mathrm{wt}=w}$ has all its traces in K. □

11. End of the proof of Theorem 6.1

Each of the lisse sheaves

$$\mathcal{H}yp_{\psi}(\mathsf{Char}(MAB); \mathsf{Char}(A, \sigma^{-\beta}) \sqcup \mathsf{Char}(B, \sigma^{-\alpha})) \otimes \mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)}$$

is mixed of weight ≤ 2. Their **Cancel**'s are, arithmetically, the lisse sheaves on $\mathbb{G}_m/E_1$

$$\mathcal{K}l(M, A, B, \sigma^{-\beta}, \sigma^{-\alpha})(-A - B) \otimes \mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)},$$

for σ nontrivial in $\mathsf{Char}(M)$, and

$$\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})(-A - B + 1) \otimes \mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)}.$$

Each of these is pure of weight 2.

Theorem 11.1. *Consider the $M - 1$ lisse sheaves*

$$\mathcal{K}l(M, A, B, \sigma^{-\beta}, \sigma^{-\alpha})(-A - B + 1) \otimes \mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)},$$

for σ nontrivial in $\mathsf{Char}(M)$, and

$$\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})(-A - B + 2) \otimes \mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)},$$

each viewed on $\mathbb{G}_m/\mathbb{F}_p(\mu_M)$. Then we have the following results.

(i) *If $M = 2$ (which forces p to be odd), each sheaf has all its traces in K, the unique quadratic extension of $\mathbb{Q}$ inside $\mathbb{Q}(\zeta_p)$.*
(ii) *If $M = q + 1$, each sheaf indexed by σ nontrivial in $\mathsf{Char}(M)$ has all its traces in $\mathbb{Q}(\sigma)$. The remaining one has traces in $\mathbb{Q}$.*
(iii) *Each of these M sheaves has finite arithmetic and geometric monodromy groups.*
(iv) *In both cases, each of the above M sheaves satisfies ($\mathbf{S+}$); moreover, if $n = a + b \geq 3$ then each has an arithmetic and geometric monodromy group which is almost quasisimple.*

Proof. (i) and (ii) are immediate on combining Lemmas 8.3 and 8.4 with Theorem 10.1, and (iii) was proven in Theorems 4.7 and 5.1.

Next, let $\mathcal{H}$ be any of the above M sheaves and H be the arithmetic or geometric monodromy group of $\mathcal{H}$. Then the first statement in (iv) is already proved in Corollaries 4.12 and 5.4. Being geometric, it applies to $G_{\rm geom}$. But as noted in [KT7, Lemma 1.3], it applies a fortiori to the larger $G_{\rm arith}$ as well. This implies by [GT, Proposition 2.8] that, if H is not almost quasisimple, then $\mathrm{rank}(\mathcal{H}) = r^m$ for some prime r and H contains an extraspecial r-group R of order r^{2m+1} that acts irreducibly on $\mathcal{H}$. Suppose we are in the latter case. Then a generator z of $\mathbf{Z}(R)$ acts on $\mathcal{H}$ via multiplication by ζ_r.

In the case $M = 2$, q is necessarily odd (otherwise q^a+1, q^b+1 are both odd) we have $r^m = \mathrm{rank}(\mathcal{H}) = (q^n \pm 1)/2$. As $n \geq 3$ and $2 \nmid q$, $(q^n \pm 1)/2$ is divisible by some primitive prime divisor $\ell > 2$ by [Zs], whence $r = \ell > 2$. On the other hand, $\zeta_r \in \mathbb{Q}(\zeta_p)$ by (i) and so $r = p$, a contradiction.

Assume now that $M = q+1$, whence a, b are odd and $n = a+b \geq 4$. If $\mathcal{H}$ is the hypergeometric one, then $\mathrm{rank}(\mathcal{H}) = q(q^{n-1}+1)/(q+1)$ is the product of two relatively prime integers, each ≥ 2, and so it cannot be equal to r^m. Hence $\mathcal{H}$ is Kloosterman, and $\mathrm{rank}(\mathcal{H}) = (q^n-1)/(q+1) = r^m$; in particular, $(n,q) \neq (6,2)$. As $n \geq 4$, this forces again by [Zs] that r is a primitive prime divisor of $q^n - 1$, in particular, $r \neq 2$ and $r \nmid (q+1)$ (otherwise r would divide q^2-1). This implies $\zeta_r \notin \mathbb{Q}(\sigma)$, contradicting (ii). □

Remark 11.2. Suppose we are in the situation $p = 2, M = 1$. Then there is only the hypergeometric sheaf, and its rank is q^n. So we cannot conclude its arithmetic and geometric monodromy groups are almost quasisimple in this case.

The direct sum of our M sheaves is the weight 2 part of the local system whose trace function at a point $v \in E^\times$, is given by

$$v \mapsto \sum_{x,w \in E^\times} \psi_E(MABxw - v^{-\alpha}Ax^{q^b+1} - v^{\beta}Bw^{q^a+1}).$$

If we replace ψ by $t \mapsto \psi(t/MAB)$, this trace becomes

$$v \mapsto \sum_{x,w \in E^\times} \psi_E(xw - v^{-\alpha}x^{q^b+1} - v^{\beta}w^{q^a+1}),$$

simply because $MAB = A = B$ in $\mathbb{F}_p$.

Let us admit momentarily the truth of the following theorem.

Theorem 11.3. *Fix integers $d \geq 3, d \geq 2$, both of which are prime to p. Consider the parameter space $S/\mathbb{F}_p$ of pairs of one-variable polynomials*

(f_d, g_e) of degrees d and e respectively. We may view S as the space

$$\mathbb{G}_m \times (\mathbb{A}^1)^d \times \ \mathbb{G}_m \times (\mathbb{A}^1)^e$$

of coefficients of f and g. On $(\mathbb{A}^2)_S$, with "coordinates" (x, w, f, g), we have the Artin–Schreier sheaf

$$\mathcal{L}_{\psi(f(x)+g(w)+xw)}.$$

Denote by $\pi : (\mathbb{A}^2)_S \to S, \quad (x, w, f, g) \mapsto (f, g)$ the projection onto S. The higher direct images $R^i\pi_!(\mathcal{L}_{\psi(f(x)+g(w)+xw)})$ vanish for $i \neq 2$, and

$$R^2\pi_!(\mathcal{L}_{\psi(f(x)+g(w)+xw)})$$

is lisse of rank $(d-1)(e-1)$ and pure of weight 2.

We apply this with $d = q^b+1, e = q^a+1$, and with $\mathbb{G}_m$ embedded into S by $(-v^{-\alpha}x^{q^b+1}, -v^{\beta}w^{q^a+1})$. Then we find that on $\mathbb{G}_m$, there is a lisse local system of rank q^{a+b} which is pure of weight 2, whose trace function is

$$v \mapsto \sum_{x,w \in E} \psi_E(xw - v^{-\alpha}x^{q^b+1} - v^{\beta}w^{q^a+1}).$$

This is the weight 2 part of the local system given by the same formula, but with x, w both restricted to lying in $\mathbb{G}_m$. Indeed, the difference is the sum of the terms with $x = 0$, which is a one-variable sum over w which is pure of weight 1, the sum of the terms with $w = 0$, which is a one-variable sum over x which is pure of weight 1, minus the single term with $x = w = 0$, which is pure of weight 0.

Here is a more geometric way of saying this. Consider the universal situation: we have the inclusion of the open set

$$j : U := (\mathbb{G}_m \times \mathbb{G}_m)_S \subset (\mathbb{A}^2)_S,$$

with complement Z_S, Z being the locus $xw = 0$ in $\mathbb{A}^2$. Denoting by π_U and π_Z the projections onto S, a piece of the long exact excision sequence is

$$\begin{aligned} R^1(\pi_Z)_!(\mathcal{L}_{\psi(f(x)+g(y)+xy)}) &\to R^2(\pi_U)_!(\mathcal{L}_{\psi(f(x)+g(y)+xy)}) \\ &\to R^2(\pi)_!(\mathcal{L}_{\psi(f(x)+g(y)+xy)}) \to 0, \end{aligned}$$

the final 0 being $R^2(\pi_Z)_!(\mathcal{L}_{\psi(f(x)+g(y)+xy)})$, which vanishes fibre by fibre. By Deligne's main theorem [De, 3.3.1], the first term is mixed of weight ≤ 1, and the second term is mixed of weight ≤ 2. Therefore the third term is indeed the weight 2 quotient of the second term.

It remains only to prove Theorem 11.3.

Proof. To show that the $R^i\pi_!(\mathcal{L}_{\psi(f(x)+g(w)+xw)})$ vanish for $i \neq 2$, it suffices, thanks to proper base change, to do so point by point. To show that the $R^2\pi_!(\mathcal{L}_{\psi(f(x)+g(w)+xw)})$ is lisse of rank $(d-1)(e-1)$, we use the fact that it is a "sheaf of perverse origin", so it suffices to show that at each point the stalk has constant rank $(d-1)(e-1)$. Once we know the $R^2\pi_!$ is lisse, to show it pure of weight 2, it suffices to show punctual purity of weight 2. So what must be shown is that for any two polynomials f, g of degrees d, e respectively over some finite field $E/\mathbb{F}_p$, the cohomology groups

$$H^i_c(\mathbb{A}^2/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(f(x)+g(w)+xw)})$$

vanish for $i \neq 2$, and the H^2_c has dimension $(d-1)(e-1)$ and is pure of weight 2.

Write the sum

$$\sum_{x,w} \psi(f(x) + g(w) + xw)$$

as

$$\sum_{w} \psi(g(w)) FT_\psi(\mathcal{L}_{\psi(f(x)})(w),$$

and view it as the trace of Frobenius on $H^1_c((\mathbb{A}^1/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(g(w)} \otimes FT_\psi(\mathcal{L}_{\psi(f(x)})$ to see its asserted purity.

More precisely, apply the Leray spectral sequence for $\mathrm{pr}_2 : (x, w) \mapsto w$. Then by the projection formula we have

$$R^i(\mathrm{pr}_2)_!(\mathcal{L}_{\psi(f(x)+g(w)+xw)}) = \mathcal{L}_{\psi(g(w))} \otimes R^i(\mathrm{pr}_2)_!(\mathcal{L}_\psi(f(x) + xw)).$$

The second tensor factor vanishes for $i \neq 1$, and for $i = 1$ it is the Fourier Transform $FT_\psi(\mathcal{L}_{\psi(f(x))})$. Thus we have

$$H^i_c(\mathbb{A}^2/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(f(x)+g(w)+xw)}) = H^{i-1}_c(\mathbb{A}^1/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(g(w)} \otimes FT_\psi(\mathcal{L}_{\psi(f(x))}).$$

One knows [K3, Theorem 17] that $FT_\psi(\mathcal{L}_{\psi(f(x)})$ is lisse on the $\mathbb{A}^1$ of w, of rank $d-1$, with all its ∞-slopes $d/(d-1)$. Because $d \geq 3$, these ∞-slopes are < 2. But $\mathcal{L}_{\psi(g(w)}$ has ∞-slope $e \geq 2$, and thus $\mathcal{L}_{\psi(g(w)} \otimes FT_\psi(\mathcal{L}_{\psi(f(x)})$ is lisse of rank $d-1$ with all ∞-slopes e. The total wildness gives the vanishing of this last cohomology group except possibly in degree one. The Euler–Poincaré formula then shows the dimension is the asserted $(d-1)(e-1)$.

The total wildness, together with Deligne's main theorem [De, 3.2.3] gives the purity of weight 2. □

Theorem 11.4 (Theorem 6.1 made precise)**.** *Over the field $E_1 := \mathbb{F}_p(\mu_{MAB})$, the direct sum*

$$\mathcal{H}yp(M, A, B, \mathbb{1}, \mathbb{1})(-A-B+2) \bigoplus_{\mathbb{1}\neq\sigma\in\mathsf{Char}(M)} \mathcal{K}l(M, A, B, \sigma^{-\beta}, \sigma^{-\alpha})(-A-B+1),$$

when further twisted by $\mathsf{Gauss}(M, A, B)^{-\deg(E/E_1)}$, is pure of weight zero and its trace function at $v \in E^\times$, E/E_1 a finite extension, is given by

$$v \mapsto (1/\#E) \sum_{x,w\in E} \psi_E(MABxw - v^{-\alpha}Ax^{q^b+1} - v^\beta Bw^{q^a+1}).$$

Definition 11.5. Fix $\alpha, \beta \in \mathbb{Z}$ such that $\alpha A - \beta B = 1$. Let us denote by

$$\mathcal{W}(M, A, B)$$

the arithmetically semisimple local system on $\mathbb{G}_m/\mathbb{F}_p$ whose trace function at $v \in E^\times$, $E/\mathbb{F}_p$ a finite extension, is given by

$$v \mapsto \frac{1}{\#E} \sum_{x,w\in E} \psi_E\big(MABxw - v^{-\alpha}Ax^{q^b+1} - v^\beta Bw^{q^a+1}\big).$$

The pullback of $\mathcal{W}(M, A, B)$ to $\mathbb{G}_m/E_1$ remains arithmetically semisimple (because $\pi_1(\mathbb{G}_m/E_1)$ is a subgroup of finite index in $\pi_1(\mathbb{G}_m/\mathbb{F}_p)$), so this pullback is the above direct sum (as both are arithmetically semisimple and have the same trace functions).

Let us recall the underlying finiteness theorem, cf. [KRLT1, Proposition 2.1 and Remark 2.2].

Theorem 11.6. *Let k be a finite field of characteristic $p > 0$, U/k a smooth, geometrically connected k-scheme, $\ell \neq p$, and $\mathcal{G}$ an arithmetically semisimple $\overline{\mathbb{Q}_\ell}$-local system on U which is pure of weight 0 (for all embeddings of $\overline{\mathbb{Q}_\ell}$ into $\mathbb{C}$). Then G_{arith} is finite if and only if all traces of $\mathcal{G}$ are algebraic integers.*

To show the integrality of traces of $\mathcal{W}(M, A, B)$, we can apply the van der Geer–van der Vlugt argument, cf. [KT2, Section 5] which uses [vG-vV], to show that $\mathcal{W}(M, A, B)$ has finite G_{arith}. The key point is that for any finite extension $E/\mathbb{F}_p$ and any $v \in E^\times$, the $\mathbb{F}_p$-valued function on $E \times E$ given by

$$F(x,w) := \mathrm{Trace}_{E/\mathbb{F}_p}\big(MABxw - v^{-\alpha}Ax^{q^b+1} - v^{\beta}Bw^{q^a+1}\big)$$

is a quadratic form on $E \times E$ viewed as an $\mathbb{F}_p$ vector space, with associated bilinear form

$$\langle (x,w),(X,W)\rangle := F(x+X, w+W) - F(x,w) - F(X,W).$$

The resulting finiteness of G_{arith} for $\mathcal{W}(M,A,B)$ gives another proof of Theorems 4.7 and 5.1.

Let us prove now some basic rationality results.

Theorem 11.7. *We have the following results.*

(i) *If p is odd, then for any finite extension $E/\mathbb{F}_p$, and any $v \in E^\times$, $\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))$ lies in the ring of integers of the subfield of $\mathbb{Q}(\zeta_p)$ fixed by the subgroup of squares in $\mathbb{F}_p^\times$. If q is even, all these traces lie in $\mathbb{Z}$.*

(i-bis) *For any prime p, for any finite extension $E/\mathbb{F}_{p^2}$, and for any $v \in E^\times$, $\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))$ lies in $\mathbb{Z}$.*

(i-ter) *If q is a square, then for any finite extension $E/\mathbb{F}_q$ and any $v \in E^\times$, $\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))$ lies in $\mathbb{Z}$.*

(ii) *If ab is odd, then for any finite extension $E/\mathbb{F}_{q^2}$, and any $v \in E^\times$, $\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))$ lies in $\mathbb{Z}$.*

Proof. The first assertion is that for any $t \in \mathbb{F}_p^\times$, if we replace ψ by the character $\psi_{t^2} : x \mapsto \psi(t^2x)$, the trace does not change. This is obvious, by the substitution $(x,w) \mapsto (tx,tw)$. If q is even, then ψ takes values in ±1, so the traces lie in $\mathbb{Q}$, and are integral, so lie in $\mathbb{Z}$.

For (i-bis), notice that any $t \in \mathbb{F}_p^\times$ becomes a square in $\mathbb{F}_{p^2}$, say $t = s^2$ with $s \in \mathbb{F}_{p^2}^\times$. Then the substitution $(x,w) \mapsto (sx,sw)$ gives the invariance of the sum under the entire group $\mathbb{F}_p^\times$.

Statements (i-ter) and (ii) result trivially from (i-bis). □

The arguments of van der Geer–van der Vlugt lead to the following theorem.

Theorem 11.8. *We have the following results.*

(i) *If q is odd, then for any finite extension $E/\mathbb{F}_q$, and any $v \in E^\times$,*

$$|\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))|^2$$

is a power q^m of q with $0 \le m \le 2a+2b$. If q is even, the values are either a power q^m, $0 \le m \le 2a+2b$, or 0.

(i-bis) *For any subfield $k \subseteq \mathbb{F}_q$, and any $v \in k^\times$,*

$$|\mathrm{Trace}(Frob_{v,k}|\mathcal{W}(M,A,B))|^2$$

is a power of $\#k$ (or possibly 0 *if q is even).*

(i-ter) *If a, b are both odd, then for any subfield $k \subseteq \mathbb{F}_{q^2}$, and any $v \in k^\times$,*

$$|\mathrm{Trace}(Frob_{v,k}|\mathcal{W}(M,A,B))|^2$$

is either 1 *or $\#k$ (or possibly* 0 *if q is even).*

(ii) *If $a+b$ is even and q is odd, then for any finite extension $E/\mathbb{F}_{q^2}$, and any $v \in E^\times$,* $\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))$ *is $\pm q^m$, with $0 \le m \le a+b$. If $a+b$ is even and q is even, the values are either $\pm q^m$, $0 \le m \le a+b$, or* 0.

Proof. (a) Let $E/\mathbb{F}_p$ be a finite extension. Fix $v \in E^\times$. Denote

$$F(x,w) := \mathrm{Trace}_{E/\mathbb{F}_p}(MABxw - v^{-\alpha}Ax^{q^b+1} - v^{\beta}Bw^{q^a+1}).$$

Then

$$\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B)) = (1/\#E)\sum_{(x,w)\in E\times E}\psi(F(x,w)), \tag{11.8.1}$$

hence

$$|\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M,A,B))|^2 = \frac{1}{\#(E\times E)}\sum_{\substack{(x,w)\in E\times E\\(X,W)\in E\times E}}\psi(F(x,w)-F(X,W)).$$

With the substitution $(x,w)\mapsto(x+X, w+W), (X,W)\mapsto(X,W)$, the above sum becomes

$$(1/\#(E\times E))\sum_{(x,w)\in E\times E,(X,W)\in E\times E}\psi\big(\langle(x,w),(X,W)\rangle\big)\psi(F(x,w))$$
$$= \sum_{(x,w)\in E\times E}\bigg(\psi(F(x,w))\cdot\frac{1}{\#(E\times E)}\sum_{(X,W)\in E\times E}\psi\big(\langle(x,w),(X,W)\rangle\big)\bigg).$$

The inner sum

$$(1/\#(E\times E))\sum_{(X,W)\in E\times E}\psi\big(\langle(x,w),(X,W)\rangle\big)$$

vanishes unless (x, w) is orthogonal to every element of $E \times E$, in which case this inner sum is 1. Let us denote by Null(E) this null space. So we have

$$|\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M, A, B))|^2 = \sum_{(x,w)\in \mathrm{Null}(E)} \psi(F(x, w)).$$

Note that, if q is odd then $F(x, w)$ vanishes on the null space, simply because $F(x, w) = (1/2)\langle(x, w), (x, w)\rangle$. So we get

$$|\mathrm{Trace}(Frob_{v,E}|\mathcal{W}(M, A, B))|^2 = \#\mathrm{Null}(E), \quad q \text{ odd}.$$

If q is even, then $F(x, w)$ is an additive function on the null space. If this function is identically zero on Null(E), we again get #Null(E). If it is nonzero, then we are summing a nontrivial character over the null space, and we get 0.

(b) Now let us write down explicitly the null space. The null space does not change if we replace $F(x, w)$ by a nonzero $\mathbb{F}_p$-multiple. Using M as the multiple, we deal instead with

$$F(x, w) := \mathrm{Trace}_{E/\mathbb{F}_p}(xw - Mv^{-\alpha}x^{q^b+1} - Mv^{\beta}w^{q^a+1}).$$

Let us consider the slightly more general case of

$$F_{s,t}(x, w) := \mathrm{Trace}_{E/\mathbb{F}_p}(xw - sx^{q^b+1} - tw^{q^a+1}), \tag{11.8.2}$$

with both $s, t \in E^{\times}$. Then the associated bilinear form $\langle(x, w), (X, W)\rangle$ is the $\mathrm{Trace}_{E/\mathbb{F}_p}$ of

$$\begin{aligned}&(x + X)(w + W) - s(x + X)^{q^b+1} - t(w + W)^{q^a+1}\\ &- (xw - sx^{q^b+1} - tw^{q^a+1}) - (XW - sX^{q^b+1} - tW^{q^a+1})\\ =\ & xW + wX - sxX^{q^b} - sx^{q^b}X - twW^{q^a} - tw^{q^a}W,\end{aligned}$$

which has the same $\mathrm{Trace}_{E/\mathbb{F}_p}$ as

$$\begin{aligned}&xW + wX - (sx)^{1/q^b}X - sx^{q^b}X - (tw)^{1/q^a}W - tw^{q^a}W\\ =\ & (x - (tw)^{1/q^a} - tw^{q^a})W + (w - (sx)^{1/q^b} - sx^{q^b})X.\end{aligned}$$

Thus (x, w) lies in the null space if and only if (x, w) satisfies the two equations

$$x = (tw)^{1/q^a} + tw^{q^a}, \ w = (sx)^{1/q^b} + sx^{q^b}. \tag{11.8.3}$$

From this description of the null space, when $E \supseteq \mathbb{F}_q$ we see that it is an $\mathbb{F}_q$ vector space, with $(x, w) \mapsto (\lambda x, \lambda w)$ as the scalar multiplication by $\lambda \in \mathbb{F}_q$. When $E \subseteq \mathbb{F}_q$, it is a vector space over E. Moreover, if ab is odd and $E \supseteq \mathbb{F}_{q^2}$, then it is an $\mathbb{F}_{q^2}$ vector space, with $(x, w) \mapsto (\lambda x, \lambda^q w)$ as the scalar multiplication by $\lambda \in \mathbb{F}_{q^2}$. If $a+b$ is even and both a, b are even, then we are in the situation for $(a/2, b/2)$ and $q_0 := q^2$, and the null space is an $\mathbb{F}_{q_0}$-vector space, i.e., an $\mathbb{F}_{q^2}$ vector space. The cardinality of the null space, being the square absolute value of a Frobenius trace, is at most the square of the rank q^{a+b} of $\mathcal{W}(M, A, B)$, simply because $\mathcal{W}(M, A, B)$ has finite G_{arith}. Thus the null space has $\mathbb{F}_q$ dimension at most $2a + 2b$.

To get statement (i-ter), notice that because a, b are odd, for any element $z \in E \subset \mathbb{F}_{q^2}$, we have

$$z^{q^a} = z^q = z^{1/q} = z^{1/q^a} \text{ and } z^{q^b} = z^q = z^{1/q} = z^{1/q^b}.$$

Also, because $z = z^{q^2}$ we have $z + z^q = (z + z^q)^q$. Thus when $E \subset \mathbb{F}_{q^2}$, the equations for $\text{Null}(E)$ are

$$x = (tw)^q + tw^q, w = (sx)^q + sx^q, \text{ i.e., } x = (t + t^q)w^q, w = (s + s^q)x^q.$$

So if (x, w) is in the Null space, then

$$x = (t + t^q)(s + s^q)^q x^{q^2} = (t + t^q)(s + s^q)x,$$

and for such an x, the pair $(x, w := (s + s^q)x^q)$ satisfies $x = (t + t^q)w^q$, simply because $w^q = (s + s^q)^q x^{q^2} = (s + s^q)x$.

If $(t+t^q)(s+s^q) = 1$, then the Null space is isomorphic to E by projection onto its x coordinate, so has cardinality $\#E$. If $(t+t^q)(s+s^q) \neq 1$, then the Null space is just the single point $(0, 0)$ of cardinality 1.

To get statement (ii), we need only observe that when $a + b$ is even, then $\text{Trace}(Frob_{v,E}|\mathcal{W}(M, A, B))$ is an integer. When ab is odd, this is (ii) of Theorem 11.7. When a and b are both even, then we are in the situation for $(a/2, b/2)$ and q^2, and we apply (i-bis) of Theorem 11.7. □

In the case $2 \nmid ab$, we can further strengthen Theorem 11.8:

Theorem 11.9. *Suppose ab is odd. Then for any finite extension E of $\mathbb{F}_{q^2}$ and for any $v \in E^\times$, $\text{Trace}\big(Frob_{v,E}|\mathcal{W}(M, A, B)\big)$ is $(-q)^m$ for some m with $0 \leq m \leq a + b$.*

Proof. By Theorem 11.7(ii) and Theorem 11.8(ii), we have that

$$\varphi(v) := \text{Trace}\big(Frob_{v,E}|\mathcal{W}(M, A, B)\big)$$

is $\pm q^m$ for $0 \leq m = m(v) \leq a + b$ or 0. To prove that it is actually some $(-q)^m$, it suffices to show that $\varphi(v) \equiv 1(\bmod\,(q+1))$. To do this, we use the bijective map $(x, w) \mapsto (\varrho x, \varrho^{-1} w)$ on $E \times E \smallsetminus \{(0,0)\}$ for a fixed $\varrho \in \mathbb{F}_{q^2}^\times$ of order $q+1$; in fact, any orbit under this map on this set has length $q+1$. Note that, because both a, b are odd, we have

$$F(x,w) = F(\varrho x, \varrho^{-1} w), \quad F(0,0) = 0.$$

Thus with $\#E = q^{2d}$, (11.8.1) implies that

$$q^{2d} \varphi(v) = 1 + (q+1)\alpha$$

for some algebraic integer $\alpha \in \mathbb{Z}[\zeta_p]$. Now $\alpha = (q^{2d}\varphi(v) - 1)/(q+1)$ is rational, whence $\alpha \in \mathbb{Z}$. But $q^{2d} \equiv 1(\text{mod } (q+1))$, hence

$$\varphi(v) \equiv q^{2d} \varphi(v) \equiv 1(\text{mod } (q+1)),$$

as desired. □

Whatever the parity of a, b, we have the following strengthening of Theorem 11.8; see also Remark 15.8:

Theorem 11.10. *Let q be a power of an odd prime p, $E/\mathbb{F}_q$ a finite extension, and $f(x), g(x) \in E[x]$ polynomials of the form*

$$f(x) = \sum_{i=0}^{n} a_i x^{q^i+1}, \quad g(x) := \sum_{i=0}^{m} b_i x^{q^i+1},$$

with n, m strictly positive integers, and a_n, b_m nonzero. Denote by Gauss_E the quadratic Gauss sum

$$\mathsf{Gauss}_E := \mathsf{Gauss}(\psi_E, \chi_2).$$

Then we have the following results.

(i) *The sum*

$$S_f := (1/\mathsf{Gauss}_E) \sum_{x \in E} \psi_E(f(x))$$

is equal to $\pm(\mathsf{Gauss}_{\mathbb{F}_q})^d$ for some integer d with $0 \leq d \leq 2n$.

(ii) *For any* $t \in E$*, and with*

$$F(x,y) := txy + f(x) + g(y),$$

the sum

$$S_F := (1/\#E) \sum_{x,y \in E} \psi_E(F(x,y))$$

is equal to $\pm(\mathsf{Gauss}_{\mathbb{F}_q})^d$ *for some integer* d *with* $0 \le d \le 2(n+m)$*.*

In particular, S_f^2, S_F^2 *are each nonzero integers, which are* $\pm$ *powers of* q*.*

Proof. We begin with (i). We have the $\mathbb{F}_p$-valued symmetric bilinear form $(x.y)_f$ on $E \times E$ given by

$$(x,y)_f := f(x+y) - f(x) - f(y).$$

Its null space $\mathrm{Null}_f(E)$, the set of $x \in E$ such that $(x,y)_f = 0$ for all $y \in E$, is an $\mathbb{F}_q$ vector space of dimension $\le 2n$, defined by the equation

$$f(x) + \sum_{i=0}^{n} (a_i x)^{1/q^i} = 0.$$

The van der Geer–van der Vlugt argument gives

$$|S_f|^2 = \#\mathrm{Null}_f(E) = q^{\dim_{\mathbb{F}_q}(\mathrm{Null}_f(E))}.$$

It will be more convenient to work with the “non-normalized” sum

$$S_{0,f} := \mathsf{Gauss}_E \times S_f = \sum_{x \in E} \psi_E(f(x)).$$

Indeed, as $(-\mathsf{Gauss}_E) = (-\mathsf{Gauss}_{\mathbb{F}_q})^{\deg(E/\mathbb{F}_q)}$, it suffices to prove that $S_{0,f}$ is $\pm$ a power of $\mathsf{Gauss}_{\mathbb{F}_q}$.

Let us denote by $S_{0,f}(-)$ the complex conjugate sum

$$S_{0,f}(-) := \sum_{x \in E} \psi_E(-f(x)).$$

Suppose first that q is 1 mod 4. Then $i \in \mathbb{F}_q$, and $f(ix) = -f(x)$. So in this case $S_{0,f}(-) = S_{0,f}$, by the substitution $x \mapsto ix$, and hence

$$|S_{0,f}|^2 = S_{0,f} S_{0,f}(-) = S_{0,f}^2,$$

proving that $S_{0,f}^2$ is a nonnegative power of q, and hence a power of $\mathsf{Gauss}_{\mathbb{F}_q}$.

Suppose next that q is 3 mod 4. Then the sum $S_{0,f}$ lies in the field $\mathbb{Q}(\mathsf{Gauss}_{\mathbb{F}_q}) = \mathbb{Q}(\mathsf{Gauss}_{\mathbb{F}_p}) = \mathbb{Q}(\sqrt{-p})$ (because q is an odd power of p and p is 3 mod 4). We claim that the ratio $S_{0,f}/S_{0,f}(-)$ is a unit in the ring of integers of $\mathbb{Q}(\sqrt{-p})$. From the equality

$$S_{0,f}S_{0,f}(-) = \text{a power of } q$$

we see that $S_{0,f}$ and $S_{0,f}(-)$ are units at all finite places of residue characteristic other than p. As they are Galois conjugate in $\mathbb{Q}(\sqrt{-p}) \subset \mathbb{Q}(\zeta_p)$, which has a unique place over p, $S_{0,f}$ and $S_{0,f}(-)$ have the same p-adic ord at this place. Being complex conjugates, they have the same absolute value at the unique archimedean place. Therefore their ratio is a unit.

If $p \neq 3$, the only units in the ring of integers of $\mathbb{Q}(\sqrt{-p})$ are ± 1.

If $p = 3$, then $\mathbb{Q}(\sqrt{-p}) = \mathbb{Q}(\zeta_3)$, and the units are now the sixth roots of unity. However, we observe that because each exponent $q^i + 1$ is even, $f(-x) = f(x)$, so we have

$$S_{0,f} \in 1 + 2\mathbb{Z}[\zeta_p].$$

To see this, choose a subset $V \subset E^\times$ of representatives of the quotient group $E^\times/(\pm 1)$ and writing

$$S_{0,f}(x) = 1 + \sum_{x \in V}(\psi_E(f(x)) + \psi_E(f(-x))) = 1 + 2\sum_{x \in V}\psi(f(x)).$$

Similarly, $S_{0,f}(-) \in 1 + 2\mathbb{Z}[\zeta_p]$. Thus for some unit u in the ring of integers of $\mathbb{Q}(\sqrt{-p})$, we have $S_{0,f} = uS_{0,f}(-)$. Reducing mod the ideal (2) in $\mathbb{Z}[\zeta_p]$, we see that the unit u must lie in $1 + 2\mathbb{Z}[\zeta_p]$. Among the sixth roots of unity, only ± 1 lie in $1 + 2\mathbb{Z}[\zeta_3]$. Indeed, if u has order 3, then $(u-1)/2$ would lie in $\mathbb{Z}[\zeta_3]$, which is nonsense because its norm down to $\mathbb{Q}$ is $3/4$. And if u_6 has order 6, then $u_6 = -u$ for some u of order 3, so $(-u-1)/2$ would lie in $\mathbb{Z}[\zeta_3]$, again nonsense because its norm down to $\mathbb{Q}$ is $1/4$.

So in all cases when q is 3 mod 4, we have $S_{0,f} = \pm S_{0,f}(-)$. If we have $S_{0,f} = S_{0,f}(-)$, then just as in the case when q is 1 mod 4, we have

$$|S_{0,f}|^2 = S_{0,f}S_{0,f}(-) = S_{0,f}^2,$$

proving that $S_{0,f}^2$ is a nonnegative power of q, and hence $\pm S_{0,f}$ is a power of $\mathsf{Gauss}_{\mathbb{F}_q}$. Suppose now that $S_{0,f} = -S_{0,f}(-)$ is purely imaginary. Then if we write $S_{0,f} = A + B(1+\sqrt{-p})/2$ with $A, B \in \mathbb{Z}$, we have $A + B/2 = 0$. Thus $B = -2A$, and $S_{0,f} = A + B(1+\sqrt{-p})/2 = A - A(1+\sqrt{-p}) = -A\sqrt{-p}$. This already shows that the square of $S_{0,f}$ is an integer, namely $-pA^2$. But this square has absolute value a power of q, so A^2 is itself a nonnegative power of p, hence A is a nonnegative power of p, so $\pm A$ is a power of $\mathsf{Gauss}_{\mathbb{F}_p}$, and hence $\pm S_{0,f}$ is a power of $\mathsf{Gauss}_{\mathbb{F}_p}$, say $\pm S_{0,f} = (\mathsf{Gauss}_{\mathbb{F}_p})^d$. Then we recover d as the log to the base $\sqrt{p}$ of $|S_{0,f}|$. But in absolute value, $|S_{0,f}|^2$ is a power of q, so $\pm|S_{0,f}|$ is a power of $\mathsf{Gauss}_{\mathbb{F}_q} = \pm(\mathsf{Gauss}_{\mathbb{F}_p})^{\deg(\mathbb{F}_q/\mathbb{F}_p)}$. Comparing absolute values, we see that d is a multiple of $\deg(\mathbb{F}_q/\mathbb{F}_p)$. Thus $\pm S_{0,f}$ is a power of $\mathsf{Gauss}_{\mathbb{F}_q}$, as asserted.

For (ii), we argue as follows. Suppose first that $t = 0$, so that

$$F(x,y) = f(x) + g(y).$$

Then $S_F = \chi_{2,E}(-1)S_f S_g$, and the assertion is immediate from (i), applied to f and to g. If $t \neq 0$, the change of variable $(x,y) \mapsto (x/t, y)$ reduces us to the case when $t = 1$ (with f replaced by $f(x/t)$). We have the $\mathbb{F}_p$-valued symmetric bilinear form $((x.y),(X,Y))_F$ on $E^2 \times E^2$ given by

$$((x,y),(X,Y))_F := F((x,y)+(X,Y)) - F(x,y) - f(X,Y).$$

Its null space $\mathrm{Null}_F(E^2)$, the set of $(x,y) \in E^2$ with $((x,y),(X,Y))_F = 0$ for all $(X,Y) \in E^2$, is an $\mathbb{F}_q$ vector space of dimension $\leq 2n+2m$, defined by the two equations

$$y + \sum_{i=0}^{n} (a_i x)^{1/q^i} = 0, \quad x + \sum_{i=0}^{m} (b_i y)^{1/q^i} = 0.$$

The van der Geer–van der Vlugt argument gives

$$|S_F|^2 = \#\mathrm{Null}_F(E^2) = q^{\dim_{\mathbb{F}_q}(\mathrm{Null}_F(E^2))}.$$

We now proceed exactly as in the proof of (i). We consider instead the "non-normalized" sum

$$S_{0,F} := (\#E) \times S_F = \sum_{x,y \in E} \psi_E(F(x,y)),$$

and its complex conjugate

$$S_{0,F}(-) := (\#E) \times S_F = \sum_{x,y \in E} \psi_E(-F(x,y)).$$

When q is 1 mod 4, the substitution $(x, y) \mapsto (ix, iy)$ carries $F(x, y)$ to $-F(x, y)$, and so $S_{0,F}(-) = S_{0,F}$. Hence $S_{0,F}^2 = |S_{0,F}|^2$ is a power of q, and so $\pm S_{0,F}$ is a power of $\mathsf{Gauss}_{\mathbb{F}_q}$.

When q is 3 mod 4, we use the same arguments as in (i). We take care of the extra possible units in $\mathbb{Z}[\zeta_3]$ by observing that

$$S_{0,F} \in 1 + 2\mathbb{Z}[\zeta_p]$$

to rule out units other than ±1. We see this by observing that the sum is invariant under $(x, y) \mapsto (-x, -y)$, an action which fixes the origin $(0, 0)$, but which on $E^2 \smallsetminus \{(0, 0)\}$ has all orbits of size 2. We then treat the two cases $S_{0,F} = \pm S_{0,F}(-)$ exactly as in the proof of (i). □

Corollary 11.11. *Let q be a power of an odd prime p, and E a subfield of $\mathbb{F}_q$. Let*

$$f(x) = \sum_{i=0}^{n} a_i x^{q^i+1} \in E[x], \quad g(x) := \sum_{i=0}^{m} b_i x^{q^i+1} \in E[x]$$

with $n, m \in \mathbb{Z}_{>0}$, and a_n, b_m nonzero. Let $t \in E$, and let

$$F(x, y) := txy + f(x) + g(y).$$

Then the sums S_f, S_F formed over E as in Theorem 11.10 are each ± a power of Gauss_E.

Proof. Apply Theorem 11.10 over the ground field $E = \mathbb{F}_{q_0}$, remembering that q is a power of q_0. □

Remark 11.12. In characteristic $p = 2$, the sums S_f and S_F of Theorem 11.10 can both vanish. For example, over $\mathbb{F}_4$, $S_f = 0$ for $f(x) = x^3 + x^5$, with the convention that $\mathsf{Gauss}_{\mathbb{F}_4} = 2$. And over $\mathbb{F}_{16}$, we have $S_F = 0$ for $F(x, y) = xy + v^{13}x^5 + vy^3$, for v any generator of the cyclic group $\mathbb{F}_{16}^\times$. This phenomenon will be studied in [KT8] in more detail.

12. A pullback result for $\mathcal{W}(M, A, B)$

The main result of this section is the following theorem about a well chosen Kummer pullback of the local system $\mathcal{W}(M, A, B)$ introduced in Definition 11.5.

Theorem 12.1. *The Kummer pullback* $[MAB]^\star\mathcal{W}(M,A,B)$ *of* $\mathcal{W}(M,A,B)$ *by* $v \mapsto v^{MAB}$ *(or more precisely, its extension across* 0 *by* $j_\star$*, for the inclusion* $j : \mathbb{G}_m \to \mathbb{A}^1$*) is lisse on* $\mathbb{A}^1$ *and pure of weight* 0*.*

Proof. Recall that the trace function of $\mathcal{W}(M,A,B)$ at $v \in E^\times$, $E/\mathbb{F}_p$ a finite extension, is given by

$$v \mapsto (1/\#E)\sum_{x,w\in E}\psi_E(MABxw - v^{-\alpha}Ax^{q^b+1} - v^\beta Bw^{q^a+1}).$$

If we replace ψ by the nontrivial additive character $t \mapsto \psi(t/MAB)$, this formula becomes

$$v \mapsto (1/\#E)\sum_{x,w\in E}\psi_E(xw - v^{-\alpha}x^{q^b+1} - v^\beta w^{q^a+1}),$$

simply because both MA, MB are 1 mod p. After the pullback by the map $v \mapsto v^{MAB}$, the trace function becomes

$$\begin{aligned} v &\mapsto (1/\#E)\sum_{x,w\in E}\psi_E(xw - v^{-\alpha MAB}x^{q^b+1} - v^{\beta MAB}w^{q^a+1}) \\ &= (1/\#E)\sum_{x,w\in E}\psi_E(xw - v^{-\alpha A(q^b+1)}x^{q^b+1} - v^{\beta B(q^a+1)}w^{q^a+1}). \end{aligned}$$

After the change of variable $x \mapsto v^{\alpha A}x, w \mapsto v^{-\beta B}w$, this becomes

$$\begin{aligned} v &\mapsto (1/\#E)\sum_{x,w\in E}\psi_E(v^{\alpha A-\beta B}xw - x^{q^b+1} - w^{q^a+1}) \\ &= (1/\#E)\sum_{x,w\in E}\psi_E(vxw - x^{q^b+1} - w^{q^a+1}), \end{aligned}$$

simply because $\alpha A - \beta B = 1$.

We will show in Theorem 12.2 below that this trace function, stripped of the $1/\#E$ factor, is the trace function of a sheaf on $\mathbb{A}^1$ which is lisse and pure of weight 2. All such sheaves are geometrically semisimple (by purity) and have isomorphic semisimplifications (by Chebotarev), hence are all geometrically isomorphic. Any such is geometrically isomorphic to $\mathcal{W}(M,A,B)$ on $\mathbb{G}_m$, so must agree geometrically with $j_\star\mathcal{W}(M,A,B)$ on $\mathbb{A}^1$. □

To show this, let us consider the following slightly more general situation, similar to that of Theorem 11.3.

Theorem 12.2. *Fix integers $d \geq 3, e \geq 2$, both of which are prime to p. Fix one-variable polynomials $f(x) \in \mathbb{F}_p[x]$ and $g(w) \in \mathbb{F}_p[w]$ of respective degrees d and e. On $\mathbb{A}^3/\mathbb{F}_p$, with coordinates (v, x, w), form the Artin–Schreier sheaf*

$$\mathcal{L}_{\psi(f(x)+g(w)+vxw)}.$$

Denote by

$$\mathrm{pr}_1 : \mathbb{A}^3 \mapsto \mathbb{A}^1$$

the first projection $(v, x, w) \mapsto v$. Then

$$R^i\pi_! := R^i\pi_!(\mathcal{L}_{\psi(f(x)+g(w)+vxw)})$$

vanishes for $i \neq 2$, and $R^2\pi_!$ is lisse on $\mathbb{A}^1$ of rank $(d-1)(e-1)$ and pure of weight two, with trace function given at $v \in E$ for E a finite extension of $\mathbb{F}_p$ by

$$v \mapsto \sum_{x,w \in E} \psi_E(vxw + f(x) + g(w)).$$

Proof. For $i \neq 2$, the asserted vanishing can be checked fibre by fibre. Over $\mathbb{G}_m$, the substitution $x \mapsto x/v, w \mapsto w$ reduces us to a particular case of the vanishing established in Theorem 11.3. Over 0, we have

$$\begin{aligned} R^i\pi_!|_{v=0} &= H^i_c(\mathbb{A}^2/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(f(x)+g(w))})) \\ &= \oplus_{j+k=i} H^j_c(\mathbb{A}^1/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(f)}) \otimes H^k_c(\mathbb{A}^1/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(g)}). \end{aligned}$$

The asserted vanishing for $i \neq 2$ results from the (standard) fact that the H^j and H^j here each vanish unless $j = k = 1$. Because the $R^2\pi_!$ is a sheaf of perverse origin, it is lisse on $\mathbb{A}^1$ of rank $(d-1)(e-1)$ if and only each stalk has dimension $(d-1)(e-1)$. Over $\mathbb{G}_m$, this results from Theorem 11.3 (after the same change of variable $x \mapsto x/v, w \mapsto w$). Over 0, it results from the (standard) fact that $H^1_c(\mathbb{A}^1/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(f)})$ has dimension $d-1$, and $H^k_c(\mathbb{A}^1/\overline{\mathbb{F}_p}, \mathcal{L}_{\psi(g)})$ has dimension $e-1$. Once we know that $R^2\pi_!$ is pure of weight 2 on $\mathbb{G}_m$ and lisse at 0, it is automatically pure of weight 2 on $\mathbb{A}^1$, cf. [De, 1.8.10]. The formula for the trace is immediate from the Lefschetz trace formula, and the vanishing of the $R^i\pi_!$ for $i \neq 2$. □

13. Determinants

We now return to the consideration of the M lisse sheaves discussed in Theorem 11.1, except that we do an additional Tate twist to be in weight 0.

Thus

$$\mathcal{K}l(M,A,B,\sigma^{-\beta},\sigma^{-\alpha})(-A-B+1)\otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)},$$

for σ nontrivial in $\mathsf{Char}(M)$, and

$$\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})(-A-B+2)\otimes \mathsf{Gauss}(M,A,B)^{-\deg(E/E_1)}.$$

Each of them, by Theorems 4.7 and 5.1, has finite G_{arith}. We viewed them as lisse sheaves on $\mathbb{G}_m/E_1$, for E_1 the field $\mathbb{F}_p(\mu_{MAB})$. However, each has a descent to $\mathbb{G}_m/\mathbb{F}_p(\mu_M)$, as follows. Each of them is the highest weight quotient (now weight zero) of the lisse sheaf on $\mathbb{G}_m/E_1$ whose trace function is

$$\sum_{x,w\in E^\times}(1/\#E)\psi_E(MABx-v^{-\alpha}Ax^{MB}/w^B-v^{\beta}Bw^A)\sigma(w).$$

This sheaf has an obvious descent to $\mathbb{G}_m/\mathbb{F}_p(\mu_M)$ (just so the characters σ of order dividing M are defined). Its highest weight quotient is the desired descent. [Unfortunately, we do not know an explicit formula for its trace function.] Let us call these descended sheaves

$$\mathcal{G}_\sigma.$$

Strictly speaking, we should remember that their definition made use of chosen (α,β) with $\alpha A-\beta B=1$, and denote them

$$\mathcal{G}_{\sigma,\alpha,\beta}.$$

If (α,β) is one such, then so is $(\alpha+B,\beta+A)$.

Lemma 13.1. *If the integers α, A, β, B satisfy $\alpha A-\beta B=1$, then we have* $\gcd(\alpha+\beta,A+B)=1$.

Proof. If not, there exists a prime r which divides both $\alpha+\beta$ and $A+B$. So modulo r,

$$\alpha A-\beta B\equiv \alpha A-(-\alpha)(-A)=0,$$

a contradiction. □

Corollary 13.2. *Given relatively prime integers A, B and a real constant $X>0$, there exist integers α,β with $\alpha A-\beta B=1$ such that either $\alpha+\beta=\pm 1$ or $\alpha+\beta$ is a prime P with $P>X$. In particular, given an integer $D>1$, there exist such α,β with* $\gcd(\alpha+\beta,D)=1$.

Proof. Because $\gcd(A,B)=1$, there are integers α_0,β_0 with $\alpha_0 A-\beta_0 B=1$. If $A+B=0$, then $(A,B)=\pm(1,-1)$, and then $\alpha+\beta=\pm 1$. If $A+B\neq 0$, we argue as follows. For every integer n, the pair $(\alpha_n,\beta_n):=(\alpha_0+nB,\beta_0+nA)$ is another such pair. Then

$$\alpha_n+\beta_n=(\alpha_0+\beta_0)+n(A+B).$$

By the previous Lemma 13.1, $\gcd(\alpha_0+\beta_0,A+B)=1$. Now apply Dirichlet's theorem to the sequence $\alpha_n+\beta_n$ for positive n if $A+B>0$, or to the sequence of negative n if $A+B<0$. □

Theorem 13.3. *For σ of order dividing M, the geometric determinant of $\mathcal{G}_{\sigma,\alpha,\beta}$ is*

$$\mathcal{L}_{\chi_2^{MAB-1-(A-1)-(B-1)}}\mathcal{L}_{\sigma^{\alpha+\beta}},$$

with the understanding that if $p=2$, then $\chi_2:=\mathbb{1}$.

Proof. By [K2, 8.11.6], the geometric determinant in each case is the product of the "upstairs" characters. One has the general formula

$$\prod_{\chi\in\mathsf{Char}(A,\rho)}\mathcal{L}_\chi=\mathcal{L}_{\chi_2^{A-1}}\otimes\mathcal{L}_\rho.$$

Therefore the geometric determinant of $\mathcal{G}_\sigma$ is $\mathcal{L}_{\chi_2^{MAB-1-(A-1)-(B-1)}}\mathcal{L}_{\sigma^{\alpha+\beta}}$. [If we are in characteristic 2, then each of M,A,B is odd, and the determinant is just $\mathcal{L}_{\sigma^{\alpha+\beta}}$.] □

Corollary 13.4. *Choose α,β with $\alpha A-\beta B=1$ and $\gcd(\alpha+\beta,M)=1$ (possible by Corollary 13.2). Then there exists characters σ of order dividing M such that the geometric determinant of $\mathcal{G}_{\sigma,\alpha,\beta}$ has order M.*

Proof. If p is odd, then M is even, and $\sigma\mapsto\chi_2\sigma$ is a bijection of $\mathsf{Char}(M)$. On the other hand, $\sigma\mapsto\sigma^{\alpha+\beta}$ is another such bijection. If $p=2$ or if the exponent of χ_2 in the geometric determinant of $\mathcal{G}_{\sigma,\alpha,\beta}$ is even, simply take $\mathcal{G}_{\sigma,\alpha,\beta}$ with σ of full order M. If p is odd and the geometric determinant of $\mathcal{G}_{\sigma,\alpha,\beta}$ is $\chi_2\sigma^{\alpha+\beta}$, use the fact that the composite map

$$\sigma\mapsto\sigma^{\alpha+\beta}\mapsto\chi_2\sigma^{\alpha+\beta}$$

is a bijection of $\mathsf{Char}(M)$, and take $\mathcal{G}_{\sigma_1,\alpha,\beta}$, for any σ_1 whose image under this map is a character of full order M. □

From Theorem 13.3, we get the following corollary.

Corollary 13.5. *For any α, β with $\alpha A - \beta B = 1$, and any σ of order dividing M, the Kummer pullback $[M]^\star \mathcal{G}_{\sigma,\alpha,\beta}$ has geometrically trivial determinant.*

From Lemma 4.3, we see that we have

Lemma 13.6. *For σ of order dividing M, we have*

$$\mathcal{L}_\sigma \otimes \mathcal{G}_{\sigma,\alpha,\beta} \cong \mathcal{G}_{\sigma,\alpha-B,\beta-A}.$$

This "indeterminacy" in the "definition" of $\mathcal{G}_\sigma$ can be "corrected" by considering the Kummer pullback

$$[M]^\star \mathcal{G}_\sigma,$$

since $[M]^\star$ kills the $\mathcal{L}_\sigma$ factor.

Theorem 13.7. *Each sheaf $[M]^\star \mathcal{G}_\sigma$ has geometrically trivial determinant.*

Proof. From the explicit formulas for the geometric determinants in Theorem 13.3, it is clear that they become trivial after $[M]^\star$. Indeed, in odd characteristic, M is even (since both q^a+1, q^b+1 are even), and hence $[M]^\star$ kills both $\mathcal{L}_{\chi_2}$ and any power of $\mathcal{L}_\sigma$. In any characteristic, $[M]^\star$ kills any power of $\mathcal{L}_\sigma$. □

Remark 13.8. Presumably (?) we should hope that each $[M]^\star \mathcal{G}_\sigma$ already has arithmetically trivial determinant over the small field $\mathbb{F}_p$ or $\mathbb{F}_{q^2}$, without any extension of scalars being needed.

14. Some general results on $G_{\rm geom}$ and $G_{\rm arith}$

First we recall the following result concerning the image of the wild inertia group $P(\infty)$ in $G_{\rm geom}$:

Proposition 14.1 ([KT7, Proposition 4.8])**.** *Let $\mathcal{H}$ be an (irreducible) hypergeometric sheaf of type (D, m) in characteristic p, with $D > m$ and with finite geometric monodromy group $G = G_{\rm geom}$. Then the following statements hold for the image Q of $P(\infty)$ in G:*

(i) *If $\mathcal{H}$ is not Kloosterman, i.e. if $m > 0$, then $Q \cap \mathbf{Z}(G) = 1$.*
(ii) *Suppose $\mathcal{H}$ is Kloosterman and $D > 1$. Then $Q \not\leq \mathbf{Z}(G)$. If $p \nmid D$, then $Q \cap \mathbf{Z}(G) = 1$. If $p | D$ then either $Q \cap \mathbf{Z}(G) = 1$ or $Q \cap \mathbf{Z}(G) \cong C_p$.*
(iii) *If $D > 1$, then $1 \neq Q/(Q \cap \mathbf{Z}(G)) \hookrightarrow G/\mathbf{Z}(G)$ and p divides $|G/\mathbf{Z}(G)|$.*

(iv) *If $D - m \geq 2$, the determinant of G is a p'-group. If moreover $p \nmid D$, then $\mathbf{Z}(G)$ is a p'-group.*

(v) *Suppose $p = 2$. Then the trace of any element $g \in G$ on the sheaf $\mathcal{H}$ is* 2*-rational* (i.e. lies in a cyclotomic field $\mathbb{Q}(\zeta_N)$ for some odd integer N)*; in particular, the* 2*-part of $|\mathbf{Z}(G)|$ is at most* 2*.*

Lemma 14.2. *Let $X/\mathbb{F}_q$ be smooth and geometrically connected, k a topological field, and V a finite dimensional continuous k-representation of $\pi_1(X)$. Denote by $G_{\rm arith} \leq {\rm GL}(V)$ the image of $\pi_1^{\rm arith}(X) := \pi_1(X)$, and by*

$$G_{\rm geom} \lhd G_{\rm arith}$$

the image of $\pi_1^{\rm geom}(X) := \pi_1(X/\overline{\mathbb{F}_q})$. Let $E/\mathbb{F}_q$ be a finite extension. Then for any points $v_1, v_2 \in X(E)$, the $G_{\rm arith}$-conjugacy classes of $Frob_{v_1,E}$ and $Frob_{v_2,E}$ lie the same $G_{\rm geom}$-coset in $G_{{\rm arith},k}$.

Proof. Let us explain this in the universal case. The key point is that we have the short exact sequence of fundamental groups [Gr2, Exp. IX, Thm. 6.1]

$$1 \to \pi_1^{\rm geom}(X) \to \pi_1^{\rm arith}(X) \xrightarrow{\rm deg} {\rm Gal}(\overline{\mathbb{F}_q}/\mathbb{F}_q) \to 1.$$

When we identify ${\rm Gal}(\overline{\mathbb{F}_q}/\mathbb{F}_q)$ with the profinite completion of $\mathbb{Z}$ by decreeing that $x \mapsto x^q$ has degree -1, then each $Frob_{v_i,E}$ has degree $\deg(E/\mathbb{F}_q)$ in ${\rm Gal}(\overline{\mathbb{F}_q}/\mathbb{F}_q)$. Hence for any elements $g_i \in \pi_1^{\rm arith}(X)$ which lie in the conjugacy classes of $Frob_{v_i,E}$, the “ratio” $g_1^{-1}g_2$ has degree 0, i.e., lies in $\pi_1^{\rm geom}(X)$, which is precisely the subgroup of $\pi_1^{\rm arith}(X)$ consisting of elements of degree 0. □

Next we prove some general facts concerning pullbacks of local systems.

Lemma 14.3. *Given a local system $\mathcal{F}$ on $X/\mathbb{F}_q$, and an $\mathbb{F}_q$-morphism $f : Y \to X$ of $\mathbb{F}_q$-schemes. Then for any finite extension $k/\mathbb{F}_q$, and any point $v \in Y(k)$, we have*

$$Frob_{v,k}|_{f^\star\mathcal{F}} = Frob_{f(v),k}|_{\mathcal{F}}.$$

Proof. Let us explain this in terms of representations of fundamental groups. When X, Y are each connected, and we pick appropriate base points, f induces a homomorphism of fundamental groups $f_\star : \pi_1(Y/\mathbb{F}_q) \to \pi_1(X/\mathbb{F}_q)$ which maps the conjugacy class of $Frob_{v,k}$ in $\pi_1(Y/\mathbb{F}_q)$ to the conjugacy class of $Frob_{f(v),k}$ in $\pi_1(X/\mathbb{F}_q)$. The local system $\mathcal{F}$ is a representation $\rho_{\mathcal{F}}$ of $\pi_1(X/\mathbb{F}_q)$, and its pullback $f^\star\mathcal{F}$ is the representation $\rho_{\mathcal{F}} \circ f_\star$ of $\pi_1(Y/\mathbb{F}_q)$. □

Theorem 14.4. *Let $N \in \mathbb{Z}_{\geq 1}$ and let G be a finite group with a normal subgroup S such that $G/S \cong C_N$. Let k be a finite field in which N is invertible and which contains the N^{th} roots of unity. Let $\mathcal{W}$ be a local system on $\mathbb{G}_m/k$ which has geometric monodromy group G_{geom} and arithmetic monodromy group $G_{\text{arith},k}$, with $G_{\text{geom}} = G_{\text{arith},k} \cong G$. Then the $[N]^\star$ Kummer pullback $\mathcal{W}_N$ of $\mathcal{W}$ has geometric and arithmetic monodromy group $G_{\text{geom},\mathcal{W}_N} = G_{\text{arith},k,\mathcal{W}_N} = S$.*

Proof. From the point of view of Galois theory, the fact that the system $\mathcal{W}$ has $G_{\text{arith},k} = G_{\text{geom}} = G$ means that we have a finite Galois extension $L/k(t)$ with $\text{Gal}(L/k(t)) = G$, which is linearly disjoint from the extension $\overline{k}/k$, i.e., $L_{\overline{k}}/\overline{k}(t)$ continues to have $\text{Gal}(L_{\overline{k}}/\overline{k}(t)) = G$.

When we form the $[N]$ pullback, we replace the Galois extension $L/k(t)$ by its compositum with the finite Galois extension $k(t^{1/N})/k(t)$. [It is Galois because k contains the N^{th} roots of unity.] This new extension has Galois group $G_{\text{arith},k,\mathcal{W}_N}$. Similarly, when we replace the Galois extension $L_{\overline{k}}/\overline{k}(t)$ by its compositum with the finite Galois extension $\overline{k}(t^{1/N})/\overline{k}(t)$, this new extension has Galois group $G_{\text{geom},\mathcal{W}_N}$.

Consider a homomorphism

$$\theta : G \twoheadrightarrow \mu_N(k)$$

with $\text{Ker}(\theta) = S$. This surjective homomorphism means that there is a subfield

$$\overline{k}(t) \subset K \subset L_{\overline{k}},$$

with $K/\overline{k}(t)$ Galois, with $\text{Gal}(K/\overline{k}(t)) \cong \mu_N(\mathbb{F}_q)$. But this extension $K/\overline{k}(t)$ is the function field of a $\mu_N(\mathbb{F}_q)$-covering of $\mathbb{G}_m/\overline{k}$. The only such covering is the $[N]$ Kummer covering. Thus the intermediate field K must be $K = \overline{k}(t^{1/N})$; we have

$$\overline{k}(t) \subset \overline{k}(t^{1/N}) \subset L_{\overline{k}}.$$

This in turn means that the compositum of the extension $L_{\overline{k}}/\overline{k}(t)$ with $\overline{k}(t^{1/N})/\overline{k}(t)$ is just the extension $L_{\overline{k}}/\overline{k}(t^{1/N})$. Its Galois group is the index N normal subgroup of G on which θ is trivial, i.e. its Galois group is S.

Now let us consider the interaction of the homomorphism θ with the extension $L/k(t)$. Its existence means that there is a subfield

$$k(t) \subset K_0 \subset L,$$

with $K_0/k(t)$ Galois, with group $\mu_N(\mathbb{F}_q)$. This extension $K_0/k(t)$ is the function field of a $\mu_N(k)$-covering of $\mathbb{G}_m/k$, which when we extend scalars to $\overline{k}$ becomes the $[N]$ Kummer covering. In general, for any field k in which N is invertible and which contains the $N^{\rm th}$ roots of unity, the $\mu_N(k)$-coverings of $\mathbb{G}_m/k$ are classified by the cokernel $k[t,1/t]^\times$ modulo the subgroup of $N^{\rm th}$ powers, cf. [Gr2, Cor. 6.5, Exp. XI]. The group of units $k[t,1/t]^\times$ is $k^\times t^{\mathbb{Z}}$. Since geometrically our covering is adjoining $t^{1/N}$, our covering must be $k((\alpha t)^{1/N})$, for some $\alpha \in k^\times$. This means that if we take αt instead of t as the parameter of $\mathbb{G}_m/k$, then $K_0/k(t)$ is the extension $k(t^{1/N})/k(t)$. Thus the compositum of $L/k(t)$ with $k(t^{1/N})/k(t)$ is just the extension $L/k(t^{1/N})$. Its Galois group is the index N subgroup of G on which θ is trivial, i.e., its Galois group is S. □

For possible later reference, we state the following corollary, which is immediate from the proof of Theorem 14.4.

Corollary 14.5. *With G, S, N as in Theorem 14.4, let k be an algebraically closed field in which N is invertible and which contains the $N^{\rm th}$ roots of unity. Let $\mathcal{W}$ be a local system on $\mathbb{G}_m/k$ which has geometric monodromy group $G_{\rm geom} \cong G$. Then $\mathcal{W}_N := [N]^\star\mathcal{W}$ has geometric monodromy group $G_{{\rm geom},\mathcal{W}_N} = S$.*

Here is another version, which deals with Kummer pullbacks in fair generality. Let $\overline{k}$ be an algebraically closed field of characteristic $p > 0$, and let G be a finite group which is a quotient of $\pi_1(\mathbb{G}_m/\overline{k})$. One knows by [Abh, Proposition 6(III)] that the quotient of G by the subgroup $\mathbf{O}^{p'}(G)$ generated by its Sylow p-subgroups is a cyclic group of order prime to p; this is simply the statement that the prime to p quotient of $\pi_1(\mathbb{G}_m/\overline{k})$ is pro-cyclic, in fact non-canonically isomorphic to $\prod_{\ell\neq p}\mathbb{Z}_\ell$. Let us denote by $n(G)$ this order:

$$n(G) := |G/\mathbf{O}^{p'}(G)|.$$

Then the normal subgroups $H \lhd G$ such that G/H has order prime to p are precisely those containing $\mathbf{O}^{p'}(G)$. Because $G/\mathbf{O}^{p'}(G)$ is cyclic of order $n(G)$, such a subgroup $H \lhd G$ with G/H of order d has $d \nmid n(G)$, and H is thus the unique normal subgroup $G_d \lhd G$ such that $G/G_d = d$ is cyclic of order d, and we have

$$n(G_d) = n(G)/d.$$

Theorem 14.6. *Let G be a finite group. Let k be a finite field of characteristic p, and $\mathcal{W}$ a local system on $\mathbb{G}_m/k$ with*

$$G_{{\rm arith},k} = G_{\rm geom} = G.$$

Let N be a prime to p integer, and let

$$N_0 := \gcd(N, n(G)).$$

Suppose that k contains the N_0^{th} roots of unity, i.e. $N_0|(\#k-1)$. Then for $\mathcal{W}_N := [N]^\star\mathcal{W}$, we have

$$G_{\text{arith},k,\mathcal{W}_N} = G_{\text{geom},\mathcal{W}_N} = G_{N_0}.$$

Proof. Write $N = N_0N_1$, with $\gcd(N_1, n(G)/N_0) = 1$. Then $\mathcal{W}_N = [N_1]^\star\mathcal{W}_{N_0}$. By Theorem 14.4 applied to $\mathcal{W}_{N_0}$, we have

$$G_{\text{arith},k,\mathcal{W}_{N_0}} = G_{\text{geom},\mathcal{W}_{N_0}} = G_{N_0},$$

and $n(G_{N_0}) = n(G)/N_0$. So we are reduced to treating universally the case when $\gcd(N, n(G)) = 1$. Then $G_{\text{geom},\mathcal{W}_N} \lhd G$ is a normal subgroup of index dividing N. But there are none other than G itself. Therefore $G_{\text{geom},\mathcal{W}_N} = G$. As $G_{\text{arith},k,\mathcal{W}_N} \leq G_{\text{arith},k,\mathcal{W}} = G$ but $G_{\text{arith},k,\mathcal{W}_N} \geq G_{\text{geom},\mathcal{W}_N} = G$, we have $G_{\text{arith},k,\mathcal{W}_N} = G$ as well. □

We now discuss G_{arith} for a geometrically irreducible $\overline{\mathbb{Q}_\ell}$-adic hypergeometric sheaf $\mathcal{H}$ on $\mathbb{G}_m/\mathbb{F}_q$ whose G_{geom} is finite. To make clear the underlying structure, we will consider the more general case of a smooth, geometrically connected variety $X/\mathbb{F}_q$, and a geometrically irreducible $\overline{\mathbb{Q}_\ell}$-adic sheaf $\mathcal{F}$ on $X/\mathbb{F}_q$ whose G_{geom} is finite. One knows that $\det(\mathcal{F})$ is geometrically of finite order (e.g., because **its** G_{geom} is a semisimple group inside GL_1, cf. [De, 1.3.9]).

Lemma 14.7. *There exists an ℓ-adic unit $C \in \overline{\mathbb{Q}_\ell}^\times$ such that the sheaf $\det(\mathcal{F}) \otimes C^{-\deg/k}$ is arithmetically of finite order. Moreover, any such C is determined up to multiplication by a root of unity.*

Proof. To see this, choose an integer $M \geq 1$ such that $\det(\mathcal{F})^{\otimes M}$ is geometrically trivial. This means precisely that arithmetically

$$\det(\mathcal{F})^{\otimes M} \cong D^{\deg/k}$$

for some ℓ-adic unit $D \in \overline{\mathbb{Q}_\ell}^\times$. Then for any C with $C^M = D$, the sheaf $\det(\mathcal{F}) \otimes C^{-\deg/k}$ has arithmetic order dividing M.

It is obvious that if C works, then so does ζC for any root of unity ζ. Conversely, if C' works, then both $\det(\mathcal{F}) \otimes C^{-\deg/k}$ and $\det(\mathcal{F}) \otimes (C')^{-\deg/k}$ are arithmetically of finite order, so their ratio $(C/C')^{\deg/k}$ is arithmetically of finite order, i.e., C/C' is a root of unity. □

Corollary 14.8. *There exists an ℓ-adic unit $G \in \overline{\mathbb{Q}_\ell}^\times$ such that $\mathcal{F} \otimes G^{-\deg/k}$ has finite arithmetic determinant, and this condition determines G up to multiplication by a root of unity.*

Proof. For $\mathcal{F}$ of rank D, any D^{th} root of the C of Lemma 14.7 does the job, and G does the job if and only if G^D is some root of unity times C. □

Lemma 14.9. *Suppose $\mathcal{F}$ has finite G_{geom}. Then $\mathcal{F} \otimes G^{-\deg/k}$ has finite G_{arith} if and only if its arithmetic determinant is finite.*

Proof. G_{arith} cannot be finite if its determinant fails to be finite. To see that G_{arith} is finite if its determinant is finite and G_{geom} is finite, use the fact that G_{arith} normalizes G_{geom}. Denote by N the order of the finite group $\text{Aut}(G_{\text{geom}})$. Then for $\gamma \in G_{\text{arith}}$, γ^N commutes with every element of G_{geom}. As G_{geom} is an irreducible subgroup of GL_D with $D := \text{rank}(\mathcal{F})$, each γ^N is a scalar. But as G_{arith} has a determinant of finite order, say M, each γ^N is a root of unity of order dividing MD. Thus $\text{Lie}(G_{\text{arith}})$ is killed by NMD, so $\text{Lie}(G_{\text{arith}}) = 0$ and hence G_{arith} is finite. □

Let us recall the following criterion for finite arithmetic and geometric monodromy, cf. [KRLT1, 2.1, 2.2].

Proposition 14.10. *Suppose we have $(\mathbb{F}_q, \ell, X)$ as above, with $\mathcal{G}$ a lisse $\overline{\mathbb{Q}_\ell}$ sheaf on X. Suppose further that $\mathcal{G}$ is pure of weight zero (for all embeddings of $\overline{\mathbb{Q}_\ell}$ into $\mathbb{C}$). Consider the following four conditions.*

(a) *G_{arith} is finite.*
(b) *All traces of $\mathcal{G}$ are algebraic integers. More precisely, for every finite extension $L/\mathbb{F}_q$, and for every point $x \in X(L)$, $\text{Trace}(Frob_{L,x}|\mathcal{G})$ is an algebraic integer.*
(c) *G_{geom} is finite.*
(d) *$\det(\mathcal{G})$ is arithmetically of finite order.*

Then we have the implications

$$(\text{a}) \implies (\text{b}) \implies (\text{c}),\ (\text{b}) \implies (\text{d}).$$

If $\mathcal{F}$ is geometrically irreducible, we have (a) $\iff$ (b) $\iff$ (c). *If $\mathcal{F}$ is arithmetically semisimple, we have* (a) $\iff$ (b).

Proposition 14.11. *Suppose we have $(\mathbb{F}_q, \ell, X)$ as above, with $\mathcal{G}$ a lisse $\overline{\mathbb{Q}_\ell}$ sheaf on X which is geometrically irreducible, and pure of integer weight w.*

Suppose that for some monomial in Gauss sums over $\mathbb{F}_q$, i.e., an expression of the form

$$A = \pm \prod_{\chi \in \mathsf{Char}(q-1)} \left(-\mathsf{Gauss}(\psi_{\mathbb{F}_q}, \chi)\right)^{n_\chi},$$

with exponents $n_\chi \in \mathbb{Z}$, the constant field twist $\mathcal{G} \otimes A^{-\deg/\mathbb{F}_q}$ has algebraic integer traces, and hence has finite arithmetic monodromy group, denoted $G_{\mathrm{arith},A}$. Suppose further that p is odd. Then the p-primary part of the finite cyclic group $\mathbf{Z}(G_{\mathrm{arith},A})$ is independent of the choice of monomial A in Gauss sums over $\mathbb{F}_q$ for which $\mathcal{G} \otimes A^{-\deg/\mathbb{F}_q}$ has algebraic integer traces.

Proof. By Chebotarev, every element in $G_{\mathrm{arith},A}$ is the image of some Frobenius $Frob_{L,x}$. The given representation of $G_{\mathrm{arith},A}$ is irreducible (because it is already irreducible on the subgroup G_{geom}). So the Frobenii which land in $\mathbf{Z}(G_{\mathrm{arith},A})$ are precisely those for which $Frob_{L,x}|_{\mathcal{G}}$ is a scalar, call it $\alpha(L,x)$. Then in $G_{\mathrm{arith},A}$, this Frobenius gives the central scalar $\alpha(L,x)/A^{\deg(L/\mathbb{F}_q)}$. If we use a different monomial in Gauss sums, say A_1 for which $\mathcal{G} \otimes A_1^{-\deg/\mathbb{F}_q}$ has algebraic integer traces, then this same Frobenius gives the central scalar $\alpha(L,x)/A_1^{\deg(L/\mathbb{F}_q)}$. So what must be shown is that the ratio A/A_1 is a root of unity of order prime to p.

Since both $\mathcal{G} \otimes A^{-\deg/\mathbb{F}_q}$ and $\mathcal{G} \otimes A_1^{-\deg/\mathbb{F}_q}$ have arithmetic determinants of finite order, it results from Corollary 14.8 that the ratio A/A_1 is a root of unity. Now A/A_1 is itself a monomial in Gauss sums, so the assertion results from the following lemma. □

Lemma 14.12. *Suppose p is odd, $\mathbb{F}_q/\mathbb{F}_p$ a finite extension, and A a monomial in Gauss sums over $\mathbb{F}_q$ which is a root of unity. Then A has order prime to p.*

Proof. Each Gauss sum over $\mathbb{F}_q$ lies in $\mathbb{Q}(\zeta_p, \zeta_{q-1})$. Thus A is a root of unity in this field. We will show that in fact it lies in the subfield $\mathbb{Q}(\zeta_{q-1})$, whose only roots of unity are μ_{q-1} (remember $q-1$ is even). For this, it suffices to show that A is invariant under $\mathrm{Gal}(\mathbb{Q}(\zeta_p, \zeta_{q-1})/\mathbb{Q}(\zeta_{q-1}))$. This is the group $\mathbb{F}_p^\times$, with σ_a, $a \in \mathbb{F}_p^\times$, mapping ζ_p to ζ_p^a and fixing ζ_{q-1}. The claimed invariance holds for A if and only if it holds for $-A$, so we may assume

$$A = \prod_{\chi \in \mathsf{Char}(q-1)} \left(-\mathsf{Gauss}(\psi_{\mathbb{F}_q}, \chi)\right)^{n_\chi}.$$

When we apply σ_a to A, we get

$$\begin{aligned}\sigma_a(A) &= \prod_{\chi \in \mathsf{Char}(q-1)} \left(-\mathsf{Gauss}(\psi_{a,\mathbb{F}_q}, \chi)\right)^{n_\chi} \\ &= \prod_{\chi \in \mathsf{Char}(q-1)} \left(-\overline{\chi}(a)\mathsf{Gauss}(\psi_{a,\mathbb{F}_q}, \chi)\right)^{n_\chi} = \Lambda(a)A\end{aligned}$$

for Λ the character $\prod_{\chi \in \mathsf{Char}(q-1)} \overline{\chi}^{n_\chi}$. Suppose now that the order of A is not prime to p. The roots of unity in $\mathbb{Q}(\zeta_p, \zeta_{q-1})$ form the group

$$\mu_{p(q-1)} = \mu_p \times \mu_{q-1}.$$

Then A^{q-1} would be a p^{th} root of unity, and a prime to p power of A^{q-1} would be ζ_p. Such a power is itself a monomial in Gauss sums, so we would have

$$A = \zeta_p.$$

Then $\sigma_a(A) = \zeta_p^a = \zeta^{a-1}A$, but also $\sigma_a(A) = \Lambda(a)A$. Thus $\zeta_p^{a-1} = \Lambda(a)$. The left side lies in μ_p, the right side lies in μ_{q-1}. Thus both are 1. In particular, $\Lambda(a) = 1$, and A lies in $\mathbb{Q}(\zeta_{q-1})$. □

We now turn to the special case of geometrically irreducible hypergeometric sheaves $\mathcal{H}$ of type (D, m) with $D > m \geq 0$ on $\mathbb{G}_m/\mathbb{F}_q$. Thus we have

$$\mathcal{H} = \mathcal{H}yp(\chi_1, \ldots, \chi_D; \rho_1, \ldots, \rho_m)$$

with each χ_i and each ρ_j a (possibly trivial) character of $\mathbb{F}_q^\times$, such that for all i, j, $\chi_i \neq \rho_j$.

Proposition 14.13. *Suppose $D - m \geq 2$. Define*

$$A := \det(Frob_{1,\mathbb{F}_q}|\mathcal{H}).$$

Then we have the following results.

(i) *A is a monomial in Gauss sums.*
(ii) *For any B with $B^D = A$, the constant field twist $\mathcal{H} \otimes B^{-\deg/\mathbb{F}_q}$ has finite arithmetic determinant $\mathcal{L}_\Lambda$, for $\Lambda := \prod_i \chi_i$, of order dividing $q - 1$.*
(iii) *Suppose $p \nmid D$, and that $\mathcal{H} \otimes B^{-\deg/\mathbb{F}_q}$ on $\mathbb{G}_m/\mathbb{F}_q$, has finite arithmetic monodromy group G_{arith}. Then $\mathbf{Z}(G_{\text{arith}})$ has order prime to p.*

(iv) *Suppose that $\mathcal{H}$ has a descent $\mathcal{H}_0$ to $\mathbb{G}_m/k$, for some subfield k of $\mathbb{F}_q$, in the sense that for some monomial J in Gauss sums over $\mathbb{F}_q$, the pullback of $\mathcal{H}_0$ to $\mathbb{G}_m/\mathbb{F}_q$ is arithmetically isomorphic to $\mathcal{H}\otimes J^{\deg/\mathbb{F}_q}$. Suppose further that for some monomial G in Gauss sums, $\mathcal{H}_0\otimes G^{-\deg/k}$ has finite $G_{\mathrm{arith},\mathcal{H}_0}$. If $p\nmid D$ and $p\nmid \deg(\mathbb{F}_q/k)$, then $\mathbf{Z}(G_{\mathrm{arith},\mathcal{H}_0})$ has order prime to p.*

Proof. When $D-m\geq 2$, one has the arithmetic determinant formula [K2, 8.12.2]

$$\det(\mathcal{H})\cong\mathcal{L}_\Lambda\otimes A^{\deg/\mathbb{F}_q},$$

with $\Lambda:=\prod_i\chi_i$ and with

$$A=\Lambda((-1)^{D-1})q^{D(D-1)/2}\prod_{i,j}\big(-\mathsf{Gauss}(\overline{\psi_{\mathbb{F}_q}},\chi_i/\rho_j)\big).$$

Recall that q is, up to sign, itself the square of the quadratic Gauss sum, to see that A is indeed a monomial in Gauss sums. This formula makes (ii) obvious. To show (iii), let γ be a scalar in G_{arith}. As $\det(G_{\mathrm{arith}})$ lies in μ_{q-1}, we see that $\gamma^D=\det(\gamma)$ has order dividing $q-1$, so γ has order dividing $D(q-1)$, which is prime to p.

To show (iv), we argue as follows. Let us write

$$d:=\deg(\mathbb{F}_q/k).$$

Because $\mathcal{H}_0\otimes G^{-\deg/k}$ on $\mathbb{G}_m/k$ has finite $G_{\mathrm{arith},\mathcal{H}_0,G}$, so does its pullback to $\mathbb{G}_m/\mathbb{F}_q$. This pullback is $\mathcal{H}\otimes J^{\deg/\mathbb{F}_q}\otimes(G^d)^{-\deg/\mathbb{F}_q}$, which is a constant field twist of $\mathcal{H}$ by a monomial in Gauss sums, namely by G^d/J. Let us denote its G_{arith} as $G_{\mathrm{arith},\mathcal{H},G^d/J}$. Thus $G_{\mathrm{arith},\mathcal{H},G^d/J}$ is a subgroup of $G_{\mathrm{arith},\mathcal{H}_0,G}$ of index dividing $d:=\deg(\mathbb{F}_q/k)$, which is prime to p. So if $G_{\mathrm{arith},\mathcal{H}_0,G}$ contained a scalar of nontrivial p power order, then γ^d would be a scalar of nontrivial p power order in $G_{\mathrm{arith},\mathcal{H},G^d/J}$. So it suffices to show that the center of $G_{\mathrm{arith},\mathcal{H},G^d/J}$ is prime to p. We know this to be true for $G_{\mathrm{arith},\mathcal{H},B}$ by part (iii). So it suffices to show that the ratio $B/(G^d/J)$, a priori a root of unity by Lemma 14.8, has order prime to p. Since $p\nmid D$, it suffices to show that the D^{th} power of this ratio has order prime to p. But this D^{th} power is a monomial in Gauss sums, namely AJ^D/G^{dD}, hence has order prime to p by Lemma 14.12. □

Theorem 14.14. *Suppose $\mathcal{H}$ is a geometrically irreducible hypergeometric sheaf $\mathcal{H}$ of type (D,m) with $D>m\geq 0$ on $\mathbb{G}_m/\mathbb{F}_q$. Suppose G_{geom} is finite.*

Then for $\mathsf{Gauss}(\psi_{\mathbb{F}_q}, \chi_2)$ *the quadratic Gauss sum, with the convention that when* q *is even, we "define"* $-\mathsf{Gauss}(\psi_{\mathbb{F}_q}, \chi_2) := \sqrt{q}$,

$$\mathsf{G} := \big(-\mathsf{Gauss}(\psi_{\mathbb{F}_q}, \chi_2)\big)^{D+m-1}, \text{ and } C := \mathsf{G}^D,$$

$\det(\mathcal{H}) \otimes C^{-\deg/\mathbb{F}_q}$ *is arithmetically of finite order, and* $\mathcal{H} \otimes \mathsf{G}^{-\deg/\mathbb{F}_q}$ *has finite* $G_{\rm arith}$.

Proof. In view of Lemma 14.9, the two assertions are equivalent. Let us write simply ψ for $\psi_{\mathbb{F}_q}$. Define

$$A = \Lambda((-1)^{D-1})q^{D(D-1)/2}\prod_{i,j}(-\mathsf{Gauss}(\overline{\psi}, \chi_i/\rho_j)).$$

By [K2, 8.12.2], $\det(\mathcal{H}) \otimes A^{-\deg/\mathbb{F}_q}$ is arithmetically of finite order. The weight of A is

$$D(D-1) + mD = D(D+m-1).$$

We must show that A is (some root of unity)×G^D. For this, it suffices to show that for every p-adic ${\rm ord}_q$ on $\mathbb{Q}(\zeta_p, \zeta_{q-1})$ (normalized to have ${\rm ord}_q(q) = 1$), we have

$${\rm ord}_q(A) \geq {\rm ord}_q(\mathsf{G}^D) = D(D+m-1)/2.$$

For then A/G^D is an algebraic integer in $\mathbb{Q}(\zeta_p, \zeta_{q-1})$ (since G^{2D} divides $q^{D(D+m-1)}$), all of whose complex absolute values are 1, and hence A/G^D is a root of unity.

For a character τ of $\mathbb{F}_q^\times$, and a chosen ${\rm ord}_q$, let us write

$$V(\tau) := {\rm ord}_q(\mathsf{Gauss}(\psi, \tau)).$$

Then

$${\rm ord}_q(A) = D(D-1)/2 + \sum_{i,j} V(\chi_i/\rho_j)$$

and the asserted inequality becomes

$$\sum_{i,j} V(\chi_i/\rho_j) \geq mD/2.$$

Let B have $B^D = A$. Because $G_{\rm geom}$ is finite, $\mathcal{H} \otimes B^{-\deg/\mathbb{F}_q}$ has finite $G_{\rm arith}$. Therefore for each $t \in \mathbb{G}_m(\mathbb{F}_q) = \mathbb{F}_q^\times$, if we denote

$$\mathcal{H}(t) := {\rm Trace}(Frob_{t,\mathbb{F}_q}|\mathcal{H}),$$

we have

$$\mathrm{ord}_q(\mathcal{H}(t)) \geq (1/D)\mathrm{ord}_q(A), \quad \mathrm{ord}_q(\sigma(t)\mathcal{H}(t)) \geq (1/D)\mathrm{ord}_q(A)$$

for every character σ of $\mathbb{F}_q^\times$. Thus for every such σ, we have

$$\mathrm{ord}_q\Bigl(\sum_{t \in \mathbb{F}_q^\times} \sigma(t)\mathcal{H}(t)\Bigr) \geq (1/D)\mathrm{ord}_q(A).$$

But one knows [K2, 8.2.8] that $\sum_{t \in \mathbb{F}_q^\times} \sigma(t)\mathcal{H}(t)$ is equal to

$$\bigl(\prod_i (-\mathsf{Gauss}(\psi, \chi_i\sigma))\bigr)\bigl(\prod_j (-\mathsf{Gauss}(\overline{\psi}, \overline{\rho_i\sigma}))\bigr).$$

Hence we have the inequality

$$\sum_i V(\chi_i\sigma) + \sum_j (V(\overline{\rho_j\sigma}) \geq (1/D)[D(D-1)/2 + \sum_{i,j} V(\chi_i/\rho_j)]$$

for every σ.

Apply this with σ successively taken to be $1/\rho_j$, and add the resulting m inequalities. We get

$$\sum_{i,j} V(\chi_i/\rho_j) + \sum_{j,k} V(\overline{\rho_j/\rho_k}) \geq (m/D)[D(D-1)/2 + \sum_{i,j} V(\chi_i/\rho_j)].$$

For each $\tau \neq \mathbb{1}$, we have

$$V(\tau) + V(\overline{\tau}) = 1,$$

since the product of the corresponding Gauss sums is $\pm q$. Therefore

$$\sum_{j,k} V(\overline{\rho_j/\rho_k}) = m(m-1)/2.$$

Writing Σ for $\sum_{i,j} V(\chi_i/\rho_j)$, we have

$$\Sigma + m(m-1)/2 \geq m(D-1)/2 + (m/D)\Sigma,$$

i.e.,

$$(1 - m/D)\Sigma \geq \frac{m(D-m)}{2}, \quad \text{i.e. } \frac{D-m}{D}\Sigma \geq \frac{m(D-m)}{2}, \quad \text{i.e. } \Sigma \geq \frac{mD}{2},$$

as asserted. [It is only at this very last step that we use the hypothesis that $D-m>0$.] □

Corollary 14.15. *Suppose $\mathcal{H}$ is a geometrically irreducible hypergeometric sheaf $\mathcal{H}$ of type (D,m) with $D>m\geq 0$ on $\mathbb{G}_m/\mathbb{F}_q$. Then G_{geom} is finite if and only if for*

$$\mathsf{G} := \big(-\mathsf{Gauss}(\psi_{\mathbb{F}_q},\chi_2)\big)^{D+m-1},$$

again with the convention that when $2|q$ we "define" $-\mathsf{Gauss}(\psi_{\mathbb{F}_q},\chi_2) := \sqrt{q}$, the constant field twist $\mathcal{H}\otimes\mathsf{G}^{-\deg/\mathbb{F}_q}$ has finite G_{arith}.

15. Determination of monodromy groups: the case $M=2$

In this section we assume that

$$(15.0.1)\qquad 2|ab,\;\; \gcd(a,b)=1,\;\; n=a+b,\;\; p>2,\;\; q=p^f.$$

In particular, $M=2$, $A=(q^a+1)/2$, $B=(q^b+1)/2$, $\gcd(A,B)=1$. Fix $\alpha,\beta\in\mathbb{Z}$ such that $\alpha A-\beta B=1$ and $2\nmid(\alpha+\beta)$ using Corollary 13.2. With this choice of parameters, the principal objects of this section are the following local systems on $\mathbb{G}_m/\mathbb{F}_p$ and $\mathbb{A}^1/\mathbb{F}_p$, cf. Definition 11.5 and Theorem 12.1.

Definition 15.1. Let us denote by

$$\mathcal{W}(a,b)$$

the arithmetically semisimple local system on $\mathbb{G}_m/\mathbb{F}_p$ whose trace function at $v\in E^\times$, $E/\mathbb{F}_p$ a finite extension, is given by

$$v\mapsto \frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(xw-v^{-\alpha}x^{q^b+1}-v^{\beta}w^{q^a+1}\big).$$

This is $\mathcal{W}(M,A,B)$ introduced in Definition 11.5, but with ψ replaced by $t\mapsto\psi(t/MAB)=\psi(2t)$.

It results from Corollaries 9.2 and 9.3 that $\mathcal{W}(M,A,B)$ is the direct sum

$$\mathcal{W}(M,A,B)=\mathcal{K}l_0\oplus\mathcal{H}_0$$

of descents (in the sense of the beginning of §13), from $\mathbb{G}_m/\mathbb{F}_p(\mu_{MAB})$ to $\mathbb{G}_m/\mathbb{F}_p$, of the Kloosterman sheaf

$$\begin{aligned}&\mathcal{K}l(2, A, B, \chi_2^{\beta}, \chi_2^{\alpha})(-A-B+1)\\ &= \mathcal{K}l_{\psi}\big(\mathsf{Char}(2AB) \smallsetminus (\mathsf{Char}(A, \chi_2^{\beta}) \sqcup \mathsf{Char}(B, \chi_2^{\alpha})\big)(-A-B+1),\end{aligned}$$

see (4.2.1), and the hypergeometric sheaf

$$\begin{aligned}&\mathcal{H}yp(2, A, B, \mathbb{1}, \mathbb{1})(-A-B+2)\\ &= \mathcal{H}yp_{\psi}\big(\mathsf{Char}(2AB) \sqcup \{\mathbb{1}\} \smallsetminus (\mathsf{Char}(A) \sqcup \mathsf{Char}(B)); \mathbb{1}\big)(-A-B+2),\end{aligned}$$

see (5.0.1) which went into the definition of $\mathcal{W}(M, A, B)$, the descents being the relevant systems $\big(\mathrm{gr}_{\mathsf{wt}=2}(R^2(\mathrm{pr}_1)(\mathcal{F}_{\chi,\rho}))\big)(1)$.

Definition 15.2. The Kummer pullback

$$\mathcal{W}^{\star}(a, b) := [MAB]^{\star}\mathcal{W}(a, b)$$

is a lisse sheaf on $\mathbb{A}^1/\mathbb{F}_p$, with trace function at $v \in E$, $E/\mathbb{F}_p$ a finite extension, given by

$$v \mapsto \frac{1}{\#E} \sum_{x,w \in E} \psi_E\big(vxw - x^{q^b+1} - w^{q^a+1}\big).$$

In general, the local system $\mathcal{W}^{\star}(a, b)$ on $\mathbb{A}^1/\mathbb{F}_p$ makes sense for q any power of any prime p, and any positive integers a, b. By Theorem 12.2, $\mathcal{W}^{\star}(a, b)$ is lisse of rank q^{a+b} and pure of weight zero. In this section, our interest is in the case when hypothesis (15.0.1) holds. In the next section, our interest will be in the case when hypothesis (16.0.1) holds.

The explicit trace formulas allow us to prove:

Lemma 15.3. *Given the hypothesis (15.0.1), the following statements hold.*

(i) *Let E be any subfield of $\mathbb{F}_q$. Then the squared absolute value of the trace at $v = 2$ of $\mathcal{W}^{\star}(a, b)$ is $\#E$. Furthermore, the squared absolute value of the trace at $v = 4$ of $\mathcal{W}(a, b)$ is $\#E$.*

(ii) *If $p = 3$ and $E = \mathbb{F}_q$, the square of the trace at $v = 1$ of $\mathcal{W}(a, b)$ is $(-1)^{(q-1)/2}q$.*

Proof. (i) By Definition 15.2, the trace at $v = 2$ on $\mathcal{W}^\star(a,b)$ is

$$\begin{aligned}\frac{1}{\#E}\sum_{x,w\in E}\psi_E(2xw - x^2 - w^2) &= \frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(-(x-w)^2\big)\\ &= \sum_{y\in E}\psi_E(-y^2)\\ &= \sum_{y\in E}\psi_{-1,E}(y^2),\end{aligned}$$

a Gauss sum over E. Hence its squared absolute value is $\#E$. For the statement in $\mathcal{W}(a,b)$ form, just recall $\mathcal{W}^\star(a,b) = [MAB]^\star\mathcal{W}(a,b)$, and note that $MAB \equiv 2(\mathrm{mod}\ (q-1))$, whence $2^{MAB} = 4$ in E, and we are done by using Lemma 14.3.

(ii) By Definition 15.1, the trace at $v = 1$ is

$$\begin{aligned}\frac{1}{\#E}\sum_{x,w\in E}\psi_E(xw - x^2 - w^2) &= \frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(-(x+w)^2\big)\\ &= \sum_{y\in E}\psi_E(-y^2)\\ &= \sum_{y\in E}\psi_{-1,E}(y^2),\end{aligned}$$

a Gauss sum over $E = \mathbb{F}_q$. Hence its square is $(-1)^{(q-1)/2}q$. □

Proposition 15.4. *Given the hypothesis (15.0.1), suppose that for each $\epsilon = \pm$, there is a hypergeometric sheaf $\mathcal{H}_\epsilon$ of rank $(q^n - \epsilon)/2$ in characteristic p with finite geometric monodromy group G_ϵ, which is almost quasisimple. Assume furthermore that $G_\epsilon^{(\infty)}$ is irreducible on $\mathcal{H}_\epsilon$ and that $q^n > 49$. Then, for some $\gamma = \pm$, either $G_\gamma^{(\infty)}$ is a cover of some A_N, or $G_\gamma^{(\infty)}$ is a quotient of $\mathrm{Sp}_{2m}(p^a)$ for some integers $m, a \geq 1$ such that $p^{ma} = q^n$.*

Proof. Let S_ϵ denote the (unique) non-abelian composition factor of G_ϵ, so that $S_\epsilon \lhd G_\epsilon/\mathbf{Z}(G_\epsilon) \leq \mathrm{Aut}(S_\epsilon)$. As G_ϵ is almost quasisimple, its layer $E(G_\epsilon)$ (i.e. the largest semisimple normal subgroup) is equal to the last term $G_\epsilon^{(\infty)}$ in its derived series. Next, since $\mathcal{H}_\epsilon$ is hypergeometric, a generator of $I(0)$ has a simple spectrum on $\mathcal{H}_\epsilon$, whence G_ϵ satisfies the condition ($\star$) of [KT7, §6]. Also, the condition $q^n > 49$ implies that $D_\epsilon := \mathrm{rank}(\mathcal{H}_\epsilon) > 24$. Note that, since $2D_\epsilon + \epsilon$ is a prime power (namely q^n), $D_\epsilon \neq 28$. Hence, by [KT7, Theorem 6.4], S_ϵ is not any of 26 sporadic simple groups. We will now assume that neither $G_+^{(\infty)}$ nor $G_-^{(\infty)}$ is a cover of an alternating group, whence both S_+ and S_- are simple groups of Lie type in characteristic r_+ and r_-, respectively. Now we can apply [KT7, Theorem 6.6] to conclude that there is some power s_ϵ of r_ϵ such that either $S_\epsilon = \mathrm{PSL}_2(s_\epsilon)$, or $E(G_\epsilon)$ is a quotient of $\mathrm{SL}_{m_\epsilon}(s_\epsilon)$, $\mathrm{SU}_{m_\epsilon}(s_\epsilon)$, or $\mathrm{Sp}_{2m_\epsilon}(s_\epsilon)$, and it acts on $\mathcal{H}_\epsilon$ via one of its Weil representations. As $D_\epsilon > 24$, we have $r_+ = p = r_-$ by [KT7, Theorem 7.4]. If furthermore $S_\epsilon = \mathrm{PSp}_{2m_\epsilon}(s_\epsilon)$ with $s_\epsilon m_\epsilon = q^n$ then the statement follows with $\gamma = \epsilon$.

Consider the case $S_\epsilon = \mathrm{PSU}_{m_\epsilon}(s_\epsilon)$ with $m_\epsilon \geq 2$, where we have

$$D_\epsilon = (s_\epsilon^{m_\epsilon} + (-1)^{m_\epsilon} s_\epsilon)/(s_\epsilon + 1) \text{ or } (s_\epsilon^{m_\epsilon} - (-1)^{m_\epsilon})/(s_\epsilon + 1).$$

As $p = r_\epsilon \nmid D_\epsilon$, we must have that $D_\epsilon = (s_\epsilon^{m_\epsilon} - (-1)^{m_\epsilon})/(s_\epsilon + 1)$. Now, if $\epsilon = (-1)^{m_\epsilon}$, then p divides

$$q^n = 2D_\epsilon + \epsilon = 2\frac{s_\epsilon^{m_\epsilon} - (-1)^{m_\epsilon}}{s_\epsilon + 1} + (-1)^{m_\epsilon} = \frac{2s_\epsilon^{m_\epsilon} + (-1)^{m_\epsilon} s_\epsilon - (-1)^{m_\epsilon}}{s_\epsilon + 1},$$

a contradiction as $p|s_\epsilon$. Recall that $n \geq 3$. Hence, if $\epsilon = -(-1)^{m_\epsilon}$, then p^3 divides

$$q^n = 2D_\epsilon + \epsilon = 2\frac{s_\epsilon^{m_\epsilon} - (-1)^{m_\epsilon}}{s_\epsilon + 1} - (-1)^{m_\epsilon} = \frac{2s_\epsilon^{m_\epsilon} - (-1)^{m_\epsilon}(s_\epsilon + 3)}{s_\epsilon + 1},$$

again a contradiction.

It remains to consider the case $S_\epsilon = \mathrm{PSL}_{m_\epsilon}(s_\epsilon)$ with $m_\epsilon \geq 2$, and

$$D_\epsilon = (s_\epsilon^{m_\epsilon} - s_\epsilon)/(s_\epsilon - 1) \text{ or } (s_\epsilon^{m_\epsilon} - 1)/(s_\epsilon - 1)$$

for both $\epsilon = \pm$. As $p = r_\epsilon \nmid D_\epsilon$, we must have that $D_\epsilon = (s_\epsilon^{m_\epsilon} - 1)/(s_\epsilon - 1)$. Now, if $\epsilon = -$, then p divides

$$q^n = 2D_\epsilon + \epsilon = 2\frac{s_\epsilon^{m_\epsilon} - 1}{s_\epsilon - 1} - 1 = \frac{2s_\epsilon^{m_\epsilon} - s_\epsilon - 1}{s_\epsilon - 1},$$

a contradiction as $p|s_\epsilon$. Thus the statement follows with $\gamma = -$. □

Remark 15.5. Note that in the case $q = 3$ of Proposition 15.4, the main result of [KT5] produces a hypergeometric sheaf in characteristic $p = 3$ of rank $(3^n - 1)/2$ and with the geometric monodromy group being a quotient of $\mathrm{GL}_n(3)$.

Next we prove a variation of [KT6, Theorem 6.4]:

Theorem 15.6. *Let $q = p^f$ be a power of a prime $p > 2$, $n \in \mathbb{Z}_{\geq 1}$, and let $L := \mathrm{Sp}_{2n}(q)$ with $(n, q) \neq (1, 3)$. Suppose that $\Phi : G \to \mathrm{GL}_{q^n}(\mathbb{C})$ is a faithful representation of a finite group $G \rhd L$ with the following properties:*

(a) *Φ is a sum of two representations, Φ^+ of degree $(q^n - 1)/2$ and Φ^- of degree $(q^n + 1)/2$;*
(b) *For all $g \in G$, $\mathrm{Tr}(\Phi(g)) \in \mathbb{K} := \mathbb{Q}(\sqrt{(-1)^{(p-1)/2}p})$;*
(c) *$\Phi|_L$ is a total Weil representation; and*
(d) *For all $g \in G$, $|\mathrm{Tr}(\Phi(g))|^2$ is always a power of p.*

Then the following statements hold.

(i) *$\mathbf{C}_G(L) = \mathbf{Z}(G) = C \times \mathbf{Z}(L)$, where $\mathbf{Z}(L) = \langle \boldsymbol{j} \rangle \cong C_2$, and either*

(α) *$|C| \leq 2$, or*

(β) *$p = 3$ divides $|\det(\Phi^\epsilon(G))|$ for each $\epsilon = \pm$, $2 \nmid f$, and furthermore $C \in \{C_3, C_6\}$.*

In all cases, C can be chosen to act via scalars in Φ.

(ii) *Embed L in $\Gamma := \mathrm{Sp}_{2nf}(p)$ and extend $\Phi|_L$ to a total Weil representation $\Gamma \to \mathrm{GL}_{q^n}(\mathbb{C})$ (which we also denote by Φ) using* [KT6, Lemma 6.1]*. Then there exist a divisor $e|f$ and a standard subgroup $H := L \rtimes C_e$ of Γ such that*

$$\mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C}))\Phi(G) = \mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C}))\Phi(H).$$

Proof. (a) Since $\Phi|_L$ is a total Weil representation, the central involution $\boldsymbol{j}$ of L satisfies $\Phi(\boldsymbol{j}) = \kappa \cdot \mathrm{diag}(\mathrm{Id}, -\mathrm{Id})$ for some $\kappa = \pm$. Hence, for any $g \in G$ we have by (b) that

$$\mathrm{Tr}(\Phi(g)) + \kappa\mathrm{Tr}(\Phi(\boldsymbol{j}g)) = 2\mathrm{Tr}(\Phi^+(g)),\ \mathrm{Tr}(\Phi(g)) - \kappa\mathrm{Tr}(\Phi(\boldsymbol{j}g)) = 2\mathrm{Tr}(\Phi^-(g))$$

both belong to $\mathbb{K}$. Thus $\mathrm{Tr}(\Phi^\epsilon(g)) \in \mathbb{K}$ for each $\epsilon = \pm$. Now statement (i) follows from [KT6, Lemma 6.3].

(b) Note that any element in $\mathbf{N}_\Gamma(L)$ preserves the equivalence class of each of the Weil representations $\Phi^\epsilon|_L$, hence it can only induce a field automorphism of L (modulo $\mathrm{Inn}(L)$). The subgroup of all the field automorphisms of L is cyclic of order f, see [GLS, Theorem 2.5.12]. Thus we may assume that there is some $e|f$ such that G induces a cyclic subgroup of field automorphisms of L of order e. Thus the action of G via conjugation on L induces the same automorphism subgroup as of a standard subgroup

$H := L \rtimes \langle\sigma\rangle \cong \mathrm{Sp}_{2n}(q) \rtimes C_e$ of Γ. As $\mathbf{C}_G(L) = \mathbf{Z}(G) = C\mathbf{Z}(L)$, we can write

$$G = \langle CL, g\rangle, \tag{15.6.1}$$

where $g \in G$ induce (via conjugation) the same automorphism of L as of σ. It follows that $\Phi^\epsilon(g)\Phi^\epsilon(\sigma)^{-1}$ centralizes $\Phi^\epsilon(L)$, and so by Schur's lemma we have

$$\Phi^+(g) = \alpha\Phi^+(\sigma),\ \ \Phi^-(g) = \beta\Phi^-(\sigma) \tag{15.6.2}$$

for some $\alpha, \beta \in \mathbb{C}^\times$. As σ has order e, we obtain that

$$\mathbf{Z}(G) \ni \Phi(g^e) = \mathrm{diag}(\alpha^e \cdot \mathrm{Id}, \beta^e \cdot \mathrm{Id}). \tag{15.6.3}$$

On the other hand, $\mathbf{Z}(G)$ has exponent $2d$, where $d := \gcd(p, 3)$. It follows that

$$\alpha^{2de} = \beta^{2de} = 1. \tag{15.6.4}$$

Recall that Φ^ϵ is irreducible over both L and $H = L \rtimes C_e$. Hence by [Is, Lemma (8.14)(c)], for each $\epsilon = \pm$ and for the coset σL we can find $h^\epsilon \in L$ such that

$$\mathrm{Tr}(\Phi^\epsilon(\sigma h^\epsilon)) \neq 0. \tag{15.6.5}$$

Now using (15.6.2) we have

$$\mathrm{Tr}(\Phi^+(gh^+)) = \mathrm{Tr}(\alpha\Phi^+(\sigma)\Phi^+(h^+)) = \mathrm{Tr}(\alpha\Phi^+(\sigma h^+)) = \alpha\mathrm{Tr}(\Phi^+(\sigma h^+)).$$

But $\mathrm{Tr}(\Phi^+(\sigma h^+)) \in \mathbb{K}$ by [Gro, Lemma 13.5], and $\mathrm{Tr}(\Phi^+(gh^+)) \in \mathbb{K}$ as shown in (i). Together with (15.6.5), this shows that $\alpha \in \mathbb{K}$. The same argument applied to h^- shows that $\beta \in \mathbb{K}$. On the other hand, the only roots of unity in $\mathbb{K}$ are ± 1 if $p > 3$, and $\pm\zeta_3^i$, $0 \leq i \leq 2$; in particular, they are $(2d)^{\mathrm{th}}$ roots of unity. Hence, (15.6.4) now implies that $|\alpha|$ and $|\beta|$ both divide $\gcd(2de, 2d) = 2d$.

We have shown that $(\beta/\alpha)^{2d} = 1$. As $2 \nmid d = \gcd(p, 3)$, replacing g by $g\boldsymbol{j}$ if necessary, we obtain that in fact $(\beta/\alpha)^d = 1$. Consider the case $\alpha = \beta$. Then $\Phi(g) = \alpha\Phi(\sigma)$ by (15.6.2). Recalling that $\Phi(C)$ consists of scalar matrices, we see from (15.6.1) that

$$\mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C}))\Phi(G) = \Phi(\mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C})))\Phi(H),$$

as stated.

(c) It remains to consider the case $\alpha \neq \beta$, whence $p = 3$ and $\gamma := \beta/\alpha$ is a primitive cubic root of unity. First we consider the case $3 \nmid e$. Hence (15.6.3) implies that $\mathbf{Z}(G) = \mathbf{Z}(L)C$ contains $\alpha^e \cdot \mathrm{diag}(\mathrm{Id}, \gamma^e \cdot \mathrm{Id})$. This is impossible, since $\Phi(C)$ consists of scalar matrices and $\Phi(\mathbf{Z}(L)) = \langle \kappa \cdot \mathrm{diag}(\mathrm{Id}, -\mathrm{Id})\rangle$.

Next we consider the case $2 \nmid e$. In this case, we can use the same arguments given in parts (i) and (ii) of the proof of [KT6, Theorem 6.4] to show that we can choose g so that $\alpha = \beta$, and the statement follows again.

In the general case, write $e = e_1e_2$ with e_1 being the 3-part of e, and so $2 \nmid e_1$ and $3 \nmid e_2$. Correspondingly, we can also write $g = g_1g_2$ and $\sigma = \sigma_1\sigma_2$, with σ_1 being the 3-part of σ, and g_i inducing the same automorphism of L as of σ_i. Note that $G = \langle G_1, G_2\rangle$ and $H = \langle H_1, H_2\rangle$, where $G_i := \langle CL, g_i\rangle$ and $H_i = L \rtimes \langle \sigma_i\rangle$ for $i = 1, 2$. The above two cases then yield

$$\mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C}))\Phi(G_i) = \Phi(\mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C})))\Phi(H_i)$$

for $i = 1, 2$, whence the statement follows for G. □

Now we can prove the main result concerning the symplectic groups:

Theorem 15.7. *Let $q = p^f$ be a power of a prime $p > 2$, and let $n = a + b$ with $a, b \in \mathbb{Z}_{\geq 1}$, $2|ab$, and $\gcd(a, b) = 1$. Then the following statements hold.*

(a) *Over any finite extension k of $\mathbb{F}_q$, the local system $\mathcal{W}(a, b)$ introduced in Definition 15.1 has geometric and arithmetic monodromy groups*

$$G_{\mathrm{geom}} = G_{\mathrm{arith},k} = \langle \boldsymbol{t}\rangle \times \mathrm{Sp}_{2n}(q),$$

where $\mathrm{Sp}_{2n}(q)$ acts on $\mathcal{W}(a, b)$ via one of its total Weil representations and $\boldsymbol{t}$ acts as the scalar -1 on $\mathcal{W}(a, b)$.

(b) *Let $\mathcal{H}_{even}(a, b)$ and $\mathcal{H}_{odd}(a, b)$ denote the two irreducible subsheaves of even, respectively odd, rank of $\mathcal{W}(a, b)$. Then their geometric and arithmetic monodromy groups are $\mathrm{Sp}_{2n}(q)$ in an even-dimensional irreducible Weil representation, respectively $C_2 \times \mathrm{PSp}_{2n}(q)$ in an odd-dimensional irreducible Weil representation.*

(c) *Over any subfield $k = \mathbb{F}_{q^{1/d}}$ of $\mathbb{F}_q$, the arithmetic monodromy group $G_{\mathrm{arith},k}$ of $\mathcal{W}(a, b)$ over k satisfies $G_{\mathrm{arith},k} = (\langle \boldsymbol{t}\rangle \times \mathrm{Sp}_{2n}(q)) \cdot C_d$, and induces a subgroup C_d of outer field automorphisms of $\mathrm{Sp}_{2n}(q)$. Moreover, $\mathbf{Z}(G_{\mathrm{arith},k}) = \mathbf{Z}(G_{\mathrm{geom}}) = \langle \boldsymbol{t}\rangle \times \mathbf{Z}(\mathrm{Sp}_{2n}(q)) \cong C_2^2$, and*

$$G_{\mathrm{arith},k}/\mathbf{Z}(G_{\mathrm{arith},k}) \cong \mathrm{PSp}_{2n}(q) \rtimes C_d \cong \mathrm{PSp}_{2n}(q) \rtimes \mathrm{Gal}(\mathbb{F}_q/k).$$

(d) *Over any finite extension k of $\mathbb{F}_q$, the local system*

$$\widetilde{\mathcal{W}}(a,b) := \mathcal{W}(a,b) \otimes \mathcal{L}_{\chi_2}$$

has geometric and arithmetic monodromy groups

$$\tilde{G}_{\mathrm{geom}} = \tilde{G}_{\mathrm{arith},k} = \mathrm{Sp}_{2n}(q).$$

(e) *Over any finite extension k of $\mathbb{F}_q$, the local system $\mathcal{W}^\star(a,b)$ introduced in Definition 15.2 has geometric and arithmetic monodromy groups*

$$G^\star_{\mathrm{geom}} = G^\star_{\mathrm{arith},k} = (G_{\mathrm{geom}})^{(\infty)} = \mathrm{Sp}_{2n}(q).$$

(f) *Over any subfield $k = \mathbb{F}_{q^{1/d}}$ of $\mathbb{F}_q$, the arithmetic monodromy groups of $\widetilde{\mathcal{W}}(a,b)$ and of $\mathcal{W}^\star(a,b)$ are isomorphic to $\mathrm{Sp}_{2n}(q) \cdot C_d$, both inducing a subgroup C_d of outer field automorphisms of $\mathrm{Sp}_{2n}(q)$. Moreover, each group X of these two has $\mathbf{Z}(X) = \mathbf{Z}(\mathrm{Sp}_{2n}(q)) \cong C_2$, and*

$$X/\mathbf{Z}(X) \cong \mathrm{PSp}_{2n}(q) \rtimes C_d \cong \mathrm{PSp}_{2n}(q) \rtimes \mathrm{Gal}(\mathbb{F}_q/k).$$

Proof. (i) Let $\Phi : G_{\mathrm{arith},k} \to \mathrm{GL}_{q^n}(\mathbb{C})$ denote the corresponding representation of $G_{\mathrm{arith},k}$ on $\mathcal{W} := \mathcal{W}(a,b)$. By Theorem 11.1, $\Phi \cong \Phi^+ \oplus \Phi^-$, where $\deg(\Phi^\epsilon) = (q^n - \epsilon)/2$ and each of $\Phi^\epsilon(G_{\mathrm{arith},k})$ and $\Phi^\epsilon(G_{\mathrm{geom}})$ is an irreducible almost quasisimple group for $\epsilon = \pm$. As $G_{\mathrm{arith},k}/G_{\mathrm{geom}}$ is cyclic, it follows from [GT, Lemma 2.5] that

$$L := (G_{\mathrm{arith},k})^{(\infty)} = (G_{\mathrm{geom}})^{(\infty)}$$

and $\Phi^\epsilon(L)$ is irreducible, quasisimple.

By Theorem 11.8(i), $\mathrm{Tr}(\Phi(g)) \neq 0$ for all $g \in G_{\mathrm{arith},k}$. Applying [KT6, Proposition 6.7], we conclude that L is quasisimple. Now, as the two irreducible summands of $\mathcal{W}$ are hypergeometric in characteristic p with finite monodromy, we see that G_{geom} contains a p'-element g with simple spectrum of order divisible by MAB.

Assume in addition that $q^n > 49$. Then we can apply Proposition 15.4 to $\Phi^\epsilon(G_{\mathrm{geom}})$. If L is a cover of A_N, then, since $\deg(\Phi^\epsilon) > 24$, we see by Theorem 6.2 and Lemma 9.1 of [KT7] that $N - 1 = \deg(\Phi^+) = \deg(\Phi^-)$, which is impossible. Hence $\Phi^\epsilon(L)$ is a quotient of some $\mathrm{Sp}_{2m_\epsilon}(p^{a_\epsilon})$ with $m_\epsilon a_\epsilon = nf$. Now, using Theorem 11.8(i) and [KT6, Theorem 6.5], we have that

$$L \cong \mathrm{Sp}_{2n/d}(q^d) \text{ for some divisor } d|n, \tag{15.7.1}$$
$$\text{and } \Phi|_L \text{ is a total Weil representation.}$$

Now we consider the remaining case $(n,q) = (3,3)$, where we have $\{\deg(\Phi^+), \deg(\Phi^-)\} = \{13, 14\}$. Using [HM], we see that the quasisimple group L that is irreducible in both Φ^+ and Φ^- either satisfies (15.7.1), or $L \cong \mathrm{SL}_2(13)$. We also note by [KRLT2, Lemma 3.1] that $P(\infty)$ acts on $\mathcal{K}l_0$ as an elementary abelian of order 3^3 and its image intersects $\mathbf{Z}(\Phi^+(G))$ trivially by Proposition 14.1(ii). It follows that the image Q of $P(\infty)$ in G has order divisible by 3^3 and in fact 3^3 divides $|G/\mathbf{Z}(G)|$ which is a divisor of $|\mathrm{Aut}(L/\mathbf{Z}(L))|$. This rules out the latter possibility $L \cong \mathrm{SL}_2(13)$, and thus (15.7.1) always hold.

Now, using Theorem 11.1(i) and [KT6, Lemma 6.3], and taking G to be either G_{geom} or $G_{\mathrm{arith},k}$, we have that

$$\mathbf{C}_G(L) = \mathbf{Z}(G) = C \times \mathbf{Z}(L), \tag{15.7.2}$$

for a cyclic scalar subgroup C, where $|C| \leq 2$ or $p = 3$ and $|C| = 3$, 6.

(ii) Recall that G contains an element g of order divisible by

$$MAB = (q^a + 1)(q^b + 1)/2.$$

Without loss of generality, we may assume that $a > b$, whence $a \geq 2$. It follows from [Zs] that $|G|$ is divisible by a primitive prime divisor ℓ of $p^{2af} - 1$; in particular,

$$\ell \geq 2af + 1 > \max(4, nf) \geq df, \tag{15.7.3}$$

and so ℓ is coprime to $|\mathbf{C}_G(L)|$ because of (15.7.2). As $L \lhd G$, it follows that ℓ divides $|\mathrm{Aut}(L)| = |L| \cdot df$. Together with (15.7.3), this implies that ℓ divides $|L|$. Hence we can find some $1 \leq i \leq n/d$ such that ℓ divides $q^{2di} - 1 = p^{2dif} - 1$. The choice of ℓ now yields that $2af$ divides $2dif$, i.e. $a|di$. But $a > n/2$ and $di \leq n$, so we must have that $a = di$, and so $d|a$. As $d|n = a + b$ by (15.7.1), we also have that $d|b$. Since $\gcd(a,b) = 1$ by (15.0.1), we conclude that $d = 1$. Thus $G \rhd L \cong \mathrm{Sp}_{2n}(q)$.

In the case $G = G_{\mathrm{geom}}$, any central element acts on the two individual subsheaves of rank $(q^n \pm 1)/2$ as an element of p'-order by [KT7, Proposition 7.1], whence $|C| \leq 2$. On the other hand, by Corollary 13.4, some hypergeometric summand of $\mathcal{W}(a,b)$ has nontrivial geometric determinant $\mathcal{L}_{\chi_2}$, hence G_{geom} cannot be perfect. It follows that $G_{\mathrm{geom}} = \langle \boldsymbol{t} \rangle \times L$ with $C = \langle \boldsymbol{t} \rangle \cong C_2$.

(iii) Applying Theorem 11.7(i), (i-bis), and Theorem 11.8(i), and using the results of (ii), we can now deduce from [KT6, Theorem 6.4] that $G_{\mathrm{arith},k} = C_{\mathrm{arith},k} \times L$, where either

(α) $C_{\mathrm{arith},k} = \langle \boldsymbol{t} \rangle \cong C_2$, or
(β) $C_{\mathrm{arith},k} = \langle \boldsymbol{t} \rangle \times \langle \boldsymbol{z} \rangle \cong C_6$, $p = 3$, and $2 \nmid f$.

Suppose we are in the case of (β). As $G_{\mathrm{arith},\mathbb{F}_q} \geq G_{\mathrm{arith},k}$, it follows that $C_{\mathrm{arith},\mathbb{F}_q} \cong C_6$, and that $G_{\mathrm{arith},\mathbb{F}_q}/G_{\mathrm{geom}} \cong C_3$, where $C_3 = \langle \boldsymbol{z} \rangle$ with $\boldsymbol{z}$ acting via as the scalar ζ_3. Thus, modulo G_{geom}, any element $Frob_{v,k}$ of $G_{\mathrm{arith},\mathbb{F}_q}$ is $\boldsymbol{z}^{\deg(k/\mathbb{F}_q)}$; in particular, the element $g := Frob_{1,\mathbb{F}_q}$ over $\mathbb{F}_q$ in $G_{\mathrm{arith},\mathbb{F}_q}$ is $\boldsymbol{z}h$ for some $h \in G_{\mathrm{geom}}$. Recall from (ii) that $G_{\mathrm{geom}} = \langle \boldsymbol{t} \rangle \times \mathrm{Sp}_{2n}(q)$, with $\boldsymbol{t}$ acting as -1 and $\mathrm{Sp}_{2n}(q)$ acting via one of its total Weil representations. Hence, by [GMT, Lemma 2.3], we have that $m := \big(\mathrm{Tr}(\Phi(h))\big)^2 \in \mathbb{Z}_{\neq 0}$, and so $\big(\mathrm{Tr}(\Phi(g))\big)^2 = m\zeta_3^2 \notin \mathbb{Z}$. On the other hand, by Lemma 15.3(ii), the square of the trace at $v = 1$ over $\mathbb{F}_q$ is $\pm q$, a nonzero integer, a contradiction.

Thus (α) must hold for all $k \supseteq \mathbb{F}_q$, and statement (a) is proved completely.

Statement (b) now follows, by inspecting the image of $C_2 \times \mathrm{Sp}_{2n}(q)$ in individual irreducible Weil representations.

(iv) To prove (c), we apply Theorem 15.6 to $\tilde{G} := G_{\mathrm{arith},k}$ to obtain a divisor $e|f$ and a standard subgroup

$$H \cong \mathrm{Sp}_{2n}(q) \rtimes C_e \leq \mathrm{Sp}_{2ne}(q^{1/e}) \leq \mathrm{Sp}_{2nf}(p) \tag{15.7.4}$$

such that

$$\mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C}))\Phi(\tilde{G}) = \mathbf{Z}(\mathrm{GL}_{q^n}(\mathbb{C}))\Phi(H). \tag{15.7.5}$$

By [KT3, Theorem 3.5], there exists $h \in H$ such that $|\mathrm{Tr}(\Phi(h))|^2 = q^{1/e}$. Using (15.7.5), we can write $\Phi(h) = \gamma\Phi(g)$ for some $g \in \tilde{G}$ and $\gamma \in \mathbb{C}^\times$. As g and h both have finite order, γ is a root of unity and so $|\gamma| = 1$. It follows that $|\mathrm{Tr}(\Phi(g))|^2 = |\mathrm{Tr}(\Phi(h)|^2 = q^{1/e}$. Theorem 11.8(i-bis) applied to $\mathcal{W}(a,b)$ over $\mathbb{F}_{q^{1/d}}$ implies that $q^{1/e}$ is a power of $q^{1/d}$, i.e. $e|d$.

On the other hand, by Lemma 15.3(i), there exists $g' \in \tilde{G}$ such that $|\mathrm{Tr}(\Phi(g'))|^2 = q^{1/d}$. Using (15.7.5), we can again write

$$\Phi(g') = \gamma'\Phi(h')$$

for some $h' \in H$ and $\gamma' \in \mathbb{C}^\times$ with $|\gamma'| = 1$. It follows that

$$|\mathrm{Tr}(\Phi(h'))|^2 = |\mathrm{Tr}(\Phi(g')|^2 = q^{1/d}.$$

Note that H embeds in $\mathrm{Sp}_{2ne}(q^{1/e}) \leq \Gamma$ (as a standard subgroup), see (15.7.4). Hence [GMT, Lemma 2.3] applied to $\mathrm{Sp}_{2ne}(q^{1/e})$ implies that $q^{1/d}$ is a power of $q^{1/e}$, i.e. $d|e$.

We have shown that $d = e$. This implies that $G_{\text{arith},k}$ induces the subgroup C_d of outer field automorphisms of $G_{\text{arith},\mathbb{F}_q}/C_{\text{arith},\mathbb{F}_q} \cong L$. On the other hand, the index of $G_{\text{arith},\mathbb{F}_q}$ in $G_{\text{arith},k} = G_{\text{arith},\mathbb{F}_{q^{1/d}}}$ divides d. This can happen only when $G_{\text{arith},k} = G_{\text{arith},\mathbb{F}_q} \cdot C_d$, and that

$$\mathbf{C}_{G_{\text{arith},k}}(L) = \mathbf{Z}(G_{\text{arith},k}) = \mathbf{Z}(G_{\text{arith},\mathbb{F}_q}) = \langle \boldsymbol{t} \rangle \times \mathbf{Z}(L) \cong C_2^2.$$

Hence we can identify the quotient $G_{\text{arith},k}/\mathbf{Z}(G_{\text{arith},k})$ with the subgroup $\mathrm{PSp}_{2n}(q) \rtimes \mathrm{Gal}(\mathbb{F}_q/k)$ inside $\mathrm{Aut}(L)$, proving (c).

(v) Now we prove (d). Recall that $\mathbf{Z}(L) = \langle \boldsymbol{j} \rangle$, where $\boldsymbol{j}$ acts as -1 on the even-rank summand $\mathcal{H}_{even}$ of $\mathcal{W}(a,b)$ and trivially on the odd-rank summand $\mathcal{H}_{odd}$ of $\mathcal{W}(a,b)$. On the other hand, $\Phi(\boldsymbol{t}) = -\mathrm{Id}$, hence $\boldsymbol{tj}$ acts trivially on $\mathcal{H}_{even}$ and as -1 on $\mathcal{H}_{odd}$. Since $G_{\text{arith},k} = \langle \boldsymbol{tj} \rangle \times \mathrm{Sp}_{2n}(q)$ and $\mathrm{Sp}_{2n}(q)$ is perfect, it follows that $\mathcal{H}_{odd}$ has arithmetic determinant $\mathcal{L}_{\chi_2}$ and $\mathcal{H}_{even}$ has trivial arithmetic determinant. Hence, both $\mathcal{H}_{odd} \otimes \mathcal{L}_{\chi_2}$ and $\mathcal{H}_{even} \otimes \mathcal{L}_{\chi_2}$ have trivial arithmetic determinants.

Next, tensoring with $\mathcal{L}_{\chi_2}$ changes the trace at $v \in E^\times$ by a factor of $\chi_2(v) = \pm 1$. In particular, it does not change the absolute value of the trace at any $v \in E^\times$. Furthermore, the $[2]^\star$ Kummer pullbacks of $\mathcal{W}(a,b)$ and $\widetilde{\mathcal{W}}(a,b)$ are isomorphic, and so $\tilde{G}_{\text{geom}}$ has a normal subgroup X of index at most 2, which is also a normal subgroup of G_{geom} of index at most 2. Furthermore, as usual $\tilde{G}_{\text{arith},k}/\tilde{G}_{\text{geom}}$ is cyclic. It follows that

$$(\tilde{G}_{\text{arith},k})^{(\infty)} = (\tilde{G}_{\text{geom}})^{(\infty)} = X^{(\infty)} = L \cong \mathrm{Sp}_{2n}(q).$$

Applying [KT6, Theorem 6.4] to $\tilde{G}_{\text{arith},k}$ and arguing as in (ii), we conclude that $\tilde{G}_{\text{arith},k} = \tilde{C} \times L$, where $\tilde{C} = \langle \tilde{c} \rangle$ and either $\tilde{c} \in \langle \boldsymbol{t} \rangle$, or $p = 3$ and moreover $\tilde{c} \in \langle \boldsymbol{t}, \boldsymbol{z} \rangle$. [Note that condition (b) of [KT6, Theorem 6.4] is seen to be satisfied by applying Galois automorphisms to the two irreducible constituents of different dimensions.] As shown above, $\tilde{c}$ has trivial determinant acting on the two subsheaves of rank $(q^n \pm 1)/2$, and this rules out the case where $p = 3$ but $\tilde{c} \notin \langle \boldsymbol{t} \rangle$. As $\boldsymbol{t}$ has determinant -1 on the odd-rank subsheaf, the case $\tilde{c} = \boldsymbol{t}$ is also impossible. Thus $\tilde{c} = 1$ and $\tilde{G}_{\text{arith},k} = \tilde{G}_{\text{geom}} = L$.

(vi) For (e), we note that $\mathcal{W}^\star(a,b)$ is also arithmetically isomorphic to the $[MAB]^\star$ Kummer pullback of $\widetilde{\mathcal{W}}(a,b)$. Hence $G^\star_{\text{geom}}$ is a normal subgroup of $\tilde{G}_{\text{geom}} = \mathrm{Sp}_{2n}(q)$, with cyclic quotient, and that $G^\star_{\text{arith},k}$ is a subgroup of $\tilde{G}_{\text{arith},k} = \tilde{G}_{\text{geom}}$. It follows that $G^\star_{\text{geom}} = G^\star_{\text{arith},k} = \mathrm{Sp}_{2n}(q)$.

For (f), recall that $\tilde{G}_{\text{arith},\mathbb{F}_p}$ contains $\tilde{G}_{\text{geom}} = \tilde{G}_{\text{arith},\mathbb{F}_q} = \mathrm{Sp}_{2n}(q)$ as a normal subgroup with cyclic quotient of order e that divides $f := \deg(\mathbb{F}_q/\mathbb{F}_p)$.

We now look at the element $g := Frob_{4,\mathbb{F}_p} \in \tilde{G}_{\mathrm{arith},\mathbb{F}_p}$. For any divisor c of f, by Lemma 15.3 the squared absolute value of the trace of $g^c = Frob_{4,\mathbb{F}_{p^c}}$ on $\mathcal{W}(a,b)$, and so on $\widetilde{\mathcal{W}}(a,b)$ as well, is p^c. On the other hand, by (d) and [GMT, Lemma 2.3], the squared absolute value of the trace of any element in $G^\star_{\mathrm{geom}} \leq G_{\mathrm{geom}}$ on $\mathcal{W}^\star(a,b)$ is a power of $q = p^f$. It follows that $g^c \notin \tilde{G}_{\mathrm{geom}}$ whenever c is a proper divisor of f. Hence we conclude that $e = f$. Next, suppose that $\tilde{G}_{\mathrm{arith},\mathbb{F}_p}$ induces a group of order e' of outer (field) automorphisms of $\tilde{G}_{\mathrm{geom}} = \mathrm{Sp}_{2n}(q)$; in particular, $e'|f$. Using Theorem 15.6 and [GMT, Lemma 2.3] (applied to $\mathrm{Sp}_{2ne'}(q^{1/e'}) \geq \mathrm{Sp}_{2n}(q) \rtimes C_{e'}$), we get that the squared absolute value p of the trace of $g = Frob_{4,\mathbb{F}_p}$ on $\widetilde{\mathcal{W}}(a,b)$ is a power of $q^{1/e'} = p^{f/e'}$. It follows that $e' = f$.

Now, if $\mathbb{F}_{q^{1/d}}$ is a subfield of $\mathbb{F}_q$, then $\tilde{G}_{\mathrm{arith},\mathbb{F}_{q^{1/d}}}/\tilde{G}_{\mathrm{geom}}$ is cyclic of order dividing d and $\tilde{G}_{\mathrm{arith},\mathbb{F}_{q^{1/d}}}$ has index at most f/d in $\tilde{G}_{\mathrm{arith},\mathbb{F}_p} = \tilde{G}_{\mathrm{geom}} \cdot C_f$, whence $\tilde{G}_{\mathrm{arith},\mathbb{F}_{q^{1/d}}} = \tilde{G}_{\mathrm{geom}} \cdot C_d$, inducing the subgroup C_d of outer field automorphisms of $\tilde{G}_{\mathrm{geom}}$. It follows that

$$\mathbf{C}_{\tilde{G}_{\mathrm{arith},\mathbb{F}_{q^{1/d}}}}(\tilde{G}_{\mathrm{geom}}) = \mathbf{Z}(\tilde{G}_{\mathrm{arith},\mathbb{F}_{q^{1/d}}}) = \mathbf{Z}(\tilde{G}_{\mathrm{geom}}) \cong C_2,$$

and we can identify the quotient $\tilde{G}_{\mathrm{arith},\mathbb{F}_{q^{1/d}}}/\mathbf{Z}(\tilde{G}_{\mathrm{geom}})$ with the subgroup $\mathrm{PSp}_{2n}(q) \rtimes \mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_{q^{1/d}})$ of $\mathrm{Aut}(\tilde{G}_{\mathrm{geom}})$.

The arithmetic monodromy group of $\mathcal{W}^\star(a,b)$ over $\mathbb{F}_{q^{1/d}}$ can be determined entirely similarly, utilizing Lemma 15.3 for $Frob_{2,\mathbb{F}_p}$. □

Remark 15.8. As mentioned above, [GMT, Lemma 2.3] shows that the square of a total Weil character of $\mathrm{Sp}_{2n}(q)$, q any odd prime power, takes values ± powers of q. This phenomenon is explained in full generality by Theorem 11.10.

16. Determination of monodromy groups: the case $M = q + 1$ and $n \geq 4$

In this section we assume that

$$2 \nmid ab,\ \ \gcd(a,b) = 1,\ \ n = a + b \geq 4,\ \ p \text{ any prime},\ \ q = p^f, \tag{16.0.1}$$

in particular, $M = q + 1$, $A = (q^a + 1)/(q + 1)$, $B = (q^b + 1)/(q + 1)$, $\gcd(A, B) = 1$. Fix $\alpha, \beta \in \mathbb{Z}$ such that

$$\alpha A - \beta B = 1 \text{ and } \alpha + \beta \text{ coprime to } M \tag{16.0.2}$$

using Corollary 13.2. With this choice of parameters, the principal objects of this section are the following local systems on $\mathbb{G}_m/\mathbb{F}_p$ and $\mathbb{A}^1/\mathbb{F}_p$, cf. Definition 11.5 and Theorem 12.1.

Definition 16.1. Let us denote by

$$\mathcal{W}_\alpha(a,b) = \mathcal{W}(a,b) := \mathcal{W}(M,A,B)$$

the arithmetically semisimple local system on $\mathbb{G}_m/\mathbb{F}_p$ whose trace function at $v \in E^\times$, $E/\mathbb{F}_p$ a finite extension, is given by

$$v \mapsto \frac{1}{\#E} \sum_{x,w \in E} \psi_E\big(xw - v^{-\alpha}x^{q^b+1} - v^\beta w^{q^a+1}\big).$$

This system $\mathcal{W}(M,A,B)$ is the descent (cf. the beginning of §13) from $\mathbb{G}_m/\mathbb{F}_p(\mu_{MAB})$ to $\mathbb{G}_m/\mathbb{F}_p$ of the direct sum of the Kloosterman sheaves

$$\begin{aligned} &\mathcal{K}l(M,A,B,\sigma^{-\beta},\sigma^{-\alpha})(-A-B+1) \\ &= \mathcal{K}l_\psi\big(\mathsf{Char}(MAB) \smallsetminus (\mathsf{Char}(A,\sigma^{-\beta}) \sqcup \mathsf{Char}(B,\sigma^{-\alpha}))\big)(-A-B+1) \end{aligned}$$

with $\mathbb{1} \neq \sigma \in \mathsf{Char}(q+1)$, see (4.2.1), and the hypergeometric sheaf

$$\begin{aligned} &\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})(-A-B+2) \\ &= \mathcal{H}yp_\psi\big(\mathsf{Char}(MAB) \sqcup \{\mathbb{1}\} \smallsetminus (\mathsf{Char}(A) \sqcup \mathsf{Char}(B)); \mathbb{1}\big)(-A-B+2), \end{aligned}$$

each summand being the relevant $(\mathrm{gr}_{\mathsf{wt}=2}(R^2(\mathrm{pr}_1)(\mathcal{F}_{\chi,\rho})))(1)$, see (5.0.1). Its Kummer pullback

$$\mathcal{W}^\star(a,b) := [MAB]^\star\mathcal{W}(M,A,B)$$

is a lisse sheaf on $\mathbb{A}^1$, with trace function at $v \in E$, $E/\mathbb{F}_p$ a finite extension, given by

$$v \mapsto \frac{1}{\#E} \sum_{x,w \in E} \psi_E\big(vxw - x^{q^b+1} - w^{q^a+1}\big). \tag{16.1.1}$$

Definition 16.2. When $2 \nmid q$, we also consider the local system

$$\widetilde{\mathcal{W}}(a,b) := \mathcal{W}(a,b) \otimes \mathcal{L}_{\chi_2},$$

where χ_2 is the quadratic character. By Theorem 13.3, the geometric determinant of $\mathcal{W}(a,b)$ is $\mathcal{L}_{\chi_2}$ and the geometric determinant of $\widetilde{\mathcal{W}}(a,b)$ is trivial.

First we prove an analogue of Lemma 15.3:

Lemma 16.3. *Given any odd integers $a, b \geq 1$, the following statements hold.*

(i) *Suppose $p > 2$. Then for any subfield E of $\mathbb{F}_{q^2}$, the squared absolute value of the trace of $Frob_{v,E}$ at $v = 2$ on $\mathcal{W}^\star(a,b)$ as defined in (16.1.1) is $\#E$. If in addition $\gcd(a,b) = 1$, then $\mathcal{W}^\star(a,b) = [MAB]^\star\mathcal{W}(a,b)$, and hence the squared absolute value of the trace of $Frob_{v,E}$ at $v = 4$ on $\mathcal{W}(a,b)$ is $\#E$.*

(ii) *Suppose $p = 2$. Then for any subfield $E = \mathbb{F}_{q^{2/c}}$ of $\mathbb{F}_{q^2}$, the trace of $Frob_{v,E}$ at $v = 0$ on $\mathcal{W}^\star(a,b)$ as defined in (16.1.1) is $\#E$ if $2 \nmid c$ and 0 if $2|c$.*

Proof. (i) First we prove the statement in its $\mathcal{W}^\star(a,b)$ form. By Definition 16.1, the trace at $v = 2$ is

$$\frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(2xw - x^{q^b+1} - w^{q^a+1}\big) = \frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(2xw - x^{q+1} - w^{q+1}\big).$$

Following part (b) of the proof of Theorem 11.8 and taking $s = t := 1/2$, we see that the squared absolute value of this trace is $\#\mathrm{Null}(E)$, where

$$\mathrm{Null}(E) = \Big\{(x,w) \in E^2 \mid x = (w/2)^{1/q} + (w/2)^q,\ w = (x/2)^{1/q} + (x/2)^q\Big\},$$

cf. (11.8.3). We must show that the pair $(x, w := (x/2)^{1/q} + (x/2)^q)$ lies in $\mathrm{Null}(E)$ for any $x \in E$. Assuming $x \in E \subseteq \mathbb{F}_{q^2}$ and $w = (x/2)^{1/q} + (x/2)^q$, we have that

$$(w/2)^{1/q} + (w/2)^q = \big((x/2)^{1/q^2} + (x/2) + (x/2) + (x/2)^{q^2}\big)/2 = 2x/2 = x.$$

Thus $\#\mathrm{Null}(E) = \#E$, and the claim follows for $\mathcal{W}^\star(a,b)$. For $\mathcal{W}(a,b)$, note that $MAB \equiv 2(\mathrm{mod}\ (p-1))$ and so $2^{MAB} = 4$ in E, whence we are done by Lemma 14.3.

(ii) First we show that

$$\sum_{x\in E}\psi_E(x^{q^a+1}) = \begin{cases} \#E, & \text{if } 2 \nmid c, \\ 0, & \text{if } 2|c. \end{cases} \tag{16.3.1}$$

Write $q = p^f$ with $p = 2$. To say that E is a subfield of $\mathbb{F}_{q^2}$ is to say that $c|2f$. If c is odd then $c|f$. Putting $r := p^{f/c}$, we have $E = \mathbb{F}_{r^2}$. In this case, as both a, c are odd, we have $q^a + 1 = r^{ac} + 1 \equiv r + 1 (\mathrm{mod}\ (r^2 - 1))$. Then for $x \in E = \mathbb{F}_{r^2}$, $x^{q^a+1} = x^{r+1} \in \mathbb{F}_{r^2}$, and hence

$$\mathrm{Tr}_{\mathbb{F}_{r^2}/\mathbb{F}_r}(x^{r+1}) = x^{r+1} + x^{r^2+r} = 2x^{r+1} = 0.$$

Thus for $x \in E = \mathbb{F}_{r^2}$,

$$\begin{aligned}\psi_E(x^{q^a+1}) &= \psi\Big(\mathrm{Tr}_{\mathbb{F}_s/\mathbb{F}_2}(x^{q^a+1})\Big) = \psi\Big(\mathrm{Tr}_{\mathbb{F}_r/\mathbb{F}_2}(\mathrm{Tr}_{\mathbb{F}_{r^2}/\mathbb{F}_r}(x^{r+1}))\Big)\\ &= \psi\big(\mathrm{Tr}_{\mathbb{F}_r/\mathbb{F}_2}(0)\big) = \psi(0) = 1.\end{aligned}$$

Hence $\sum_{x\in E} \psi_E(x^{q^a+1}) = \#E$ as claimed.

If c is even, then $2f/c$ divides f, so that E is a subfield of $\mathbb{F}_q$. Therefore, $x^{q^a+1} = x^2$ for any $x \in E$, and so

$$\sum_{x\in E} \psi_E(x^{q^a+1}) = \sum_{x\in E} \psi_E(x^2) = \sum_{x\in E} \psi_E(x) = 0.$$

Now, the trace at $v = 0$ in question is

$$\frac{1}{\#E} \sum_{x,w\in E} \psi_E\big(x^{q^b+1} + w^{q^{a+1}}\big) = \frac{1}{\#E} \sum_{x\in E} \psi_E(x^{q^b+1}) \sum_{w\in E} \psi_E(w^{q^a+1}),$$

and the statement follows from (16.3.1). □

Lemma 16.4. *Let Z be a finite abelian group, $q = p^f$ a prime power, and let $\lambda_0, \lambda_1, \dots, \lambda_q \in \mathrm{Irr}(Z)$.*

(i) *Suppose $\Lambda := \sum_{i=0}^q \lambda_i$ vanishes on $Z \smallsetminus \{1\}$. Then $|Z|$ divides $q + 1$.*

(ii) *Suppose there is some element $z \in Z$ such that $\Lambda = \sum_{i=0}^q \lambda_i$ vanishes on $Z \smallsetminus \{1, z\}$ and $\Lambda(z) = -(q + 1)$. Then $|Z|$ divides $2(q + 1)$.*

(iii) *Suppose $2|n \geq 4$, $\lambda_0^2 = 1_Z$, $(n, q) \neq (4, 2)$, and that*

$$\Sigma := \lambda_0 + D \sum_{i=0}^q \lambda_i,$$

with $D := (q^n - 1)/(q + 1)$, takes values only in

$$\{-q^n, 0, \pm p^i \mid 0 \leq i \leq nf - 1\}$$

on $Z \smallsetminus \{1\}$. Then either $|Z|$ divides $q + 1$, or Z contains an element z with $\lambda_i(z) = -1$ for all $0 \leq i \leq q$. In the latter case, $|Z|$ divides $2(q + 1)$.

(iv) *Suppose* $(n,q)=(4,2)$, $\lambda_0^2=1_Z$, *and that*

$$\Sigma := \lambda_0 + D\sum_{i=0}^{q}\lambda_i,$$

with $D:=(q^n-1)/(q+1)=5$, *takes values only in*

$$\bigl\{0, \pm q^i \mid 0 \leq i \leq n-1\bigr\}$$

on $Z \smallsetminus \{1\}$. *Then either* $|Z|$ *divides* $q+1$, *or* $Z=\{1,z\}\cong C_2$ *with* $\lambda_1(z)=\lambda_2(z)=-\lambda_0(z)$.

Proof. (i) Note that

$$[\Lambda, 1_Z]_Z = \frac{1}{|Z|}\sum_{x\in Z}\Lambda(x) = \frac{q+1}{|Z|}$$

is an integer, whence the statement follows.

(ii) Let α be the linear character of the cyclic subgroup $\langle z\rangle$ sending z to -1. Since Z is abelian, we can find a linear extension β of α to Z. Now

$$[\beta,\Lambda]_Z = \frac{1}{|Z|}\sum_{x\in Z}\beta(x)\overline{\Lambda}(x) = \frac{(q+1)\beta(1)-(q+1)\beta(z)}{|Z|} = \frac{2(q+1)}{|Z|}$$

is an integer, whence the statement follows.

(iii) Consider any $1\neq x\in Z$. By the assumption, $\lambda_0(x)=\pm 1$, and $\Sigma(x)=0$, $-q^n$, or $\pm p^j$ for some $0\leq j\leq nf-1$. Now

$$\mathbb{Z} \ni \Sigma(x)-\lambda_0(x) = D\cdot\Lambda(x),$$

and so $\Lambda(x)=(\Sigma(x)-\lambda_0(x))/D$ is both rational and an algebraic integer, whence

$$\text{(16.4.1)} \qquad D \text{ divides } \Sigma(x)-\lambda_0(x).$$

We will now show that

$$\text{(16.4.2)} \qquad \text{Either } \Sigma(x)=\lambda_0(x) \text{ or } \Sigma(x)=-q^n.$$

Indeed, if $(n,q)\neq(6,2)$, then $p^{nf}-1$ admits a primitive prime divisor ℓ by [Zs]; if $(n,q)=(6,2)$, we take $\ell:=D=21$. In either case, $\ell|D$ and so

ℓ divides $\Sigma(x) - \lambda_0(x)$ by (16.4.1); furthermore, $\ell \geq 5$. Now if $\Sigma(x) = 0$ or $-\lambda_0(x)$, then $|\Sigma(x) - \lambda_0(x)| \leq 2 < \ell$, a contradiction. Next, suppose that $\Sigma(x) = \pm\lambda_0(x)p^j$ with $1 \leq j \leq nf - 1$. Then ℓ divides $p^j \mp 1$. If in addition $(n, q) = (6, 2)$, then $0 \leq j \leq 5$, so $\ell = 21$ cannot divide $p^j - 1$, again a contradiction. Consider now the case $(n, q) \neq (6, 2)$. Then $\ell|(p^{2j} - 1)$ implies by the choice of ℓ that $nf|2j$. However, $1 \leq j < nf$, so we must have $j = nf/2$. In this case,

$$1 \leq |\Sigma(x) - \lambda_0(x)| \leq p^j + 1 = p^{nf/2} + 1 = q^{n/2} + 1 < (q^n - 1)/(q + 1) = D$$

(using $(n, q) \neq (4, 2)$), and this contradicts (16.4.1).

Now, if $\Sigma(x) \neq -q^n$ for all $1 \neq x \in Z$, then by (16.4.2) we have that $\Sigma(x) = \lambda_0(x)$ and $\Lambda(x) = 0$ for all $1 \neq x \in Z$, whence the statement follows from (i).

Consider the case $\Sigma(x) = -q^n$ for some $1 \neq x \in Z$. Then by (16.4.1) we must have that $\lambda_0(x) = -1$, and so

$$\sum_{i=0}^{q}(-\lambda_i(x)) = -\Lambda(x) = (\lambda_0(x) - \Sigma(x))/D = q + 1,$$

implying that all roots of unity $-\lambda_i(x)$ must be 1. Note that Σ is faithful by assumption, and fix an element $z \in Z$ with $\Sigma(z) = -q^n$, which implies that $\lambda_i(z) = -1$ for all i. In this case, $\lambda_i(xz^{-1}) = 1$ for all i, and so $\Sigma(xz^{-1}) = q^n$ and $x = z$ by faithfulness of Σ. We have shown that $\Lambda(x) = -(q + 1)$ for $x = z$, and $\Lambda(x) = 0$ for all $x \in Z \smallsetminus \{1, z\}$, and so the statement follows from (ii).

(iv) We continue to argue as in (iii) and note that (16.4.1) still holds. In particular, this rules out the possibilities $\Sigma(x) = -\lambda_0(x)$, 0, ± 2, $-\lambda_0(x)$, and ± 8 for $1 \neq x \in Z$. Thus $\Sigma(x) \in \{\lambda_0(x), -4\lambda_0(x)\}$ when $1 \neq x \in Z$. Now if $\Sigma(x) \neq -4\lambda_0(x)$ for all $1 \neq x \in Z$, then $\Sigma(x) = \lambda_0(x)$ and $\Lambda(x) = 0$, whence $|Z|$ divides $q + 1$ by (i).

Suppose that $\Sigma(x) = -4\lambda_0(x)$ for some $1 \neq x \in Z$. Then

$$\sum_{i=0}^{2}\lambda_i(x) = \Lambda(x) = (\Sigma(x) - \lambda_0(x))/D = -\lambda_0(x),$$

and so $(\lambda_1(x)/\lambda_0(x)) + (\lambda_2(x)/\lambda_0(x)) = -2$. As $\lambda_i(x)$'s are roots of unity, we must have that

$$\lambda_1(x) = \lambda_2(x) = -\lambda_0(x).$$

Now, fix an element $z \in Z$ with $\Sigma(z) = -4\lambda_0(z)$, for which we then have $\lambda_1(z) = \lambda_2(z) = -\lambda_0(z)$. Then,

$$\lambda_0(xz^{-1}) = \lambda_1(xz^{-1}) = \lambda_2(xz^{-1}),$$

and so $\Sigma(xz^{-1}) = 16\lambda_0(xz^{-1}) \in \{\pm 16\}$, whence $x = z$ by the assumption. Thus, when $y \in Z$ we have that $\Lambda(y)$ is equal to 3 if $y = 1$, $-\lambda_0(z)$ if $y = z$, and 0 otherwise. As in (ii), let α be the linear character of the cyclic subgroup $\langle z \rangle$ sending z to $\lambda_0(z)$, and consider a linear extension β of α to Z. Now

$$[\beta, \Lambda]_Z = \frac{1}{|Z|}\sum_{y \in Z} \beta(y)\overline{\Lambda}(y) = \frac{\beta(1)\Lambda(1) + \beta(z)\overline{\Lambda}(z)}{|Z|} = \frac{2}{|Z|}$$

is an integer, and so $|Z|$ divides 2. Since $z \neq 1$, we have $Z = \{1, z\} \cong C_2$, as stated. □

For any prime power q and any $n \geq 2$, recall that the finite unitary group $\mathrm{GU}(W) = \mathrm{GU}_n(q)$, with $W := \mathbb{F}_{q^2}^n$, admits a *total Weil representation* of degree q^n over $\mathbb{C}$, with character

$$\zeta_{n,q}(g) = (-1)^n(-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g - 1_W)} \tag{16.4.3}$$

for any $g \in \mathrm{GU}_n(q)$, see e.g. [TZ2, (9)]. Fix primitive $(q+1)^{\text{th}}$ roots of unity $\boldsymbol{\varrho} \in \mathbb{C}^\times$ and $\varrho \in \mathbb{F}_{q^2}^\times$. Then $\zeta_{n,q} = \sum_{i=0}^{q} \zeta_{n,q}^i$ is the sum of $q+1$ irreducible *Weil characters* of $\mathrm{GU}_n(q)$, with

$$\zeta_{n,q}^i(g) = \frac{(-1)^n}{q+1}\sum_{l=0}^{q} \boldsymbol{\varrho}^{il}(-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g - \varrho^l \cdot 1_W)} \tag{16.4.4}$$

being the character of the irreducible summand of the total Weil representation of $\mathrm{GU}_n(q)$, on which the generator $\boldsymbol{z} := \varrho \cdot \mathrm{Id}$ acts as the scalar $\boldsymbol{\varrho}^i$, see [TZ2, Lemma 4.1]. More intrinsically, $\mu_{q+1}(\mathbb{F}_{q^2})$ acts on $\mathrm{GU}_n(q)$ by $(\xi, g) \mapsto \xi g$. For each $\mathbb{C}$-valued character χ of $\mu_{q+1}(\mathbb{F}_{q^2})$, the corresponding Weil character $\zeta_{\chi,n}$ is the χ-isotypical component of $\zeta_{n,q}$:

$$\zeta_{\chi,n}(g) = \frac{(-1)^n}{q+1}\sum_{\xi \in \mu_{q+1}(\mathbb{F}_{q^2})} \overline{\chi}(\zeta)(-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g\xi - 1_W)}.$$

If $2|q$ or if $n \geq 3$, then the restrictions ζ_n^i of $\zeta_{n,q}^i$ to $\mathrm{SU}_n(q)$, $0 \leq i \leq q$, are pairwise distinct *irreducible Weil characters* of $\mathrm{SU}_n(q)$, see [TZ2, Lemma 4.7].

Formula (16.4.4) also makes sense for $n = 1$, except that $\zeta^0_{1,q}$ becomes the zero class function on $\mathrm{GU}_1(q)$. With this convention, we note the following branching formulas, which generalize [KT4, (2.0.3)]:

Lemma 16.5. (i) *Let $n = m + l$ with $m, l \in \mathbb{Z}_{\geq 1}$. Then the restriction of $\zeta^i_{n,q}$ to the natural subgroup $\mathrm{GU}_m(q) \times \mathrm{GU}_l(q)$ of $\mathrm{GU}_n(q)$ is*

$$\sum_{\substack{0 \leq r,s \leq q,\\ (q+1)|(r+s-i)}} \zeta^r_{m,q} \boxtimes \zeta^s_{l,q}.$$

(ii) *Let $T = \langle t \rangle$ be a cyclic maximal torus of order $q^n - (-1)^n$ of $\mathrm{GU}_n(q)$, and let β be a generator of the character group $\mathrm{Irr}(T)$. Then the restriction of $\zeta^i_{n,q}$ to T is*

$$\sum_{\substack{0 \leq r < q^n-(-1)^n,\\ (q+1)|(r-i)}} \beta^r + (-1)^n \delta_{i,0} 1_T.$$

Proof. (i) Formula (16.4.3) shows that the restriction of $\zeta_{n,q}$ to the subgroup $\mathrm{GU}_m(q) \times \mathrm{GU}_l(q)$ is $\zeta_{m,q} \boxtimes \zeta_{l,q}$. Now write

$$\boldsymbol{z} = \mathrm{diag}(\boldsymbol{z}_m, \boldsymbol{z}_l)$$

with $\boldsymbol{z}_m = \varrho \cdot \mathrm{Id} \in \mathbf{Z}(\mathrm{GU}_m(q))$ and $\boldsymbol{z}_l = \varrho \cdot \mathrm{Id} \in \mathbf{Z}(\mathrm{GU}_l(q))$. The desired formula then follows by looking up the $\boldsymbol{\varrho}^i$-eigenspace of $\boldsymbol{z}$ in $V_m \boxtimes V_l$, with V_m affording the $\mathrm{GU}_m(q)$-character $\zeta_{m,q}$ and V_l affording the $\mathrm{GU}_l(q)$-character $\zeta_{l,q}$.

(ii) Note that no nontrivial power t^i has eigenvalue 1 on $\mathbb{F}^n_{q^2}$, hence $\zeta_{n,q}(t^i) = (-1)^n$ for $1 \neq t^i \in T$ by (16.4.3), and thus

$$\zeta_{n,q}|_T = \sum_{j=0}^{q^n-(-1)^n-1} \beta^j + (-1)^n 1_T.$$

We can choose t in such a way that $\boldsymbol{z} = t^{(q^n-(-1)^n)/(q+1)}$, and then deduce the stated formula by looking up the $\boldsymbol{\varrho}^i$-eigenspace for $\boldsymbol{z}$ in $\zeta_{n,q}|_T$. □

The *total Weil character* $\sum^q_{i=0} \zeta^i_n$ of $\mathrm{SU}_n(q)$ can be characterized as follows:

Theorem 16.6. *Let p be any prime and q be any power of p. Let $L = \mathrm{SU}_n(q)$ with $n \geq 3$ and $(n, q) \neq (3, 2)$. Suppose ψ is a (not necessarily irreducible) complex character of L such that*

(a) $\psi(1) = q^n$;
(b) $\psi(g) \in \{0, \pm q^i \mid 0 \leq i \leq n\}$ *for all* $g \in L$; *and*
(c) *every irreducible constituent of ψ is among the $q+1$ irreducible Weil characters ζ_n^u, $0 \leq u \leq q$, of L.*

Then ψ is the total Weil character, that is, $\psi = \sum_{u=0}^{q} \zeta_n^u$.

Proof. (i) By assumption (c),

$$\psi = \sum_{u=0}^{q} a_u \zeta_n^u,$$

where $a_u \in \mathbb{Z}_{\geq 0}$. Setting $\kappa := (-1)^n$ and comparing the degrees, we obtain

$$(a_0 - 1)\kappa = \frac{q^n - \kappa}{q+1}\left(q + 1 - \sum_{u=0}^{q} a_u\right);$$

in particular, $a_0 - 1$ is divisible by $(q^n - \kappa)/(q+1)$. On the other hand,

$$\begin{aligned} -1 \leq a_0 - 1 &\leq \frac{\psi(1)}{\zeta_n^0(1)} - 1 = \frac{q^n}{(q^n + q\kappa)/(q+1)} - 1 \\ &\leq \frac{q^3}{q^2 - q} - 1 = \frac{q^2 - q + 1}{q - 1} < \frac{q^3 - 1}{q + 1} \leq \frac{q^n - \kappa}{q + 1}, \end{aligned}$$

since $n \geq 3$ and $(n, q) \neq (3, 2)$. It follows that

$$a_0 = 1, \ \sum_{u=1}^{q} a_u = q. \tag{16.6.1}$$

(ii) Now, view L as $\mathrm{SU}(W)$, where the space $W = \mathbb{F}_{q^2}^n$ is endowed with an L-invariant non-degenerate Hermitian form, and consider the subgroup $H \cong \mathrm{SU}_3(q)$ of L that acts trivially on a non-degenerate $(n-3)$-dimensional subspace of W. An easy induction on $n \geq 3$ using Lemma 16.5(i) and (16.6.1) shows that

$$\psi_H = \sum_{u=0}^{q} b_u \zeta_3^u, \text{ where } b_u := q^{n-3} + \kappa(1 - a_u), \tag{16.6.2}$$

in particular,

$$b_0 = q^{n-3}, \ \sum_{u=1}^{q} b_u = q^{n-2}. \tag{16.6.3}$$

Also, let $d := \gcd(2, q+1)$, $\varepsilon = \zeta_{q+1}$ be a primitive $(q+1)^{\rm th}$ root of unity, and set

$$\Sigma_k := \sum_{u=1}^{q} b_u \varepsilon^{uk}$$

for any $k \in \mathbb{Z}$, which in fact depends only on $k (\bmod\ (q+1))$. Then (16.6.3) implies that

$$\Sigma_0 = q^{n-2}, \ \ |\Sigma_k| \leq q^{n-2}. \tag{16.6.4}$$

(iii) Here we consider the case $q \geq 4$. This ensures that $q-1$ does not divide $q+1$. We will use the character table (and the notation for various conjugacy classes) of H as displayed in [Geck, Table 3.1]. Consider any $k \in \mathbb{Z}$ with $(q+1)/2 \nmid k$. Evaluating ψ at an element of the class $C_6^{(k,-k,0)}$, we have by (b) that

$$\Sigma'_k := 2b_0 + \Sigma_0 + \Sigma_k + \Sigma_{-k} = q^{n-2} + 2q^{n-3} + \Sigma_k + \overline{\Sigma_k} \tag{16.6.5}$$

belongs to

$$\mathcal{V} := \{0, \pm q^i \mid 0 \leq i \leq n\}.$$

Next, as $q \geq 4$, by adding $q+1$ to k if necessary, which does not change Σ_k, we may assume that $(q-1) \nmid k$. Evaluating ψ at an element of the class $C_7^{(k)}$ and using (b) again, we have that $\Sigma_k \in \mathcal{V}$.

Now, if $|\Sigma_k| \leq q^{n-4}$, then

$$q^{n-1} > q^{n-2} + 2q^{n-3} + 2q^{n-4} \geq |\Sigma'_k| \geq q^{n-2} + 2q^{n-3} - 2q^{n-4} > q^{n-2},$$

contradicting (16.6.5). On the other hand, if $\Sigma_k = q^{n-2}$, respectively, $-q^{n-2}$, q^{n-3}, then $\Sigma'_k = q^{n-3}(3q+2)$, $q^{n-3}(2-q)$, $q^{n-3}(q+4)$, respectively, which again contradicts (16.6.5). Together with (16.6.4), this leaves only one possibility that $\Sigma_k = -q^{n-3}$. Now using (16.6.2), we deduce that

$$\sum_{k=0}^{q} a_u \varepsilon^{uk} = 0$$

if $1 \leq k \leq q$ and $k \neq (q+1)/2$. Thus the polynomial

$$f(t) := \sum_{u=0}^{q} a_u t^u \in \mathbb{Z}[t]$$

has ε^k with $1 \leq k \leq q$, $k \neq (q+1)/2$ as roots. Also, $f(1) = \sum_{u=0}^{q} a_u = q+1$ by (16.6.1). If $2|q$, it follows that $f(t)$ is divisible by $(t^{q+1}-1)/(t-1)$, and so $f(t) = \sum_{u=0}^{q} t^u$. If $2 \nmid q$, we have that $f(t)$ is divisible by $(t^{q+1}-1)/(t^2-1)$, whence $f(t) = (at+b)(t^{q-1}+t^{q-3}+\ldots+t^2+1)$ with $a, b \in \mathbb{Q}$. Evaluating at $t=1$ we obtain $a+b=2$. Next, $b = f(0) = a_0 = 1$, and so $a=1$, whence $f(t) = \sum_{u=0}^{q} t^u$ again. In other words, $a_u = 1$ for all u, as stated.

(iv) Assume now that $q=2$. Note that condition (b) implies that ψ is real-valued. However, $\overline{\zeta_n^1} = \zeta_n^2$. It follows from (16.6.1) that $a_1 = a_2 = 1$, as stated.

Finally, we consider the case $q=3$. Then ζ_n^i is real-valued when $i = 0, 2$ and $\overline{\zeta_n^1} = \zeta_n^3$. Again using (16.6.1) and assuming that ψ is not the total Weil character, we must then have that $\psi = \zeta_n^0 + 3\zeta_n^2$, i.e.

$$(a_0, a_1, a_2, a_3) = (1, 0, 3, 0).$$

Now using (16.6.2) and evaluating ψ at an involution $g \in H$, we obtain

$$\psi(g) = 3^{n-2} - 8(-1)^{n-3},$$

which does not belong to $\mathcal{V}$, a contradiction. □

Remark 16.7. The total Weil character $\sum_{i=0}^{q} \zeta_n^i$ of $\mathrm{SU}_n(q)$ is characterized in Theorem 16.6 as the unique character, whose irreducible constituents are among the $q+1$ Weil characters ζ_n^u, $0 \leq u \leq q$, and which takes values only among $0, \pm q^l$, $0 \leq l \leq n$. One may wonder if an analogous characterization can be found for the total Weil character $\zeta_{n,q} = \sum_{i=0}^{q} \zeta_{n,q}^i$ of $\mathrm{GU}_n(q)$:

Is $\zeta_{n,q}$ *the only character of* $\mathrm{GU}_n(q)$, *whose irreducible constituents are among the* $(q+1)^2$ *Weil characters* $\zeta_{n,q}^i \lambda^j$, $0 \leq i, j \leq q$ (where λ is a fixed linear character of order $q+1$ of $\mathrm{GU}_n(q)$, see [TZ2, (10)]), *and which takes values only among* $0, \pm q^l$, $0 \leq l \leq n$?

Suppose $2 \nmid q$ and let $\tilde{\chi}_2 = \lambda^{(q+1)/2}$ denote the unique quadratic character of $\mathrm{GU}_n(q)$. Then certainly $\tilde{\chi}_2 \zeta_{n,q}$ also satisfies the same properties, and in fact, this is the character obtained when we embed $\mathrm{GU}_n(q)$ in $\mathrm{Sp}_{2n}(q)$ and restrict a total Weil character of $\mathrm{Sp}_{2n}(q)$ to $\mathrm{GU}_n(q)$, see e.g. [KT3, Theorem 3.1].

However, there are other sums of irreducible Weil characters of $\mathrm{GU}_n(q)$ that also share the same properties. For instance, consider any $1 \le e \le q$ and the character $\sum_{i=0}^{q} \zeta_{n,q}^i \lambda^{ei}$. We may assume that $\lambda(g) = \boldsymbol{\varrho}^d$ whenever $\det(g) = \varrho^d$, $0 \le d \le q$. For such an element $g \in \mathrm{GU}_n(q)$, by (16.4.4) we have

$$\begin{aligned}\sum_{i=0}^{q} \zeta_{n,q}^i \lambda^{ei}(g) &= \frac{(-1)^n}{q+1} \sum_{i=0}^{q} \boldsymbol{\varrho}^{edi} \sum_{l=0}^{q} \boldsymbol{\varrho}^{il} (-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g - \varrho^l \cdot 1_W)} \\ &= \frac{(-1)^n}{q+1} \sum_{l=0}^{q} (-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g - \varrho^l \cdot 1_W)} \cdot \sum_{i=0}^{q} \boldsymbol{\varrho}^{i(l+de)} \\ &= \frac{(-1)^n}{q+1} \sum_{l=0}^{q} (-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g - \varrho^l \cdot 1_W)} \cdot (q+1)\delta_{l,-de} \\ &= (-1)^n (-q)^{\dim_{\mathbb{F}_{q^2}} \mathrm{Ker}(g - \varrho^{-de} \cdot 1_W)};\end{aligned}$$

in particular, $\sum_{i=0}^{q} \zeta_{n,q}^i \lambda^{ei}$ takes values only among $(-1)^n(-q)^l$, $0 \le l \le q$, as $\zeta_{n,q}$ does.

Another way of describing the character $\sum_{i=0}^{q} \zeta_{n,q}^i \lambda^{ei}$ is this. For each $a \in \mathbb{Z}/(q+1)\mathbb{Z}$, the map of $\mathrm{GU}_n(q)$ to itself given by

$$\gamma_e : g \mapsto g \cdot \det(g)^e \tag{16.7.1}$$

is an endomorphism of $\mathrm{GU}_n(q)$; furthermore, if $\det(g) = \varrho^d$ then

$$\mathrm{Ker}(g - \varrho^{-de} \cdot 1_W) = \mathrm{Ker}(\gamma_e(g) - 1_W).$$

For any representation Φ of $\mathrm{GU}_n(q)$,

$$g \mapsto \Phi(\gamma_e(g)) = \Phi(g \cdot \det(g)^e))$$

is another representation of $\mathrm{GU}_n(q)$. Applying this construction to the total Weil representation with character $\zeta_{n,q}$, we get a new representation whose character is $\sum_{i=0}^{q} \zeta_{n,q}^i \lambda^{ei}$. We also note that, for $e \in \mathbb{Z}/(q+1)\mathbb{Z}$, γ_e is an automorphism of $\mathrm{GU}_n(q)$ precisely when $ne+1$ is invertible in $\mathbb{Z}/(q+1)\mathbb{Z}$.

It has recently been proved in [Lee] that the two aforementioned kinds of characters are the only characters that can share these properties.

Fix a primitive MAB^{th} roots of unity $\varepsilon \in \overline{\mathbb{F}_q}^{\times}$ and $\boldsymbol{\varepsilon} \in \mathbb{C}^{\times}$, and set

$$\xi := \varepsilon^B,\ \nu := \varepsilon^A,\ \boldsymbol{\xi} := \boldsymbol{\varepsilon}^B,\ \boldsymbol{\nu} := \boldsymbol{\varepsilon}^B,$$

so that $\varrho = \xi^A = \nu^B$ and $\boldsymbol{\varrho} = \boldsymbol{\xi}^A = \boldsymbol{\nu}^B$. With this, we can prove the following characterization of the total Weil character $\zeta_{n,q}$ of $\mathrm{GU}_n(q)$, cf. (16.4.3).

Theorem 16.8. *Given the hypothesis* (16.0.1)*, and let*

$$\Phi : G := \mathrm{GU}_n(q) \to \mathrm{GL}_{q^n}(\mathbb{C})$$

be a faithful complex representation that satisfies the following conditions:

(a) $\Phi = \oplus_{j=0}^{q}\Phi_j$, *where* Φ_j *is irreducible of degree* $(q^n-1)/(q+1)+\delta_{j,0}$, *and* Φ_0 *is self-dual if* $1 \in \{a,b\}$;

(b) *There is an element* $g \in G$ *such that* $\Phi_j(g)$ *has spectrum*

$$\left\{\boldsymbol{\varepsilon}^i \mid 0 \le i \le MAB-1,\ (\boldsymbol{\varepsilon}^i)^A \neq \boldsymbol{\varrho}^{\beta j},\ (\boldsymbol{\varepsilon}^i)^B \neq \boldsymbol{\varrho}^{\alpha j}\right\}$$

for $0 \le j \le q$, *and that* $G = \langle [G,G], g\rangle$.

Then there exists an automorphism γ *of* G *such that* $\mathrm{Tr}\big(\Phi(\gamma(h))\big) = \zeta_{n,q}(h)$ *for all* $h \in G$.

Proof. (i) By hypothesis, g has both order and central order equal to

$$MAB = (q^a+1)(q^b+1)/(q+1),$$

and $\Phi_j(g)$ has simple spectrum for $0 \le j \le q$. Applying [KT7, Theorem 8.3], we see that $g\mathbf{Z}(G)$ generates a cyclic maximal torus in $G/\mathbf{Z}(G)$, and, after a suitable conjugation, we may assume that

$$g = \mathrm{diag}\big(\xi^c, \xi^{-qc}, \xi^{q^2c}, \dots, \xi^{(-q)^{a-1}c}, \nu^d, \nu^{-qd}, \nu^{q^2d}, \dots, \nu^{(-q)^{b-1}d}\big)$$

with $c \in \mathbb{Z}/(q^a+1)\mathbb{Z}$ and $d \in \mathbb{Z}/(q^b+1)\mathbb{Z}$. Since g generates G modulo $[G,G]$, $\det(g) = \varrho^{c+d}$ has order $q+1$. Replacing ε by another generator of μ_{MAB} to change ϱ to another element of order $q+1$, we may therefore assume that $c+d \equiv 1 (\mathrm{mod}\ (q+1))$. Now, the condition that g has central order $q^{n-1}+1$ is equivalent to that $1 = \xi^{ic}/\nu^{id} = \varepsilon^{i(cB-dA)}$ if and only if $MAB|i$, i.e.

$$\gcd(c,A) = \gcd(d,B) = \gcd(cB-dA, q+1) = 1. \tag{16.8.1}$$

(ii) The element g belongs to the standard subgroup $\mathrm{GU}_a(q) \times \mathrm{GU}_b(q)$ of G. Hence we can apply Lemma 16.5(i) to $\mathrm{GU}_a(q) \times \mathrm{GU}_b(q)$, and then apply Lemma 16.5(ii) to

$$\begin{aligned} x &:= \mathrm{diag}\big(\xi, \xi^{-q}, \xi^{q^2}, \ldots, \xi^{(-q)^{a-1}}\big) \in \mathrm{GU}_a(q), \\ y &:= \mathrm{diag}\big(\nu, \nu^{-q}, \nu^{q^2}, \ldots, \nu^{(-q)^{b-1}}\big) \in \mathrm{GU}_b(q), \end{aligned}$$

to find that the spectrum of $g = \mathrm{diag}(x, y)$ in a Weil representation with character $\zeta^j_{n,q}$ is the left-hand-side of

$$\left\{ \boldsymbol{\xi}^{rc}\boldsymbol{\nu}^{sd} \;\middle|\; \begin{array}{l} 0 < r \leq q^a, \\ 0 < s \leq q^b, \\ r + s \equiv j (\mathrm{mod}\ M) \end{array} \right\} = [\sqrt[MAB]{1}] \smallsetminus \big([\sqrt[A]{\boldsymbol{\varrho}^{cj}}] \cup [\sqrt[B]{\boldsymbol{\varrho}^{dj}}] \big), \tag{16.8.2}$$

where we denote $[\sqrt[N]{t}] := \{z \in \mathbb{C} \mid z^N = t\}$ for any $t \in \mathbb{C}$. To show that the left-hand-side and the right-hand-side of (16.8.2) are equal, suppose that $\boldsymbol{\xi}^{rc}\boldsymbol{\nu}^{sd} = \boldsymbol{\xi}^{r'c}\boldsymbol{\nu}^{s'd}$ with

$$0 \leq r, r' \leq q^a,\ 0 \leq s, s' \leq q^b,\ r + s \equiv r' + s' \equiv j (\mathrm{mod}\ M). \tag{16.8.3}$$

Then

$$\boldsymbol{\xi}^{(r-r')c} = \boldsymbol{\nu}^{(s'-s)d}, \tag{16.8.4}$$

and so $B(q+1)(r-r')c$ divides $A(q+1) = \mathrm{ord}(\boldsymbol{\xi})$. Since

$$\gcd(A, B) = \gcd(c, A) = 1$$

(see (16.0.1) and (16.8.2)), we can write $r - r' = Au$ for some $u \in \mathbb{Z}$. Likewise, we have $s - s' = Bv$ for some $v \in \mathbb{Z}$, and now, from (16.8.3) and (16.8.4) we obtain

$$cu + dv = 0,\ \ Au + Bv = 0$$

in $\mathbb{Z}/(q+1)\mathbb{Z}$. The determinant $cB - dA$ of this system is invertible in $\mathbb{Z}/(q+1)\mathbb{Z}$ by (16.8.1), hence $u, v \in (q+1)\mathbb{Z}$, i.e. $r = r'$ and $s = s'$. Now we can readily check that, when (r, s) satisfies (16.8.3) with $s = 0$, $\boldsymbol{\xi}^{rc}\boldsymbol{\nu}^{sd}$ runs over $[\sqrt[A]{\boldsymbol{\varrho}^{cj}}]$, and when (r, s) satisfies (16.8.3) with $r = 0$, $\boldsymbol{\xi}^{rc}\boldsymbol{\nu}^{sd}$ runs over $[\sqrt[B]{\boldsymbol{\varrho}^{dj}}]$, and this establishes the equality in (16.8.2).

(iii) Noting $A \equiv a$ and $B \equiv b$ modulo $q+1$ and using (16.0.1), we have that $\alpha n - (\alpha + \beta)b = 1$ and so $(\alpha + \beta)b = \alpha n - 1$ in $\mathbb{Z}/(q+1)\mathbb{Z}$. Recalling $c + d = 1$ in $\mathbb{Z}/(q+1)\mathbb{Z}$ and using (16.0.2) and (16.8.1), we then see that

$$\begin{aligned} (\alpha+\beta)(cb - da) \ &= (\alpha+\beta)\big(b(1-d) - ad\big) \ &&= (\alpha+\beta)(b - nd) \\ &= \alpha n - 1 - (\alpha+\beta)nd &&= n\big(\alpha - (\alpha+\beta)d\big) - 1 \end{aligned}$$

is coprime to $q+1$. Thus

$$\gcd(1+ne, q+1) = 1 \tag{16.8.5}$$

for $e := (\alpha+\beta)d - \alpha = \beta d - \alpha c$. This implies by Remark 16.7 that the map γ_e of (16.7.1) is an automorphism of G. Hence we can replace g by

$$\begin{aligned}\gamma_e(g) &= \boldsymbol{z}^e g \\ &= \mathrm{diag}\big(\xi^{c'}, \xi^{-qc'}, \xi^{q^2c'}, \ldots, \xi^{(-q)^{a-1}c'}, \nu^{d'}, \nu^{-qd'}, \nu^{q^2d'}, \ldots, \nu^{(-q)^{b-1}d'}\big),\end{aligned}$$

where

$$c' := c+Ae = c+A(\beta d-\alpha c) = c(1-\alpha A)+d\beta A = -c\beta B+d\beta A = \beta(dA-cB)$$

and

$$d' := d+Be = d+B(\beta d-\alpha c) = d(1+\beta B)-c\alpha B = d\alpha A-c\alpha B = \alpha(dA-cB).$$

Setting $t := dA - cB$, we have that

$$\gcd(t, q+1) = 1 \tag{16.8.6}$$

by (16.8.1). Now, (16.8.2) applied to c' and d' shows that the spectrum of g in a Weil representation with character $\zeta_{n,q}^j$ becomes

$$[\sqrt[MAB]{1}] \smallsetminus \big([\sqrt[A]{\boldsymbol{\varrho}^{t\beta j}}] \cup [\sqrt[B]{\boldsymbol{\varrho}^{t\alpha j}}]\big). \tag{16.8.7}$$

(iv) Now we will determine the character φ_j of Φ_j. Any irreducible constituent of the restriction $(\Phi_j)|_L$ to $L := [G,G] \cong \mathrm{SU}_n(q)$ has degree dividing $\deg(\Phi_j)$, hence, by [TZ1, Theorem 4.1], it must be equal to $\deg(\Phi_j)$ and in fact $(\Phi_j)|_L$ is an irreducible Weil character of L. Thus $(\varphi_j)|_L = (\zeta_{n,q}^{r_j})|_L$ for some $0 \le r_j \le q$; in fact, $r_j = 0$ if and only if $j = 0$ (by degree comparison). Now, applying [TZ2, Lemma 4.7], we see that

$$\varphi_j = \zeta_{n,q}^{r_j}\lambda^{s_j}$$

with $0 \le s_j \le q$, where $\lambda \in \mathrm{Irr}(\mathrm{GU}_n(q))$ sends $x \in \mathrm{GU}_n(q)$ to $\boldsymbol{\varrho}^d$ whenever $\det(x) = \varrho^d$. Since g now has $\det(g) = \varrho^{1+ne}$, it follows from (16.8.7) that $\Phi_j(g)$ has spectrum

$$[\sqrt[MAB]{1}] \smallsetminus \big(\boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[A]{\boldsymbol{\varrho}^{t\beta r_j}}] \cup \boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[B]{\boldsymbol{\varrho}^{t\alpha r_j}}]\big).$$

But, according to (b) the spectrum of $\Phi_j(g)$ is $[\sqrt[MAB]{1}] \smallsetminus \big([\sqrt[A]{\boldsymbol{\varrho}^{\beta j}}] \cup [\sqrt[B]{\boldsymbol{\varrho}^{\alpha j}}]\big)$. It follows that

$$(16.8.8) \qquad \boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[A]{\boldsymbol{\varrho}^{t\beta r_j}}] \cup \boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[B]{\boldsymbol{\varrho}^{t\alpha r_j}}] = [\sqrt[A]{\boldsymbol{\varrho}^{\beta j}}[\cup[\sqrt[B]{\boldsymbol{\varrho}^{\alpha j}}].$$

Since $n \geq 4$, we may assume that $a > b$ and hence $A > B$. In this case, the set $\boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[A]{\boldsymbol{\varrho}^{t\beta r_j}}]$ of size A cannot be contained in the set $[\sqrt[B]{\boldsymbol{\varrho}^{\alpha j}}]$ of size B. Therefore, there exists some θ that belongs to both $\boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[A]{\boldsymbol{\varrho}^{t\beta r_j}}]$ and $[\sqrt[A]{\boldsymbol{\varrho}^{\beta j}}]$. Now, both these two sets become $\theta \cdot [\sqrt[A]{1}]$, and so they are equal:

$$(16.8.9) \qquad \boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[A]{\boldsymbol{\varrho}^{t\beta r_j}}[= [\sqrt[A]{\boldsymbol{\varrho}^{\beta j}}].$$

Equating the products of all elements in each set (and using $2 \nmid A$), we get

$$(16.8.10) \quad \boldsymbol{\varrho}^{A(1+ne)s_j}\boldsymbol{\varrho}^{t\beta r_j} = \boldsymbol{\varrho}^{\beta j}, \text{ i.e. } A(1+ne)s_j + t\beta r_j = \beta j \text{ in } \mathbb{Z}/(q+1)\mathbb{Z}.$$

Assume in addition that $1 \leq j \leq q$. Then (16.8.8) is an equality of two disjoint unions of two subsets, so (16.8.9) implies

$$(16.8.11) \qquad \boldsymbol{\varrho}^{(1+ne)s_j} \cdot [\sqrt[B]{\boldsymbol{\varrho}^{t\alpha r_j}}] = [\sqrt[B]{\boldsymbol{\varrho}^{\alpha j}}].$$

Again equating products over all elements in each set, we obtain

$$(16.8.12) \qquad B(1+ne)s_j + t\alpha r_j = \alpha j \text{ in } \mathbb{Z}/(q+1)\mathbb{Z}.$$

The system of linear equations (16.8.10) and (16.8.12), in two variables s_j and r_j, has determinant $(1+ne)t(\alpha A - \beta B) = (1+ne)t$, an invertible element in $\mathbb{Z}/(q+1)\mathbb{Z}$ by (16.8.5) and (16.8.6). Hence it has a unique solution $s_j = 0$, $r_j = j/t$.

Assume now that $j = 0$. Then $r_0 = 0$ as noted above. If $b > 1$, then we have $B > 1$, and (16.8.8) and (16.8.9) imply that

$$\boldsymbol{\varrho}^{(1+ne)s_j} \cdot \big([\sqrt[B]{1}] \smallsetminus \{1\}\big) = [\sqrt[B]{1}] \smallsetminus \{1\}.$$

In particular, for some $1 \neq \theta \in [\sqrt[B]{1}]$ we have $\boldsymbol{\varrho}^{(1+ne)s_j}\theta \in [\sqrt[B]{1}]$, whence (16.8.12) also holds, and we can conclude as above that $s_0 = 0$. Suppose $b = 1$. Then (16.8.10) implies that $A(1+ne)s_0 = 0$ and so $(n-1)s_0 = 0$ in $\mathbb{Z}/(q+1)\mathbb{Z}$. We also have in this case that the character $\zeta^0_{n,j}\lambda^{s_0}$ of the self-dual representation Φ_0 is real, whence λ^{s_0} is real, i.e. $2s_0 = 0$ in $\mathbb{Z}/(q+1)\mathbb{Z}$. As $2|n$, we conclude that $s_0 = 0$.

Thus we have shown that $\varphi_j = \zeta_{n,q}^{j/t}$ for $0 \le j \le q$. Hence the character of Φ is

$$\sum_{j=0}^{q} \zeta_{n,q}^{j/t} = \sum_{j=0}^{q} \zeta_{n,q}^{j} = \zeta_{n,q},$$

as stated. □

Proposition 16.9. *Given the hypothesis* (16.0.1), *suppose that for some* $\delta = 0$ *or* 1, *there is a hypergeometric sheaf* $\mathcal{H}$ *of rank* $D = (q^n-1)/(q+1)+\delta$ *in characteristic* p *with finite geometric monodromy group* G, *which is almost quasisimple. Assume furthermore that* $G^{(\infty)}$ *is irreducible on* $\mathcal{H}$ *and that the following conditions hold.*

(α) *If* $(n,q) = (4,2)$, *then* $G/\mathbf{Z}(G)$ *contains an element* g *of order* 9, *and furthermore* $G^{(\infty)}$ *admits only real-valued traces on* $\mathcal{H}$.

(β) *If* $(n,q) = (4,3)$, *then* $G/\mathbf{Z}(G)$ *contains an element* g *of order* 28 *and an elementary abelian subgroup* $Q \cong C_3^4$.

(γ) *If* $(n,q) = (6,2)$, *then* $G/\mathbf{Z}(G)$ *contains an element* g *of order* 33 *and an elementary abelian subgroup* $Q \cong C_2^6$.

Then one of the following statements holds.

(i) $G^{(\infty)}$ *is a cover of some* A_N *with* $N \ge 8$.

(ii) $G^{(\infty)}$ *is a quotient of* $\mathrm{SU}_n(q)$.

(iii) $q = 2$, $\gamma = 0$, *and* $G^{(\infty)}$ *is a quotient of* $\mathrm{SL}_{n/2}(4)$.

Proof. Let S denote the (unique) non-abelian composition factor of G, so that $S \lhd G/\mathbf{Z}(G) \le \mathrm{Aut}(S)$. As G is almost quasisimple, $E(G) = G^{(\infty)}$. Next, since $\mathcal{H}$ is hypergeometric, a generator of $I(0)$ has a simple spectrum on $\mathcal{H}$, whence G satisfies the condition ($\star$) of [KT7, §6].

(A) First we consider the generic case, that is, where $D \ge 23$. Note that the given rank D cannot be equal to 23, 24, or 28, since $n \ge 4$. As $D \ge 23$, it follows from [KT7, Theorem 6.4] that S is not any of 26 sporadic simple groups. We will now assume that $G^{(\infty)}$ is not a cover of an alternating group, whence S is a simple groups of Lie type in some characteristic r. Now we can apply [KT7, Theorem 6.6] to conclude that there is some power s of r such that either $S = \mathrm{PSL}_2(s)$, or $E(G)$ is a quotient of $\mathrm{SL}_m(s)$, $\mathrm{SU}_m(s)$, or $\mathrm{Sp}_{2m}(s)$, and it acts on $\mathcal{H}$ via one of its Weil representations. Furthermore, as $D \ge 23$, we have $r = p$ by [KT7, Theorem 7.3].

(a) Consider the case $S = \mathrm{PSL}_m(s)$ with $m \ge 2$, and

$$D = (s^m - s)/(s-1) \text{ or } (s^m-1)/(s-1).$$

If $D = (q^n+q)/(q+1)$, then $p|D$, whence $(q^n+q)/(q+1) = (s^m-s)/(s-1)$. Comparing the p-part, we obtain $s = q$ and

$$s^{n-2} - \ldots + s^2 - s + 1 = \frac{s^{n-1}+1}{s+1} = \frac{s^{m-1}-1}{s-1} = s^{m-1} + \ldots + s^2 + s + 1.$$

Since $n \geq 4$, it follows that $m \geq 3$, and $-s+1 \equiv s+1 (\mathrm{mod}\, s^3)$, which is impossible.

Suppose $D = (q^n-1)/(q+1)$, i.e. $\gamma = 0$. Then $p|(D+1)$, whence $D = (s^m-1)/(s-1) \equiv 1 (\mathrm{mod}\, p)$ and so $p = 2$. If moreover $s = 2$, then $2^m = D+1 = (q^n+q)/(q+1)$, and so, by comparing the 2-part, we obtain $q = 2^m = D$, which is impossible since $n \geq 4$. Thus $s \geq 4$. If $q \geq 4$, then $D = (q^n-1)/(q+1) \equiv q-1 \equiv -1 (\mathrm{mod}\, 4)$ and, at the same time, $D = (s^m-1)/(s-1) \equiv s+1 \equiv 1 (\mathrm{mod}\, 4)$, a contradiction. If $q = 2$ and $s \geq 8$, then $D = (q^n-1)/(q+1) \equiv -q^2+q-1 \equiv -3 (\mathrm{mod}\, 8)$ and $D = (s^m-1)/(s-1) \equiv s+1 \equiv 1 (\mathrm{mod}\, 8)$, again a contradiction. Thus $(q,s) = (2,4)$ and $n = 2m$, leading to (iii).

(b) Next we consider the case $S = \mathrm{PSp}_{2m}(s)$ with $m \geq 1$ and $2 \nmid s$, and $D = (s^m \pm 1)/2$. In particular, $p \nmid D$, hence

$$D = (q^n-1)/(q+1) \equiv -1 (\mathrm{mod}\, p).$$

Now $2D \equiv -2 (\mathrm{mod}\, p)$, so we must have $D = (s^m+1)/2$ and $p = 3$. Comparing the p-part of $(q^n+q)/(q+1) = (s^m+3)/2$, we get $q = 3$ and $3^n - 2s^m = 3$, a contradiction, as $s > 3$ is a 3-power.

(c) It remains to consider the case $S = \mathrm{PSU}_m(s)$ with $m \geq 2$, and

$$D = (s^m + (-1)^m s)/(s+1) \text{ or } (s^m - (-1)^m)/(s+1).$$

If $D = (q^n+q)/(q+1)$, then $p|D$, whence

$$(q^n+q)/(q+1) = (s^m + (-1)^m s)/(s+1).$$

Comparing the p-part, we obtain $s = q$ and $m = n$, as stated in (ii).

Suppose $D = (q^n-1)/(q+1)$. Then $p \nmid D$, whence

$$D = (s^m - (-1)^m)/(s+1) \equiv (-1)^{m-1} (\mathrm{mod}\, p). \tag{16.9.1}$$

If moreover $2|m$, then we get $(q^n+q)/(q+1) = D+1 = (s^m+s)/(s+1)$, and so $q = s$ by comparing the p-part, whence we also get $m = n$, again leading to (ii). Assume $2 \nmid m$. Then using (16.9.1) and $p|(q^n+q)/(q+1) = D+1$, we see that $p = 2$. Now, if $q, s \geq 4$, then

$$D = (q^n - 1)/(q+1) \equiv q - 1 \equiv -1(\bmod 4)$$

and

$$D = (s^m + 1)/(s+1) \equiv -s + 1 \equiv 1(\bmod 4),$$

a contradiction. Thus either q or s equals to 2. Since

$$(q^n - 1)/(q+1) = (s^m - 1)/(s+1),$$

we also get

$$s + q + 2 = q^n(s+1) - s^m(q+1)$$

is divisible by 8, whence $\{s, q\} = \{2, 4\}$. Now, if $(q, s) = (2, 4)$, then we have $(2^n - 1)/3 = (4^m + 1)/5$ and $5 \cdot 2^n - 3 \cdot 4^m = 8$ with $n \geq 4$ and $m \geq 3$, a contradiction. Finally, if $(q, s) = (4, 2)$, then $(4^n - 1)/5 = (2^m + 1)/3$ and $3 \cdot 4^n - 5 \cdot 2^m = 8$ with $n \geq 4$ and $m \geq 5$, again a contradiction.

(B) Now we consider the remaining cases where $D \leq 22$, that is, where either $(n, q) = (4, 2)$ and $D = 5, 6$, or $(n, q) = (4, 3)$ and $D = 20, 21$, or $(n, q) = (6, 2)$ and $D = 21, 22$.

In the first case, by assumption (α), $G/\mathbf{Z}(G) \leq \mathrm{Aut}(S)$ contains an element g of order 9. This rules out all possible covers $G^{(\infty)}$ of S that can have irreducible representations of degree 5 or 6 by [HM]: $S = \mathsf{A}_{5,6,7}$, $\mathrm{PSL}_2(5, 7, 9, 11, 13)$, $\mathrm{PSL}_3(4)$, $\mathrm{SU}_3(3)$, and J_2, leaving out only the possibilities that $G^{(\infty)} = \mathrm{SU}_4(2)$ or $6_1 \cdot \mathrm{PSU}_4(3)$. The latter case is also ruled out for the reason that $G^{(\infty)}$ would then admit traces $6\zeta_3$.

Next suppose that $D = 20$. By assumption (β), $G/\mathbf{Z}(G) \leq \mathrm{Aut}(S)$ contains an element g of order 28 and a subgroup $Q \cong C_3^4$. This rules out all possible covers $G^{(\infty)}$ of S that can have irreducible representations of degree 20 by [HM]: $S = \mathsf{A}_{7,8}$, $\mathrm{PSL}_2(19, 41)$, $\mathrm{PSL}_3(4)$, $\mathrm{PSU}_3(5)$, and $\mathrm{SU}_4(2)$, leaving out only the possibility that $G^{(\infty)}$ is a quotient of $\mathrm{SU}_4(3)$.

Suppose now that $D = 21$. By assumptions (β) and (γ), either the group $G/\mathbf{Z}(G) \leq \mathrm{Aut}(S)$ contains an element g of order 28 and a subgroup $Q \cong C_3^4$, or $G/\mathbf{Z}(G) \leq \mathrm{Aut}(S)$ contains an element g of order 33 and a subgroup $Q \cong C_2^6$. This rules out all possible covers $G^{(\infty)}$ of S that can have irreducible representations of degree 21 by [HM]: $S = \mathsf{A}_{7,8,9}$, $\mathrm{PSL}_2(41)$, $\mathrm{PSL}_2(43)$, $\mathrm{PSL}_3(4)$, $\mathrm{SU}_3(3)$, $\mathrm{PSU}_3(5)$, $\mathrm{Sp}_6(2)$, M_{22}, and J_2, leaving out only the possibilities that $G^{(\infty)} = \mathrm{PSU}_4(3)$ when $q = 3$ and $G^{(\infty)} = \mathrm{SU}_6(2)$ when $q = 2$.

Finally, let $D = 22$. By assumption (γ), $G/\mathbf{Z}(G) \leq \mathrm{Aut}(S)$ contains an element g of order 33. This rules out all possible covers $G^{(\infty)}$ of S that can have irreducible representations of degree 22 by [HM]: $S = \mathrm{PSL}_2(23)$, $\mathrm{PSL}_2(43)$, M_{22}, HS, and McL, leaving out only the case $G^{(\infty)} = \mathrm{PSU}_6(2)$. □

Proposition 16.10. *Let q be a prime power, $2|n \in \mathbb{Z}_{\geq 4}$, and let L be a perfect finite group with a faithful representation $\Phi : L \to \mathrm{GL}_{q^n}(\mathbb{C})$ that satisfies the following conditions:*

(a) *$\Phi = \oplus_{i=0}^{q} \Phi_i$ is a sum of $q+1$ irreducible constituents, of degree*

$$\deg(\Phi_i) = (q^n - 1)/(q+1) + \delta_{i,0};$$

(b) *Each $L_i := \Phi_i(L)$ is quasisimple, with simple quotient $S_i = L_i/\mathbf{Z}(L_i)$ being either $\mathrm{PSU}_n(q)$ or an alternating group A_{N_i} with $N_i \geq 8$; and*

(c) *$|\mathrm{Tr}(\Phi(g))|$ is always a q-power for all $g \in L$.*

Then $L \cong \mathrm{SU}_n(q)$, and Φ is the total Weil representation.

Proof. (i) First we will construct certain elements in $\mathrm{SU}_n(q)$ and A_N with $N \geq 8$.

Let ρ denote the smallest irreducible character of A_N of degree $N-1$ and labeled by the partition $(N-1, 1)$, and choose $g_1 \in \mathsf{A}_N$ to be a single $(N-2)$-cycle if $2 \nmid N$ and a disjoint product of two $(N-2)/2$-cycles if $2|N$; this ensures that $\rho(g_1) = 1$. Similarly, choose $g_2 \in \mathsf{A}_N$ to be a single $(N-3)$-cycle if $2|N$ and a disjoint product of two $(N-3)/2$-cycles if $2 \nmid N$; this ensures that $\rho(g_2) = 2$.

Next, if $(n, q) \neq (6, 2)$, by [Zs] there exists a primitive prime divisor ℓ of $p^{nf} - 1 = q^n - 1$ (which will then be coprime to $q+1$) and an element $h \in \mathrm{SU}_n(q)$ of order ℓ. Then the character formula [TZ2, Lemma 4.1] for the irreducible Weil characters ζ_n^i of $\mathrm{SU}_n(q)$, of degree $(q^n+q)/(q+1)$ when $i = 0$ and $(q^n-1)/(q+1)$ when $0 < i \leq q$, shows that

$$\zeta_n^i(h) = \delta_{i,0}.$$

The same conclusion holds in the case $(n, q) = (6, 2)$, by taking $\ell = 7$, see the character table of $\mathrm{SU}_6(2)$ [GAP].

(ii) Now we will use [KT6, Proposition 6.7] and modify its proof to our case. First, conditions (a) and (b) imply by [KT6, Proposition 6.7] that

$$L = R_1 * R_2 * \ldots * R_m \tag{16.10.1}$$

is a central product of quasisimple groups, each being a cover of some A_{N_i} or $\mathrm{PSU}_n(q)$.

We aim to show that $m = 1$, that is, L is quasisimple. Assume the contrary: $m > 1$. In accordance with (16.10.1) we can express

$$\Phi_i = \Psi_{i,1} \boxtimes \Psi_{i,2} \boxtimes \ldots \boxtimes \Psi_{i,m}$$

as an outer tensor product of $\Psi_{i,k} \in \mathrm{Irr}(R_k)$, $1 \leq k \leq m$. It follows that $L_i = \Phi_i(L)$ is a central product $\Psi_{i,1}(R_1)*\Psi_{i,2}(R_2)*\ldots*\Psi_{i,m}(R_m)$ of (normal) subgroups. Since L_i is quasisimple and since each R_k is also quasisimple, we conclude that all but one $\Psi_{i,k}$ are trivial, say for all $k \neq k_i$. This implies that

$$L_i = \Phi_i(L) = \Psi_{i,k_i}(R_{k_i}) = \Phi_i(R_{k_i}).$$

On the other hand, the faithfulness of Φ implies that each R_j with $1 \leq j \leq m$ must be acting nontrivially in some Φ_i. So we can partition $\{\Phi_0, \Phi_1, \ldots, \Phi_q\}$ into a disjoint union $\mathcal{X}_1 \sqcup \mathcal{X}_2 \sqcup \ldots \sqcup \mathcal{X}_m$ of non-empty subsets such that for each $1 \leq t \leq m$ and for all $\Phi_i \in \mathcal{X}_t$ we have

$$L_i = \Phi_i(L) = \Phi_i(R_t) \tag{16.10.2}$$

but $\Phi_i(R_{j'})$ is trivial for all $j' \in \{1, 2, \ldots, m\} \smallsetminus \{t\}$. Relabeling the R_j's (and interchanging their order in (16.10.1)) if necessary, we may assume that $\Phi_0 \in \mathcal{X}_1$. Furthermore, since $\deg(\Phi_i) \neq 8, 14$, Theorem 6.2 and Lemma 9.1 of [KT7] imply that if R_t is a cover of A_{N_t} with $N_t \geq 8$ in (16.10.2), then $L_i \cong \mathsf{A}_{N_t}$ and $(\Phi_i)|_{\mathsf{A}_{N_t}}$ is the smallest representation of degree $N_t - 1$. Likewise, [KT7, Theorem 6.6] implies that if R_t is a cover of $\mathrm{PSU}_n(q)$ in (16.10.2), then L_i is a quotient of $\mathrm{SU}_n(q)$ and the $\mathrm{SU}_n(q)$-character afforded by Φ_i is one of the $q + 1$ irreducible Weil characters ζ_n^l, $0 \leq l \leq q$.

Following the proof of [KT6, Proposition 6.7], first we consider the case where

$$\begin{aligned} &\text{for each } 1 \leq t \leq m, \text{ there exists } x_t \in R_t \\ &\text{such that } \mathrm{Tr}(\Phi_i(x_t)) = 0 \text{ for all } \Phi_i \in \mathcal{X}_t. \end{aligned} \tag{16.10.3}$$

Setting $y := x_1 x_2 \ldots x_t$, we see that

$$\mathrm{Tr}(\Phi_i(y)) = \mathrm{Tr}(\Phi_i(x_t)) = 0$$

for all $\Phi_i \in \mathcal{X}_t$. It follows that $\mathrm{Tr}(\Phi(y)) = \sum_{i=0}^q \mathrm{Tr}(\Phi_i(y)) = 0$, contradicting (c).

Next we consider the case R_1 is a cover of some A_{N_1}. Since $\Phi_0 \in \mathcal{X}_1$, we must have

$$N_1 - 1 = \deg(\Phi_0) = (q^n + q)/(q+1).$$

It follows that R_1 cannot have an irreducible character of degree

$$N_1 - 2 = (q^n - 1)/(q+1),$$

and so $\mathcal{X}_1 = \{\Phi_0\}$. It also follows that, for each $t \geq 2$, $\mathcal{X}_t$ consists only of some of the Φ_i of the same degree $(q^n-1)/(q+1)$. Now the elements constructed in (i) guarantee that (16.10.3) holds, and so we are done as above.

We have shown that R_1 is a cover of $\mathrm{PSU}_n(q)$. If, moreover, $\mathcal{X}_1 = \{\Phi_0\}$, then we again see that, for each $t \geq 2$, $\mathcal{X}_t$ consists only of some of the Φ_i of the same degree $(q^n-1)/(q+1)$, whence (16.10.3) holds, and we are done as above. So we may assume that

$$\mathcal{X}_1 \supsetneqq \{\Phi_0\}. \tag{16.10.4}$$

Now we consider the case where some R_j is a cover of some A_{N_j}. As mentioned above, this can happen only when

$$N_j - 1 = \deg(\Phi_i) = (q^n - 1)/(q+1)$$

(for some $i > 0$). Thus we may assume that there is some

$$1 \leq s \leq q$$

such that exactly s representations Φ_i with $i > 0$ occur in (16.10.2) with R_t a cover of A_{N_t}. For any such (quasisimple) R_t, and for any $\Phi_i \in \mathcal{X}_t$, $\Phi_i(R_t) \cong \mathsf{A}_{N_t}$. As $\Phi = \oplus_{i=0}^{q}\Phi_i$ is faithful and $\Phi_{i'}$ is trivial on R_t for all $i' \notin \mathcal{X}_t$, we conclude that $R_t \cong \mathsf{A}_{N_t}$. For any such R_t, we fix an element $g_{t,1} \in R_t$ of type g_1 and an element $g_{t,2} \in R_t$ of type g_2 exhibited in (i).

Each of the remaining R_t is a cover of $\mathrm{PSU}_n(q)$. As mentioned above, the restriction of each $\Phi_i \in \mathcal{X}_t$ is obtained from an irreducible Weil representation of $\mathrm{SU}_n(q)$. Using the faithfulness of Φ, we can view R_t as a quotient of $\mathrm{SU}_n(q)$. For such an R_t, fix an element $g_{t,1} = g_{t,2} \in R_t$ of order ℓ as in (i).

Now, in accordance with (16.10.1) we consider the elements

$$g = g_{1,1}g_{2,1}\cdots g_{m,1},\ g' = g_{1,2}g_{2,2}\cdots g_{m,2}$$

in L. Their construction and the considerations in (i) imply that

$$\mathrm{Tr}(\Phi(g)) = 1+s, \quad \mathrm{Tr}(\Phi(g')) = 1+2s.$$

By (c), both $1+s$ and $1+2s$ are p-powers, and this is impossible since $s \geq 1$.

We have shown that each R_t, $1 \leq t \leq m$, is a cover of $\mathrm{PSU}_n(q)$, hence a quotient of $\mathrm{SU}_n(q)$. Now, in accordance with (16.10.1) we consider the element

$$g'' = h_2 h_3 \ldots h_m,$$

where $h_i \in R_i$ has order ℓ as in (ii) – note that the R_1-component is trivial. Now the considerations in (i) together with (16.10.4) show that

$$q^n > \mathrm{Tr}(\Phi(g'')) = \sum_{\Phi_i \in \mathcal{X}_1} \deg(\Phi_i) > 2(q^n-1)/(q+1) > q^{n-1},$$

again contradicting (c).

(iii) We have shown that L is quasisimple. If L is a cover of A_N, then we see that

$$N-1 = \deg(\Phi_0) = \deg(\Phi_1),$$

which is impossible. Hence each $\Phi_i(L)$ is a quotient of $\mathrm{SU}_n(q)$, and so we can view L as a quotient of $\mathrm{SU}_n(q)$ by a central subgroup, by the faithfulness of Φ. Applying Theorem 16.6 and using the faithfulness of the total Weil character, we conclude that $L = \mathrm{SU}_n(q)$, and it acts in Φ via its total Weil representation. □

Now we can prove the main result concerning unitary groups:

Theorem 16.11. *Let $q = p^f$ be a power of a prime p, and let $n = a+b \geq 4$ with $a, b \in \mathbb{Z}_{\geq 1}$, $2 \nmid ab$, and $\gcd(a,b) = 1$. Then the following statements hold for the arithmetic monodromy groups $G_{\mathrm{arith},k}$, respectively $\tilde{G}_{\mathrm{arith},k}$, and geometric monodromy groups G_{geom}, respectively $\tilde{G}_{\mathrm{geom}}$, of the local systems $\mathcal{W}(a,b)$ and $\widetilde{\mathcal{W}}(a,b)$, introduced in Definitions 16.1 and 16.2, respectively, over any finite extension k of $\mathbb{F}_{q^2}$.*

(a) *$G_{\mathrm{arith},k} = G_{\mathrm{geom}} \cong \mathrm{GU}_n(q)$, and $(G_{\mathrm{geom}})^{(\infty)} \cong \mathrm{SU}_n(q)$ acts on $\mathcal{W}(a,b)$ via its total Weil representation. Furthermore, we can identify G_{geom} with $\mathrm{GU}_n(q)$ in such a way that the action of $\mathrm{GU}_n(q)$ on $\mathcal{W}(a,b)$ affords the total Weil character $\zeta_{n,q}$.*

(b) *Let $\mathcal{H}_i$ be any of the $q+1$ hypergeometric constituents of $\mathcal{W}(a,b)$. Then $\mathcal{H}_i$ has arithmetic and geometric monodromy groups $G^i_{\mathrm{arith},k} = G^i_{\mathrm{geom}}$, $G^i_{\mathrm{geom}}/\mathbf{Z}(G^i_{\mathrm{geom}}) \cong \mathrm{PGU}_n(q)$, and $\mathbf{Z}(G^i_{\mathrm{geom}})$ is cyclic of order dividing $q+1$.*

(c) *$\tilde{G}_{\mathrm{arith},k} = \tilde{G}_{\mathrm{geom}} \cong \mathrm{GU}_n(q)$, and $(\tilde{G}_{\mathrm{geom}})^{(\infty)} \cong \mathrm{SU}_n(q)$ acts on $\widetilde{\mathcal{W}}(a,b)$ via its total Weil representation. Furthermore, we can identify $\tilde{G}_{\mathrm{geom}}$ with $\mathrm{GU}_n(q)$ in such a way that the action of $\mathrm{GU}_n(q)$ on $\widetilde{\mathcal{W}}(a,b)$ affords the total Weil character $\zeta_{n,q}\chi_2$, with χ_2 denoting the linear character of order 2 of $\mathrm{GU}_n(q)$.*

(d) *The local system $\mathcal{W}^\star(a,b)$ introduced in Definition 16.1 has geometric monodromy group and arithmetic monodromy group*

$$G^\star_{\mathrm{arith},k} = G^\star_{\mathrm{geom}} = \mathrm{SU}_n(q).$$

Proof. (i) Let $\Phi : G := G_{\mathrm{arith},k} \to \mathrm{GL}_{q^n}(\mathbb{C})$ denote the corresponding representation of $G_{\mathrm{arith},k}$ on $\mathcal{W} := \mathcal{W}(a,b)$. By Theorem 11.1, $\Phi \cong \oplus_{i=0}^q \Phi_i$, where $\deg(\Phi_i) = (q^n-1)/(q+1) + \delta_{i,0}$, and each of $\Phi_i(G_{\mathrm{arith},k})$ and $\Phi_i(G_{\mathrm{geom}})$ is an irreducible almost quasisimple group for $0 \le i \le q$. As $G_{\mathrm{arith},k}/G_{\mathrm{geom}}$ is cyclic, it follows from [GT, Lemma 2.5] that

$$L := (G_{\mathrm{arith},k})^{(\infty)} = (G_{\mathrm{geom}})^{(\infty)}$$

and $\Phi_i(L)$ is irreducible, quasisimple. Also, by Theorem 11.9, we have that

$$\text{(16.11.1)} \qquad \mathrm{Tr}(\Phi(x)) \text{ is a power of } (-q) \text{ for all } x \in G_{\mathrm{arith},k}.$$

Next, the $q+1$ irreducible summands $\mathcal{H}_i$ of $\mathcal{W}$ are hypergeometric in characteristic p with finite monodromy. Recalling the construction of these sheaves, we see that $G_{\mathrm{geom}} \lhd G$ contains a p'-element g (namely, a generator of the image of $I(0)$), of order $MAB = (q^a+1)(q^b+1)/(q+1)$, with simple spectrum consisting of at least $MAB - A - B = (q^{a+b}-1)/(q+1) < MAB/2$ eigenvalues. Let N_0 denote the order of $g\mathbf{Z}(G)$ in $G/\mathbf{Z}(G)$. Then we have $N_0|MAB$ (as $g^{MAB} = 1$) and $N_0 > MAB/2$ (since the spectrum of g consists of all N_0^{th} roots of some fixed root of unity, but g has more than $MAB/2$ distinct eigenvalues). It follows that $N_0 = MAB$.

We can also check that the assumptions (α)–(γ) of Proposition 16.9 hold in the cases where $(q^n-1)/(q+1) \le 23$, that is, where $(n,q) = (4,2)$, $(4,3)$, and $(6,2)$. Indeed, we can see by Proposition 14.1 that the image Q of $P(\infty)$ acting on any $\mathcal{H}_i$ intersects $\mathbf{Z}(G^i_{\mathrm{geom}})$ trivially, and so $Q \hookrightarrow G^i_{\mathrm{geom}}/\mathbf{Z}(G^i_{\mathrm{geom}})$; furthermore, Q is elementary abelian of order 2^4, 3^4, and 2^6 in these cases by [KRLT2, Lemma 3.1]. Finally, the sheaf $\mathcal{H}_0$ of rank $(q^n+q)/(q+1)$ is self-dual.

Now we can apply Proposition 16.9. In the case of 16.9(iii), we have that $q=2$, $G^{(\infty)}$ is a quotient of $\mathrm{SL}_{n/2}(4)$, and

$$N = (2^a+1)(2^b+1)/3 > (4^{n/2}-1)/3,$$

contradicting [KT7, Theorem 6.6(ii)]. Hence, we conclude that each $\Phi_i(L)$ is either a cover of some A_N or a quotient of $\mathrm{SU}_n(q)$. Now using (16.11.1) and applying Proposition 16.10, we obtain that $L = \mathrm{SU}_n(q)$, and it acts on $\mathcal{W}$ via its total Weil representation.

(ii) In this part of the proof, let H denote either $G_{\mathrm{arith},k}$ or G_{geom}. Since each of $(\Phi_i)|_L$ extends to $H \rhd L$, but only inner-diagonal automorphisms of $\mathrm{SU}_n(q)$ can fix each of the $q+1$ Weil characters ζ_n^i, we see that H can only induce inner-diagonal automorphisms of L. As $\mathbf{C}_H(L) = \mathbf{Z}(H)$, it follows that $\mathrm{PSU}_n(q) \leq H/\mathbf{Z}(H) \leq \mathrm{PGU}_n(q)$, and the same holds for the, arithmetic or geometric, monodromy group K_i of each of the $q+1$ individual hypergeometric sheaves $\mathcal{H}_i$ (as K_i is just the image of H acting on $\mathcal{H}_i$). Since K_i has its $I(0)$ being cyclic of order $MAB = (q^a+1)(q^b+1)/(q+1)$, by [KT7, Theorem 8.3] we must have that $K_i/\mathbf{Z}(K_i) \cong \mathrm{PGU}_n(q)$, and so

$$H/\mathbf{Z}(H) \cong \mathrm{PGU}_n(q). \tag{16.11.2}$$

Now let λ_i be the central character of $\mathbf{Z}(H)$ acting on $\mathcal{H}_i$, $0 \leq i \leq q$. Recall that Φ has integer traces by Theorem 11.9, and so it is self-dual. But Φ_0 is the unique irreducible constituent of Φ of degree $D+1$, hence Φ_0 is self-dual; in particular, λ_0^2 is trivial. Now, Theorem 11.9 implies that $\Sigma := \lambda_0 + D\sum_{i=0}^{q}\lambda_i$ satisfies all the hypotheses of Lemma 16.4; moreover, (16.11.1) rules out the existence of the trace $-q^n$. Hence, by Lemma 16.4, either $\mathbf{Z}(G_{\mathrm{geom}}) \leq \mathbf{Z}(G_{\mathrm{arith},k})$ has order dividing $q+1$, or $(n,q) = (4,2)$ and $\mathbf{Z}(G_{\mathrm{geom}}) \leq \mathbf{Z}(G_{\mathrm{arith},k}) \leq C_2$. Suppose we are in the latter case; in particular, $\mathbf{Z}(G_{\mathrm{geom}}) \leq C_2$. By (16.11.2), $G_{\mathrm{geom}}/\mathbf{Z}(G_{\mathrm{geom}}) \cong \mathrm{SU}_4(2)$ is simple, and so $G_{\mathrm{geom}} \in \{\mathrm{SU}_4(2), \mathrm{Sp}_4(3), C_2 \times \mathrm{SU}_4(2)\}$. On the other hand, by Corollary 13.4, at least one of the sheaves $\mathcal{H}_i$ has geometric determinant of order $M=3$ and so G_{geom} projects onto C_3, a contradiction. Therefore, we have shown that

$$\mathbf{Z}(G_{\mathrm{geom}}) \leq \mathbf{Z}(G_{\mathrm{arith},k}) \text{ has order dividing } q+1; \tag{16.11.3}$$

in particular, $\mathbf{Z}(G^i_{\mathrm{arith},k})$ is cyclic of order dividing $q+1$.

(iii) Recall that $\mathcal{W}^\star(a,b)$ is the $[MAB]^\star$ Kummer pullback of $\mathcal{W}(a,b)$. Hence $G_{\rm geom}/G^\star_{\rm geom}$ is a cyclic group of order dividing MAB; also, $G^\star_{\rm geom}$ has no nontrivial p'-quotient. On the other hand, as shown above, the subgroup $L = (G_{\rm geom})^{(\infty)} = \mathrm{SU}_n(q)$ is a quasisimple normal subgroup of $G_{\rm geom}$, and furthermore, by (16.11.2), $|G_{\rm geom}/L| = |G_{\rm geom}|/|\mathrm{PGU}_n(q)| = |\mathbf{Z}(G_{\rm geom})|$ divides $q+1$, which is coprime to p. It follows that

$$G^\star_{\rm geom} = L = \mathrm{SU}_n(q). \tag{16.11.4}$$

(iv) We now have that $d := |G_{\rm geom}/G^\star_{\rm geom}| = |G_{\rm geom}/L| = |\mathbf{Z}(G_{\rm geom})|$ divides $q+1$. Furthermore, by Corollary 13.4, some hypergeometric summand of $\mathcal{W}(a,b)$, of rank $(q^n-1)/(q+1)$, has geometric determinant $\mathcal{L}_\nu$ with ν of order exactly $M = q+1$. [We note that when $2 \nmid q$, the respective summand of $\widetilde{\mathcal{W}}(a,b)$ will have the same geometric determinant $\mathcal{L}_\nu$, since $\chi_2^{(q^n-1)/(q+1)} = \mathbb{1}$.] This implies that the order d of the quotient $G_{\rm geom}/G^\star_{\rm geom}$ is divisible by $q+1$. We conclude that $d = q+1$, and

$$G_{\rm geom}/G^\star_{\rm geom} \cong C_{q+1}, \quad |\mathbf{Z}(G_{\rm geom})| = q+1. \tag{16.11.5}$$

To determine $G_{{\rm arith},k}$, we note by (16.11.2) that

$$|G_{{\rm arith},k}/L| = |G_{{\rm arith},k}|/|\mathrm{PGU}_n(q)| = |\mathbf{Z}(G_{{\rm arith},k})|$$

which divides $q+1$ by (16.11.3). On the other hand, $G_{{\rm arith},k}$ contains the normal subgroup $G_{\rm geom}$ of order $(q+1)\cdot|L|$. It follows that $G_{{\rm arith},k} = G_{\rm geom}$.

(v) In this part of the proof, we establish the abstract group isomorphism $H := G_{\rm geom} \cong \mathrm{GU}(W) \cong \mathrm{GU}_n(q)$ with $W := \mathbb{F}_{q^2}^n$. First, using (16.11.5) and $L = G^\star_{\rm geom} \cong \mathrm{SU}_n(q)$, we can write

$$H = \langle L, g \rangle \tag{16.11.6}$$

for some element $g \in H$. We can view L as the commutator subgroup of $\mathrm{GU}(W) \cong \mathrm{GU}_n(q)$, and then fix some extension of $(\Phi_j)|_L$ to $\mathrm{GU}_n(q)$, with character $\tilde{\zeta}_{j,n}$ specified in [KT3, (3.1.2)], which we also denote by Φ_j. As mentioned in (iv), $\mathbf{C}_H(L) = \mathbf{Z}(H)$, and H induces the full group of inner-diagonal automorphisms of L, which is the one induced by elements by $\mathrm{GU}_n(q)$ acting on L via conjugation. It follows that we can find an element $h \in \mathrm{GU}_n(q)$ such that

(16.11.7) g and h induce the same automorphism of $L = \mathrm{SU}_n(q)$;

furthermore, changing g to another representative in its coset gG if necessary, we can make sure that

$$h = \mathrm{diag}(\varrho, 1, 1, \ldots, 1) \tag{16.11.8}$$

for some $\varrho \in \mathbb{F}_{q^2}^\times$ of order $q+1$, and so

$$\mathrm{ord}(h) = q+1, \quad L \cap \langle h \rangle = 1. \tag{16.11.9}$$

The choice (16.11.7) of h ensures that $\Phi_j(g)\Phi_j(h)^{-1}$ centralizes $\Phi_j(L)$, whence

$$\Phi_j(g) = \alpha_j \Phi_j(h) \tag{16.11.10}$$

for some $\alpha_j \in \mathbb{C}^\times$. In fact, α_j is a root of unity because both g and h have finite order.

Recall by [KT3, (3.1.2)] (evaluated at h) that $0 \neq \mathrm{Tr}(\Phi_j(h)) \in \mathbb{Q}(\zeta_{q+1})$. On the other hand, since σ is chosen to have order $q+1$, we have that $\mathrm{Tr}(\Phi_j(g)) \in \mathbb{Q}(\zeta_{q+1})$ by Theorem 11.1. Hence the root of unity α_j belongs to $\mathbb{Q}(\zeta_{q+1})$ by (16.11.10). If $2|(q+1)$ then it follows that

$$\alpha_j^{q+1} = 1 \tag{16.11.11}$$

for all j. In the case $2|q$, we have $\alpha_j^{2(q+1)} = 1$. Replacing g by g^2 and h by h^2, which still fulfills (16.11.7)–(16.11.10) and which replaces each α_j by α_j^2, we then see that (16.11.11) holds in this case as well. Together with (16.11.9) and (16.11.10), this implies that $\Phi_j(g)^{q+1} = \mathrm{Id}$ for all j, whence $\Phi(g)^{q+1} = \mathrm{Id}$ and $g^{q+1} = 1$ by faithfulness of Φ. Recalling (16.11.5) and (16.11.6), we must then have that

$$\mathrm{ord}(g) = q+1, \quad L \cap \langle g \rangle = 1. \tag{16.11.12}$$

Thus $H = L \rtimes \langle g \rangle$ and $\mathrm{GU}_n(q) = L \rtimes \langle h \rangle$ are two split extensions of the group $L \cong \mathrm{SU}_n(q)$ by C_{q+1}. Now using (16.11.7), (16.11.9), and (16.11.12), one can readily check that the map $sg^i \mapsto sh^i$, $s \in L$ and $0 \leq i \leq q$, yields a group isomorphism $\iota : H \cong \mathrm{GU}_n(q)$.

(vi) Now, applying Theorem 14.4 to the system $\mathcal{W} := \mathcal{W}(a,b)$ and $N := M$, we see that $\mathcal{W}_M := [M]^\star\mathcal{W}(a,b)$ has arithmetic monodromy group $G_{\mathrm{arith},k,\mathcal{W}_M} = \mathrm{SU}_n(q)$. It follows that the arithmetic monodromy group $G^{\star}_{\mathrm{arith},k}$ of $\mathcal{W}^{\star}(a,b) = [AB]^{\star}\mathcal{W}_M$ is contained in $\mathrm{SU}_n(q) = G^{\star}_{\mathrm{geom}}$, whence $G^{\star}_{\mathrm{arith},k} = G^{\star}_{\mathrm{geom}} = \mathrm{SU}_n(q)$.

To determine $G_{\mathrm{arith},k}$, we note by (16.11.2) that

$$|G_{\mathrm{arith},k}/L| = |G_{\mathrm{arith},k}|/|\mathrm{PGU}_n(q)| = |\mathbf{Z}(G_{\mathrm{arith},k})|$$

which divides $q+1$ by (16.11.3). On the other hand, $G_{\mathrm{arith},k}$ contains the normal subgroup G_{geom} of order $(q+1)\cdot|L|$. It follows that $G_{\mathrm{arith},k} = G_{\mathrm{geom}}$.

(vii) Next we identify the character of G_{geom} on $\mathcal{W}(a,b)$. Let $\langle g_0\rangle$ denote the image of $I(0)$ in $H = G_{\mathrm{geom}}$. Then we can relabel Φ_j so that the spectrum of $\Phi_j(g_0)$ equals

$$\Big\{\boldsymbol{\varepsilon}^i \mid 0 \le i \le MAB-1,\ (\boldsymbol{\varepsilon}^i)^A \ne \boldsymbol{\varrho}^{\beta j},\ (\boldsymbol{\varepsilon}^i)^B \ne \boldsymbol{\varrho}^{\alpha j}\Big\},$$

and furthermore Φ_0 is self-dual. Note that, since H/L is cyclic, $\langle L, g_0\rangle$ is normal in H and so contains the normal closure of $\langle g_0\rangle$ in H. But the normal closure of $\langle g_0\rangle$ in H equals H by [KT7, Theorem 4.1], hence $\langle L, g_0\rangle = H$. Now we can apply Theorem 16.8 to obtain an automorphism γ of H such that $\mathrm{Tr}\big(\Phi(\gamma(x))\big) = \zeta_{n,q}(x)$ for all $x \in H$. Thus, adjusting the identification $\iota : H \cong \mathrm{GU}_n(q)$ by γ, we see that $H \cong \mathrm{GU}_n(q)$ acts on $\mathcal{W}(a,b)$ with the total Weil character $\zeta_{n,q}$.

(viii) Now we again assume $p > 2$ and turn our attention to $\widetilde{\mathcal{W}}(a,b)$. The arguments in (i), (ii) also apply to $\tilde{G}_{\mathrm{arith},k}$ and $\tilde{G}_{\mathrm{geom}}$. The only difference is that instead of (16.11.1) we can now say only that all traces are $\pm q^m$, $0 \le m \le n$.

Hence, when we apply Lemma 16.4(iii), we cannot (yet) rule out the existence of the trace $-q^n$, and so, instead of (16.11.3), we now have

$$\mathbf{Z}(G_{\mathrm{geom}}) \le \mathbf{Z}(G_{\mathrm{arith},k}) \text{ has order dividing } 2(q+1).$$

But now we note that the sheaf $[M]^{\star}\widetilde{\mathcal{W}}(a,b)$ is arithmetically isomorphic to $[M]^{\star}\mathcal{W}(a,b) = \mathcal{W}_M$. Hence $L = \mathrm{SU}_n(q) = G_{\mathrm{geom},\mathcal{W}_M} = G_{\mathrm{arith},k,\mathcal{W}_M}$ is a normal subgroup of $\tilde{G}_{\mathrm{geom}}$ of index dividing M and a subgroup of $\tilde{G}_{\mathrm{arith},k}$ of index dividing M. With this extra information, the arguments in (iv), (v) can now be repeated verbatim to show that $\tilde{G}_{\mathrm{geom}} \cong \mathrm{GU}_n(q)$; in particular, $|\tilde{G}_{\mathrm{geom}}/L| = M$. As $\tilde{G}_{\mathrm{arith},k} \ge \tilde{G}_{\mathrm{geom}}$ and $[\tilde{G}_{\mathrm{arith},k} : L]|M$, we conclude that $\tilde{G}_{\mathrm{arith},k} = \tilde{G}_{\mathrm{geom}}$.

To identify the character $\tilde{\varphi}$ of $\tilde{G}_{\mathrm{geom}}$ acting on $\widetilde{\mathcal{W}}(a,b)$, let $\langle g_0\rangle$ denote the image of $I(0)$ in $\tilde{H} := \tilde{G}_{\mathrm{geom}}$. Again applying [KT7, Theorem 4.1], we see that g_0 generates $\tilde{H}$ modulo $[\tilde{H}, \tilde{H}]$; in particular, $\chi_2(g_0) = -1$. Note that tensoring with $\mathcal{L}_{\chi_2}$ has the effect of multiplying the eigenvalues of g_0 by -1. It follows that, the eigenvalues of g_0 in a representation of $\tilde{H}$ affording the character $\tilde{\varphi}\chi_2$ are the same as the eigenvalues of g_0 acting on $\mathcal{W}(a,b)$. By the result of (vii), we know that $\tilde{\varphi}\chi_2 = \zeta_{n,q}$, hence $\tilde{\varphi} = \zeta_{n,q}\chi_2$ as stated in (c). □

The final result of this section determines the arithmetic monodromy groups of $\mathcal{W}(a,b)$, $\widetilde{\mathcal{W}}(a,b)$, and $\mathcal{W}^\star(a,b)$.

Theorem 16.12. *Let $q = p^f$ be a power of a prime p, and let $n = a+b \geq 4$ with $a,b \in \mathbb{Z}_{\geq 1}$, $2 \nmid ab$, and $\gcd(a,b) = 1$. Then over any subfield $k = \mathbb{F}_{q^{2/d}}$ of $\mathbb{F}_{q^2}$ the following statements hold.*

(i) *The arithmetic monodromy group $G_{\mathrm{arith},k}$ of $\mathcal{W}(a,b)$, respectively $\tilde{G}_{\mathrm{arith},k}$ of $\widetilde{\mathcal{W}}(a,b)$, is $\mathrm{GU}_n(q) \cdot C_d$, which in each case induces a subgroup of outer field automorphisms of $\mathrm{SU}_n(q)$ of order d. Furthermore,*

$$G_{\mathrm{arith},k}/\mathbf{Z}(\mathrm{GU}_n(q)) \cong \tilde{G}_{\mathrm{arith},k}/\mathbf{Z}(\mathrm{GU}_n(q)) \cong \mathrm{PGU}_n(q) \rtimes \mathrm{Gal}(\mathbb{F}_{q^2}/k).$$

(ii) *The arithmetic monodromy group $G^\star_{\mathrm{arith},k}$ of the local system $\mathcal{W}^\star(a,b)$ is $\mathrm{SU}_n(q) \cdot C_d$, and induces a subgroup of outer field automorphisms of $\mathrm{SU}_n(q)$ of order d, modulo the inner-diagonal automorphisms of $\mathrm{SU}_n(q)$.*

Proof. (i) First we determine the order of cyclic quotients $G_{\mathrm{arith},k}/G_{\mathrm{geom}}$ and $G^\star_{\mathrm{arith},k}/G^\star_{\mathrm{geom}}$.

Suppose that $p > 2$. Recall that $G_{\mathrm{arith},\mathbb{F}_p}$ contains $G_{\mathrm{geom}} = G_{\mathrm{arith},\mathbb{F}_{q^2}}$ as a normal subgroup with cyclic quotient of order e that divides the integer $2f := \deg(\mathbb{F}_{q^2}/\mathbb{F}_p)$. We now look at the element $g := Frob_{4,\mathbb{F}_p} \in G_{\mathrm{arith},\mathbb{F}_p}$. For any divisor c of $2f$, by Lemma 16.3 the absolute value of the trace of $g^c = Frob_{4,\mathbb{F}_{p^c}}$ on $\mathcal{W}(a,b)$ is $p^{c/2}$. On the other hand, by Theorem 16.11(a), the absolute value of the trace of any element in G_{geom} on $\mathcal{W}(a,b)$ is a power of $q = p^f$. It follows that $g^c \notin G_{\mathrm{geom}}$ whenever c is a proper divisor of $2f$. Hence we conclude that $e = 2f$. Now, since $k = \mathbb{F}_{q^{2/d}}$ is a subfield of $\mathbb{F}_{q^2}$, then $G_{\mathrm{arith},k}/G_{\mathrm{geom}}$ is cyclic of order dividing d and $G_{\mathrm{arith},k}$ has index at most $2f/d$ in $G_{\mathrm{arith},\mathbb{F}_p} = G_{\mathrm{geom}} \cdot C_{2f}$, whence $G_{\mathrm{arith},k} = G_{\mathrm{geom}} \cdot C_d$.

The structure of $G^\star_{\mathrm{arith},k}/G^\star_{\mathrm{geom}}$ can be determined entirely similarly, utilizing Lemma 16.3 for $Frob_{2,\mathbb{F}_p}$.

Next we assume $p = 2$ and consider the element $h \in G^\star_{\mathrm{arith},\mathbb{F}_2}$ provided by $Frob_{0,\mathbb{F}_2}$. By Lemma 16.3(ii), when $c|2f$ the trace of $h^{2f/c} = Frob_{0,\mathbb{F}_{q^{2/c}}}$

on $\mathcal{W}^\star(a,b)$ is 0 if $2|c$, and $q^{2/c}$ if $2 \nmid c$. In particular, if $c > 1$ this trace is not a power of $-q$, and so $h^{2f/c} \notin G^\star_{\mathrm{geom}}$ by the result of Theorem 16.11(d). Thus $h^{2f} \in G^\star_{\mathrm{geom}}$ but $h^{2f/c} \notin G^\star_{\mathrm{geom}}$ for any $1 < c|2f$. It follows that $|G^\star_{\mathrm{arith},\mathbb{F}_2}/G^\star_{\mathrm{geom}}| = 2f$, and more generally $|G^\star_{\mathrm{arith},k}/G^\star_{\mathrm{geom}}| = d$, as above.

To show that $|G_{\mathrm{arith},\mathbb{F}_2}/G_{\mathrm{geom}}| = 2f$, we note that $G^\star_{\mathrm{arith},\mathbb{F}_2}$ is a subgroup of $G_{\mathrm{arith},\mathbb{F}_2}$. Thus the element $h \in G^\star_{\mathrm{arith},\mathbb{F}_2}$ also lies in $G_{\mathrm{arith},\mathbb{F}_2}$ and moreover the representation of $G_{\mathrm{arith},\mathbb{F}_2}$ on $\mathcal{W}(a,b)$ restricts to the representation of $G^\star_{\mathrm{arith},\mathbb{F}_2}$ on $\mathcal{W}^\star(a,b)$. So viewing h as lying in $G_{\mathrm{arith},\mathbb{F}_2}$, for each divisor c of $2f$ with $c > 1$, the trace of $h^{2f/c}$ on $\mathcal{W}(a,b)$ is not a $(-q)$-power and so $h^{2f/c} \notin G_{\mathrm{geom}}$. Hence $|G_{\mathrm{arith},\mathbb{F}_2}/G_{\mathrm{geom}}| = 2f$, and we can conclude as above.

(ii) Let Φ denote the representation of $G_{\mathrm{arith},\mathbb{F}_p}$ on $\mathcal{W}(a,b)$, with character, say, φ. Next we show that $G_{\mathrm{arith},\mathbb{F}_p}$ cannot contain any element z which acts as the scalar -1 on $\mathcal{W}(a,b)$. Assume the contrary. First, by Theorem 16.11(a), no element in G_{geom} can have trace $-q^n$ on $\mathcal{W}(a,b)$, hence $z \notin G_{\mathrm{geom}}$. Now, if $p > 2$, then, as shown in (i), $G_{\mathrm{arith},\mathbb{F}_p} = \langle g, G_{\mathrm{geom}}\rangle$. Hence we can find $0 \le j \le 2f-1$ such that $z \in g^j G_{\mathrm{geom}}$. As $z^2 \in G_{\mathrm{geom}}$ but $z \notin G_{\mathrm{geom}}$, we have $g^{2j} \in G_{\mathrm{geom}}$ with $j > 0$, which implies $j = f$ by (i). Thus $g^f = zg_0$ for some $g_0 \in G_{\mathrm{geom}} = \mathrm{GU}_n(q)$. As $\Phi(z) = -\mathrm{Id}$, we then obtain that $\varphi(g^f) = -\varphi(g_0)$. But this is a contradiction, since

$$|\varphi(g^f)| = \big|\mathrm{Trace}\big(Frob_{4,\mathbb{F}_q}|\mathcal{W}(a,b)\big)\big| = \sqrt{q}$$

as mentioned in (i), whereas $\varphi(g_0)$ is a power of $-q$ by Theorem 16.11(a). Similarly, if $p = 2$, then, as shown in (i), $G_{\mathrm{arith},\mathbb{F}_2} = \langle h, G_{\mathrm{geom}}\rangle$. Hence we can again find $0 \le j \le 2f-1$ such that $z \in h^j G_{\mathrm{geom}}$. As $z^2 \in G_{\mathrm{geom}}$ but $z \notin G_{\mathrm{geom}}$, we have $h^{2j} \in G_{\mathrm{geom}}$ with $j > 0$, which implies $j = f$ again by (i). Thus $h^f = zg_0$ for some $g_0 \in G_{\mathrm{geom}} = \mathrm{GU}_n(q)$. As $\Phi(z) = -\mathrm{Id}$, we then obtain that $\varphi(h^f) = -\varphi(g_0)$. But this is a contradiction, since

$$\varphi(h^f) = \mathrm{Trace}\big(Frob_{0,\mathbb{F}_q}|\mathcal{W}^\star(a,b)\big) = 0$$

as mentioned in (i), whereas $\varphi(g_0)$ is a power of $-q$, in particular nonzero, by Theorem 16.11(a).

(iii) Now we study the subgroup

$$Z_d := \mathbf{C}_{G_{\mathrm{arith},\mathbb{F}_{q^{2/d}}}}(\mathrm{SU}_n(q))$$

for any $d|2f$, and aim to show that

$$Z_{2f} \cong C_{q+1}. \tag{16.12.1}$$

Recall by Theorem 16.11(a) that the restriction of Φ to $\mathrm{SU}_n(q)$ is a sum of $q+1$ pairwise non-isomorphic irreducible Weil representations. It follows that Z_d fixes each of these $q+1$ summands, and acts via scalars on each of them, inducing a linear character λ_i, $0 \leq i \leq q$. In particular, Z_d is a finite abelian group. We can label these characters so that λ_0 corresponds to the unique hypergeometric summand $\mathcal{H}_0$ (of rank $(q^n+q)/(q+1)$) of $\mathcal{W}(a,b)$. We claim that

$$(16.12.2)\quad \varphi(x) \in \{0, \pm p^i \mid 0 \leq i \leq nf-1\},\ \lambda_0(x) = \pm 1,\ \text{ for all } 1 \neq x \in Z_d$$

if $p=2$, or if $p>2$ but $d|f$. Indeed, by Theorem 11.7 (i), respectively (i-bis), $\varphi(y)$ is an integer for any $y \in G_{\mathrm{arith},k}$. In particular, the representation of $G_{\mathrm{arith},k}$ on $\mathcal{H}_0$ is self-dual, and so $\lambda_0(x) = \pm 1$. Furthermore, by Theorem 11.8(i-ter), $|\varphi(x)|^2$ is either 0 or a power of p, hence the integer $\varphi(x)$ itself is also 0 or ± a power of p. Moreover, $\varphi(x) \neq -q^n$ by (ii), and $\varphi(x) = q^n$ implies $x=1$ by faithfulness of φ. Hence (16.12.2) follows.

Assume now that $p=2$. Note that

$$(16.12.3)\qquad Z_{2f} \geq Z_1 = \mathbf{C}_{\mathrm{GU}_n(q)}(\mathrm{SU}_n(q)) = C_{q+1}.$$

Also, (16.12.2) implies that Z_{2f} satisfies the assumptions in Lemma 16.4. If $(n,q) \neq (4,2)$, then Lemma 16.4(iii) implies that $|Z_{2f}|$ divides $q+1$. Together with (16.12.3), this implies (16.12.1). Suppose $(n,q)=(4,2)$. Then (16.12.3) and Lemma 16.4(iv) again imply that $|Z_{2f}|$ divides $q+1$, and so (16.12.1) follows again.

(iv) Here we assume that $p>2$. Using (16.12.2) and Lemma 16.4(iii), we obtain that $|Z_f|$ divides $q+1$. Since $Z_f \geq Z_1 = \mathbf{C}_{\mathrm{GU}_n(q)}(\mathrm{SU}_n(q)) = C_{q+1}$, we conclude that

$$(16.12.4)\qquad Z_f = C_{q+1} = \mathbf{Z}(G_{\mathrm{geom}}).$$

Assume now that (16.12.1) does not hold, i.e. $Z_{2f} > Z_f$. As $G_{\mathrm{arith},\mathbb{F}_{p^2}}$ has index 2 in $G_{\mathrm{arith},\mathbb{F}_p}$ by (i), we have that $Z_{2f}G_{\mathrm{arith},\mathbb{F}_{p^2}} = G_{\mathrm{arith},\mathbb{F}_p}$, whence

$$|Z_{2f}| = |G_{\mathrm{arith},\mathbb{F}_p}/G_{\mathrm{arith},\mathbb{F}_{p^2}}| \cdot |Z_f| = 2(q+1).$$

It follows that

$$(16.12.5)\qquad Z_{2f} = \langle t, Z_f \rangle$$

for some 2-element t, say of order 2^e for some $e \in \mathbb{Z}_{\geq 1}$. Recall that t acts as a scalar α_i on each of the $q+1$ subsheaves $\mathcal{H}_i$ of $\mathcal{W}(a,b)$, hence $\alpha_i \in \mathbb{Q}(\zeta_{2^e})$.

Next, by Theorem 16.11(a), the trace of each element $y \in G_{\rm geom}$ on each $\mathcal{H}_i$ is its trace in some Weil representation with character $\zeta^{i'}_{n,q}$, hence belonging to $\mathbb{Q}(\zeta_{q+1})$ by (16.4.4). Now, by (16.12.4) and (16.12.5), any element $x \in Z_{2f}$ is $t^c y$ for some $c \in \mathbb{Z}$ and some $y \in G_{\rm geom}$, so we get $\varphi(x) \in \mathbb{Q}(\zeta_{2^e}, \zeta_{q+1})$. On the other hand, $\varphi(x) \in \mathbb{Q}(\zeta_p)$ by Theorem 11.7(i). Thus

$$\varphi(x) \in \mathbb{Q}(\zeta_{2^e}, \zeta_{q+1}) \cap \mathbb{Q}(\zeta_p) = \mathbb{Q},$$

i.e. $\varphi(x) \in \mathbb{Z}$. Furthermore, $|\varphi(x)|^2$ is a p-power by Theorem 11.8(i-ter), so we conclude that $\varphi(x)$ is ± a p-power. Next, recall from Theorem 16.11(a) that $G_{\rm geom}$ acts on $\mathcal{H}_0$ via its Weil character $\zeta^0_{n,q}$ which is trivial at $\mathbf{Z}(G_{\rm geom})$. It follows that $\lambda_0(u) = 1$ for all $u \in Z_f$. As $t^2 \in Z_f$, we must have that $\lambda_0(t)^2 = \lambda_0(t^2) = 1$, i.e. $\lambda_0(t) = \pm 1$. Thus λ_0 takes values ± 1 on Z_{2f}. We have therefore shown that (16.12.2) holds for $d = 2f$ as well. Now we can again apply Lemma 16.4(iii) to see that the equality $|Z_{2f}| = 2(q+1)$ must imply the existence of some element $z \in Z_{2f}$ that acts as the scalar -1 on $\mathcal{W}(a,b)$, which is impossible by (ii).

(v) We have shown that (16.12.1) holds, that is, $Z_{2f} = C_{q+1} = \mathbf{Z}(G_{\rm geom})$. Together with the result of (i), it implies that, while acting via conjugation on $\mathrm{SU}_n(q)$, $G_{{\rm arith},\mathbb{F}_p}$ induces a subgroup of automorphisms of order $2f|\mathrm{PGU}_n(q)|$, which is exactly $|\mathrm{Aut}(\mathrm{SU}_n(q))|$. Hence $G_{{\rm arith},\mathbb{F}_p}$ induces the full group C_{2f} of outer field automorphisms of $\mathrm{SU}_n(q)$ (modulo inner-diagonal automorphisms), whereas $\mathrm{GU}_n(q)$ induces the full group of inner-diagonal automorphisms of $\mathrm{SU}_n(q)$. Since $G_{{\rm arith},k} \geq \mathrm{GU}_n(q)$, it follows that $G_{{\rm arith},k}$ induces the full group C_d of outer field automorphisms of $\mathrm{SU}_n(q)$. Using (16.12.1) again, we can identify $G_{{\rm arith},k}/\mathbf{Z}(G_{\rm geom})$ with the subgroup $\mathrm{PGU}_n(q) \rtimes \mathrm{Gal}(\mathbb{F}_{q^2}/k)$ of $\mathrm{Aut}(\mathrm{SU}_n(q))$.

Next, the generator g when $p > 2$ and h when $p = 2$ of $G_{{\rm arith},\mathbb{F}_p}$ modulo $G_{\rm geom}$, induces an outer field automorphism of $\mathrm{SU}_n(q)$ of order $2f$ modulo inner-diagonal automorphisms of $\mathrm{SU}_n(q)$. As $G^\star_{{\rm arith},\mathbb{F}_p}$ is also generated by g, respectively by h, modulo $G^\star_{\rm geom} = \mathrm{SU}_n(q)$, we obtain the statement for $G^\star_{{\rm arith},k}$ as well.

(vi) To identify the arithmetic monodromy group of $\widetilde{\mathcal{W}}(a,b)$ over $\mathbb{F}_{q^{2/d}}$ (when $p > 2$), we note that the absolute value of the trace of $g^c = Frob_{4,\mathbb{F}_{p^c}}$ on $\widetilde{\mathcal{W}}(a,b)$ is still $p^{c/2}$ when $c|2f$, whereas the absolute value of the trace of any element in $\tilde{G}_{\rm geom}$ is ± a power of q but never $-q^n$, by Theorem 16.11(c). Now we can repeat the arguments in (i) verbatim to obtain that

$$\tilde{G}_{{\rm arith},k} = \tilde{G}_{\rm geom} \cdot C_d.$$

Next, note that, since $2|n$, the determinant on $\mathbb{F}_{q^2}^n$ of any central element y of $\mathrm{GU}_n(q)$ is a square in μ_{q+1}, hence $\chi_2(y)=1$ and so, by Theorem 16.11(a), (c), y still acts trivially on the hypergeometric summand of $\widetilde{\mathcal{W}}(a,b)$. Now, applying Theorem 11.8(i-ter) to $\tilde{G}_{\mathrm{arith},\mathbb{F}_{p^2}}$ and repeating the arguments of (ii)–(iv), we obtain that $\tilde{G}_{\mathrm{arith},\mathbb{F}_p}$ induces the full group C_{2f} of outer field automorphisms of $\mathrm{SU}_n(q)$, and so we are done with $\tilde{G}_{\mathrm{arith},k}$ as well. □

17. Determination of monodromy groups: the case $M=q+1$ and $n=2$

In this section we assume that

(17.0.1) $$p \text{ any prime},\ \ q=p^f,\ \ M=q+1,\ \ A=B=1.$$

Fix $\alpha,\beta\in\mathbb{Z}$ such that $\alpha A-\beta B=1$ and $\alpha+\beta$ coprime to M, i.e.

(17.0.2) $$\alpha=\beta+1 \text{ and } \gcd(1+2\beta,q+1)=1.$$

With this choice of parameters, the principal objects of this section are the local systems

$$\mathcal{W}_\alpha(1,1)=\mathcal{W}(1,1):=\mathcal{W}(M,A,B)$$

on $\mathbb{G}_m/\mathbb{F}_p$ and

$$\mathcal{W}^\star_\alpha(1,1)=\mathcal{W}^\star(1,1):=[MAB]^\star\mathcal{W}(M,A,B)$$

on $\mathbb{A}^1/\mathbb{F}_p$ as introduced in Definition 16.1; moreover, we can and will view α as an integer modulo $q+1$. In particular, $\mathcal{W}_\alpha(1,1)$ is the arithmetically semisimple local system on $\mathbb{G}_m/\mathbb{F}_p$ whose trace function at $v\in E^\times$, $E/\mathbb{F}_p$ a finite extension, is given by

$$v\mapsto \frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(xw-v^{-\alpha}x^{q+1}-v^\beta w^{q+1}\big).$$

It is the descent (cf. the beginning of §13) from $\mathbb{G}_m/\mathbb{F}_{q^2}$ to $\mathbb{G}_m/\mathbb{F}_p$ of the direct sum of the Kloosterman sheaves

$$\mathcal{K}l(M,A,B,\sigma^{-\beta},\sigma^{-\alpha})(-1)=\mathcal{K}l_\psi\big(\mathsf{Char}(q+1)\smallsetminus\{\sigma^{-\beta},\sigma^{-\alpha}\}\big)(-1),$$

with $\mathbb{1}\neq\sigma\in\mathsf{Char}(q+1)$, see (4.2.1), and the hypergeometric sheaf

$$\mathcal{H}yp(M,A,B,\mathbb{1},\mathbb{1})=\mathcal{H}yp_\psi\big(\mathsf{Char}(q+1)\smallsetminus\{\mathbb{1}\};\mathbb{1}\big),$$

see (5.0.1). Its Kummer pullback $\mathcal{W}^\star_\alpha(1,1) = [q+1]^\star\mathcal{W}_\alpha(1,1)$ is a lisse sheaf on $\mathbb{A}^1$, with trace function at $v \in E$, $E/\mathbb{F}_p$ a finite extension, given by

$$v \mapsto \frac{1}{\#E}\sum_{x,w\in E}\psi_E\big(vxw - x^{q+1} - w^{q+1}\big).$$

First we prove a unitary analogue of [KT5, Lemma 7.1]:

Lemma 17.1. *Let $V = \mathbb{C}^{q^2}$ and let $\Phi : G \to \mathrm{GL}(V)$ be a faithful representation such that*

(a) $\mathrm{Tr}(\Phi(g)) \in \{1, -q, q^2\}$ *for all* $g \in G$.
(b) $\Phi \cong \oplus_{i=0}^q \Phi_i$, *where the* $\Phi_i \in \mathrm{Irr}(G)$ *are pairwise inequivalent.*

Then $|G| = |\mathrm{GU}_2(q)|$.

Proof. Let

$$r := \#\{g \in G \mid \mathrm{Tr}(\Phi(g)) = -q\},\ \ s := \#\{g \in G \mid \mathrm{Tr}(\Phi(g)) = 1\},$$

so that $|G| = r+s+1$ by (a). The assumption (b) implies for $\varphi := \mathrm{Tr}(\Phi)$ that

$$0 = [\varphi, 1_G]_G = \frac{q^2 - rq + s}{r+s+1},\quad q+1 = [\varphi,\varphi]_G = \frac{q^4 + rq^2 + s}{r+s+1}.$$

Solving for r and s, we obtain that $r = q^3 - 1$, $s = q^4 - q^2 - q$, and so $|G| = (q^2-1)(q^2+q) = |\mathrm{GU}_2(q)|$. □

The total Weil character $\zeta_{2,q}$ of $\mathrm{GU}_2(q)$, cf. (16.4.3), decomposes as $\sum_{i=0}^q \zeta_{i,2}$, with $\zeta_{i,2} \in \mathrm{Irr}(\mathrm{GU}_2(q))$ of degree $q - 1 + \delta_{i,0}$ and pairwise distinct. The larger-degree character $\zeta_{0,2}$ restricts to the Steinberg character St of $L = \mathrm{SL}_2(q) \cong \mathrm{SU}_2(q)$. Furthermore, if $1 \le i \le q/2$ then $\zeta_{i,2}$ and $\zeta_{q-1-i,2}$ restrict to the same irreducible character (denoted θ_i in [Do, §38]) of $L = \mathrm{SL}_2(q)$, and those $\lfloor q/2\rfloor$ characters are pairwise distinct. If $2 \nmid q$, then $(\zeta_{(q-1)/2,2})|_L$ is the sum of two distinct irreducible characters (denoted η_1, η_2 in [Do, §38]) of degree $(q-1)/2$. We will refer to these characters θ_i, and also η_1, η_2 when $2 \nmid q$, as *irreducible Weil characters* of $\mathrm{SU}_2(q)$, and $(\zeta_{2,q})|_L$ as the *total Weil character* of L, now viewed as $\mathrm{SU}_2(q)$.

Now we prove an analogue of Theorem 16.6, which characterizes the total Weil representation of $\mathrm{SU}_2(q)$.

Theorem 17.2. *Let p be any prime, q be any power of p, $q \ge 4$, and let $L = \mathrm{SL}_2(q)$. Suppose φ is a reducible complex character of L such that*

(a) $\varphi(1) = q^2$;
(b) $\varphi(g) \in \{1, -q, q^2\}$ *for all* $g \in L$;
(c) *every irreducible constituent of* φ *is among the irreducible Weil characters* St, θ_i, $0 \leq i \leq q/2$, *and also* η_1, η_2 *when* $2 \nmid q$, *of* L.

Then φ *is the total Weil character* $(\zeta_{2,q})|_L$ *of* L.

Proof. (i) We will use the character tables of $\mathrm{SL}_2(q)$, Theorem 38.1 of [Do] for $2 \nmid q$ and Theorem 38.2 of [Do] for $2|q$. Write

$$\varphi = \begin{cases} a \cdot \mathsf{St} + \sum_{i=1}^{(q-1)/2} b_i\theta_i + c_1\eta_1 + c_2\eta_2, & 2 \nmid q, \\ a \cdot \mathsf{St} + \sum_{i=1}^{q/2} b_i\theta_i, & 2|q, \end{cases} \tag{17.2.1}$$

with coefficients $a, b_i, c_i \in \mathbb{Z}_{\geq 0}$. Evaluating φ at an element x of order $q-1$, we see by (b) that $\varphi(y) = a$ is a $(-q)$-power with $0 \leq a \leq \varphi(1)/\mathsf{St}(1) = q$, which is possible only when $a = 1$. As before, let $\boldsymbol{\varrho}$ denote a primitive $(q+1)^{\text{th}}$ root of unity in $\mathbb{C}$.

First suppose that $2|q$. Then $\sum_i b_i = (q^2 - q)/(q-1) = q$ by degree comparison in (17.2.1). Next, we fix an element $y \in L$ of order $q+1$, and for $1 \leq l \leq q/2$ we have

$$\varphi(y^l) = -1 - \sum_{i=1}^{q/2} b_i\big(\boldsymbol{\varrho}^{il} + \boldsymbol{\varrho}^{-il}\big).$$

It follows that

$$\sum_{l=1}^{q/2} \varphi(y^l) = -q/2 - \sum_{i=1}^{q/2} b_i\Bigg(\sum_{l=1}^{q/2}\big(\boldsymbol{\varrho}^{il} + \boldsymbol{\varrho}^{-il}\big)\Bigg) = -q/2 + \sum_{i=1}^{q/2} b_i = q/2.$$

As each value $\varphi(y^l)$ is either 1 or $-q$, we must have that $\varphi(y^l) = 1$ for all $1 \leq l \leq q/2$. Thus, the polynomial

$$f(t) = \sum_{i=1}^{q/2} b_i\big(t^{q+1-i} + t^i\big) + 2 \in \mathbb{Q}[t]$$

of degree q has all $\boldsymbol{\varrho}^l$, $1 \leq l \leq q$ as roots. Since $f(1) = 2\sum_{i=1}^{q/2} b_i + 2 = 2q+2$, we conclude that $f(t) = 2(t^{q+1} - 1)/(t-1)$, i.e. $b_i = 2$ for all i, and so $\varphi = (\zeta_{2,q})|_L$, as stated.

(ii) Assume now that $2 \nmid q$. Then $\sum_i b_i + (c_1+c_2)/2 = (q^2-q)/(q-1) = q$ by degree comparison in (17.2.1). Evaluating φ at an element $u \in L$ of order p and another element $v \in L$ of order p that is not conjugate to u, we obtain

$$\varphi(u) = -\sum_i b_i - \frac{c_1+c_2}{2} + \sqrt{\epsilon q}\frac{c_1-c_2}{2},\ \varphi(v) = -\sum_i b_i - \frac{c_1+c_2}{2} - \sqrt{\epsilon q}\frac{c_1-c_2}{2},$$

where $\epsilon := (-1)^{(q-1)/2}$. Thus $\varphi(u)+\varphi(v) = -2q$. As each of $\varphi(u)$, $\varphi(v)$ is either 1 or $-q$, we must have that $\varphi(u) = -q = \varphi(v)$, whence $c_1 = c_2 =: c$, and so

$$\sum_{i=1}^{(q-1)/2} b_i + c = q.$$

Next we evaluate φ at the central involution $\boldsymbol{j}$ of L:

$$\varphi(\boldsymbol{j}) = q + (q-1)\sum_i b_i(-1)^i - c\epsilon(q-1).$$

In particular,

$$q^2 - \varphi(\boldsymbol{j}) = \varphi(1) - \varphi(\boldsymbol{j}) = 2(q-1)\Big(\sum_{2\nmid i} b_i + \frac{1+\epsilon}{2}c\Big)$$

is divisible by $2(q-1)$. On the other hand, $\varphi(\boldsymbol{j}) \in \{1, -q, q^2\}$ and $q \geq 4$, so $\varphi(\boldsymbol{j}) \neq -q$, and either

$$\varphi(\boldsymbol{j}) = q^2,\ \sum_{2\nmid i} b_i + \frac{1+\epsilon}{2}c = 0,\ \sum_{2|i} b_i + \frac{1-\epsilon}{2}c = q, \tag{17.2.2}$$

or

$$\varphi(\boldsymbol{j}) = 1,\ \sum_{2\nmid i} b_i + \frac{1+\epsilon}{2}c = \frac{q+1}{2},\ \sum_{2|i} b_i + \frac{1-\epsilon}{2}c = \frac{q-1}{2}. \tag{17.2.3}$$

As above, we fix an element $y \in L$ of order $q+1$, and for $1 \leq l \leq (q-1)/2$ we then have

$$\varphi(y^l) = -1 - \sum_{i=1}^{(q-1)/2} b_i\big(\boldsymbol{\varrho}^{il} + \boldsymbol{\varrho}^{-il}\big) - 2c(-1)^l.$$

It follows that

$$\begin{aligned}\sum_{l=1}^{(q-1)/2} \varphi(y^l) &= -(q-1)/2 - \sum_{i=1}^{(q-1)/2} b_i\Bigl(\sum_{l=1}^{(q-1)/2}\bigl(\boldsymbol{\varrho}^{il}+\boldsymbol{\varrho}^{-il}\bigr)\Bigr) - 2c\sum_{l=1}^{(q-1)/2}(-1)^l\\ &= -(q-1)/2 - \sum_{i=1}^{(q-1)/2} b_i\bigl(-1-(-1)^i\bigr) + c(1-\epsilon)\\ &= -(q-1)/2 + 2\bigl(\sum_{2|i} b_i + c(1-\epsilon)/2\bigr).\end{aligned}$$

In the case of (17.2.2), $\sum_{l=1}^{(q-1)/2}\varphi(y^l) = -(q-1)/2 + 2q > (q-1)/2$, a contradiction. Hence (17.2.3) holds, and we have that

$$\sum_{l=1}^{(q-1)/2} \varphi(y^l) = -(q-1)/2 + (q-1) = (q-1)/2.$$

As each value $\varphi(y^l)$ is either 1 or $-q$, we must have that $\varphi(y^l) = 1$ for all $1 \le l \le (q-1)/2$. Thus, the polynomial

$$g(t) = \sum_{i=1}^{(q-1)/2} b_i\bigl(t^{q+1-i}+t^i\bigr) + 2ct^{(q+1)/2} + 2 \in \mathbb{Q}[t]$$

of degree q admits each of $\boldsymbol{\varrho}^l \neq \pm 1$ with $0 \le l \le q$ as a root, and so $g(t) = (at+b)(t^{q+1}-1)/(t^2-1)$ for some $a, b \in \mathbb{Q}$. Since $b = g(0) = 2$ and $(a+b)(q+1)/2 = g(1) = 2\sum_{i=1}^{(q-1)/2} b_i + 2c + 2 = 2q+2$, we conclude that $a = b = 2$, $g(t) = 2(t^{q+1}-1)/(t-1)$, i.e. $b_i = 2$ for all i and $c_1 = c_2 = 1$, and so $\varphi = (\zeta_{2,q})|_L$, as stated. □

A characterization of the total Weil character $\zeta_{2,q}$ of $\mathrm{GU}_2(q)$, cf. (16.4.3), is given in the next result, which is an analogue of Theorem 16.8:

Theorem 17.3. *Let q be any prime power, $\boldsymbol{\varrho} := \zeta_{q+1}$, and let*

$$\Phi : G := \mathrm{GU}_2(q) \to \mathrm{GL}_{q^2}(\mathbb{C})$$

be a faithful complex representation that satisfies the following conditions:

(a) *$\Phi = \oplus_{j=0}^{q}\Phi_j$ with Φ_j being irreducible of degree $q-1+\delta_{j,0}$;*
(b) *There is an element $g \in G$ such that the matrix $\Phi_j(g)$ has spectrum $\{\boldsymbol{\varrho}^i \mid 0 \le i \le q,\ i \neq 0, j\}$ when $0 \le j \le q$, and that $G = \langle [G,G], g\rangle$.*

Then there exists an automorphism γ of G such that $\mathrm{Tr}\big(\Phi(\gamma(h))\big) = \zeta_{2,q}(h)$ *for all $h \in G$.*

Proof. The spectra of $\Phi_j(g)$ show that g has both order and central order $q+1$ in G. Thus, for a fixed $\varrho \in \mathbb{F}_{q^2}^\times$ of order $q+1$, after a suitable conjugation, we may assume that $g = \mathrm{diag}(\varrho^c, \varrho^d)$ with $c, d \in \mathbb{Z}/(q+1)\mathbb{Z}$. Since g generates G modulo $[G, G]$, $\det(g)$ has order $q+1$. Changing ϱ to another element of order $q+1$, we may therefore assume that $c + d = 1$. Now, the condition that g has central order $q+1$ is equivalent to that $\gcd(1-2c, q+1) = 1$. As noted in Remark 16.7, since $1 + 2(c-1) = 2c-1$ is coprime to $q+1$, the map γ_{c-1} of (16.7.1) is an automorphism of G. Hence we can replace g by $\gamma_{c-1}(g) = \mathrm{diag}(\varrho, 1)$ and thus assume that

$$g = \mathrm{diag}(\varrho, 1). \tag{17.3.1}$$

We will use the character table of G as given in [E]. In particular, the character φ_0 of Φ_0 is denoted $\chi_q^{(t_0)}$ therein, and by (b) we have

$$-1 = \varphi_0(g) = \chi_q^{(t_0)}(g) = -\boldsymbol{\varrho}^{t_0},$$

whence $t_0 = 0$. Furthermore, the character φ_j of Φ_j, $1 \le j \le q$, is denoted $\chi_{q-1}^{(t_j,u_j)}$ therein for some $t_j, u_j \in \mathbb{Z}/(q+1)\mathbb{Z}$ with $t_j \ne u_j$ (and one has $\chi_{q-1}^{(t_j,u_j)} = \chi_{q-1}^{(u_j,t_j)}$). Using (b) and (17.3.1), we then obtain

$$-1 - \boldsymbol{\varrho}^{ij} = \varphi_j(g^i) = \chi_{q-1}^{(t_j,u_j)}(g^i) = -\boldsymbol{\varrho}^{t_j i} - \boldsymbol{\varrho}^{u_j i}$$

for $1 \le i \le q$. Viewing $0 \le t_j, u_j \le q$ and setting

$$f_j(x) := x^{t_j} + x^{u_j} - x^j - 1 \in \mathbb{Q}[x],$$

we see that f_j has degree at most q and vanishes at all ϱ^i, $1 \le i \le q$. It follows that $f_j(x)$ is identically zero, i.e. $\{t_j, u_j\} = \{0, j\}$.

We have shown that the character of Φ is $\chi_q^{(0)} + \sum_{j=1}^q \chi_{q-1}^{(j,0)}$. Direct check shows that the latter character is $\zeta_{2,q}$, and so we are done. For later use, we also note that, for the central element $\boldsymbol{z} := \mathrm{diag}(\varrho, \varrho)$, we have $\chi_{q-1}^{(j,0)}(\boldsymbol{z}) = (q-1)\boldsymbol{\varrho}^j$, i.e.

$$\Phi_j(\boldsymbol{z}) = \boldsymbol{\varrho}^j \cdot \mathrm{Id}. \tag{17.3.2}$$

□

Lemma 17.4. *Denote by $\mathcal{H}$ the hypergeometric component of $\mathcal{W}(1,1)$, with the choice $(\alpha,\beta)=(1,0)$ of (α,β) with $\alpha A-\beta B=\alpha-\beta=1$. Denote by $\mathcal{H}_{0,0}$ the lisse sheaf of weight 0 on $\mathbb{G}_m/\mathbb{F}_p$ which is the weight 0 quotient of the lisse sheaf on $\mathbb{G}_m/\mathbb{F}_p$ which is mixed of weight ≤ 0 and whose trace function is given at $v\in E^\times$ for $E/\mathbb{F}_p$ a finite extension by*

$$v\mapsto (1/\#E)\sum_{(x,w)\in E^\times\times E^\times}\psi_E(x-v^{-1}x^{q+1}/w-w).$$

Denote by $\mathcal{F}_{\mathbb{1},0}$ the lisse sheaf on $\mathbb{G}_m/\mathbb{F}_p$ whose trace function is given at $v\in E^\times$ for $E/\mathbb{F}_p$ a finite extension by

$$v\mapsto (1/\#E)\sum_{x\in E,w\in E^\times}\psi_E(x-v^{-1}x^{q+1}/w-w).$$

Then we have the following results.

(i) *$\mathcal{H}_{0,0}$ is geometrically isomorphic to $\mathcal{H}$.*
(ii) *$\mathcal{F}_{\mathbb{1},0}$ is pure of weight zero, and its pullback to $\mathbb{G}_m/\mathbb{F}_q$ is arithmetically isomorphic to the lisse sheaf $\mathcal{F}_{\mathbb{1}}$ of* [KT5, Section 4] *with $n=2$ there.*
(iii) *$\mathcal{H}_{0,0}$ is arithmetically isomorphic to $\mathcal{F}_{\mathbb{1},0}$.*

Proof. For the first assertion, $\mathcal{H}$ is geometrically the **Cancel** of the sheaf $\mathcal{H}yp(\mathsf{Char}(q+1);\mathbb{1},\mathbb{1})$, cf. Corollary 9.3 (ii), whose trace function is that of $\mathcal{H}_{0,0}$ (up to a constant field twist), cf. Corollary 8.2. For the second assertion, the trace function of $\mathcal{F}_{\mathbb{1},0}$, restricted to extensions $E/\mathbb{F}_q$, is identical to that of $\mathcal{F}_{\mathbb{1}}$, cf. [KT5, Section 4] in the case $n=2$. For the third assertion, we know by (ii) that

$$v\mapsto (1/\#E)\sum_{x\in E,w\in E^\times}\psi_E(x-v^{-1}x^{q+1}/w-w)$$

is pure of weight zero. We must show that it is the weight zero quotient of

$$v\mapsto (1/\#E)\sum_{(x,w)\in E^\times\times E^\times}\psi_E(x-v^{-1}x^{q+1}/w-w).$$

Equivalently, we must show that their difference is mixed of weight ≤ -1.

But their difference is

$$v \mapsto (1/\#E) \sum_{x=0, w\in E^\times} \psi_E(x - v^{-1}x^{q+1}/w - w)$$
$$= (1/\#E) \sum_{w\in E^\times} \psi_E(-w) = -1/\#E. \qquad \square$$

The main result of this section is the following theorem, which complements Theorem 16.11:

Theorem 17.5. *Let $q = p^f \geq 4$ be a power of a prime p. Then the following statements hold for the geometric and arithmetic monodromy groups $G_{\rm geom}$ and $G_{{\rm arith},k}$ of the local system $\mathcal{W}(1,1)$ over any finite extension k of $\mathbb{F}_{q^2}$.*

(a) *$G_{{\rm arith},k} = G_{\rm geom} \cong {\rm GU}_2(q)$. Furthermore, we can identify $G_{\rm geom}$ with ${\rm GU}_2(q)$ in such a way that the action of ${\rm GU}_2(q)$ on $\mathcal{W}(1,1)$ affords the total Weil character $\zeta_{2,q}$.*

(b) *Let $\mathcal{H}_i$ be any of the $q+1$ hypergeometric constituents of $\mathcal{W}(1,1)$. Then $\mathcal{H}_i$ has arithmetic and geometric monodromy groups $G^i_{{\rm arith},k} = G^i_{\rm geom}$, $G^i_{\rm geom}/\mathbf{Z}(G^i_{\rm geom}) \cong {\rm PGU}_2(q)$, and $\mathbf{Z}(G^i_{\rm geom})$ is cyclic of order dividing $q+1$.*

(c) *Over any subfield $\mathbb{F}_{q^{2/d}}$ of $\mathbb{F}_{q^2}$, the arithmetic monodromy group $G_{{\rm arith},\mathbb{F}_{q^{2/d}}}$ of $\mathcal{W}(1,1)$ is ${\rm GU}_2(q)\cdot C_d$, and induces a subgroup of outer field automorphisms of ${\rm SU}_2(q)$ of order $d/\gcd(2,d)$. Furthermore, $\mathbf{C}_{G_{{\rm arith},\mathbb{F}_{q^{2/d}}}}({\rm SU}_2(q))$ has order $(q+1)\cdot\gcd(2,d)$, and*

$$G_{{\rm arith},\mathbb{F}_{q^{2/d}}}/\mathbf{C}_{G_{{\rm arith},\mathbb{F}_{q^{2/d}}}}({\rm SU}_2(q)) \cong {\rm PGU}_2(q) \rtimes {\rm Gal}(\mathbb{F}_{q^2}/\mathbb{F}_{q^{2\gcd(2,d)/d}}).$$

(d) *The local system $\mathcal{W}^\star(1,1)$ has geometric monodromy group and arithmetic monodromy group $G^\star_{{\rm arith},k} = G^\star_{\rm geom} = (G_{\rm geom})^{(\infty)} \cong {\rm SU}_2(q)$, with ${\rm SU}_2(q)$ acting via its total Weil representation. Furthermore, over any subfield $\mathbb{F}_{q^{2/d}}$ of $\mathbb{F}_{q^2}$, the arithmetic monodromy group of $\mathcal{W}^\star(1,1)$ is ${\rm SU}_2(q)\cdot C_d$, and induces a subgroup of outer automorphisms of ${\rm SU}_2(q)$ of order $d/\gcd(2,d)$, modulo the inner-diagonal automorphisms of ${\rm SU}_2(q)$.*

Proof. (i) Let $\Phi : G := G_{{\rm arith},k} \to {\rm GL}_{q^n}(\mathbb{C})$ denote the corresponding representation of $G_{{\rm arith},k}$ on $\mathcal{W} := \mathcal{W}(1,1)$. By Theorem 11.1, $\Phi \cong \oplus_{i=0}^q \Phi_i$, where $\deg(\Phi_i) = q - 1 + \delta_{i,0}$. Now, by Theorem 11.9 we have

$$\text{(17.5.1)} \qquad {\rm Tr}(\Phi(u)) = 1, -q, \text{ or } q^2, \text{ for all } u \in G_{{\rm arith},k}.$$

It follows from Lemma 17.1 that $|G_{\mathrm{geom}}| = |\mathrm{GU}_2(q)| = |G_{\mathrm{arith},k}|$, and so

(17.5.2) $$G := G_{\mathrm{geom}} = G_{\mathrm{arith},k} \text{ has order equal to } |\mathrm{GU}_2(q)|.$$

Next, note that the hypergeometric summand $\mathcal{H}_0$ of rank q is precisely the sheaf $\mathcal{H}_{\mathbb{1}}$ considered in [KT5, §1], hence

(17.5.3) $$\Phi_0(G) = G^0_{\mathrm{geom}} \cong \mathrm{PGL}_2(q) \cong \mathrm{PGU}_2(q)$$

by [KT5, Corollary 8.2].

(ii) Next we take $L := G^{(\infty)}$. Then L has $S := \mathrm{PSU}_2(q)$ as a composition factor, and so we can write $|L| = e \cdot |\mathrm{PSU}_2(q)| = eq(q^2-1)/\gcd(2, q-1)$ for some $e \in \mathbb{Z}_{\geq 1}$. On the other hand, by Corollary 13.4, some hypergeometric summand of $\mathcal{W}(1,1)$ has geometric determinant $\mathcal{L}_\nu$ with ν of order exactly $M = q+1$, whence $q+1$ divides $|G/[G,G]|$. It now follows from (17.5.2) that $q+1$ divides $|G/L| = (q+1) \cdot \gcd(2, q-1)/e$, i.e.

(17.5.4) $$e | \gcd(2, q-1).$$

Since $|G/L|$ divides the integer $(q+1) \cdot \gcd(2, q-1)$ which is prime to p, we have that $L \geq \mathbf{O}^{p'}(G)$. On the other hand, $G/\mathbf{O}^{p'}(G)$ is cyclic for $G = G_{\mathrm{geom}}$ by [Abh, Proposition 6(III)], therefore $\mathbf{O}^{p'}(G) \geq L$. Thus $L = \mathbf{O}^{p'}(G)$, and so the integer $n(G)$ defined prior to Theorem 14.6 is $(q+1) \cdot \gcd(2, q-1)/e$, a multiple of $q+1$ by (17.5.4). Now applying Theorem 14.6 to the sheaf $\mathcal{W}^\star(1,1) = [q+1]^\star\mathcal{W}$, we see that $G^\star_{\mathrm{arith},k} = G^\star_{\mathrm{geom}}$; moreover, $G^\star_{\mathrm{geom}}$ has index $q+1$ in G and contains L as a normal subgroup of index $\gcd(2, q-1)/e$. But $\mathcal{W}^\star(1,1)$ is a local system on $\mathbb{A}^1$, so $G^\star_{\mathrm{geom}}$ has no nontrivial p'-quotient. Thus we conclude that $e = \gcd(2, q-1)$, and $|L| = |\mathrm{SU}_2(q)|$. Recall that L is perfect and has $S = \mathrm{PSU}_2(q)$ as a composition factor. If $2|q$, we must have that $L \cong \mathrm{SU}_2(q)$. If $2 \nmid q$, then L admits a normal subgroup L_1 of order 2 such that $L/L_1 \cong S$. In this case, $L_1 \leq \mathbf{Z}(L)$, and so $L \cong \mathrm{SU}_2(q)$ as well. Thus we have shown that

(17.5.5) $$G^\star_{\mathrm{arith},k} = G^\star_{\mathrm{geom}} = L \cong \mathrm{SU}_2(q).$$

Moreover, the geometric determinant $\mathcal{L}_\nu$ mentioned above now implies that

(17.5.6) $$G/L \cong C_{q+1}, \;\; G = \langle L, g \rangle.$$

(iii) More generally, let us consider the kernel K of Φ_0. By (17.5.2) and (17.5.3),

(17.5.7) $$|K| = |G|/|\mathrm{PGU}_2(q)| = q+1.$$

Next, $K \cap L$ is the kernel of the representation $(\Phi_0)|_L$ of degree q. First we note that any representation $(\Phi_i)|_L$ cannot be trivial, as otherwise $\Phi_i(G)$ would have order dividing

$$|G/L| = q+1 < (q-1)^2 \leq \mathrm{rank}(\mathcal{H}_i)^2,$$

contradicting the irreducibility of G on $\mathcal{H}_i$. In particular, this holds for $(\Phi_0)|_L$. Now if $2|q$, then $L \cong S$ is simple, and so $K \cap L = 1 = \mathbf{Z}(L)$. It follows that $|KL| = |G|$, and so $G \cong K \times L$ and $K \cong G/L \cong C_{q+1}$ by (17.5.6), whence there exists $\iota : G \cong \mathrm{GU}_2(q)$. We also note that, since the smallest degree of nontrivial irreducible representations of $\mathrm{SU}_2(q)$ is $q-1$, all $(\Phi_i)|_L$ are irreducible, and afford characters St or θ_i, whence $\Phi|_L$ is the total Weil representation by (17.5.1) and Theorem 17.2.

(iv) In this and the next parts of the proof we will assume $2 \nmid q$. If $(\Phi_0)|_L$ is reducible, then each irreducible constituent of it has degree $\leq q/3 < (q-1)/2$, contrary to the fact that every nontrivial irreducible representation of $L \cong \mathrm{SL}_2(q)$ has degree $\geq (q-1)/2$. Hence $(\Phi_0)|_L$ is an irreducible representation of degree q, i.e. its Steinberg representation, and so $K \cap L = \mathbf{Z}(L)$. As $[K, L] \leq K \cap L$, in both cases we now have $[[K, L], L] = 1$, and so by the Three Subgroups Lemma and by the perfectness of L we have that $[K, L] = [K, [L, L]]$ is contained in $[[K, L], L] = 1$, i.e.

$$K \leq \mathbf{C}_G(L). \tag{17.5.8}$$

We have also shown that each irreducible constituent of $(\Phi_i)|_L$ is of degree q (hence it is the Steinberg representation) if $i = 0$, or of degree $q-1$ or $(q-1)/2$ if $1 \leq i \leq q$ and thus affords the character θ_i or η_j, in the notation of Theorem 17.2. Together with (17.5.1), Theorem 17.2 applied to L implies that $\Phi|_L$ is the total Weil representation of $L = \mathrm{SU}_2(q)$, as stated in (a). In particular, the character of $\Phi|_L$ contains exactly two irreducible constituents of degree $(q-1)/2$, namely η_1 and η_2.

By Corollary 4.12, for any $1 \leq i \leq q$, $\mathcal{H}_i$ satisfies the condition **(S+)**, except for the sheaf $\mathcal{K}l(M, A, B, \sigma^{-\beta}, \sigma^{-\alpha})$ with $\sigma^{\alpha-\beta} = \chi_2$, equivalently, $\sigma = \chi_2$ (recall that $\alpha - \beta = 1$). We will choose our labeling so that this sheaf is $\mathcal{H}_{(q+1)/2}$. Hence, if $i \neq (q+1)/2$ then the normal subgroup L of G acts irreducibly on $\mathcal{H}_i$ by [GT, Lemma 2.5].

Suppose for a moment that $K \neq \mathbf{C}_G(L)$. By (17.5.7) and (17.5.8), we then have $|\mathbf{C}_G(L)| \geq 2(q+1)$. On the other hand, $\mathbf{C}_G(L) \cap L = \mathbf{Z}(L)$ has order 2. Hence $|\mathbf{C}_G(L)L| \geq (q+1)|L| = G$, and so $G = \mathbf{C}_G(L) * L$, a central product with $\mathbf{C}_G(L) \cap L = \mathbf{Z}(L) = K \cap L$. It follows that $|\mathbf{C}_G(L)| = 2(q+1)$, and

$$G/(K \cap L) \cong \mathbf{C}_G(L)/\mathbf{Z}(L) \times L/\mathbf{Z}(L),$$

a direct product of a group of order $q+1$ and the simple group $S \cong \mathrm{PSU}_2(q)$. Thus $G/(K \cap L)$ cannot map onto $G/K = \Phi_0(G) \cong \mathrm{PGU}_2(q)$, contrary to (17.5.3). Thus we have shown that

$$K = \mathbf{C}_G(L). \tag{17.5.9}$$

(v) We can view L as the commutator subgroup of $\mathrm{GU}_2(q)$. Recalling $G/\mathbf{C}_G(L) \cong \mathrm{PGU}_2(q)$ from (17.5.3) and (17.5.9), we now see that G induces the full group of inner-diagonal automorphisms of L, which is the one induced by elements by $\mathrm{GU}_2(q)$ acting on L via conjugation. It follows that we can find an element $h \in \mathrm{GU}_2(q)$ such that

$$g \text{ and } h \text{ induce the same automorphism of } L = \mathrm{SU}_2(q); \tag{17.5.10}$$

furthermore, changing g to another representative in its coset gG if necessary, we can ensure that

$$h = \mathrm{diag}(\varrho, 1) \tag{17.5.11}$$

for some $\varrho \in \mathbb{F}_{q^2}^\times$ of order $q+1$, and so

$$\mathrm{ord}(h) = q+1, \quad L \cap \langle h \rangle = 1. \tag{17.5.12}$$

Next, as shown in (iv), if $j \neq 0, (q+1)/2$ then $(\Phi_j)|_L$ is irreducible, of degree $q-1$. Each such representation extends to a representation $\tilde{\Phi}_j$ of $\mathrm{GU}_2(q)$. Moreover, as one can check using the character table of $\mathrm{GU}_2(q)$ [E, §6],

$$0 \neq \mathrm{Tr}(\tilde{\Phi}_j(h)) \in \mathbb{Q}(\boldsymbol{\varrho}) = \mathbb{Q}(\zeta_{q+1}) \tag{17.5.13}$$

(indeed, any irreducible representation of degree $q-1$ of $\mathrm{GU}_2(q)$ is reducible over $\mathrm{SU}_2(q)$ if and only its trace at h is zero). Furthermore, the choice (17.5.10) of h, and again the irreducibility of $(\Phi_j)|_L$ established in (iv) ensure that $\Phi_j(g)\tilde{\Phi}_j(h)^{-1}$ centralizes $\Phi_j(L)$, whence

$$\Phi_j(g) = \alpha_j \tilde{\Phi}_j(h) \tag{17.5.14}$$

for some $\alpha_j \in \mathbb{C}^\times$. In fact, α_j is a root of unity because both g and h have finite order. Also, since σ in Definition 16.1 is chosen to have order dividing

$q+1$, $\mathrm{Tr}(\Phi_j(g)) \in \mathbb{Q}(\zeta_{q+1})$ by Theorem 11.1. Hence the root of unity α_j belongs to $\mathbb{Q}(\zeta_{q+1})$ by (17.5.14). As $2|(q+1)$, it follows that

$$\alpha_j^{q+1} = 1$$

for all $j \neq 0, (q+1)/2$. Together with (17.5.12) and (17.5.14), this implies that $\Phi_j(g)^{q+1} = \mathrm{Id}$ for all $j \neq 0, (q+1)/2$. In particular, $g^{q+1} \in G$ has trace $q-1$ on all $\mathcal{H}_j$ with $j \neq 0, (q+1)/2$. Hence

$$\begin{aligned}\bigl|\mathrm{Tr}(\Phi(g^{q+1}))\bigr| &= \Bigl|\sum_{j=0}^{q} \mathrm{Tr}(\Phi_j(g^{q+1}))\Bigr| \\ &\geq \Bigl|\sum_{j\neq 0, \frac{q+1}{2}} \mathrm{Tr}(\Phi_j(g^{q+1}))\Bigr| - \bigl|\mathrm{Tr}(\Phi_0(g^{q+1}))\bigr| - \bigl|\mathrm{Tr}(\Phi_{\frac{q+1}{2}}(g^{q+1}))\bigr| \\ &\geq (q-1)^2 - q - (q-1) = q^2 - 4q + 2 \geq q+2\end{aligned}$$

(as $q \geq 5$). It follows from (17.5.1) that $\mathrm{Tr}(\Phi(g^{q+1})) = q^2$ and so $g^{q+1} = 1$ by faithfulness of Φ. Recalling (17.5.6), we must then have that

$$\mathrm{ord}(g) = q+1, \quad L \cap \langle g \rangle = 1. \tag{17.5.15}$$

Thus $G = L \rtimes \langle g \rangle$ and $\mathrm{GU}_2(q) = L \rtimes \langle h \rangle$ are two split extensions of $L \cong \mathrm{SU}_2(q)$ by C_{q+1}. Now using (17.5.10), (17.5.12), and (17.5.15), one can readily check that the map $sg^i \mapsto sh^i$, $s \in L$ and $0 \leq i \leq q$, yields a group isomorphism $\iota : G \cong \mathrm{GU}_2(q)$.

(vi) Now we return to the general case of any prime p. Statement (b), both for $2|q$ and $2 \nmid q$, follows by applying Φ_i to $G = G_{\mathrm{geom}} = G_{\mathrm{arith},k}$.

To complete the proof of (a), let $\langle g_0 \rangle$ denote the image of $I(0)$ in the group $G = G_{\mathrm{geom}}$. First we consider the case $\alpha = 1$. Then we can relabel Φ_j so that the spectrum of $\Phi_j(g_0)$ equals $\{\boldsymbol{\varrho}^i \mid i \neq 0, j\}$. Note that, since G/L is cyclic, $\langle L, g_0 \rangle$ is normal in G and so contains the normal closure of $\langle g_0 \rangle$ in G. But the normal closure of $\langle g_0 \rangle$ in G equals G by [KT7, Theorem 4.1], hence $\langle L, g_0 \rangle = G$. Now we can apply Theorem 17.3 (and its proof) to obtain $\gamma \in \mathrm{Aut}(G)$ such that $\mathrm{Tr}\bigl(\Phi(\gamma(x))\bigr) = \zeta_{2,q}(x)$ for all $x \in G$, $\gamma(g) = \mathrm{diag}(\varrho, 1)$, cf. (17.3.1), and $\gamma(\boldsymbol{z}) = \mathrm{diag}(\varrho, \varrho)$ acts in Φ_j via the scalar $\boldsymbol{\varrho}^j$ for a generator $\boldsymbol{z}$ of $\mathbf{Z}(G)$, cf. (17.3.2). In particular, adjusting the identification ι by σ, we see that $G \cong \mathrm{GU}_2(q)$ acts on $\mathcal{W}_1(1,1)$ with the total Weil character $\zeta_{2,q}$. We also note that the local system $\mathcal{W}_1(1,1)$ gives rise to a surjection

$$\phi : \pi_1(\mathbb{G}_m/\overline{\mathbb{F}_p}) \twoheadrightarrow G,$$

and composing with Φ_j, it realizes the hypergeometric sheaf $\mathcal{H}_j$.

Next, we consider the general case of any (α, β) satisfying (17.0.2). As noted in Remark 16.7, the map γ_β defined in (16.7.1) is an automorphism of $G = \mathrm{GU}_2(q)$. Since $\gamma_\beta(g_0) = g_0 \boldsymbol{z}^\beta$, the spectrum of $\gamma_\beta(g_0)$ in Φ_j equals $\{\boldsymbol{\varrho}^i \mid i \neq \alpha j, \beta j\}$. Now we twist the representation Φ of $\mathrm{GU}_2(q)$ on $\mathcal{W}_1(1,1)$ by γ_β to obtain

$$\Psi(x) = \Phi(\gamma_\beta(x)) \text{ and } \Psi_j(x) = \Phi_j(\gamma_\beta(x))$$

for all $x \in G$. Note that γ_β does not change any unipotent element in G, hence

$$\mathrm{Tr}(\Psi_j(y)) = \mathrm{Tr}(\Phi_j(y))$$

for all p-elements $y \in G$. It follows from [KT7, Theorem 5.1] that composing ψ with Ψ_j realizes a hypergeometric sheaf $\widetilde{\mathcal{H}}_0$ of type $(q,1)$ when $j = 0$ and a Kloosterman sheaf $\widetilde{\mathcal{H}}_j$ of rank $q-1$ when $1 \leq j \leq q$. The spectrum of $\Psi_j(g_0)$ when $j > 0$ shows that

$$\widetilde{\mathcal{H}}_j = \mathcal{K}l\big(\mathsf{Char}(q+1) \smallsetminus \{\sigma^{-j\alpha}, \sigma^{-j\beta}\}\big)$$

for a fixed character σ of order $q+1$. Likewise, the "upstairs" characters of $\widetilde{\mathcal{H}}_0$ are $\mathsf{Char}(q+1) \smallsetminus \{\mathbb{1}\}$. We show that the "downstairs" character is $\mathbb{1}$. Indeed, the image of $I(\infty)$ in $G^0_{\mathrm{geom}} \cong \mathrm{PGU}_2(q)$ is an elementary abelian group of order q extended semidirectly by C_{q-1}. Now, Ψ_0 still affords the same character $\chi_q^{(0)}$ as of Φ_0, so a generator of this C_{q-1} has trace 1 in Ψ_0, showing that the "downstairs" character is $\mathbb{1}$. Thus

$$\bigoplus_{j=0}^{q} \widetilde{\mathcal{H}}_j \cong \mathcal{W}_\alpha(1,1),$$

with its geometric monodromy group acting via $\Psi = \Phi \circ \gamma_\beta$.

(vii) Note that Lemma 16.3 also holds when $a = b = 1$. Hence, the same arguments as in part (i) of the proof of Theorem 16.12, using the $a = b = 1$ case of Lemma 16.3, show that

$$G_{\mathrm{arith},\mathbb{F}_p}/G_{\mathrm{geom}} \cong G^\star_{\mathrm{arith},\mathbb{F}_p}/G^\star_{\mathrm{geom}} \cong C_{2f}, \tag{17.5.16}$$

in fact,

$$G^\star_{\mathrm{arith},\mathbb{F}_p} = \langle g^\star, G^\star_{\mathrm{geom}} \rangle, \tag{17.5.17}$$

where $g^\star = \mathit{Frob}_{2,\mathbb{F}_p}$ when $p > 2$ and $g^\star = \mathit{Frob}_{0,\mathbb{F}_2}$ when $p = 2$.

Next, as shown in Lemma 17.4, the hypergeometric summand $\mathcal{H}_0$ of $\mathcal{W}(1,1)$ is arithmetically isomorphic to the sheaf $\mathcal{H}_{\mathbb{1}}$ considered in [KT5, §1]. By [KT5, Theorem 8.3], the latter has arithmetic monodromy group $(\mathrm{GL}_2(q) \rtimes C_f)/A$ over $\mathbb{F}_p$, where A is the kernel of the action of $\mathrm{GL}_2(q) \rtimes C_f$ on $\mathcal{H}_{\mathbb{1}}$ and C_f induces the full outer automorphism group (of order f) of the simple group $\mathrm{PSL}_2(q)$; furthermore $|A| = q+1$ by [KT5, Corollary 8.2]. Thus, if B is the kernel of the action of $G_{\mathrm{arith},\mathbb{F}_p}$ on $\mathcal{H}_0$, then

$$G_{\mathrm{arith},\mathbb{F}_p}/B \cong (\mathrm{GL}_2(q) \rtimes C_f)/A,$$

and so $|B| = 2(q+1)$. We note that B centralizes $L = G^{\star}_{\mathrm{geom}} \cong \mathrm{SU}_2(q)$. Indeed, as L is perfect, $[B,L] = [B,[L,L]]$ is contained in $[B \cap L, L]$. Now $B \cap L \lhd L$, and any normal subgroup of order $\leq 2(q+1)$ of L is central in L. Hence $[B \cap L, L] = 1$, and so $[B,L] = 1$, as claimed. Also, $(\mathrm{GL}_2(q) \rtimes C_f)/A$ induces the full automorphism group $\mathrm{PGL}_2(q) \rtimes C_f$ of $\mathrm{PSL}_2(q) \cong \mathrm{PSU}_2(q)$. Hence $|\mathbf{C}_{G_{\mathrm{arith},\mathbb{F}_p}}(\mathrm{SU}_2(q))| = 2(q+1)$, and the statements in (c) for $G_{\mathrm{arith},\mathbb{F}_p}$ follow.

Furthermore, as shown in Lemma 16.3, when $j|2f$, $\mathrm{Trace}(\Phi((g^{\star})^j))$ can be a power of $-q$ only for $j = 2f$. Since $\mathrm{Trace}(\Phi(h))$ is a power of $-q$ for any $h \in G_{\mathrm{geom}}$ and $G^{\star}_{\mathrm{geom}} \leq G_{\mathrm{geom}}$, it then follows from (17.5.17) that

$$G^{\star}_{\mathrm{arith},\mathbb{F}_p} \cap G_{\mathrm{geom}} = G^{\star}_{\mathrm{geom}}.$$

Together with (17.5.16), this implies that $G_{\mathrm{arith},\mathbb{F}_p} = G^{\star}_{\mathrm{arith},\mathbb{F}_p} G_{\mathrm{geom}}$. Now, G_{geom} induces only inner-diagonal automorphisms of $\mathrm{SU}_2(q)$ whereas $G_{\mathrm{arith},\mathbb{F}_p}$ induces the full automorphism group of $\mathrm{SU}_2(q)$. It follows that $G^{\star}_{\mathrm{arith},\mathbb{F}_p}$ must induce the full group C_f of outer field automorphisms of $\mathrm{SU}_2(q)$, and thus (d) follows for $G^{\star}_{\mathrm{arith},\mathbb{F}_p}$.

Note that $G_{\mathrm{geom}} = \mathrm{GU}_2(q)$ induces the full subgroup $\mathrm{PGU}_2(q)$ of inner-diagonal automorphisms of $\mathrm{SU}_2(q)$, and the quotient $G_{\mathrm{arith},\mathbb{F}_p}/G_{\mathrm{geom}} \cong C_{2f}$ maps onto the group C_f of outer field automorphisms of $\mathrm{SU}_2(q)$, hence with kernel C_2, the unique subgroup of order 2 in it, which then must coincide with $G_{\mathrm{arith},\mathbb{F}_q}/G_{\mathrm{geom}}$. Arguing as in part (i) of the proof of Theorem 16.12, we also obtain (c) and (d) for $G_{\mathrm{arith},\mathbb{F}_{q^{2/d}}}$ and $G^{\star}_{\mathrm{arith},\mathbb{F}_{q^{2/d}}}$. □

Note that the extra $\gcd(2,d)$ factor in Theorem 17.5(c) and (d), compared to Theorem 16.12, is explained by the fact that the transpose-inverse automorphism of $\mathrm{SU}_n(q)$ becomes an inner-diagonal automorphism when $n = 2$.

References

[Abh] S. Abhyankar, Coverings of algebraic curves. *Amer. J. Math.* **79** (1957), 825–856.

[BEW] B. C. Berndt, B. C., R. J. Evans, and K. S. Williams, Gauss and Jacobi Sums. Can. Math. Soc. Series of Monographs and Advanced Texts, Wiley, New York, 1998, xii+583 pp.

[De] P. Deligne, La conjecture de Weil II. *Pub. Math. I.H.E.S.* **52** (1981), 313–428.

[Do] L. Dornhoff, Group Representation Theory. Dekker, New York, 1971.

[E] V. Ennola, On the characters of the finite unitary groups. *Ann. Acad. Scient. Fenn.* A I, no. **323** (1963).

[GAP] The GAP group, GAP – groups, algorithms, and programming. Version 4.4, 2004, http://www.gap-system.org.

[Geck] M. Geck, Irreducible Brauer characters of the 3-dimensional special unitary groups in non-describing characteristic. *Comm. Algebra* **18** (1990), 563–584.

[GLS] R. Gorenstein, R. Lyons, and R. M. Solomon, The Classification of the Finite Simple Groups, Number 3. Part I. Chapter A, vol. **40**, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1998.

[Gr1] A. Grothendieck, Formule de Lefschetz et rationalité des fonctions L. Seminaire Bourbaki 1964–65, Exposé **279**, reprinted in Dix Exposés sur la cohomologie des schémas, North-Holland, 1968.

[Gr2] A. Grothendieck, augmented de deux exposés de Mme. M. Raynaud, Revêtements Etales et Groupe Fondamental. Lecture Notes in Math. **224**, Springer-Verlag, 1971.

[Gro] B. H. Gross, Group representations and lattices. *J. Amer. Math. Soc.* **3** (1990), 929–960.

[GMT] R. M. Guralnick, K. Magaard, and P. H. Tiep, Symmetric and alternating powers of Weil representations of finite symplectic groups. *Bull. Inst. Math. Acad. Sinica* **13** (2018), 443–461.

[GT] R. M. Guralnick and P. H. Tiep, Symmetric powers and a conjecture of Kollár and Larsen. *Invent. Math.* **174** (2008), 505–554.

[HM] G. Hiss and G. Malle, Low-dimensional representations of quasi-simple groups. *LMS J. Comput. Math.* **4** (2001), 22–63.

[Is] I. M. Isaacs, Character Theory of Finite Groups. AMS-Chelsea, Providence, 2006.

[K1] N. Katz, Gauss sums, Kloosterman sums, and monodromy groups. Annals of Mathematics Studies, **116**. Princeton Univ. Press, Princeton, NJ, 1988. ix+246 pp.

[K2] N. Katz, Exponential sums and differential equations. Annals of Mathematics Studies, **124**. Princeton Univ. Press, Princeton, NJ, 1990. xii+430 pp.

[K3] N. Katz, On the monodromy groups attached to certain families of exponential sums. *Duke Math. J.* **54** (1987), 41–56, and Correction, *Duke Math. J.* **89** (1997), 201.

[K4] N. Katz, G_2 and hypergeometric sheaves. *Finite Fields Appl.* **13** (2007), 175–223.

[K5] N. Katz, From Clausen to Carlitz: low-dimensional spin groups and identities among character sums. *Mosc. Math. J.* **9** (2009), 57–89.

[KRL] N. Katz and A. Rojas-León, A rigid local system with monodromy group $2.J_2$. *Finite Fields Appl.* **57** (2019), 276–286.

[KRLT1] N. Katz, A. Rojas-León, and P. H. Tiep, Rigid local systems with monodromy group the Conway group Co_3. *J. Number Theory* **206** (2020), 1–23.

[KRLT2] N. Katz, A. Rojas-León, and P. H. Tiep, A rigid local system with monodromy group the big Conway group $2.\mathsf{Co}_1$ and two others with monodromy group the Suzuki group $6.\mathsf{Suz}$. *Trans. Amer. Math. Soc.* **373** (2020), 2007–2044.

[KT1] N. Katz, with an Appendix by P. H. Tiep, Rigid local systems on $\mathbb{A}^1$ with finite monodromy. *Mathematika* **64** (2018), 785–846.

[KT2] N. Katz and P. H. Tiep, Rigid local systems and finite symplectic groups. *Finite Fields Appl.* **59** (2019), 134–174.

[KT3] N. Katz and P. H. Tiep, Local systems and finite unitary and symplectic groups. *Adv. Math.* **358** (2019), 106859.

[KT4] N. Katz and P. H. Tiep, Moments of Weil representations of finite special unitary groups. *J. Algebra* **561** (2020), 237–255.

[KT5] N. Katz and P. H. Tiep, Rigid local systems and finite general linear groups. *Math. Z.* **298** (2021), 1293–1321.

[KT6] N. Katz and P. H. Tiep, Exponential sums and total Weil representations of finite symplectic and unitary groups. *Proc. Lond. Math. Soc.* **122** (2021), 745–807.

[KT7] N. Katz and P. H. Tiep, Monodromy groups of Kloosterman and hypergeometric sheaves. *Geom. Funct. Analysis* **31** (2021), 562–662.

[KT8] N. Katz and P. H. Tiep, Local systems, extraspecial groups, and finite unitary groups in characteristic 2 (in preparation).

[Lee] T.-Y. Lee, A question of Katz and Tiep on representations of finite general unitary groups (submitted).

[TZ1] P. H. Tiep and A. E. Zalesskii, Minimal characters of the finite classical groups. *Comm. Algebra* **24** (1996), 2093–2167.

[TZ2] P. H. Tiep and A. E. Zalesskii, Some characterizations of the Weil representations of the symplectic and unitary groups. *J. Algebra* **192** (1997), 130–165.

[vG-vV] G. van der Geer and M. van der Vlugt, Reed–Muller codes and supersingular curves. I. *Compos. Math.* **84** (1992), 333–367.

[Zs] K. Zsigmondy, Zur Theorie der Potenzreste. *Monatsh. Math. Phys.* **3** (1892), 265–284.

Nicholas M. Katz
Department of Mathematics
Princeton University
Princeton, NJ 08544
USA
E-mail address: nmk@math.princeton.edu

Pham Huu Tiep
Department of Mathematics
Rutgers University
Piscataway, NJ 08854
USA
E-mail address: tiep@math.rutgers.edu

Received July 30, 2021