COMPOSITIO MATHEMATICA

RICHARD CREW

Kloosterman sums and monodromy of a p-adic hypergeometric equation

Compositio Mathematica, tome 91, n° 1 (1994), p. 1-36
<http://www.numdam.org/item?id=CM_1994__91_1_1_0>

Compositio Mathematica **91**: 1–36, 1994.

Kloosterman sums and monodromy of a *p*-adic hypergeometric equation

RICHARD CREW*

Department of Mathematics, University of Florida, Gainesville, FL 32611 USA

Received 10 May 1990; accepted in final form 30 November 1992

Suppose that X is a separated, geometrically connected scheme over a finite field k of characteristic p, and that $\{\rho_l\}_l$ is a comparable system of semisimple l-adic representations of $\pi_1(X)$, in the sense of Serre [21, I-10]. Let $\pi_1^{\mathrm{geom}}(X)$ denote the geometric fundamental group $\pi_1(X \otimes k^{alg})$, where k^{alg} is the algebraic closure of k, and let G_l be the Zariski closure of the image of $\rho|\pi_1^{\mathrm{geom}}$. Are the G_l in any sense "the same" for all l, or for almost all l? This kind of question goes back to Serre [21], and does not yet have a complete answer, although significant results in this direction have been obtained by Larsen and Pink [18]. Suppose now that (M, F) is an overconvergent isocrystal on X. In [7], we defined a kind of differential Galois group DGal(M) attached to M, and showed that in many ways it behaves like a geometric monodromy group; for example, it satisfies an analogue of Grothendieck's global monodromy theorem [7, Th. 4.9]. Suppose now that (M, F) is *compatible* with $\{\rho_l\}_l$ in the sense that for each closed point $x \in X$, the characteristic polynomial of F_x acting on M_x is the same as the common characteristic polynomial of the $\rho_l(\mathrm{Frob}_x)$, where Frob_x is a Frobenius element at x. Is Dgal(M) in any sense "the same" as the G_l?

We study this question in the case where $X = \mathbb{G}_m = \mathbb{P}^1 - \{0, \infty\}$ and (M, F) is the p-adic hypergeometric equation of rank n studied by Dwork [10] and Sperber [22, 23, 24]; on account of its relation to the generalized Kloosterman sums, we shall call it the *Kloosterman isocrystal* $\mathrm{Kl}(n, \psi)$. The corresponding l-adic representations were constructed by Deligne [SGA $4\frac{1}{2}$ Sommes Trig. §7], and their monodromy groups G_l were determined by Katz [16]. We will calculate DGal($\mathrm{Kl}(n, \psi)$) in the case that p is odd and does not divide n, which is the case studied by Sperber (although the results of Katz suggest that the case $p = 2$ is the most interesting). We will find that DGal($\mathrm{Kl}(n, \psi)$) is indeed isomorphic to G_l, given a suitable identification of the base fields. Along the way we shall have to calculate the monodromy group of the *convergent* isocrystal $\widehat{\mathrm{Kl}}(n, \psi)$ corresponding to

* Partialy supported by the NSF and the NSA.

$\mathrm{Kl}(n,\psi)$. In contrast to $\mathrm{DGal}(\mathrm{Kl}(n,\psi))$, $\mathrm{DGal}(\widehat{\mathrm{Kl}}(n,\psi))$ is a solvable algebraic group, since $\widehat{\mathrm{Kl}}(n,\psi)$ has a filtration by sub-F-isocrystals with quotients of rank one. These rank one F-isocrystals are twists of unit-root F-isocrystals on $\mathbb{G}_m/\mathbb{F}_q$, and thus give rise to a set of p-adic characters $\psi_i^{(n)}$ of $\pi_1(\mathbb{G}_m)$. In [16] Katz asked if there were any relations between the $\psi_i^{(n)}$ other than the ones which come from the fact that $\mathrm{DGal}(\widehat{\mathrm{Kl}}(n,\psi))$ is contained in $\mathrm{SL}(n)$, and in the symplectic group $\mathrm{SP}(n)$ if n is even. We shall show that there are no other relations; in fact, it was the effort to answer this question that led me to calculate $\mathrm{DGal}(\widehat{\mathrm{Kl}}(n,\psi))$. We will show also that there are no relations between the $\psi_i^{(n)}$ for variable n prime to p, which will lead us to consider the direct sums $\mathrm{Kl}(I,\psi)$ (c.f. (2.2.3) below) of Kloosterman F-isocrystals.

The first two sections recall the basic facts about isocrystals, monodromy groups, and the results of Sperber concerning $\mathrm{Kl}(n,\psi)$. The main results of the paper are stated in §2. In §3 we define a kind of local monodromy group, and calculate the local monodromy at 0 of $\mathrm{Kl}(n,\psi)$. The bulk of §4 is devoted to showing that $\mathrm{DGal}(\mathrm{Kl}(n,\psi)$ contains a maximal torus of $\mathrm{SL}(n)$ or $\mathrm{Sp}(n)$, depending on the parity of n. Although the question is basically a p-adic analytic one, the argument is essentially Diophantine, in that it uses in an essential way the results of Katz [16] on the equidistribution of the Frobenius classes attached to the l-adic Kloosterman sheaves.

I would like to thank Steve Sperber for a number of helpful conversations about the Kloosterman equation, and Nick Katz for a thorough critical reading of the manuscript. I am also indebted to the National Science Foundation and the National Security Agency for financial support received during various stages of the work described here.

0. Notation and terminology

0.1. If K is a field of characteristic zero, we denote by Vec_K the category of K-vector spaces (not necessarily of finite dimension). If G is an affine K-groups, then $\mathrm{Rep}_K(G)$ is the category of representations of G on K-vector spaces. By *K-linear $\otimes$-category* we will mean what Saavedra [20] calls a K-linear rigid abelian ACU $\otimes$-category. We will say that an object of a $\otimes$-category *trivial* if it is isomorphic to a sum of unit objects, the latter being as in [20, 1.3]. If $\mathscr{C}$ is a K-linear Tannakian category and L is an extension of K, then we denote by $\mathscr{C}\otimes L$ the extension of scalars $(\text{ind-}\mathscr{C})_{(L)}$ [20, 1.5 and 1.5.4) of the ind-Tannakian category ind-$\mathscr{C}$.

0.2. p will always be a fixed *odd* prime. k will be a perfect field of characteristic p, and K will usually be a finite extension of the fraction field of the ring $\mathrm{W}(k)$ of Witt vectors of k. R is the ring of integers in K, and π

is a uniformizer of R. Unless otherwise stated, k-schemes will be separated and of finite type.

1. The Kloosterman F-isocrystal

1.1. In this paper we will be concerned exclusively with isocrystals on a smooth affine curve, in which case we can give a fairly simple description of what a convergent or overconvergent isocrystal is. For the general theory we refer the reader to [3] and to the summaries in [1, 2, 7]. Let X/k be a smooth affine curve. We can find a formally smooth formal R-scheme $\mathfrak{X}/R$ lifting X/k, and we denote the corresponding affinoid rigid-analytic space by $\mathfrak{X}^{an}$. Then a *convergent isocrystal* on X/K is a locally free $\mathcal{O}_{\mathfrak{X}^{an}}$-module endowed with a convergent connection in the sense of [1, 4.1]; in other words, a convergent isocrystal on X/K can be identified with a connection on a locally free $\mathcal{O}_{\mathfrak{X}^{an}}$-module with the property that the corresponding differential system has at every point of $x \in \mathfrak{X}^{an}$ a full set of formal solutions converging in the open unit disk around x (for the affinoid norm). The category of convergent isocrystals on X/K is independent of the choice of the lifting $\mathfrak{X}/R$ of X/K, and is moreover natural in X/K, and of local nature on X. We will denote it by $\mathrm{Isoc}(X/K)$.

The condition of *overconvergence* relates to the behavior of the connection at the "boundary" of $\mathfrak{X}^{an}$. More precisely, suppose that Y/k is another smooth affine curve containing X such that $Y - X$ is a single point x_0. We can find a lifting $\mathfrak{X} \hookrightarrow \mathfrak{Y}$ of $X \hookrightarrow Y$, with $\mathfrak{X}^{an} \hookrightarrow \mathfrak{Y}^{an}$ the corresponding inclusion of rigid-analytic spaces. If t is a local section of $\mathcal{O}_{\mathfrak{Y}}$ reducing to a local parameter at x_0, then $\mathfrak{Y}^{an} - \mathfrak{X}^{an}$ is just the open unit disk $|t| < 1$. Suppose we are given a convergent isocrystal on X/K, i.e. a locally free $\mathcal{O}_{\mathfrak{X}^{an}}$-module $\mathcal{M}$ endowed with an integrable connection ∇. Then $(\mathcal{M}, \nabla)$ is *overconvergent at* x if the following two conditions hold:

(1.1.1) $(\mathcal{M}, \nabla)$ can be extended to an admissible open neighborhood $U \subset \mathfrak{Y}^{an}$ of $\mathfrak{X}^{an}$ of the form $|t| > r$ for some $r < 1$ (such an open is called a *strict neighborhood* of $\mathfrak{X}^{an}$ in $\mathfrak{Y}^{an}$);

(1.1.2) Given any positive $s < 1$, there is an r such that at any point of the annulus $r < |t| < 1$ there is a full set of formal horizontal sections of ∇ whose radius of convergence is at least s.

If these conditions hold for one choice of Y and liftings $\mathfrak{X}$, $\mathfrak{Y}$, then they hold for any such choice. Furthermore, it is clear that the validity of these conditions is a local question around x_0. We shall call a choice of $\mathfrak{X} \hookrightarrow \mathfrak{Y}$ and an extension of $(\mathcal{M}, \nabla)$ to a strict neighborhood of $\mathfrak{X}^{an}$ in $\mathfrak{Y}^{an}$ *overconvergence data* around x_0. Then if X/k is a smooth curve and $\bar{X}/k$ is a smooth compactification of X, then an *overconvergent isocrystal* on X/K

is a convergent isocrystal on X/K together with overconvergence data around each point of $\bar{X} - X$. The category of overconvergent isocrystals on X/K, which we denote by $\mathrm{Isoc}^{\dagger}(X/K)$, is of local nature on X and functorial in X/K.

The evident forgetful functor

$$\begin{aligned} \mathrm{Isoc}^{\dagger}(X/K) &\to \mathrm{Isoc}(X/K) \\ M &\mapsto \hat{M} \end{aligned} \tag{1.1.3}$$

will be called the *completion* functor, as in [7]. It is faithful, and is conjectured but not known to be fully faithful. In §2 we shall see that the restriction of the completion functor to the tensor category generated by the Kloosterman F-isocrystals is fully faithful. Another basic but unanswered question is whether the restriction functors

$$\mathrm{Isoc}^{\dagger}(X/K) \to \mathrm{Isoc}^{\dagger}(U/K)$$
$$\mathrm{Isoc}(X/K) \to \mathrm{Isoc}(U/K)$$

are fully faithful. Again, we shall see that these functors are fully faithful on the $\otimes$-category generated by Kloosterman F-isocrystals of rank prime to p (Corollary 2.5).

If $F: X \to X$ is a power of the absolute Frobenius morphism and M is a convergent isocrystal on X/K, then a *Frobenius structure* on M is an isomorphism $\Phi: F^*M \xrightarrow{\sim} M$, and the pair (M, Φ) is called a *convergent F-isocrystal.* The same definition in the overconvergent category yields the notion of an *overconvergent F-isocrystal.* If we represent M by a locally free sheaf $\mathscr{M}$ on X endowed with a convergent connection ∇, and if φ is a lifting of F to X, then a Frobenius structure on $(\mathscr{M}, \nabla)$ is a horizontal isomorphism $\Phi: \varphi^*\mathscr{M} \xrightarrow{\sim} \mathscr{M}$, and $(\mathscr{M}, \nabla, \Phi)$ defines a structure of a convergent isocrystal. Similarly, if (M, ∇) defines an overconvergent isocrystal, then one needs an isomorphism $\Phi: \varphi^*\mathscr{M} \xrightarrow{\sim} \mathscr{M}$ that is extendible into a strict neighborhood of $\mathfrak{X}^{an}$, i.e. an "overconvergent" Frobenius structure, in order to make $(\mathscr{M}, \nabla)$ into an overconvergent F-isocrystal. We denote by F-$\mathrm{Isoc}(X/K)$, resp. F-$\mathrm{Isoc}^{\dagger}(X/K)$ the category of convergent resp. overconvergent F-isocrystals on X/K.

If the maximal ideal of R has divided powers, and if F-$\mathrm{Cris}(X/R)$ denotes the category of F-crystals on X/R, then there is a functor.

$$F\text{-}\mathrm{Cris}(X/R) \to F\text{-}\mathrm{Isoc}(X/K) \tag{1.1.4}$$

that is fully faithful up to isogeny, and essentially surjective up to Tate twists if X/k is smooth [3, Th. 2.3.12]. The F-isocrystals in the essential image of 1.1.4 are not necessarily overconvergent. In the case when X is smooth and affine, the functor 1.1.4 is the "evident" one: if, say, $\mathfrak{X}/R$ is a formally smooth lifting of X/k, then an F-crystal on X/R is a coherent $\mathcal{O}_{\mathfrak{X}}$-module $\mathcal{M}$ endowed with an integrable nilpotent connection; the corresponding connection on $\mathcal{M}^{an}$ can be shown to be convergent, and thus defines a convergent isocrystal on X/K.

Recall now that p is a fixed *odd* prime (section 0.2). We take $k = \mathbb{F}_q$, and let K_0 be the extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$; furthermore we let π be a solution in $\mathbb{Q}_p$ of $\pi^{p-1} = -p$, and set $K = K_0[\pi]$. There is a unique character $\psi_0 : \mathbb{F}_p \to K^X$ such that $\psi_0(1) \equiv 1 + \pi \mod \pi^2$, and we set $\psi = \psi_0 \circ \mathrm{Tr}_{\mathbb{F}_1/\mathbb{F}_p}$. Finally we take $X = \mathbb{G}_m = \mathrm{Spec}(k[x, x^{-1}])$, $\mathfrak{X} = \mathrm{Spf}(R\{\{x, x^{-1}\}\})$, so that $\mathfrak{X}^{an} = \mathrm{Max}(K\langle\langle x, x^{-1}\rangle\rangle)$, and we consider the connection on the trivial sheaf on $\mathfrak{X}^{an}$ of rank n given by the following matrix of 1-forms:

$$\nabla = -\begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & \ddots & \\ & & & \ddots & 1 \\ \pi^n x & & & & 0 \end{pmatrix} \otimes \frac{dx}{x}. \tag{1.1.5}$$

1.2. PROPOSITION. *If p does not divide n, then the connection 1.1.5 defines an overconvergent isocrystal on $\mathbb{G}_m/K$.*

Proof. As we shall recall below (in the paragraph following (1.3.1)), the connection (1.1.5) is actually the underlying connection of an F-isocrystal in the essential image of (1.1.4), and is therefore convergent. To show overconvergence, we must examine the local behavior of (1.1.5) at 0 and infinity. First, we embed $X \hookrightarrow \mathbb{P}^1_k$ in the evident way; then the analytic space U corresponding to $\mathbb{G}_{m/K} = \mathrm{Spec}(K\text{-}[x, x^{-1}])$ is a strict neighborhood of X to which (1.1.5) extends. It remains to check (1.1.2). At 0 this is not difficult; in the punctured disk $0 < |x| < 1$, the connection (1.1.5) is equivalent to the connection

$$\nabla = N_n \otimes \frac{dx}{x} \quad \text{where } N_n = -\begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix}. \tag{1.2.1}$$

In fact the matrix $P(x)$ (c.f. [21, §4.2]) giving the equivalence can be chosen so that $P(0) = I$, and the entries of $P(x)$ are constant linear combinations of the hypergeometric function

$$ {}_0F_n(1,\ldots,1;\pi^n x) = \sum_{i\geqslant 0} \frac{(\pi^n x)^i}{(i!)^n} \tag{1.2.2}$$

and its derivatives ([22], 4.2.1, 4.2.7). Thus $P(x)$ converges for $|x| < 1$. Now the local horizontal sections of (1.2.1) around a point x_0 in $0 < |x| < 1$ are the columns of the matrix

$$\begin{aligned} x^{N_n}x_0^{-N_n} &= \exp(N_n \log(x/x_0)) \\ &= \exp(N_n \log(1 + (x - x_0)/x_0)). \end{aligned}$$

Since these converge in the disk $|x - x_0| < |x_0|$, we conclude that (1.1.5) satisfies (1.1.2) at $x = 0$.

Checking (1.1.2) at infinity is slightly trickier since the Turittin normal form of (1.1.5) is only defined after pulling back by $x \mapsto x^n$. Fortunately, we can check overconvergence by descent:

1.3. PROPOSITION. *Let $f: Y \to X$ be a finite étale morphism of smooth curves. Then descent data for f in* $\mathrm{Isoc}^\dagger(Y/K)$ *is effective; i.e. any overconvergent isocrystal on Y/K endowed with descent data relative to f is isomorphic, with its descent data, to the pullback of an overconvergent isocrystal on X/K.*

Proof. We will make use of the following construction: suppose that $f: Y \to X$ is finite étale, and galois with group G; suppose further that there is an embedding $X \hookrightarrow \bar{X}$ with $\bar{X}/k$ smooth and affine, and an extension $\bar{f}: \bar{Y} \to \bar{X}$ of f, with $\bar{f}$ proper and flat. Then since $\bar{X}$ is a smooth curve, $\bar{Y} \to \bar{X}$ is a local complete intersection; furthermore the action of G on Y extends to $\bar{Y}$, and $\bar{Y}^G = \bar{X}$. We can then find a formally smooth formal R-scheme $\bar{\mathfrak{X}}$ lifting X, and a lifting $\bar{\mathfrak{Y}} \to \bar{\mathfrak{X}}$ of $\bar{f}$ (not formally smooth, but a local complete intersection). Denote by $\mathfrak{X}$, $\mathfrak{Y}$ the restrictions to X, Y of $\bar{\mathfrak{X}}$, $\bar{\mathfrak{Y}}$. The action of G on Y/X lifts uniquely to $\mathfrak{Y}/\mathfrak{X}$; it does not necessarily extend to $\bar{\mathfrak{Y}}/\bar{\mathfrak{X}}$, but by [6], (3.2.2), there is an affinoid strict neighborhood U of $\mathfrak{X}^{an}$ in $\bar{\mathfrak{X}}^{an}$ such that if V is the inverse image of U in $\bar{\mathfrak{Y}}^{an}$, then $V \to U$ is étale and the action of G extends to V. We now claim that $V^G = U$. In fact, since $\bar{\mathfrak{Y}}$ is finite over $\bar{\mathfrak{X}}$, V is finite over U, and therefore V^G is finite over U. Since $\mathfrak{Y}^G = \mathfrak{X}$, and since the affinoid algebra of U is integrally closed in the affinoid algebra of $\mathfrak{X}^{an}$ by [5, Theorem 2], we have $V^G = U$.

Suppose now that M is an overconvergent isocrystal on Y endowed with descent data for a finite étale $f: Y \to X$. Without loss of generality we can assume that f is galois with group G. Since $\mathrm{Isoc}(X/K)$ is equivalent to the category of objects of Isoc(Y/K) with descent data for f by [19], 4.5, we see that the convergent isocrystal $\hat{M}$ descends to a convergent isocrystal N on X/K. We must show that N is overconvergent; i.e. if $X \hookrightarrow \bar{X}$ is a smooth compactification of X, we must show that N is overconvergent about each point $x \in \bar{X} - X$. This is a local question around each x, so we can replace $\bar{X}$ by an open neighborhood of x, and thus assume that $\bar{X}$ is affine and smooth. We can then find a proper flat G-cover $\bar{f}: \bar{Y} \to \bar{X}$ extending f, and apply the construction in the preceeding paragraph. Let U, V be as before; by shrinking U, we can assume that the overconvergent isocrystal M is represented by a locally free module M_V on V with an integrable connection, and the descent data on M is represented by a set of isomorphisms $g^*M_V \cong M_V$ for $g \in G$ satisfying a cocycle condition. By flat descent for rigid-analytic spaces (c.f. the proof of [19], 4.5) M_V descends to a module N_U with integrable connection on U, and by the result of Ogus cited above, the restriction of N_U to $\mathfrak{X}^{an}$ is N. Thus (1.1.1) is satisfied. Condition (1.1.2) follows from the inclusion $N_U = (f_* M_V)^G \hookrightarrow f_* M_V$, which is horizontal for the connections. □

1.3.1. REMARK. Suppose $f: Y \to X$ is finite étale, and M is a convergent isocrystal on X/K. If the pullback f^*M is overconvergent, then is M overconvergent too? This does not follow from 1.3, for while f^*M trivially has descent data in Isoc(Y/K), the descent data is not necessarily overconvergent, since the forgetful functor Isoc$^\dagger(Y/K) \to$ Isoc(Y/K) is not known to be fully faithful.

Returning to the proof of 1.2, we denote by f the maps $\mathbb{G}_{m,k} \to \mathbb{G}_{mk}$, $\mathbb{G}_{mK} \to \mathbb{G}_{mK}$ defined by $x \mapsto x^n$. By 1.3 it will be enough to show that the pullback connection f^*(1.1.5) is overconvergent, and can be given overconvergent descent data for f.

As regards the overconvergence of f^*(1.1.5), it suffices to check that (1.1.2) holds for f^*(1.1.5) at infinity. Now in the punctured disk $|x| > 1$, f^*(1.1.5) is equivalent to the connection given by the matrix

$$-\begin{pmatrix} -\frac{n-1}{2x} + n\pi & & & \\ & -\frac{n-1}{2x} + \zeta n\pi & & \\ & & \ddots & \\ & & & -\frac{n-1}{2x} + \zeta^{n-1} n\pi \end{pmatrix} \otimes dx \tag{1.3.1}$$

where ζ is a primitive nth root of unity (the change of basis matrix, which is the matrix denoted by $V(x)$ in [22], (5.1.2), has entries that are power series in x^{-1} convergent for $|x| > 1$ as long as p is odd and prime to n, c.f. [22], (5.1.7); the argument of [22] assumes that $p > n$, but it is in fact valid as long as n is prime to p). It is therefore enough to check the radius of convergence of the local horizontal sections of a connection matrix of the form

$$\left(\frac{n-1}{2x} - \zeta^i n\pi\right) dx.$$

In fact the local horizontal sections of this connection are the constant multiples of

$$(x/x_0)^{-(n-1)/2} \exp(\zeta^i n\pi(x - x_0))$$

which can be represented by power series in $(x - x_0)$ convergent for $|x - x_0| < 1$ if n is odd and $|x - x_0| < |x_0|$ if n is even. This is sufficient, as in $|x| > 1$ the appropriate local parameter is x^{-1}, and thus the disk $|x^{-1} - x_0^{-1}| < 1$ "actually" has radius $|x_0|^{-2}$. We conclude that (1.1.5) satisfies (1.1.2) at infinity.

Finally, we must see that $f^*(1.1.5)$ can be given overconvergent descent data for f. Now (1.1.5) can be viewed as a connection on a free module on the scheme $\mathbb{G}_m/K$, and so it has descent data for the map $f_K : \mathbb{G}_{mK} \to \mathbb{G}_{mK}$ given by $f_K(x) = x^n$. This yields descent data on the analytic space U corresponding to $\mathbb{G}_{mK}$. Since U is a strict neighborhood of $\mathbb{X}^{an}$, this gives descent data in $\mathrm{Isoc}^\dagger(\mathbb{G}_m/K)$. □

The *Kloosterman isocrystal* will be the overconvergent isocrystal on $\mathbb{G}_m/K$ of rank n whose existence is guaranteed by 1.2. We denote it by $\mathrm{Kl}(n, \psi)$, and the corresponding convergent isocrystal by $\widehat{\mathrm{Kl}}(n, \psi)$.

A Frobenius structure Φ for $\mathrm{Kl}(n, \psi)$ was first constructed by Dwork [10] when $n = 2$ and by Sperber [22, 23] in general. It is not so easy to describe Φ explicitly, but Sperber shows that when p does not divide n, it is given by a matrix all of whose entries have affinoid norm $\leqslant 1$, and has the form

$$\Phi = \begin{pmatrix} 1 & & & & \\ & q & & & \\ & & q^2 & & \\ & & & \ddots & \\ & & & & q^{n-1} \end{pmatrix} B \qquad B \equiv \begin{pmatrix} 1 & * & * & \cdots & * \\ & 1 & * & & \\ & & 1 & & \vdots \\ & & & \ddots & * \\ & & & & 1 \end{pmatrix} \bmod \pi$$

(c.f. [22], 5.46; note that [22] defines a Frobenius structure for the absolute, i.e. pth power, Frobenius; (1.1.5) is a power of this one). The matrix Φ is regular at $x = 0$, and $\Phi(0)$ is upper triangular, by [23, 0.28].

In fact, with this Frobenius structure, $\widehat{\mathrm{Kl}}(n, \psi)$ is isomorphic to the image under (1.1.2) of an F-crystal on X/R; the only non-obvious condition to check is the p-adic nilpotence of the connection matrix, which is a direct consequence of the estimates §3.2.1 of [22]. As a consequence, the theory of the slope filtration for F-crystals [14] is applicable, and from 1.3.3 one sees immediately that $\widehat{\mathrm{Kl}}(n, \psi)$ has a filtration by sub-F-isocrystals

$$0 = N_0 \subset N_1 \subset \cdots \subset N_n = \widehat{\mathrm{Kl}}(n, \psi) \tag{1.3.4}$$

whose quotients

$$M_i^{(n)} \overset{\mathrm{def}}{=} N_{i+1}/N_i \qquad 0 \leqslant i \leqslant n-1$$

are of rank one, and are such that the induced action of Φ on $M_i^{(n)}$ is purely of slope i. Thus the $M_i^{(n)}$ are twists of unit root F-isocrystals, and so by the theory of unit-root isocrystals [13, §3], [6, §2] give rise to a set of rank one characters $\psi_i^{(n)}$ of $\pi_1(\mathbb{G}_m)$. As representations, they are characterized by the relation

$$\begin{aligned}\psi_i^{(n)}(\mathrm{Frob}_x)q^{i\cdot\deg(x)} &= \Phi_x \,|\, (M_i^{(n)})_x \\ &= \text{the eigenvalue of } \Phi_x \,|\, (\widehat{\mathrm{Kl}}(n, \psi))_x \text{ of slope } i\end{aligned}$$

where Frob_x denotes the geometric Frobenius element for the closed point x.

1.4. We will need some other basic facts from [22]. We denote by $\mathcal{O}(n)$ the "Tate twist" F-isocrystal on $X/\mathbb{F}_q$, i.e. the constant isocrystal with the Frobenius structure given by multiplication by q^n. Sperber constructs a pairing (c.f. [22], 4.1.6)

$$\mathrm{Kl}(n, \psi) \otimes \mathrm{Kl}(n, \psi^{-1}) \to \mathcal{O}(n-1) \tag{1.4.1}$$

which is easily seen to be a morphism in $F\text{-Isoc}^\dagger(\mathbb{G}_m/K)$ (here $\mathcal{O}(n-1)$ denotes a twist of the trivial F-isocrystal). If n is even, then there is an isomorphism

$$\mathrm{Kl}(n, \psi) \cong \mathrm{Kl}(n, \psi^{-1}) \tag{1.4.2}$$

by means of which (1.4.1) gives rise to a symplectic pairing

$$\Lambda^2\mathrm{Kl}(n, \psi) \to \mathcal{O}(n-1) \tag{1.4.3}$$

in $F\text{-Isoc}^\dagger(\mathbb{G}_m/K)$.

We will need the explicit form of the pairing (1.4.3) in the punctured disk $U = \{x \mid 0 < |x| < 1\}$ about $0 \in \mathbb{P}^1_K$. With respect to a basis of $\mathrm{Kl}(n, \psi) \mid U$ for which the connection has the expansion (1.2.1), (1.4.3) is given by the matrix

$$\Theta = \begin{pmatrix} 0 & 0 & \cdots & (-1)^{n-1} \\ \vdots & & & \vdots \\ 0 & -1 & & 0 \\ 1 & 0 & & 0 \end{pmatrix} \tag{1.4.4}$$

[22, (4.1.5) and (4.2.8)].

For arbitrary n one sees immediately from (1.1.1) that there is an isomorphism

$$\det \mathrm{Kl}(n, \psi) \cong \mathcal{O}(n(n-1)/2) \tag{1.4.5}$$

in $F\text{-Isoc}^\dagger(\mathbb{G}_m/K)$.

1.5. The isocrystal $\mathrm{Kl}(n, \psi)$ and its Frobenius structure have a geometric construction [22, §6], [2, §3] which we now recall. Let $t_1, \ldots, t_n, x$, be affine coordinates on $\mathbb{A}^n_k$, resp. $\mathbb{A}^1_k$, and let p, f be the morphisms defined by

$$\begin{array}{ll} p : \mathbb{A}^n_k \to \mathbb{A}^1_k & \qquad f : \mathbb{A}^n_k \to \mathbb{A}^1_k \\ (t_1, \ldots, t_n) \mapsto x = t_1 \cdots t_n & \qquad (t_1, \ldots, t_n) \mapsto x = t_1 + \cdots + t_n \end{array}$$

and we let Z_0 denote the open subset $\mathbb{A}^n_k - p^{-1}(0)$. If $\psi : \mathbb{F}_p \to L^X$ is the character of $\mathbb{F}_p$ defined in the introduction, then the *Dwork F-isocrystal* $\mathscr{L}_\psi$ is the overconvergent F-isocrystal on $\mathbb{A}^1_k/K$ defined by the canonical sheaf on $\mathbb{A}^1/K$ with connection given by $\nabla = \pi \otimes dx$ and Frobenius structure given by $F(1) = \exp(\pi x^p - \pi x)$, c.f. [1, (4.2.3)]. If $g : Y \to \mathbb{A}^1_k$ is the Artin–Schreier cover $y^q - y = x$, then $\mathscr{L}_\psi$ is the ψ-isotypical part of the "rigid" direct image $g_{rig*}(\mathcal{O}_{Y/K})$ under the usual action of $\mathbb{F}_q$ on Y. By general principles, its pullback $f^*\mathscr{L}_\psi$ is overconvergent on Z_0/K. The calculations of [22, §6] can then be interpreted as saying that the relative rigid cohomology ([1, 4.3]) $R^{n-1}p_{rig*}f^*\mathscr{L}_\psi$ is isomorphic to $\mathrm{Kl}(n, \psi)$ (a similar but more general calculation is carried out in [Berth, 2, 3.2]). The Frobenius structure constructed by Sperber is the one induced by functoriality from the geometric Frobenius [2, 3.4].

Choose now a prime $l \neq p$ and an isomorphism $\bar{\mathbb{Q}}_l \cong \bar{K}$. Then the character $\psi: \mathbb{F}_q \to K^X$ can be identified with an l-adic character $\psi_l: \mathbb{F}_q \to \bar{Q}_l^X$, and we denote by $\mathscr{L}_{\psi_l}$ the l-adic Artin–Schreier sheaf on $\mathbb{A}^1/\mathbb{F}_p$. Then the l-adic Kloosterman sheaf is defined by

$$\mathrm{Kl}_l(n, \psi_l) = R^{n-1}p_* f^* \mathscr{L}_{\psi_l}$$

(this is actually a special case of the sheaves considered in [SGA $4\frac{1}{2}$ Sommes Trig. §7]). If Φ_l denotes the Frobenius correspondence on $\mathrm{Kl}_l(n, \psi_l)$, then the traces of Φ_l on the fibers of $\mathrm{Kl}_l(n, \psi_l)$ are Kloosterman sums, as are the traces of Φ on the fibers of $\mathrm{Kl}(n, \psi)$. More precisely, we have

$$\begin{aligned} \mathrm{Tr}\, \Phi^k \mid \mathrm{Kl}(n, \psi)_a &= \mathrm{Tr}\, \Phi^k \mid \mathrm{Kl}_l(n, \psi_l)_a \\ &= \sum_{\substack{x_1 \dots x_n = a \\ x_i \in \mathbb{F}_q}} \psi \circ \mathrm{Tr}_{\mathbb{F}_q}/\mathbb{F}_p(x_1 + \cdots + x_n) \end{aligned} \tag{1.5.1}$$

for any $k > 0$ and any fixed point $a \in \mathbb{G}_m(\bar{\mathbb{F}}_q)$ of the q^k-power Frobenius. The eigenvalues of the fibers of Φ_l (and, therefore, Φ) were shown by Deligne to be algebraic integers and pure of weight $n - 1$ [SGA$4\frac{1}{2}$] Sommes Trig. 7.5.

From the form of the slope filtration 1.3.4, we know that the eigenvalues of Φ on a fiber of $\mathrm{Kl}(n, \psi)$ have distinct p-adic ordinals, and are therefore distinct. It follows that the $\Phi_a \mid \mathrm{Kl}(n, \psi)_a$ are *semisimple.*

Finally, let $\rho_n: \pi_1(\mathbb{G}_m) \to \mathrm{GL}(n, \bar{\mathbb{Q}}_l)$ denote the representation corresponding to the lisse sheaf $\mathrm{Kl}_l(n, \psi_l)$. Of fundamental importance to us will be Katz's determination [16] of the Zariski closure of the geometric monodromy group of $\mathrm{Kl}_l(n, \psi_l)$, i.e. the Zariski closure G_n^l in $\mathrm{GL}(n)_{\mathbb{Q}_l}$ of the image under ρ_n of $\pi_1(\mathbb{G}_m \otimes \bar{\mathbb{F}}_q)$. Since p is odd, [16], (11.1) say that

$$G_n^l = \begin{cases} \mathrm{SL}(n) & n \text{ odd} \\ \mathrm{Sp}(n) & n \text{ even} \end{cases} \tag{1.5.2}$$

2. Global monodromy

2.1. The monodromy groups of an isocrystal are defined by means of the theory of Tannakian categories, in the same way as the differential Galois groups of [15]. Let M be a convergent (resp. overconvergent) isocrystal on X/K, whee X/k is separated, of finite type, and geometrically connected, and let x_0 be a k-point of X. Then since $\mathrm{Isoc}(x_0/K)$ (resp. $\mathrm{Isoc}^\dagger(x_0/K)$) is equivalent to the category Vec_K of K-vector spaces, pulling back by $x_0 \to X$ yields a functor

$$\omega : \mathrm{Isoc}(X/K) \to \mathrm{Vec}_K$$

$$\text{resp.} \quad \omega : \mathrm{Isoc}^\dagger(X/K) \to \mathrm{Vec}_K$$

which is faithful, exact, and compatible with the tensor-product structures on these categories; i.e. ω is a *fiber functor* in the sense of Saavedra [20]. Let [M] denote the $\otimes$-category generated in $\mathrm{Isoc}(X/K)$ resp. $\mathrm{Isoc}^\dagger(X/K)$ by M, i.e. the full subcategory consisting of all subquotients of objects of $\mathrm{Isoc}(X/K)$ resp. $\mathrm{Isoc}^\dagger(X/K)$ of the form $M^{\otimes n} \otimes \check{M}^{\otimes m}$ for all n, $m \geqslant 0$ ($\check{M}$ denotes the dual of M). Then the automorphism group scheme $\mathrm{Aut}^{\otimes}\omega \,|\, [M]$ of the restriction of ω to $[M]$ is an algebraic group which we call $\mathrm{DGal}(M)$, or $\mathrm{DGal}(M, x_0)$ if it is necessary to specify x_0. In fact, given two choices x_0, x_1 of base point there is an isomorphism $\mathrm{DGal}(M, x_0) \simeq \mathrm{DGal}(M, x_1)$, canonical up to inner automorphisms. The theory of Tannakian categories gives an equivalence of categories

$$[M] \xrightarrow{\sim} \mathrm{Rep}_K(\mathrm{DGal}(M, x_0))$$

natural in M, and under the above equivalence, the fiber functor ω is identified with the forgetful functor $\mathrm{Rep}_K(\mathrm{DGal}(M)) \to \mathrm{Vec}_K$.

We will need some elementary results and compatibilities from [7]. If M is any object of $\mathrm{Isoc}^\dagger(X/K)$, then the completion functor

$$\mathrm{Isoc}^\dagger(X/K) \to \mathrm{Isoc}(X/K)$$

induces a closed immersion

$$\mathrm{DGal}(\hat{M}) \hookrightarrow \mathrm{DGal}(M). \tag{2.1.1}$$

Let M be an object of $\mathrm{Isoc}(Y/K)$ or $\mathrm{Isoc}^\dagger(Y/K)$. If $f : X \to Y$ is separated of finite type, there is a closed immersion

$$\mathrm{DGal}(f^*M) \hookrightarrow \mathrm{DGal}(M) \tag{2.1.2}$$

and if K'/K is a finite extension, there is an isomorphism [7, (2.1.9) and (2.1.10)]

$$\mathrm{DGal}(M \otimes K') \simeq \mathrm{DGal}(M) \otimes K'. \tag{2.1.3}$$

If X and Y are smooth curves and $f: X \to Y$ is finite étale, then (2.1.2) induces an isomorphism

$$\mathrm{DGal}(f^*M)^0 \xrightarrow{\sim} \mathrm{DGal}(M)^0 \tag{2.1.4}$$

on the connected components, as one sees easily from [7], (2.1.4) and (4.5). Finally, for any smooth curve X/k and any M in $\mathrm{Isoc}(X/K)$ or $\mathrm{Isoc}^\dagger(X/K)$, there is a finite étale cover $f: X' \to X$ such that $\mathrm{DGal}(f^*M)$ is connected, so that (2.1.3) becomes an isomorphism [7, (4.6)]

$$\mathrm{DGal}(f^*M) \xrightarrow{\sim} \mathrm{DGal}(M)^0. \tag{2.1.5}$$

Suppose that X/k is smooth and geometrically connected, and (M, F) is a convergent *unit-root* F-isocrystal on X/K. Let $\pi_1^{\mathrm{geom}}(X)$ denote the geometric fundamental group $\pi_1(X \otimes k^{\mathrm{alg}})$ of X. One can associate to (M, F) a representation $\rho: \pi_1(X) \to \mathrm{GL}(V)$ on some K-vector space V (c.f. [6, §2] and [13, §3]; actually ρ can be defined over a smaller field), and $\mathrm{DGal}(M)$ is, after an extension of scalars, isomorphic to the Zariski-closure in $\mathrm{GL}(V)$ of the image of $\rho \mid \pi_1^{\mathrm{geom}}(X)$. It is not known whether a similar assertion is true for unit-root *overconvergent* F-isocrystals.

2.2. We now take $k = \mathbb{F}_q$, and let U be a dense open subscheme of $\mathbb{G}_m/k$. Let $f: X \to U$ be a finite étale cover such that $X(k) \neq \phi$, and choose a k-point x_0 of $X = \mathbb{G}_a$ which will be the implicit base point for all monodromy groups from now on, unless otherwise stated. We denote by V_n the fiber of $f^*\mathrm{Kl}(n, \psi)$ at x_0, and by

$$0 \subset W_0 \subset W_1 \subset \cdots \subset W_{n-1} = V_n \tag{2.2.1}$$

the fiber at x_0 of the pullback by f of slope filtration (1.3.4). When n is even, we denote by $\psi_n: V_n \times V_n \to K$ the fiber at x_0 of the symplectic pairing (1.4.3). Finally, we set

$$G_n = \begin{cases} \mathrm{SL}(V_n) & n \text{ odd} \\ \mathrm{Sp}(V_n, \psi_n) & n \text{ even} \end{cases} \tag{2.2.2}$$

and we denote by B_n the Borel subgroup of G_n which fixes the flag (2.2.1). Then from (1.4.3) and (1.4.5), we see that $\mathrm{DGal}(f^*\mathrm{Kl}(n, \psi)) \subseteq G_n$, and since (2.2.1) is the fiber at x_0 of a filtration of $f^*\widehat{\mathrm{Kl}}(n, \psi)$, we have $\mathrm{DGal}(\mathrm{Kl}(n, \psi)) \subseteq B_n$.

We will want to work, in fact, with several Kloosterman isocrystals at a time. For the rest of this paper we denote by I a finite set of distinct positive integers not divisible by p, and by $\mathrm{Kl}(I, \psi)$ the overconvergent F-isocrystal

$$\mathrm{Kl}(I, \psi) = \bigoplus_{n \in I} \mathrm{Kl}(n, \psi), \tag{2.2.3}$$

Let $\widehat{\mathrm{Kl}}(I, \psi)$ denote the completion of $\mathrm{Kl}(I, \psi)$, i.e. the corresponding convergent F-isocrystal. Finally, set

$$G_1 = \prod_{n \in I} G_n, \qquad B_I = \prod_{n \in I} B_n. \tag{2.2.4}$$

Then by the above considerations, we have

$$\mathrm{DGal}(\mathrm{Kl}(I, \psi)) \subseteq G_I, \qquad \mathrm{DGal}(\widehat{\mathrm{Kl}}(I, \psi)) \subseteq B_I. \tag{2.2.5}$$

Our main result is that these inclusions are equalities:

2.3. THEOREM. *Let U be a dense open subscheme of $\mathbb{G}_m/\mathbb{F}_q$ and $f: X \to U$ a finite étale cover. Then*

$$\begin{aligned} \mathrm{DGal}(f^*\mathrm{Kl}(I, \psi)) &= G_I \\ \mathrm{DGal}(f^*\widehat{\mathrm{Kl}}(I, \psi)) &= B_I \end{aligned} \tag{2.3.1}$$

Of course the most interesting case is when $U = X = \mathbb{G}_{m/\mathbb{F}_p}$, but in the proof of 2.3 it will be convenient to pass to finite étale covers, make extensions of the base field, etc. Note that it follows from 2.3, that $\mathrm{DGal}(\mathrm{Kl}(I, \psi))$ is unchanged under restriction to an open subscheme.

Now choose a prime $l \neq p$, and fix an isomorphism $\bar{K} \simeq \mathbb{Q}_l$. With $\mathrm{Kl}_l(n, \psi_l)$ as in 1.5, we set

$$\mathrm{Kl}(I, \psi_l) = \bigoplus_{n \in I} \mathrm{Kl}_l(n, \psi_l) \tag{2.3.2}$$

and let ρ_I be the representation of $\pi_1(U)$ corresponding to the lisse l-adic sheaf $f^*\mathrm{Kl}_l(I, \psi_l)$. We have

2.4. COROLLARY. $\mathrm{DGal}(f^*\mathrm{Kl}(I, \psi)) \otimes \bar{\mathbb{Q}}_l \simeq \overline{\mathrm{Im}\ \rho_I|\pi_1^{\mathrm{geom}}(X)}$.

Proof. Since $f: X \to U$ is finite étale $U \subset \mathbb{G}_m$, the map $\pi_1^{\mathrm{geom}}(X) \to \pi_1^{\mathrm{geom}}(\mathbb{G}_m)$ has finite cokernel, so it is enough to consider the case $X = \mathbb{G}_m$. By 2.3 it is enough to show that the Zariski closure of $\operatorname{Im} \rho | \pi_1^{\mathrm{geom}}(X)$ is isomorphic to $G_I \otimes \mathbb{Q}_l$.

If I consists of a single element, this is the result (1.5.2) of Katz. In general, the result for $|I| = 1$ implies that the closure of the image of $\pi_1(X \otimes \overline{\mathbb{F}}_q)$ is a subgroup of $G_I \otimes \mathbb{Q}_l$ which projects onto each factor. Since the Lie algebras are simple nonabelian and pairwise non-isomorphic, the closure of the image must in fact be the entire product, in virtue of the lemma of Goursat–Kolchin–Ribet [17], 1.8.2. □

For any $V \subseteq X$, let $\mathscr{K}\mathscr{L}_V$ be the full subcategory of $\operatorname{Isoc}^\dagger(V)$ generated by the $f^*\mathrm{Kl}(n, \psi) | V$ for all n not divisible by p, and let $\mathscr{K}\mathscr{L}_V$ be the corresponding category of convergent isocrystals. By the general theory of Tannakian categories, these are equivalent to the 2-inductive limits of the categories $\operatorname{Rep}_K(\mathrm{DGal}(f^*\mathrm{Kl}(I, \psi) | V))$ resp. $\operatorname{Rep}_K(\mathrm{DGal}(f^*\mathrm{Kl}(I, \psi) | V))$ (where the transition maps for $I \subset J$ are induced from the projections

$$\mathrm{DGal}(\mathrm{Kl}(I, \psi) | V) \to \mathrm{DGal}(\mathrm{Kl}(J, \psi) | V)).$$

2.5. COROLLARY. *For any dense open $V \subseteq X$, the forgetful functors*

$$\mathscr{K}\mathscr{L}_V \to \widehat{\mathscr{K}\mathscr{L}}_V$$

$$\mathscr{K}\mathscr{L}_X \to \mathscr{K}\mathscr{L}_V$$

$$\widehat{\mathscr{K}\mathscr{L}}_V \to \widehat{\mathscr{K}\mathscr{L}}_V$$

are fully faithful.

Proof. We treat the case of $\mathscr{K}\mathscr{L}_V \to \widehat{\mathscr{K}\mathscr{L}}_V$, since the other cases are similar. It suffices to show that for all I, the functor $[f^*\mathrm{Kl}(I, \psi) | V] \to [f^*\mathrm{Kl}(I, \psi) | V]$ is fully faithful. Now this functor corresponds by Tannaka duality to the group homomorphism $B_I \hookrightarrow G_I$, and since G_I/B_I is a projective variety, the result follows from the next lemma. □

2.6. LEMMA. *Let G be an affine K-group and $H \subset G$ be a subgroup. Then the forgetful functor $F: \operatorname{Rep}_K(G) \to \operatorname{Rep}_K(H)$ is fully faithful if and only if $\Gamma(G/H, \mathscr{O}) = K$.*

Proof. Faithfulness is automatic, so the only question is whether F is full. Since $\Gamma(G/H, \mathscr{O})$ is an inductive limit of finite-dimensional G-spaces trivial on H, it must be trivial if F is fully faithful, which shows that $\Gamma(G/H, \mathscr{O}) = K$. On the other hand, suppose that V, W are G-spaces and that $f \in \operatorname{Hom}_H(V, W)$. Then $g \mapsto g \circ f \circ g^{-1}$ defines a map $G/H \to \operatorname{Hom}_H(V, W)$. If $\Gamma(G/H, \mathscr{O}) = K$, the image of this map must be a point, so $f \in \operatorname{Hom}_G(V, W)$. □

We now let $X = \mathbb{G}_m/\mathbb{F}_q$, and consider the quotients $M_i^{(n)}$ of $\widehat{\mathrm{Kl}}(n, \psi)$ for the slope filtration (1.3.4), and the corresponding p-adic characters $\psi_i^{(n)}$ of $\pi_1(G_m)$. If $\widehat{\mathrm{Kl}}(I, \psi)^{ss}$ denotes the semisimplification of $\widehat{\mathrm{Kl}}(I, \psi)$ as an F-isocrystal, then we have

$$\mathrm{Kl}(I, \psi)^{ss} = \bigoplus_{\substack{n \in I \\ 0 \leqslant i \leqslant n}} M_i^{(n)}.$$

Furthermore $\mathrm{DGal}(\widehat{\mathrm{Kl}}(I, \psi)^{ss})$ is isomorphic to the quotient of B_I by its unipotent radical; i.e. to a maximal torus of G_I. Finally, since $\mathbb{G}_m/k$ is a smooth k-scheme, (M, F) is a convergent unit-root F-isocrystal $\mathbb{G}_m/K$, and $\rho: \pi_1(\mathbb{G}_m) \to \mathrm{GL}(V)$ is the corresponding representation of $\pi_1(\mathbb{G}_m)$, the group $\mathrm{DGal}(M)$ is isomorphic over some extension of K to the Zariski-closure of $\mathrm{Im}\, \rho \,|\, \pi_1^{\mathrm{geom}}(\mathbb{G}_m)$. Since the $M_i^{(n)}$ are twists of unit-root F-isocrystals, and the $\psi_i^{(n)}$ are the corresponding representations, we have

2.7. COROLLARY. *The Zariski-closure of the image of*

$$\bigoplus_{\substack{n \in I \\ 0 \leqslant i \leqslant n}} \psi_i^{(n)} \,|\, \pi_1^{\mathrm{geom}}(\mathbb{G}_m)$$

is isomorphic to a maximal torus of G_I.

In particular, we see that the only nontrivial relation between the characters $\psi_i^{(n)}$ for *all* n prime to p and all $0 \leqslant i < n$ are

$$\psi_0^{(n)} \psi_1^{(n)} \cdots \psi_{n-1}^{(n)} = 1$$

and, when n is even,

$$\psi_i^{(n)} \psi_{n-i-1}^{(n)} = 1.$$

Although 2.7 is formally a consequence of 2.3, our actual procedure will be something like the reverse: we will deduce from the equidistribution of the l-adic Frobenius classes in G_I that $\mathrm{DGal}(f^*\mathrm{Kl}(I, \psi))$ contains a maximal torus of G_I. A calculation of the local monodromy of $\mathrm{Kl}(I, \psi)$ around $0 \in \mathbb{P}^1$ will then imply that $\mathrm{DGal}(f^*\mathrm{Kl}(I, \psi)) = B_I$, and the equality $\mathrm{DGal}(f^*\mathrm{Kl}(I, \psi)) = G_I$ will be deduced from the previous equality and the global monodromy theorem for overconvergent F-isocrystals.

3. Local monodromy

3.1. In this section we will define a local analogue of the monodromy groups discussed in §2, and calculate the local monodromy of the convergent Kloosterman isocrystal at 0. In this section k, R, and K will be as in 0.2, but otherwise arbitrary.

Let $R((T))$ be the ring of Laurent series with coefficients in R, $R((T))\hat{}$ the p-adic completion of $R((T))$, and $A = R((T))\hat{} \otimes K$. In more concrete terms, A is the ring

$$A = \left\{ \sum_{n \in \mathbb{Z}} a_n T^n \,\middle|\, \begin{array}{c} a_n \in K,\ a_n \to 0 \text{ as } n \to -\infty \\ |a_n| < C \text{ for some constant } C \text{ and all } n \end{array} \right\}.$$

Since the valuation of K is discrete, A is a *field*. We denote by $f \mapsto f'$ the usual derivation $\Sigma\, a_n T^n \mapsto \Sigma\, n a_n T^{n-1}$. Furthermore we denote by $\Omega^1_{A/K}$ the A-vector space of dimension one with generator dT. A *connection* on an A-vector space V is a K-linear map $\nabla: V \to V \otimes \Omega^1_{A/K}$ satisfying Leibnitz's rule: $\nabla(fv) = f'v + f\nabla(v)$ for $f \in A$, $v \in V$. We denote by Diff_A the category of A-vector spaces endowed with a connection; it is evidently K-linear, and has a tensor product structure defined in the usual way. By means of the criterion [9, 1.20] one sees that Diff_A is a K-linear $\otimes$-category (c.f. 0.2) with unit element $(A, 0)$. The functor

$$\omega: \mathrm{Diff}_A \to \mathrm{Vec}_A \quad \text{given by} \quad (V, \nabla) \mapsto V \tag{3.1.1}$$

is a fiber functor defined over A, and so makes Diff_A into a (non-neutral) Tannakian category. To "neutralize" Diff_A, we can form the extension of scalars (c.f. 0.2) $\mathrm{Diff}_A \otimes A$, which is an A-linear neutral ind-Tannakian category. The objects of $\mathrm{Diff}_A \otimes A$ consist of ind-objects of Diff_A endowed with an action of A. The functor 3.1.1 extends to a functor

$$\omega: \mathrm{Diff}_A \otimes A \to \mathrm{Vec}_A \tag{3.1.2}$$

and we define

$$\pi_1^{\mathrm{diff}}(A) = \mathrm{Aut}^{\otimes}\omega \tag{3.1.3}$$

and, for any object (V, ∇) of $\mathrm{Diff}_A \otimes A$,

$$\mathrm{DGal}(V) = \mathrm{Aut}^{\otimes}\omega \,|\, [V] \tag{3.1.4}$$

where $[V]$ denotes the sub-$\otimes$-category of $\mathrm{Diff}_A \otimes A$ generated by (V, ∇). As always, there are equivalences of categories

$$\begin{aligned} [V] &\xrightarrow{\sim} \mathrm{Rep}_A(\mathrm{DGal}(V)) \\ \mathrm{Diff}_A \otimes A &\xrightarrow{\sim} \mathrm{Rep}_A(\pi_1^{\mathrm{diff}}(A)) \end{aligned} \tag{3.1.5}$$

and the inclusion $[V] \to \mathrm{Diff}_A \otimes A$ corresponds by Tannaka duality to a surjective homorphism $\pi_1^{\mathrm{diff}}(A) \to \mathrm{DGal}(V)$. The group $\pi_1^{\mathrm{diff}}(A)$ is a pro-algebraic affine A-group, and $\mathrm{DGal}(V)$ is an affine algebraic group over A.

The natural inclusion functor $\mathrm{Diff}_A \to \mathrm{Diff}_A \otimes A$ allows us to define $\mathrm{DGal}(V)$ for an object (V, ∇) of Diff_A. For such (V, ∇), $\mathrm{DGal}(V)$ can be given a relatively simple description. If V is an object of a rigid $\otimes$-category and n, m are nonnegative integers, we denote by $T^{n,m}(V)$ the object $V^{\otimes m} \otimes \check{V}^{\otimes n}$, where $\check{V}$ is the dual of V. A well-known theorem of Chevalley asserts that any algebraic subgroup G of $\mathrm{GL}(V)$ is characterized by the G-stable subspaces of the $T^{n,m}(V)$ for all $n, m \geqslant 0$. From this theorem and the above definition, it follows that for any (V, ∇) in Diff_A we have

$$\mathrm{DGal}(V) \simeq \left\{ g \in \mathrm{GL}(V) \,\middle|\, \begin{array}{l} \text{For all } n, m \geqslant 0,\ T^{n,m}(g) \text{ stabilizes} \\ \text{any } \nabla\text{-stable subspace of } T^{n,m},(V) \end{array} \right\}.$$

3.2. We will have to calculate $\mathrm{DGal}(V)$ in the case that $V = V_0 \otimes A$ for some finite-dimensional K-vector space V_0, and the connection has the form $\nabla = N \otimes dT/T$, where N is a nilpotent endomorphism of V_0. For this purpose, we introduce the auxiliary category

$$\mathrm{Vec}_K^{\mathrm{nil}} = \begin{cases} \text{Objects: finite-dimensional } K\text{-vector spaces } V_0 \text{ with a} \\ \text{nilpotent endomorphism } N: V_0 \to V_0 \\ \text{Morphisms: linear maps compatible with the endomorphism.} \end{cases}$$

The category $\mathrm{Vec}_K^{\mathrm{nil}}$ becomes a K-linear $\otimes$-category under the tensor product rule

$$(V_0, N) \otimes (V_0', N') = (V_0 \otimes V_0', N \otimes 1 + 1 \otimes N') \tag{3.2.1}$$

with identity object $(K, 0)$. There is an equivalence of categories

$$\mathrm{Rep}_K(\mathbb{G}_a) \xrightarrow{\sim} \mathrm{Vec}_K^{\mathrm{nil}} \tag{3.2.2}$$

which associates to $\rho: \mathbb{G}_a \to \mathrm{GL}(V_0)$ the pair (V_0, N) where N is the image of $1 \in K = \mathrm{Lie}(\mathbb{G}_a)$ under the derived representation $\mathrm{Lie}(\rho): \mathrm{Lie}\ \mathbb{G}_a \to \mathrm{End}(V_0)$. The K-linear functor

$$F: \mathrm{Vec}_K^{\mathrm{nil}} \to \mathrm{Diff}_A(V_0, N) \mapsto (V_0 \otimes A, N \otimes \mathrm{d}T/T) \tag{3.2.3}$$

is a $\otimes$-functor on account of (3.2.1), which gives rise to a $\otimes$-functor $\mathrm{Vec}_K^{\mathrm{nil}} \otimes A \to \mathrm{Diff}_A \otimes A$. From this we obtain a group homomorphism

$$\pi_1^{\mathrm{diff}}(A) \to \mathbb{G}_a \otimes A. \tag{3.2.4}$$

When $V = V_0 \otimes K$, we will denote by $d: V \to V \otimes \Omega^1_{A/K}$ the constant connection $d(v_0 \otimes a) = v_0 \otimes a' \mathrm{d}T$ on $V = V_0 \otimes A$; thus $\mathrm{d}v = 0$ if and only if $v \in V_0 \subset V_0 \otimes A$, i.e. if and only if v is "constant". When $(V, \nabla) = (V_0 \otimes A, N \otimes dT/T)$ is in the image of (3.2.3), we have

$$\nabla(v) = \mathrm{d}v + Nv \otimes \frac{\mathrm{d}T}{T}. \tag{3.2.5}$$

Then since N is "constant", i.e. is an endomorphism of $V_0 \subset V$, it commutes with d and ∇.

3.3. PROPOSITION. *The morphism 3.2.4 is faithfully flat.*

Proof. The homomorphism 3.2.4 is the "value" on the k-algebra A of a morphism of *liens* induced by the functor F in (3.2.3) [20, III (2.1.3) and (2.3.1)]. It then follows from [20, III (3.3.3)] that it is enough to show that F is fully faithful, and that for any object (V_0, N) of $\mathrm{Vec}_K^{\mathrm{nil}}$ and any subquotient (W, ∇) of $F(V_0, N)$ in Diff_A, there is a (V_0', N') in $\mathrm{Vec}_K^{\mathrm{nil}}$ such that $(W, \nabla) \simeq F(V_0', N')$. Once full faithfulness has been proven, we can replace "subquotient" by "subobject" in the latter item. We consider these points in order:

Full faithfulness. By the usual trick of considering internal Hom's, it is enough to show that $\mathrm{Hom}(1, (V_0, N)) = \mathrm{Hom}(1, F(V_0, N))$, where 1 denotes the unit object of $\mathrm{Vec}_K^{\mathrm{nil}}$ on the left, and Diff_A on the right. In $\mathrm{Vec}_K^{\mathrm{nil}}$ we have $1 = (A, 0)$, and $\mathrm{Hom}(1, (V_0, N)) = \mathrm{Ker}\ N \mid V_0$, while in Diff_A we have $\mathrm{Hom}(1, (V, \nabla)) = \mathrm{Ker}\ \nabla$. We must therefore show that

$$\mathrm{Ker}\ N \mid V_0 = \mathrm{Ker}\ \nabla \tag{3.3.1}$$

when $\nabla = N \otimes dT/T$ on $V = V_0 \otimes A$ and N is a nilpotent endomorphism of V_0. Clearly $\mathrm{Ker}\ N \mid V_0 \subseteq \mathrm{Ker}\ \nabla$; to show the reverse inequality, we pick

$v \in \operatorname{Ker} \nabla$ and let k be the smallest integer such that $N^k v = 0$. By (3.2.5) we have

$$\nabla(v) = \mathrm{d}v + Nv \otimes \frac{\mathrm{d}T}{T} = 0 \tag{3.3.2}$$

and since N commutes with d and ∇, we get

$$d(N^{k-1}v) = 0$$

on applying N^{k-1} to (3.3.2). Therefore $N^{k-1}v \in V_0$, and $N^{k-1}v \neq 0$ by the assumption on k. If $k > 1$, then applying N^{k-2} to 3.3.2 yields

$$d(N^{k-2}v) + N^{k-1}V \otimes \frac{\mathrm{d}T}{T} = 0. \tag{3.3.3}$$

Since $0 \neq N^{k-1}v \in V_0$, we see from (3.3.3) that T^{-1} is the derivative of an element of A (e.g. by choosing an isomorphism $V_0 \simeq K^n$, and comparing coefficients). But in fact T^{-1} is not the derivative of an element of A, so we must have $k = 1$, whence $v \in \operatorname{Ker} N \mid V_0$.

Subobjects. Let (V_0, N) be an object of $\mathrm{Vec}_K^{\mathrm{nil}}$, and (W, ∇) a subobject in Diff_A of $F(V_0, N)$. We must show that there is an N-stable K-subspace $W_0 \subseteq V_0$ such that $(W, \nabla) = F(W_0, N \mid W_0)$. By considering exterior powers, we can reduce to the case where W has rank one. If $W = Av$ for some $v \in V$, then we have

$$\nabla(v) = \mathrm{d}v + Nv \otimes \frac{\mathrm{d}T}{T} \doteq fv \otimes \frac{\mathrm{d}T}{T}$$

for some $f \in A$. As before, we choose the smallest integer k such that $N^k v = 0$; then applying N^{k-1} the above equation yields

$$d(N^{k-1}v) = fN^{k-1}v \otimes \frac{\mathrm{d}T}{T}.$$

Since $N^{k-1}v \neq 0$, writing down an isomorphism $V \simeq A^n$ shows that $T^{-1}f = g^{-1}g'$ for some $g \in A$. Then $w = g^{-1}v$ satisfies

$$\mathrm{d}w + Nw \otimes \frac{\mathrm{d}T}{T} = g^{-1}\,\mathrm{d}v - g^{-2}v\,\mathrm{d}g + g^{-1}Nv \otimes \frac{\mathrm{d}T}{T}$$

$$= g^{-1}\left(\mathrm{d}v - fv \otimes \frac{\mathrm{d}T}{T} + Nv \otimes \frac{\mathrm{d}T}{T}\right) = 0$$

i.e. $w \in \operatorname{Ker} \nabla$. By (3.3.1), we have $w \in \operatorname{Ker} N \mid V_0$, and therefore $W = W_0 \otimes A$, where $W_0 = Kw \subseteq V_0$. Since $W_0 \subseteq \operatorname{Ker} N$, we have that $(W_0, N \mid W_0)$ is a subobject of (V_0, N) such that $(W, \nabla) = F(W_0, N \mid W_0)$. □

3.4. COROLLARY. *Suppose that V_0 is a K-vector space, and ∇ is a connection on $V = V_0 \otimes A$ of the form $N \otimes \mathrm{d}T/T$, where N is a nonzero nilpotent endomorphism of V_0. Then*

$$\mathrm{DGal}(V) \simeq \mathbb{G}_a \otimes A \qquad (3.4.1)$$

and the image of the derived representation Lie $\mathbb{G}_a \otimes A \to \operatorname{End}(V)$ *is spanned by N.*

Proof. This follows from 3.3 and the construction of the equivalences (3.2.2). □

3.5. Let $j: X \hookrightarrow \bar{X}$ be an embedding of smooth curves over k, and suppose that there exists a lifting $\mathfrak{X} \hookrightarrow \bar{\mathfrak{X}}$ of j with $\mathfrak{X}, \bar{\mathfrak{X}}$ formally smooth over R. For any closed point X of x, we choose a local parameter t of X around x, and a lifting T of t to a local section of $\mathcal{O}_{\bar{\mathfrak{X}}}$; by shrinking $\bar{X}$, we can assume that T is defined on all of $\bar{\mathfrak{X}}$. Then the local ring $O_{\bar{\mathfrak{X}},x}$ is isomorphic to the power series ring $R[[T]]$, and the p-adic completion $R((T))^{\wedge}$ of the Laurent series ring $R((T))$ is isomorphic to the fiber at x of $j_*\mathcal{O}_{\bar{\mathfrak{X}}}$. Thus there is an injection

$$\Gamma(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}) = \Gamma(\bar{\mathfrak{X}}, j_*\mathcal{O}_{\mathfrak{X}}) \hookrightarrow R((T))^{\wedge} \qquad (3.5.1)$$

and as $A = R((T))^{\wedge} \otimes K$ and $\Gamma(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}) \otimes K = \Gamma(\mathfrak{X}^{an}, \mathcal{O}_{\mathfrak{X}^{an}})$, (3.5.1) yields

$$B \stackrel{\text{def}}{=} \Gamma(\mathfrak{X}^{an}, \mathcal{O}_{\mathfrak{X}^{an}}) \hookrightarrow A. \qquad (3.5.2)$$

We will use (3.5.2) to construct the local monodromy group of a convergent isocrystal on X/K about the point x. Using the identification of convergent isocrystals on X/K with locally free $\mathcal{O}_{\mathfrak{X}^{an}}$-modules with a convergent connection, we obtain a faithful $\otimes$-functor

$$\begin{aligned} I_x &: \operatorname{Isoc}(X/K) \to \operatorname{Diff}_A \\ M &= (\mathcal{M}, \nabla) \to (\mathcal{M} \otimes_B A, \nabla) \end{aligned} \qquad (3.5.3)$$

which is exact since the $\mathcal{M}$ are locally free. Then for any object M of $\mathrm{Isoc}(X/K)$, we define the local monodromy group $I(M, x)$ to be the affine algebraic group over A

$$I(M, x) = \mathrm{DGal}(I_x(M)). \tag{3.5.4}$$

The functor (3.5.3) induces, by Tannaka duality, a closed immersion

$$I(M, x) \hookrightarrow \mathrm{DGal}(M, x). \tag{3.5.5}$$

3.6. We now take $X = \mathbb{G}_m/k$, $x = 0 \in \mathbb{P}^1$, and let M be the convergent Kloosterman isocrystal $\widehat{\mathrm{Kl}}(I, \psi)$. Since the hypergeometric series (1.2.2) are bounded in the disk $|x| < 1$, the connection (1.1.5) is isomorphic over A to the connection (1.2.1). Therefore $\widehat{\mathrm{Kl}}(I, \psi) \otimes A$ is in the essential image of (3.2.3); more precisely we have $\widehat{\mathrm{Kl}}(I, \psi) \otimes A \simeq (A^n, N \otimes \mathrm{d}T/T) = F(V_0, N)$, where $V_0 = \bigoplus_{n \in I} K^n$ and N is the direct sum of the matrices N_n in (1.2.1). In particular, $N \neq 0$, and by 3.4 we have

$$I(\widehat{\mathrm{Kl}}(I, \psi), 0) \simeq \mathbb{G}_a \otimes A. \tag{3.6.1}$$

The slope filtration (1.3.4) yields a filtration of $I_0(\mathrm{Kl}(n, \psi)) = A^n$ for each n. We claim that this is the same as the filtration by the subspaces $\mathrm{Ker}\, N_n^k \subset A^n$; this follows from the fact that both N_n and $\Phi(0)$ are upper triangular (c.f. the paragraph after (1.3.3), and the matrix $P(x)$ giving the equivalence of (1.1.3) and (1.2.1) satisfies $P(0) = I$. Suppose now that $x \in \mathbb{G}_m(k)$, and ω is the corresponding fiber functor. A choice of isomorphism $I_0 \simeq \omega \otimes A$ yields an isomorphism of representations

$$\begin{array}{ccc} \mathrm{DGal}(\widehat{\mathrm{Kl}}(n, \psi), x) \otimes A & \rightarrow & \mathrm{GL}(\widehat{\mathrm{Kl}}(n, \psi)_x) \otimes A \\ \downarrow & & \downarrow \\ \mathrm{DGal}(\widehat{\mathrm{Kl}}(n, \psi), 0) & \rightarrow & \mathrm{GL}(I_0(\widehat{\mathrm{Kl}}(n, \psi))) \end{array} \tag{3.6.2}$$

and by what we have just seen, the Borel subgroup of $\mathrm{GL}(\widehat{\mathrm{Kl}}(n, \psi)_x) \otimes A$ determined by the slope filtration of $\widehat{\mathrm{Kl}}(n, \psi)$ maps under the right vertical arrow of (3.6.2) into the Borel subgroup of $\mathrm{GL}(I_0(\widehat{\mathrm{Kl}}(n, \psi))) \simeq \mathrm{GL}(n)_{/A}$ consisting of upper triangular matrices. Furthermore, if n is even, then the image of $\mathrm{Sp}(\widehat{\mathrm{KL}}(n, \psi)_x, \psi_n)$ in $\mathrm{GL}(I_0(\widehat{\mathrm{Kl}}(n, \psi)))$ is the symplectic group $\mathrm{Sp}(n, \Theta)$, where Θ is the symplectic form on A^n given by the matrix (1.4.4). Thus if we put

$$H_n = \begin{cases} \mathrm{SL}(n)_{/A} & n \text{ odd} \\ \mathrm{SP}(n,\ \Theta)_{/A} & n \text{ even} \end{cases}, \quad H_I = \prod_{n \in I} H_I$$

and let P_1 be the Borel subgroup of H_I consisting of upper triangular matrices, then the direct sum of (3.6.2) for all $n \in I$ fits into a commutative diagram

$$\begin{array}{ccccc} \mathrm{DGal}(\widehat{\mathrm{Kl}}(I,\ \psi),\ x) \otimes A & \to & B_I \otimes A & \to & G_I \otimes A \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{DGal}(\widehat{\mathrm{Kl}}(I,\ \psi),\ 0) & \to & P_I & \to & H_I. \end{array} \tag{3.6.3}$$

We must now recall some notation and a few facts about semisimple Lie groups and their Lie algebras. Let G be a semisimple group over K; let T be a maximal torus of G and $\mathscr{R}$ the associated root system. For $\alpha \in \mathscr{R}$ we denote by $\mathfrak{g}_\alpha \subset \mathrm{Lie}\ G$ the corresponding root space. Then to any Borel subgroup $B \subset G$ containing T is associated a basis $\mathscr{B}$ of $\mathscr{R}$, such that if U denotes the unipotent radical of $\mathscr{B}$, then

$$\mathrm{Lie}\ U = \bigoplus_{\alpha > 0} \mathfrak{g}_\alpha \tag{3.6.4}$$

where the sum is over the roots positive with respect to this basis.

A nilpotent element $N \in \mathrm{Lie}\ G$ is called a *principal nilpotent element* if the dimension of $\mathrm{Ker}(\mathrm{ad}\ N)$ is the rank of G [LIE, VIII, §11, Def. 3]. If B, T are as above and $N \in \mathrm{Lie}\ B$, then $N \in \mathrm{Lie}\ U$, and N is principal nilpotent if and only if the component of N in $\mathfrak{g}_\alpha$ for the decomposition (3.6.4) is nonzero for every $\alpha \in \mathscr{B}$ [LIE, VIII, §11, Prop. 10]. It is this last property which is of main interest now.

We now return to the situation of 2.2: $U \subseteq \mathbb{G}_m$ is open and dense, $f: X \to U$ is finite étale, and

$$\mathrm{DGal}(f^*\widehat{\mathrm{Kl}}(I,\ \psi)) \subseteq \mathrm{DGal}(f^*\mathrm{Kl}(I,\ \psi)) \subseteq G_I.$$

3.7. PROPOSITION. Lie $\mathrm{DGal}(f^*\widehat{\mathrm{Kl}}(I, \psi)) \otimes A$ *contains a principal nilpotent element of* Lie $G_I \otimes A$.

Proof. By (2.1.3), $f^*: \mathrm{Isoc}(U/K) \to \mathrm{Isoc}(X/K)$ induces an isomorphism

$$\mathrm{DGal}(f^*\widehat{\mathrm{Kl}}(I,\ \psi))^0 \xrightarrow{\sim} \mathrm{DGal}(\widehat{\mathrm{Kl}}(I,\ \psi))^0$$

and so it is enough to consider the case $X = U$, $x \in U(k) \subseteq \mathbb{G}_m(k)$. Choose an isomorphism $I_0 \simeq \omega \otimes A$, where ω is the fiber functor associated to x.

Then we have the commutative diagram 3.6.3, and it is enough to show that Lie GDal($\widehat{\mathrm{Kl}}(I, \psi)$, 0) contains a principal nilpotent element of Lie H_I. By 3.4 and (3.5.5), the image of Lie DGal($\widehat{\mathrm{Kl}}(I, \psi)$, 0) in Lie H_I contains the element $N = \oplus_{n \in I} N_n$, so it suffices to see that N is principal nilpotent in Lie H_I. For this, it is sufficient to see that each N_n is principal nilpotent in H_n. If T denotes the subgroup of H_n consisting of diagonal matrices, then $T \subset P_n$, where P_n is the upper triangular Borel subgroup of H_n; let $\mathscr{B} \subset \mathscr{R}$ be the corresponding basis of the root system of H_I. It is enough to check that the component of N_n in $\mathfrak{g}_\alpha$ is nonzero for each $\alpha \in \mathscr{B}$. Since T (resp. P_n) is the group of diagonal (resp. upper triangular) matrices in H_n, this follows immediately from the expression (1.2.1) for N_n. □

4. Weil groups

4.1. Most of this section is devoted to showing that DGal($\widehat{\mathrm{Kl}}(I, \psi)$)) contains a maximal torus of B_I. Once this is accomplished, Theorem 2.3 and the remaining assertions of §2 follow from the local monodromy calculation in §3. We will need to use the "Weil groups" and Frobenius elements associated to an F-isocrystal that were constructed in [7]; we begin by briefly recalling some of their properties.

Suppose that k is finite, X/k is a smooth curve, and $\mathscr{C}$ a full $\otimes$-subcategory of Isoc(X/K) or Isoc†(X/K). If $|k| = q$, then let F denote the qth power Frobenius, and we make the hypothesis that the pullback F^* induces a (K-linear) autoequivalence $F^*: \mathscr{C} \to \mathscr{C}$. Then F^* induces an automorphism $\Phi^{\mathscr{C}}_{x_0}: \mathrm{DGal}(\mathscr{C}, x_0) \xrightarrow{\sim} \mathrm{DGal}(\mathscr{C}, x_0)$, and we define the *Weil group* $W^{\mathscr{C}}_{x_0}$ associated to $\mathscr{C}$ to be the semidirect product $\mathrm{DGal}(\mathscr{C}, x_0) < \mathbb{Z}$, where the generator $1 \in \mathbb{Z}$ acts on $\mathrm{DGal}(\mathscr{C}, x_0)$ by $\Phi^{\mathscr{C}}_{x_0}$. We thus obtain an extension of K-groups

$$0 \longrightarrow \mathrm{DGal}(\mathscr{C}, x_0) \longrightarrow W^{\mathscr{C}}_{x_0} \xrightarrow{\deg} \mathbb{Z} \longrightarrow 0. \tag{4.1.1}$$

For example, we can take for $\mathscr{C}$ the entire category Isoc(X/K) of convergent isocrystals on X (c.f. [7]), and the group so obtained is an analogue of the usual Weil group of X. I do not know if one can do the same with Isoc†(X/K); however for any convergent or overconvergent F-isocrystal (M, F), the $\otimes$-category $\mathscr{C} = [M]$ generated by M satisfies the hypothesis that F^* is an equivalence, c.f. [7], and the corresponding Weil group, which we will denote by $W^M_{x_0}$, is an analogue of Deligne's construction [8], 1.3.7.1, which to any l-adic Weil sheaf $\mathscr{F}$ on X associates the pushout of the Weil group of X/k by the homomorphism $\pi_1^{\mathrm{geom}}(X) \to G$, G being the Zariski-closure of the image of $\pi_1^{\mathrm{geom}}(X)$ in the l-adic representation corresponding to $\mathscr{F}$. In general, if $\mathscr{C}$ satisfies our hypotheses and (M, F) is an F-isocrystal such that M is an object of $\mathscr{C}$, then we have (c.f. [7], 2.2.4 and 2.4) $F_{x_0}\rho(g)F_{x_0}^{-1} = \Phi_{x_0}^{M}(\rho(g))$, so that the canonical representation of $\mathrm{DGal}(\mathscr{C}, x_0)$ on M_{x_0} extends to a

$$\rho : W_{x_0}^{M} \to \mathrm{GL}(M_{x_0}) \tag{4.1.2}$$

such that the generator $1 \in \mathbb{Z} \subset W^{\mathscr{C}}$ is mapped to F_{x_0} under ρ.

Like $\mathrm{DGal}(\mathscr{C}, x_0)$, the group $W_{x_0}^{\mathscr{C}}$ does not depend on the choice of base point, up to an isomorphism that is canonical up to inner automorphism. More precisely, if x_0 and y_0 are k-points of X, and $p : \omega_{x_0} \to \omega_{y_0}$ in any isomorphism of fiber functors $\mathscr{C} \to \mathrm{Vec}_K$, then the isomorphism $\mathrm{DGal}(\mathscr{C}, x_0) \xrightarrow{\sim} \mathrm{DGal}(\mathscr{C}, y_0)$ canonically induced by p extends to a canonical *isomorphism* $P_{x_0y_0} : W_{x_0}^{\mathscr{C}} \xrightarrow{\sim} W_{y_0}^{\mathscr{C}}$, and a second isomorphism $p' : \omega_{x_0} \to \omega_{y_0}$ induces a $P'_{x_0y_0} : W_{x_0}^{\mathscr{C}} \to W_{y_0}^{\mathscr{C}}$ differing from $P_{x_0y_0}$ by an inner automorphism. Finally, one sees from the construction of $P_{x_0y_0}$ that this isomorphism is natural in X, $\mathscr{C}$, and x_0, in the following sense. Let $f : Y \to X$ be a morphism of k-schemes of finite type, y_0, $y_1 \in Y(k)$, and $x_0 = f(y_0)$, $x_1 = f(y_1)$. Suppose $\mathscr{C}_Y$, $\mathscr{C}_X$ are $\otimes$-categories of convergent or overconvergent isocrystals on Y resp. X on which the Frobenius pullback F^* induces autoequivalences. and we will assume either that

(a) $\mathscr{C}_X$, $\mathscr{C}_Y$ are both categories of convergent isocrystals on X/k resp. Y/K, and the pullback by $f : Y \to X$ induces a functor $f^* : \mathscr{C}_X \to \mathscr{C}_Y$; or

(b) the same as the above, but where $\mathscr{C}_X$, $\mathscr{C}_Y$ are now categories of overconvergent isocrystals; or

(c) $\mathscr{C}_X$ is a category of overconvergent isocrystals, $\mathscr{C}_Y$ is a category of convergent isocrystals, and f^* is the composition of the pullback by f and the forgetful functor (1.1.3).

In all these cases, f^* is an exact $\otimes$-functor. Let ω_{x_0}, ω_{y_0}, ω_{x_1}, ω_{y_1} denote the standard fiber functors; then if

$$\begin{array}{ccc} \omega_{y_0} \circ f^* & \to & \omega_{y_1} \circ f^* \\ \downarrow & & \downarrow \\ \omega_{x_0} & \to & \omega_{x_1} \end{array} \tag{4.1.3}$$

is a commutative diagram of isomorphisms, the canonical isomorphisms $P_{x_0x_1}$ and $P_{y_0y_1}$ sit in a commutative diagram

$$\begin{array}{ccc} W_{y_0}^{\mathscr{C}_Y} & \xrightarrow{P_{y_0y_1}} & W_{y_1}^{\mathscr{C}_Y} \\ \downarrow & & \downarrow \\ W_{x_0}^{\mathscr{C}_X} & \xrightarrow{P_{x_0x_1}} & W_{x_1}^{\mathscr{C}_X} \end{array} \tag{4.1.4}$$

where the vertical maps are the natural homomorphisms.

We have already noted that when $\mathscr{C} = [M]$ for some F-isocrystal (M, F), then the generator $a \in \mathbb{Z} \subset W^M$ maps under the representation (4.1.2) to the fiber F_{x_0} of the Frobenius structure at x_0, which justifies our regarding the conjugacy class of this generator as the *Frobenius class* Frob_{x_0}. More generally, to any closed point $x \in X$ of degree n we can associate a conjugacy class Frob_x in $W^{\mathscr{C}}_{x_0}(K_n)$, where K_n is the unramified extension of K of degree n. In fact if we choose a fiber functor ω_x corresponding to K_n-valued point of X with image x, and an isomorphism $\omega_{x_0} \otimes K_n \simeq \omega_x$, then we can define Frob_x to be the conjugacy class in $W^{\mathscr{C}}_{x_0}(K_n)$ corresponding to the generator $1 \in \mathbb{Z} \subset W^{\mathscr{C}}_x$ under the isomorphism $W^{\mathscr{C}}_{x_0} \otimes_K K_n \simeq W^{\mathscr{C}}_x$ induced by $\omega_{x_0} \otimes K_n \simeq \omega_x$. The class so defined is in fact independent of choice of ω_x (i.e. of the K_n-valued point of X with image x).

Suppose now that $f: Y \to X$ is a morphism of geometrically connected k-schemes of finite type, $y_0 \in Y(k)$, $x_0 = f(y_0) \in X(k)$ and $f^*: \mathscr{C}_X = \mathscr{C}_Y$ is as in one of the cases (a), (b), (c) described above. Then from the construction of the Frobenius classes, and the above remarks, we have

4.2. LEMMA. *The natural homomorphism*

$$G: \mathrm{DGal}(\mathscr{C}_X, x_0) \to \mathrm{DGal}(\mathscr{C}_Y, y_0)$$

extends to a morphism of exact sequences

$$\begin{array}{ccccccccc} 0 & \to & \mathrm{DGal}(\mathscr{C}_X, x_0) & \to & W^{\mathscr{C}_X}_{x_0} & \to & \mathbb{Z} & \to & 0 \\ & & \downarrow & & \downarrow & & \| & & \\ 0 & \to & \mathrm{DGal}(\mathscr{C}_Y, y_0) & \to & W^{\mathscr{C}_Y}_{y_0} & \to & \mathbb{Z} & \to & 0. \end{array} \qquad (4.2.1)$$

If y is any closed point of Y and $x = f(y)$, then $W^{\mathscr{C}_X} \to W^{\mathscr{C}_Y}$ maps $\mathrm{Frob}^{\mathscr{C}_X}_y$ *to* $(\mathrm{Frob}^{\mathscr{C}_Y}_x)^d$, *where d is the degree of $k(y)/k(x)$.*

We now return to the situation of 2.2: $k = \mathbb{F}_q$, U is an open subscheme of $\mathbb{G}_m/k$, and X a finite étale cover of U; we will assume that X has a k-rational point x_0 which will be the implicit base point in most of what follows. Denote by M the pullback of $\mathrm{Kl}(I, \psi)$ to X. We now make the supposition that

4.3 $\mathrm{DGal}(\hat{M})$ *is connected.*

which will be in force through 4.10.

From now on G_I, B_I, and V_n will have the meaning assigned to them in §2.1, and set $V_I = \Sigma_{n \in I} V_n$. Our first task is to construct subgroups of $\mathrm{GL}(V_I)$ which will turn out to be the groups W^M, $W^{\hat{M}}$ once we have proven

Theorem 2.3. Let W_I be the subgroup of $\Pi_{n\in I}\,\mathrm{GL}(V_n)$ consisting of elements g for which there is an integer $\deg(g)$ such that

$$\begin{aligned} \det(g \mid V_n) &= q^{\deg(g)n(n-1)/2} \quad n \text{ odd} \\ \psi_n(gv, gw) &= q^{(n-1)\deg(g)}\psi_n(v, w) \quad n \text{ even} \end{aligned} \tag{4.3.1}$$

for all $n\in I$. The integer $\deg(g)$ is of course unique, and the association $g\mapsto \deg(g)$ defines a homomorphism $W_I \to \mathbb{Z}$. In fact, if we denote by z the element of $\Pi_{n\in I}\,\mathrm{GL}(V_n)$ such that

$$z \mid V_n = q^{(n-1)/2}. \tag{4.3.2}$$

then z lies in the center of W_I, and one easily checks using the equality $\deg(z^n) = n$ that $W_I = G_I \times \langle z\rangle$. We now claim that there is a commutative diagram

$$\begin{array}{ccccccccc} 0 \to & \mathrm{DGal}(M) & \to & W^M & \to & \mathbb{Z} & \to 0 \\ & \downarrow & & \downarrow & & \| & \\ 0 \to & G_I & \to & W_I & \to & \mathbb{Z} & \to 0 \end{array} \tag{4.3.3}$$

in which the first vertical arrow is the inclusion in (2.2.5). Since W^M is a semidirect product of $\mathrm{DGal}(M)$ with the group generated by Frob_{x_0}, it is enough to check that this latter element satisfies the conditions of (4.3.1), and this follows from (1.4.3) and (1.4.5) by taking fibers at x_0.

We denote by $\hat{W}_I$ the subgroup of W_I stabilizing the flags (2.2.1) for each $n\in I$; since z stabilizes these flags, we have $\hat{W}_I = B_I \times \langle z\rangle$. Since the filtration (1.3.4) is a filtration by F-isocrystals, the element $\mathrm{Frob}^{\hat{M}}_{x_0} \in W^{\hat{M}}$, which is the fiber at x_0 of the Frobenius structure, stabilizes the flag (2.1.5), and thus lies in $\hat{W}_I$. Therefore $W^{\hat{M}}$ which is generated by $\mathrm{DGal}(\hat{M})$ and $\mathrm{Frob}^{\hat{M}}_{x_0}$, is contained in $\hat{M}_I$, and we get a commutative diagram

$$\begin{array}{ccccccccc} 0 \to & \mathrm{DGal}(\hat{M}) & \to & W^{\hat{M}} & \to & \mathbb{Z} & \to 0 \\ & \downarrow & & \downarrow & & \| & \\ 0 \to & B_I & \to & \hat{W}_I & \to & \mathbb{Z} & \to 0 \end{array} \tag{4.3.4}$$

in which the left vertical arrow is the one in (2.2.5).

Since the quotients of the slope filtration of $\hat{M}$ have rank one, the semisimplification $\hat{M}^{ss}$ of $\hat{M}$ as an isocrystal is the underlying isocrystal of the semisimplification of $(\hat{M}, \Phi)$ as an F-isocrystal. Set $S^0 = \mathrm{DGal}(\hat{M}^{ss})$, $S = W^{\hat{M}^{ss}}$; then 4.2 yields a commutative diagram

$$\begin{array}{ccccccccc} 0 & \to & \mathrm{DGal}(\hat{M}) & \to & W^{\hat{M}} & \to & \mathbb{Z} & \to & 0 \\ & & \downarrow & & \downarrow & & \| & & \\ 0 & \to & S^0 & \to & S & \to & \mathbb{Z} & \to & 0. \end{array} \tag{4.3.5}$$

The kernel of $\mathrm{DGal}(\hat{M}) \to S^0$ is contained in the unipotent radical of B_I; then since $\mathrm{DGal}(\hat{M})$ is connected, we see that S^0 is a torus and is the quotient of $\mathrm{DGal}(\hat{M})$ by its unipotent radical. In particular, S^0 is isomorphic to a maximal torus of $\mathrm{DGal}(\hat{M})$. To lighten the notation, we denote Frobenius elements in S by Frob_x^S.

4.4. LEMMA. *S is commutative. There is a splitting $S \to W^{\hat{M}}$ of the middle vertical arrow in (4.3.5) compatible with the degree map. For any closed point x of X the image of the Frobenius element Frob_x^S under $S \to W^{\hat{M}} \to \hat{W}_I \to W_I$ is the same as the image of Frob_x^M under $W^M \to W_I$.*

Proof. Any element in the Frobenius class $\mathrm{Frob}_{x_0}^{\hat{M}}$ acts by conjugation on $\mathrm{DGal}(\hat{M})$, and therefore on its quotient S^0, since the unipotent radical is a characteristic subgroup. The first assertion will follow if we show that this action on S^0 is trivial. However $\mathrm{Frob}_{x_0}^{\hat{M}}$ belongs to $\hat{W}_I$, and so respects the flags (2.1.1); thus it acts trivially on the quotient T of B_I by its unipotent radical. It therefore acts trivially on S^0, which is subgroup of T. We conclude that S is commutative, and is thus a direct product $S = S^0 \times \langle \mathrm{Frob}_{x_0}^S \rangle$. If we choose a Levi subgroup S' for S^0 in $\mathrm{DGal}(\hat{M})$, then $S' \times \langle \mathrm{Frob}_{x_0}^{\hat{M}} \rangle$ is a subgroup of $W^{\hat{W}}$ mapping isomorphically to $S = S^0 \times \langle \mathrm{Frob}_{x_0}^S \rangle$ under the map in (4.3.3), and the splitting so obtained is evidently compatible with the degree map. As to the assertion about Frobenius elements, we know by 4.2 that $\mathrm{Frob}_x^{\hat{M}} \mapsto \mathrm{Frob}_x^M$ under $W^{\hat{M}} \to W^M$, so it is enough to show that Frob_x^S and $\mathrm{Frob}_x^{\hat{M}}$ yield the same conjugacy class in $\hat{W}_I$. Now if T is a maximal torus of B_I, then every semisimple element of B_I is conjugate to a unique element of T, and as $\hat{W}_I = B_I \times \langle z \rangle$, every semisimple element of $\hat{W}_I$ is conjugate to a unique element of $T \times \langle z \rangle$. Thus the semisimple elements of $\hat{W}_I$ are determined up to conjugacy by their eigenvalues in V_I. Since Frob_x^M is semisimple, we conclude that Frob_x^S and $\mathrm{Frob}_x^{\hat{M}}$ map to the same conjugacy class in $\hat{W}_I$. □

From now on we will identify S with a subgroup of $W^{\hat{M}}$. Then S^0 is a maximal torus of $\mathrm{DGal}(\hat{M})$, and a subtorus of B_I. Furthermore we choose an isomorphism $\bar{K} \simeq \mathbb{C}$, and will henceforth work over the field of complex numbers.

Recall that z is the element of W_I defined in (4.3.1). We have

4.5. LEMMA. *There is a maximal torus $T \subset B_I$, such that for any $x \in S(\mathbb{C})$*

of degree n we have $xz^{-n} \in T(\mathbb{C})$. In particular, we have $S^0 \subseteq T$ and $S(\mathbb{C}) \subset T(\mathbb{C}) \times \langle z \rangle$.

Proof. Since z is central in $\hat{W}_I$, $S(\mathbb{C})\langle z \rangle$ is a (commutative) subgroup of $\hat{W}_I$, all elements of which act semisimply on V_I. Thus the degree zero subgroup $(S(\mathbb{C})\langle z \rangle)^0 \subset S(\mathbb{C})\langle z \rangle$ is a subgroup of $B_I(\mathbb{C})$ consisting of semisimple elements, and since B_I is connected and solvable, $S((\mathbb{C})\langle z \rangle)^0$ is contained in a maximal torus $T \subset B_I$ [4, 10.6]. Then if $f \in S(\mathbb{C})$ has degree n, we have $fz^{-n} \in T(\mathbb{C})$. □

We now fix a maximal torus $T \subset B_I$ satisfying the conditions of the previous lemma. The Lie group $T(\mathbb{C})$ has a unique maximal compact subgroup $T_{\mathbb{R}}$, which is a (real) torus of dimension equal to the dimension of T as an algebraic group. For $g \in T(\mathbb{C})$, we have $g \in T_{\mathbb{R}}$ if and only if the powers g^n for all $n \in \mathbb{Z}$ are bounded with respect to some Hermitian metric on V_I. In the same fashion $S^0(\mathbb{C})$ has a unique maximal torus $S^0_{\mathbb{R}}$, and we have $S^0_{\mathbb{R}} = S^0(\mathbb{C}) \cap T_{\mathbb{R}}$ by maximality.

Fix a Frobenius element $f \in S(\mathbb{C})$ of degree one, and define $S_{\mathbb{R}}$ to be the subgroup $S^0_{\mathbb{R}} \times \langle f \rangle$ of $S(\mathbb{C})$.

4:6. LEMMA. *$S_{\mathbb{R}}$ contains every Frobenius element of $S(\mathbb{C})$, and $S_{\mathbb{R}} \subseteq T_{\mathbb{R}} \times \langle z \rangle$.*

Proof. Let $F \in S(\mathbb{C})$ have degree k, so that $Ff^{-k} \in S^0(\mathbb{C})$. The eigenvalues of F on V_n are all of weight $n-1$, and since S is commutative, the eigenvalues of Ff^{-k} on V_I all have weight zero. By the criterion described above, we have $Ff^{-n} \in S^0_{\mathbb{R}}$, so that $F \in S_{\mathbb{R}}$. As to the second assertion, we have $S^0_{\mathbb{R}} \subseteq T_{\mathbb{R}}$, so it is enough to show that $f \in T_{\mathbb{R}}\langle z \rangle$. In fact, since f has degree one, we have $fz^{-1} \in T(\mathbb{C})$ by 4.5; but the eigenvalues of fz^{-1} on V_I have weight zero, so that $fz^{-1} \in T_{\mathbb{R}}$.

From 4.6 we get a commutative diagram

$$\begin{array}{ccccccccc} 0 \to & S^0_{\mathbb{R}} & \to & S_{\mathbb{R}} & \to & \mathbb{Z}\mu & \to 0 \\ & \downarrow & & \downarrow & & \| & \\ 0 \to & T_{\mathbb{R}} & \to & T_{\mathbb{R}} \times \langle z \rangle & \to & \mathbb{Z} & \to 0 \end{array} \tag{4.6.1}$$

in which the middle vertical arrow is the inclusion 4.6.

4.7. LEMMA. *There is a maximal compact subgroup $G^0_{\mathbb{R}}$ of $G_I(\mathbb{C})$ such that $T_{\mathbb{R}} = G_{\mathbb{R}} \cap T(\mathbb{C})$.*

Proof. Since all maximal tori of G_I are conjugate, we can replace T with the "standard" diagonal subgroup of $G_I = \Pi_{n \in I} G_n \subset \Pi_{n \in I} \mathrm{GL}(V_n)$. Then we can take $G_{\mathbb{R}} = \Pi_{n \in I} G_{n\mathbb{R}}$, where $G_{n\mathbb{R}}$ is the "standard" maximal compact subgroup of $G_n(\mathbb{C})$. Namely, if $G_n = \mathrm{SL}(n)$, then $G_{n\mathbb{R}}$ is the special unitary group SU(n), for the standard Hermitian form on $V_n = \mathbb{C}^n$. If $G_n = \mathrm{SP}(n)$, then $G_{n\mathbb{R}} = \mathrm{SU}(n) \cap \mathrm{SP}(n)$; c.f. [11, p. 340 and p. 346, Table I]. The verification that $G_{\mathbb{R}} \cap T(\mathbb{C}) = T_{\mathbb{R}}$ is immediate. □

Let $G^0_{\mathbb{R}}$ be the maximal compact subgroup satisfying the conclusions of 4.7. Define $G_{\mathbb{R}}$ to be the product $G^0_{\mathbb{R}} \times \langle z \rangle \subset W_I(\mathbb{C})$, so that there is an evident commutative diagram

$$\begin{array}{ccccccccc} 0 \to & T_{\mathbb{R}} & \to & T_{\mathbb{R}} \times \langle z \rangle & \to & \mathbb{Z} & \to 0. \\ & \downarrow & & \downarrow & & \| & \\ 0 \to & G^0_{\mathbb{R}} & \to & G_{\mathbb{R}} & \to & \mathbb{Z} & \to 0. \end{array} \tag{4.7.1}$$

We will use the above results to produce constraints on the Frobenius classes, if $S^0 \neq T$. The set of conjugacy classes $(G^0_{\mathbb{R}})^\natural$ of $G^0_{\mathbb{R}}$ can be identified with

$$T^\natural_{\mathbb{R}} \overset{\mathrm{def}}{=} T_{\mathbb{R}}/W$$

where W is the Weyl group of $G_{\mathbb{R}}$. Since z is a central element of $G_{\mathbb{R}}$, the set $G^\natural_{\mathbb{R}}$ of conjugacy classes in $G_{\mathbb{R}}$ can be identified with $T^\natural_{\mathbb{R}} \times \langle z \rangle$. Let x be a point of U_0 of degree n. Then by 4.6, the Frobenius class $\mathrm{Frob}^S_x \in S(\mathbb{C})$ is contained in $S_{\mathbb{R}}$, and by 4.4 the class Frob^M_x has the same image in $W_I(\mathbb{C})$ as Frob^S_x. Therefore $\mathrm{Frob}^M_x \in G^\natural_{\mathbb{R}}$, and in fact

$$\mathrm{Frob}^M_x \in \mathrm{Im}(f^n S^0_{\mathbb{R}} \to G^\natural_{\mathbb{R}}) \tag{4.7.2}$$

where the map $f^n S^0_{\mathbb{R}} \to G^\natural_{\mathbb{R}}$ is the composite

$$\begin{aligned} f^n S^0_{\mathbb{R}} &\to z^n T_{\mathbb{R}} \quad \text{(c.f. 4.5)} \\ &\tilde{\to} T^\natural_{\mathbb{R}} \times z^n \subset G^\natural_{\mathbb{R}}. \end{aligned}$$

Now let $l \neq p$ be a fixed prime. We saw in the proof of 2.4 that if ρ_I is the l-adic representation associated to the sum (2.3.2) of l-adic Kloosterman sheaves, then the Zariski closure of $\mathrm{Im}\,\rho_I \,|\, \pi_1^{\mathrm{geom}}(X)$ is isomorphic to G_I. Let $W(X/k)$ denote the Weil group of X/k, and fix an algebraic closure k^{alg} of k and an isomorphism $W(k^{\mathrm{alg}}/k) \simeq \mathbb{Z}$. Then the pushout of the extension

$$0 \to \pi_1^{\mathrm{geom}}(X) \to W(X/k) \to \mathbb{Z} \to 0$$

by $\pi_1^{\mathrm{geom}}(X) \to G_I$ yields an extension which we claim is isomorphic to W_I.

In fact, if we consider, for each $n \in I$, the twisted sheaf $\mathrm{Kl}_l(n, \psi_l)((n-1)/2)$ and its associated representation ρ_n, then in fact the entire image of ρ_n is contained in G_n [16, 11.0.2]. Since the Tate twist corresponds to the character $g \mapsto z^{-\deg(g)}$ of $W(X/k)$, where z is the element 4.2.4 of W_I, the assertion follows.

Now by [8] 2.2.6, the l-adic Frobenius classes lie in $G_{\mathbb{R}}$. They therefore coincide with the p-adic classes Frob_x^M, for as $G_{\mathbb{R}}^\natural \simeq T_{\mathbb{R}}^\natural \times \langle z \rangle$, a semisimple conjugacy class of $G_{\mathbb{R}}^\natural$ is determined by its eigenvalues on V_I. Denote by U^n the set of closed points of X of degree n, and by $\delta(a)$ the point-mass measure at $a \in T_{\mathbb{R}}^\natural$ of total mass one. If $x \in U^n$, then $z^{-n}\mathrm{Frob}_x^M \in T_{\mathbb{R}}^\natural$, and the identification of l-adic and p-adic Frobenius classes shows that the measure

$$\mu_n = |U^n|^{-1} \sum_{x \in U^n} \delta(z^{-n}\,\mathrm{Frob}_x^M) \tag{4.7.3}$$

is exactly the measure denoted by Y_n in [16, 3.5].

For any map $f: X \to Y$ of measure spaces, and any measure μ on X, we denote by $f_*\mu$ the direct image of the measure μ on Y, i.e. the measure on Y defined by $f_*\mu(A) = \mu(f^{-1}(A))$ for $A \subseteq Y$. Let $\mu^\natural$ denote the direct image on $(G_{\mathbb{R}}^0)^\natural = T_{\mathbb{R}}^\natural$ of Haar measure on $G_{\mathbb{R}}^0$. The equidistribution theorem for the Frobenius classes is the assertion that

$$\mu_n(A) \to \mu^\natural(A) \quad \text{as } n \to \infty \tag{4.7.4}$$

for any measurable $A \subseteq T_{\mathbb{R}}^\natural$. This of course is just the general equidistribution theorem of Deligne [8], [16, 3.6], combined with the identification of G_I with the geometric monodromy group of the l-adic Kloosterman sheaf $\mathrm{Kl}_l(I, \psi_l)$.

We will need a few basic facts about $\mu^\natural$. Denote by μ^T the normalized Haar measure on $T_{\mathbb{R}}$, and by

$$\pi: T_{\mathbb{R}} \twoheadrightarrow T_{\mathbb{R}}^\natural$$

the natural projection. The formula of Hermann Weyl [LIE IX.57, Cor. 2] says that

$$\mu^\natural = h\pi_*\mu^T$$

for a certain bounded function h on $T_{\mathbb{R}}^\natural$ which is positive on the regular conjugacy classes, i.e. on a dense open subset of $T_{\mathbb{R}}^\natural$. In particular, if $A \subseteq T_{\mathbb{R}}^\natural$ is *open* and *nonempty*, then $\mu^\natural(A) > 0$. Next, we have

4.8. LEMMA. *There is a constant $C > 0$ such that*

$$\mu^{\natural}(\pi(A)) \leqslant C\mu^{T}(A)$$

for any measurable $A \subseteq T_{\mathbb{R}}$.

Proof. We have $\pi^{-1}(\pi(A)) = \bigcup_{w\in W} w(A)$ where W is the Weyl group, and $0 \leqslant h < D$ for some $D > 0$. Since $\pi^{*}h$ is W-invariant, we have

$$\begin{aligned}\mu^{\natural}(\pi(A)) &= \int_{\pi^{-1}(\pi(A))} \pi^{*}h\,\mathrm{d}\mu^{T}\\ &\leqslant \sum_{w\in W}\int_{w(A)} \pi^{*}h\,\mathrm{d}\mu^{T}\\ &\leqslant |W|\int_{A} \pi^{*}h\,\mathrm{d}\mu^{T}\\ &\leqslant D|W|\mu^{T}(A)\end{aligned}$$

and we can take $C = D|W|$. □

If $A \subseteq T_{\mathbb{R}}$, we denote by S_A the union of the cosets of $S^0_{\mathbb{R}}$ in $T_{\mathbb{R}}$ that intersect A:

$$S_A = \bigcup_{tS^0_{\mathbb{R}}\cap A\neq\phi} tS^0_{\mathbb{R}}.$$

If $\tau: T_{\mathbb{R}} \to T_{\mathbb{R}}/S^0_{\mathbb{R}}$ is the natural projection, then

$$S_A = \tau^{-1}(\tau(A)). \tag{4.8.1}$$

The map τ is open, so S_A is open if A is. Note also that if $A \subseteq B$, then $S_A \subseteq S_B$.

One more fact about μ_n will be necessary. Put $a = fz^{-1}$; then by (4.7.2) and (4.7.3) we have

$$\operatorname{Supp}\mu_n \subseteq \pi(a^n S^0_{\mathbb{R}}). \tag{4.8.2}$$

4.9. LEMMA. *If $A \subset T^{\natural}_{\mathbb{R}}$ is a measurable subset such that $\mu_n(A) > 0$, then*

$$\operatorname{Supp}\mu_n \subseteq \pi(S_{\pi^{-1}(A)}).$$

Proof. By (4.8.2) and the definitions, we have

$$\mu_n(A) > 0 \Rightarrow \text{Supp } \mu_n \cap A \neq \phi$$
$$\Rightarrow \pi(a^n S^0_{\mathbb{R}}) \cap A \neq \phi$$
$$\Rightarrow a^n S^0_{\mathbb{R}} \cap \pi^{-1}(A) \neq \phi$$
$$\Rightarrow a^n S^0_{\mathbb{R}} \subseteq S_{\pi^{-1}(A)}$$
$$\Rightarrow \text{Supp } \mu_n \subseteq \pi(a^n S^0_{\mathbb{R}}) \subseteq \pi(S_{\pi^{-1}(A)}).$$

We now come to the key point: □

4.10. LEMMA. $S^0 = T$.

Proof. Since $S^0 \subseteq T$, $\dim_{\mathbb{R}} S^0_{\mathbb{R}} = \dim S^0$, and $\dim_{\mathbb{R}} T_{\mathbb{R}} = \dim T$, it is enough to show that $\dim_{\mathbb{R}} S^0_{\mathbb{R}} = \dim_{\mathbb{R}} T_{\mathbb{R}}$. Assume, to the contrary, that $\dim_{\mathbb{R}} S^0_{\mathbb{R}} < \dim_{\mathbb{R}} T_{\mathbb{R}}$. We shall show that there is a subset $A \subset T^\natural_{\mathbb{R}}$ such that

$$0 < \mu^\natural(\pi(S_{\pi^{-1}(A)})) < \mu^\natural(T^\natural_{\mathbb{R}}), \tag{4.10.1}$$

and arrive at a contradiction in the following manner. Put $U = T^\natural_{\mathbb{R}} - \pi(S_{\pi^{-1}(A)})$; then we have $\mu^\natural(U) > 0$ since $\mu^\natural(\pi(S_{\pi^{-1}(A)})) < \mu^\natural(T^\natural_{\mathbb{R}})$. On the other hand, since $\mu^\natural(\pi(S_{\pi^{-1}(A)})) > 0$ and $\mu_n(\pi(S_{\pi^{-1}(A)})) \to \mu^\natural(\pi(S_{\pi^{-1}(A)}))$ as $n \to \infty$, we must have $\mu_n(\pi(S_{\pi^{-1}(A)})) > 0$ for all sufficiently large n. Then by 4.9 we have that $\text{Supp } \mu_n \subseteq \pi(S_{\pi^{-1}(A)}) = T^\natural_{\mathbb{R}} - U$ for all $n \gg 0$, and therefore $\mu_n(U) = 0$ for all $n \gg 0$. Thus $\mu(U) = \lim_{n\to\infty} \mu_n(U) = 0$, a contradiction.

To show the existence of an A satisfying (4.10.1) (under the hypothesis that $\dim_{\mathbb{R}} S^0_{\mathbb{R}} < \dim_{\mathbb{R}} T_{\mathbb{R}}$, of course), we first remark that if A is open, then so is $\pi(S_{\pi^{-1}(A)})$. In fact if A and $\pi^{-1}(A)$ are open, the remark right after (4.8.1) shows that $S_{\pi^{-1}(A)}$ is open as well. But π is an open map, so $\pi(S_{\pi^{-1}(A)})$ must be open as well. If A is nonempty, then so is $\pi(S_{\pi^{-1}(A)})$, since π is surjective.

From this and (4.7.4) it follows that if A is a nonempty open subset of $T^\natural_{\mathbb{R}}$, then $\mu^\natural(\pi(S_{\pi^{-1}(A)})) > 0$. Thus it is enough to find a nonempty open set $A \subseteq T^\natural_{\mathbb{R}}$ such that $\mu^\natural(\pi(S_{\pi^{-1}(A)})) < \mu^\natural(T^\natural_{\mathbb{R}})$. Let $\mu^{T/S}$ be the normalized Haar measure on the quotient $T_{\mathbb{R}}/S^0_{\mathbb{R}}$, and recall that τ is the natural projection $T_{\mathbb{R}} \to T_{\mathbb{R}}/S^0_{\mathbb{R}}$. Since $\tau_* \mu^T$ is translation-invariant, it is a multiple of Haar measure, and so $\tau_* \mu^T = \mu^{T/S}$ on account of the normalization. Now *if* $\dim T_{\mathbb{R}} > \dim S^0_{\mathbb{R}}$, then the quotient $T_{\mathbb{R}}/S^0_{\mathbb{R}}$ is a torus of positive dimension. Then for all sufficiently small $\varepsilon > 0$, there is an open neighborhood V of the origin in $T_{\mathbb{R}}/S^0_{\mathbb{R}}$ such that $\mu^T(\tau^{-1}(V)) = \mu^{T/S}(V) = \varepsilon$. We set

$$B = \bigcap_{w \in W} w(\tau^{-1}(V))$$

and claim that

$$A = \pi(B)$$

will satisfy (4.10.1) for an appropriate choice of ε. We first show that A is open and nonempty. In fact $0 \in w\tau^{-1}(V)$ for all $w \in W$, so $A \neq \phi$, and since V and the $w\tau^{-1}(V)$ are open, B is open. Since π is an open map, we conclude that A is open.

$$\pi^{-1}(\pi(C)) = \bigcup_{w \in W} wC$$

and so, since B is W-stable,

$$\pi^{-1}(A) = \pi^{-1}(\pi(B)) = B.$$

Then since $B \subseteq \tau^{-1}(V)$, we have

$$S_{\pi^{-1}(A)} = S_B \subseteq S_{\tau^{-1}(V)} = \tau^{-1}(\tau(\tau^{-1}(V))) = \tau^{-1}(V)$$

by (4.8.1), and therefore

$$\mu^T(S_{\pi^{-1}(A)}) \leqslant \mu^T(\tau^{-1}(V)) = \mu^{T/S}(V) = \varepsilon.$$

Finally, 4.8 shows that

$$\mu^\natural(\pi(S_{\pi^{-1}(A)})) \leqslant C\mu^T(S_{\pi^{-1}(A)}) \leqslant C\varepsilon$$

where C is the positive constant in 4.8. If we now choose ε so that $C\varepsilon < \mu^\natural(T^\natural_{\mathbb{R}})$, we get the second inequality in (4.10.1). □

We can now put everything together:

Proof of 2.3. We first show that if $\mathrm{DGal}(\hat{M}) = B_I$, then $\mathrm{DGal}(M) = G_I$. By the analogue [7, Theorem 4.9] of Grothendieck's monodromy theorem for overconvergent F-isocrystals, we know that the radical of the connected component $\mathrm{DGal}(M)^0$ of $\mathrm{DGal}(M)$ must be unipotent. Given that $B_I = \mathrm{DGal}(\hat{M}) \subseteq \mathrm{DGal}(M)$, it follows that $\mathrm{DGal}(M)^0$ is a parabolic subgroup of G_I. Now the well-known description [4, 14.7 and 14.8] of the parabolic subgroups of a semisimple group shows that the only possibility for $\mathrm{DGal}(M)^0$ allowed by the monodromy theorem is G_I itself, so we must have $\mathrm{DGal}(M)^0 = G_I$. On the other hand we have $\mathrm{DGal}(M) \subseteq G_I$, and so $\mathrm{DGal}(M) = G_I$.

Next, we reduce to the case when 4.3 is valid. We recall from 2.1 that there is a finite étale cover $g: X' \to X$ such that the natural map $\mathrm{DGal}(g^*\hat{M}) \hookrightarrow \mathrm{Dgal}(\hat{M})$ induces an isomorphism

$$\mathrm{DGal}(g^*\hat{M}) \xrightarrow{\sim} \mathrm{DGal}(\hat{M})^0.$$

It may be necessary to extend the base field, in order to guarantee that X' has rational points; by (2.1.3), this is not a problem. Now we have $\mathrm{DGal}(\hat{M}) \subseteq B_I$, so if we show that $\mathrm{DGal}(g^*\hat{M}) = B_I$, then $\mathrm{DGal}(M) = B_I$ as well.

We can therefore assume that (4.2.3) holds, and make use of all of 4.4–4.10. To show that $\mathrm{DGal}(\hat{M}) = B_I$, it is enough to show that $\mathrm{DGal}(\hat{M}) \otimes A \simeq B_I \otimes A$, and for this, it is enough to show

- $\mathrm{DGal}(\hat{M}) \otimes A$ contains a maximal torus of $B_I \otimes A$, and
- $\mathrm{DGal}(\hat{M}) \otimes A$ contains the unipotent radical U_I of $B_I \otimes A$.

By 4.10, $\mathrm{DGal}(\hat{M})$ contains a maximal torus of B_I, from which the first assertion follows. As to the second assertion, we know from 3.7 that Lie $\mathrm{DGal}(\hat{M})$ contains a principal nilpotent element of Lie $G_I \otimes A$. Choose a maximal torus T of G_I contained in $\mathrm{DGal}(\hat{M})$, and let $\mathscr{R}$ be the corresponding root system of $G_I \otimes A$; finally, let $\mathscr{B}$ be the basis of R corresponding to the Borel $B_I \otimes A \supset T$; then

$$\mathrm{Lie}\ \mathrm{DGal}(\hat{M}) \otimes A \subseteq \mathrm{Lie}\ B_I \otimes A = \mathrm{Lie}\ T \oplus \bigoplus_{a>0} \mathfrak{g}_\alpha \tag{4.11.1}$$

in the notation of (3.6.4). If $N \in \mathrm{Lie}\ \mathrm{DGal}(\hat{M}) \otimes A$ is a principal nilpotent element of Lie $G_I \otimes A$, then we can write $N = \Sigma_{\alpha>0} N_\alpha$ with $N_\alpha \in \mathfrak{g}_\alpha$, and $N_\alpha \neq 0$ for all $\alpha \in \mathscr{B}$. Since $T \subset \mathrm{DGal}(\hat{M}) \otimes A$, we have $\mathfrak{g}_\alpha \subset$ Lie $\mathrm{DGal}(\hat{M}) \otimes A$ for all $a \in \mathscr{B}$. Then since $\mathfrak{g}_{\alpha+\beta} = [\mathfrak{g}_a, \mathfrak{g}_\beta]$, we have in fact $\mathfrak{g}_\alpha \subset$ Lie $\mathrm{DGal}(\hat{M}) \otimes A$ for all $\alpha > 0$, and so by (4.11.1) we have Lie $B_I \otimes A =$ Lie $\mathrm{DGal}(\hat{M}) \otimes A$. Since the groups are connected, we have $\mathrm{DGal}(\hat{M}) \otimes A = B_I \otimes A$. □

References

[1] P. Berthelot, *Géometrie rigide et cohomologie des variétés algèbriques de caracteristique p*, Journés d'analyse *p*-adique (Luminy 1982), Mémoire de la S.M.F. No. 23, suppl. au Bull. S.M.F. 114 (1986) fasc.2, 7–32.

[2] P. Berthelot, *Cohomologie rigide et théorie de Dwork: le cas des sommes exponentielles*, in Astérisque 119–120 (1984) 17–49.

[3] P. Berthelot, *Cohomologie rigide et cohomologie rigide à support propre*, to appear in Astérisque.

[4] A. Borel, *Linear algebraic groups*, 2nd ed., Springer-Verlag (1991).
[5] S. Bosch, B. Dwork, and Ph. Robba, *A rigid analytic version of M. Artin's theorem on analytic equations*, Math. Ann. 255 (1981) 395–404.
[6] R. Crew, *F-isocrystals and p-adic representations*, in *Algebraic Geometry – Bowdoin 1985*, Proc. Symp. Pure Math. 46(2) (1987) 111–138.
[7] R. Crew, *F-isocrystals and their monodromy groups*, Ann. Sc. Ec. Norm. Sup. 4^e sér. 25 (1992) 429–464.
[8] P. Deligne, *La conjecture de Weil II*, Publ. Math. IHES 52 (1980) 137–352.
[9] P. Deligne and J. Milne, *Tannakian Categories*, in Lecture Notes in Math. 900, Springer-Verlag (1982).
[10] B. Dwork, *Bessel functions as p-adic functions of the argument*, Duke Math. J. 41 (1974) 711–738.
[11] S. Helgason, *Differential geometry and symmetric spaces*, Academic Press (1962).
[12] N. Katz, *Travauxs de Dwork*, Séminaire Bourbaki 1971–2, exposé 409, in Lecture Notes in Math. 317, Springer-Verlag (1973), pp. 167–200.
[13] N. Katz, *p-adic properties of modular schemes and modular forms*, in Lecture Notes in Math. 350, Springer-Verlag (1973), 69–190.
[14] N. Katz, *Slope filtration of F-crystals*, in *Journées de géométrie algébrique de Rennes I*, Astérisque 63 (1979) 113–163.
[15] N. Katz, *On the calculation of some differential galois groups*, Inv. Math. 87 (1987) 13–61.
[16] N. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Math. Studies 116, Princeton Univ. Press (1988).
[17] N. Katz, *Exponential sums and differential equations*, Annals of Math. Studies 124, Princeton Univ. Press (1990).
[18] M. Larsen and R. Pink, *On l-independence of algebraic monodromy groups in compatible systems of representations*, preprint.
[19] A. Ogus, *F-isocrystals and De Rham cohomology II – Convergent isocrystals*, Duke J. Math. 51 (1984) 765–850.
[20] N. Saavedra R., *Categories Tannakiennes*, Lecture Notes in Math. 265, Springer-Verlag (1972).
[21] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, W. A. Benjamin (1968).
[22] S. Sperber, *p-adic hypergeometric functions and their cohomology*, Duke Math. J. 44 (1977) 535–589.
[23] S. Sperber, *Congruence properties of the hyperkloosterman sum*, Com. Math. 40 (1980) 3–33.
[24] S. Sperber, *Newton polygons for general hyperkloosterman sums*, in *Cohomologie p-adique*, Astérisque 119–120 (1984) 267–330.

[LIE] N. Bourbaki, *Groupes et algèbres de Lie*, Masson, Paris (1982).
[SGA 4½] P. Deligne et al., *Cohomologie étale*, Lecture Notes in Math. 569, Springer-Verlag (1977).