Exponential sums over finite fields and the large sieve

Corentin Perret-Gentil

Abstract. By using a variant of the large sieve for Frobenius in compatible systems developed in [Kow06a] and [Kow08], we obtain zero-density estimates for arguments of ℓ-adic trace functions over finite fields with values in some algebraic subsets of the cyclotomic integers, when the monodromy groups are known. This applies in particular to hyper-Kloosterman sums and general exponential sums considered by Katz.

Contents

1. Introduction

1.1. **Exponential sums and trace functions.** We consider exponential sums over a finite field $\mathbb{F}_q$ of characteristic $p \geqslant 5$ such as:

(1) Hyper-Kloosterman sums of rank $n \geqslant 2$ given by

$$\mathrm{Kl}_{n,q}(a) = \frac{(-1)^{n-1}}{q^{(n-1)/2}} \sum_{\substack{x_1,\dots,x_n \in \mathbb{F}_q^\times \\ x_1 \cdots x_n = a}} e\left(\frac{\mathrm{tr}(x_1 + \cdots + x_n)}{p}\right), \tag{1}$$

for $a \in \mathbb{F}_q^\times$, $\mathrm{tr} : \mathbb{F}_q \to \mathbb{F}_p$ the trace map, and $e(z) = \exp(2\pi i z)$ for any $z \in \mathbb{C}$. More generally, we also have hypergeometric sums as introduced in [Kat90, Chapter 8];

(2) General exponential sums of the form

$$\frac{-1}{q^{1/2}} \sum_{\substack{y \in \mathbb{F}_q \\ f(y), g(y), h(y) \neq \infty}} e\left(\frac{\mathrm{tr}(xf(y) + h(y))}{p}\right) \chi(g(y)), \tag{2}$$

for $x \in \mathbb{F}_q$, $f, g, h \in \mathbb{Q}(X)$ rational functions and χ a character of $\mathbb{F}_q^\times$. For example, we have Birch sums $q^{-1/2} \sum_{y \in \mathbb{F}_q} e\left(\mathrm{tr}(xy + y^3)/p\right)$, cubic exponential sums studied in particular by Livné [Liv87] and Katz;

Date: March 2017. Revised August 2018.

(3) Functions counting points on families of curves such as

$$\frac{q+1-|X_z(\mathbb{F}_q)|}{q^{1/2}} \quad (z\in\mathbb{F}_q,\ f(z)\neq 0), \tag{3}$$

where X_z is the smooth projective model of the affine hyperelliptic curve $y^2 = f(x)(x-z)$ over $\mathbb{F}_q$, for $f\in\mathbb{Z}[X]$ fixed squarefree of degree $2g\geqslant 2$.

1.1.1. *Exponential sums as algebraic integers.* Note that the three examples above all take values in the localization $\mathbb{Z}[\zeta_{4p}]_{q^{1/2}}$ (by the evaluation of quadratic Gauss sums), or less precisely in the cyclotomic field $\mathbb{Q}(\zeta_{4p})$.

It is an interesting question to investigate their properties as elements of these sets, as done by Fisher [Fis92, Fis95] or recently by the author [PG17a] for the distribution of their reductions modulo a prime ideal and short sums thereof.

1.1.2. *Trace functions.* Examples (1)–(3) are specific incarnations of *trace functions* $t:\mathbb{F}_q\to\mathbb{C}$ *arising from constructible middle-extension sheaves of* $\overline{\mathbb{Q}}_\ell$*-modules on* $\mathbb{P}^1/\mathbb{F}_q$, for ℓ a prime distinct from p, as constructed in particular by Deligne [Del77] and Katz [Kat90].

Very powerful tools are then available to study various aspects of these functions, such as Deligne's extension [Del80] of the Riemann hypothesis for varieties over finite fields to weights of étale cohomology groups of such sheaves.

For example, Katz [Kat88] obtained a "vertical Sato–Tate law" for the distribution of Kloosterman sums, through a general equidistribution theorem of Deligne, and similar results [Kat90] for families of the type (2) or (3).

1.2. **Zero-density estimates.** The goal of the present article is to obtain general zero-density estimates of the form

$$P\big(t(x)\in A\big) := \frac{|\{x\in\mathbb{F}_q : t(x)\in A\}|}{q} = o(1) \qquad (q\to+\infty) \tag{4}$$

where:

- $t:\mathbb{F}_q\to E$ is the trace function associated to a *coherent family* of sheaves over $\mathbb{F}_q$ (Definition 2.1), for E a number field.
- $A\subset E$ is an "algebraic" subset such as the set of m-powers ($m\geqslant 2$), the image of a polynomial, or more generally a set defined by a first-order formula in the language of rings.

This will apply in particular, with $E=\mathbb{Q}(\zeta_{4p})$, to Kloosterman sums (1) and exponential sums of the form (2).

1.2.1. *Families of curves.* The large sieve for Frobenius in compatible systems was developed by Kowalski in [Kow06a] and [Kow08] to obtain results of the type of Chavdarov [Cha97] on zeta functions of families of curves, such as the probability that the numerator has Galois group as large as possible.

In the notations of Example (3) above, Kowalski gets for example (see [Kow08, Section 8.8]) that

$$P\Big(f(z)\neq 0,\ |X_z(\mathbb{F}_q)|\in\mathbb{N}^{\times 2}\Big) := \frac{|\{z\in\mathbb{F}_q : f(z)\neq 0,\ |X_z(\mathbb{F}_q)|\in\mathbb{N}^{\times 2}\}|}{q} \ll gq^{1-(4g^2+2g+4)^{-1}}\log q,$$

for $\mathbb{N}^{\times 2}$ the set of squares of integers.

The large sieve bound ultimately relies on estimates of exponential sums obtained through Deligne's generalization of the Riemann hypothesis over finite fields.

Note that in the setting above, we have $E=\mathbb{Q}$.

1.2.2. *Examples of results for Kloosterman sums.* In the case of hyper-Kloosterman sums (1) of rank $n\geqslant 2$, our main results are the following:

Proposition 1.1. *Let $n\geqslant 2$ be an integer and $\varepsilon>0$. For $m\geqslant 2$ coprime to p, we have*

$$P\Big(\mathrm{Kl}_{n,q}(x)\in\mathbb{Q}(\zeta_{4p})^m\Big)\ll_{n,m,\varepsilon}\frac{p^\varepsilon\log q}{B_n q^{1/(2B_n)}}\to 0$$

when $q=p^e\to+\infty$ is coprime to n with $e\geqslant 16B_n$, where

$$B_n=\begin{cases}\frac{2n^2+n-1}{2} & : n \text{ odd}\\ \frac{2n^2+3n+4}{4} & : n \text{ even},\end{cases} \tag{5}$$

and $\mathbb{Q}(\zeta_{4p})^m$ is the set of mth powers in $\mathbb{Q}(\zeta_{4p})$. The implied constant depends only on n, m and ε.

More generally:

Proposition 1.2. *Let $n\geqslant 2$ be an integer and $\varepsilon>0$. For*

- *almost all*[1] *monic polynomials $f\in\mathbb{Z}[X]$ of fixed degree $d\geqslant 2$, and*
- *all $f\in\mathbb{Z}[X]$ of degree $d\geqslant 2$ such that the Galois group of $f(X)-y\in\mathbb{C}(y)[X]$ is equal to $\mathfrak{S}_d$,*

we have

$$P\Big(\mathrm{Kl}_{n,q}(x)\in f(\mathbb{Q}(\zeta_{4p}))\Big)\ll_{n,f,\varepsilon}\frac{p^\varepsilon\log q}{B_n q^{1/(2B_n)}}\to 0$$

when $q=p^e\to+\infty$ is coprime to n with $e\geqslant 16B_n$, for B_n as in (5). The implied constant depends only on n, f and ε.

Remarks 1.3.

(1) The bounds are uniform in p, thanks to the determination of the finite monodromy groups in [PG18], over a field of characteristic $\ell\gg_n 1$.
(2) This can further be extended to definable subsets of $\mathbb{Q}(\zeta_{4p})$ (i.e. defined by a first-order formula in the language of rings), under some technical conditions (Proposition 5.1 later on).
(3) The same bounds hold for unnormalized Kloosterman sums.

[1] Throughout, this will mean "for all but $o(h^d)$ such polynomials of height at most h, as $h\to+\infty$".

(4) Under the general Riemann hypothesis (GRH) for the Dedekind zeta function of $\mathbb{Q}(\zeta_{4p})$, one may take $\varepsilon = 0$ and $e \geqslant 4B_n + 1$.
(5) By relying on the determination of the monodromy groups over $\overline{\mathbb{Q}}_\ell$ by Katz and the results of Larsen–Pink (see Section 5.2), instead of [PG18], these results would only hold when p is fixed and $e \to +\infty$, with an implied constant depending on p.

1.3. **Strategy.** The general idea to obtain zero-density estimates of the type (4) is the following: in the setting of Section 1.2, let $\mathcal{O}$ be the ring of integers of E. It turns out (by definition of a coherent family) that there exists a set Λ of valuations of $\mathcal{O}$ (equivalently, of prime ideals) such for every $\lambda \in \Lambda$, the function $t : \mathbb{F}_q \to E$ coincides with the trace function $t_\lambda : \mathbb{F}_q \to \mathcal{O}_\lambda$ arising from a constructible middle-extension sheaf of $\mathcal{O}_\lambda$-modules on $\mathbb{P}^1/\mathbb{F}_p$. By reduction, we obtain a trace function $\tilde{t}_\lambda : \mathbb{F}_q \to \mathbb{F}_\lambda$, where $\mathbb{F}_\lambda$ is the residue field at λ.

Thus,

$$P\big(t(x) \in A\big) \leqslant \frac{|\{x \in \mathbb{F}_q : \tilde{t}_\lambda(x) \in A_\lambda \ \forall \lambda \in \Lambda\}|}{q},$$

where $A_\lambda = (A \cap \mathcal{O}_\lambda) \pmod{\lambda} \subset \mathbb{F}_\lambda$. A variant of Kowalski's large sieve for Frobenius in compatible systems, handling sheaves of $\mathcal{O}_\lambda$-modules instead of sheaves of $\mathbb{Z}_\ell$-modules, can then be used to bound this quantity in terms of local densities in the sets A_λ.

1.3.1. *Technical tools.* More precisely, the first part of the approach requires:

- The construction by Deligne and Katz of examples of the form (1) and (2) as trace functions of sheaves of $\mathcal{O}_\lambda$-modules.
- Information on monodromy groups:
  - When available, the determination of integral monodromy groups for a density one subset of the valuations, not depending on p.
  - Otherwise, results of Larsen and Pink [LP92, Lar95] to handle sheaves whose monodromy groups are known over $\overline{\mathbb{Q}}_\ell$ (e.g. by the works of Katz [Kat88, Kat90]), but not over $\mathbb{F}_\lambda$.
  - For sheaves associated with exponential sums of the form (2), conditions and/or normalizations so that arithmetic and geometric monodromy groups coincide.

To compute local densities in the sets A_λ, we will need bounds on "Gaussian sums" (see Section 3) over:

- Linear algebraic groups over $\mathbb{F}_\lambda$; these follow either from Deligne's generalization of the Riemann hypothesis over finite fields [Del80] and bounds of Katz on sums of Betti numbers [Kat01], or from explicit computations of D.S. Kim for certain finite groups of Lie type.
- Subsets of $\mathbb{F}_\lambda$ such as powers (Bourgain and others, e.g. [BC06]) or more generally definable subsets (Kowalski [Kow07], using the work of Chatzidakis–van der Dries–Macintyre [CvdDM92]).

The implied constant in a bound of the form (4) will depend on p (forcing to fix p and take $q = p^e$, $e \to +\infty$) when we rely on the results of Larsen–Pink, and will be absolute when more precise information about integral

monodromy groups is available.

When we want results with absolute implied constants, we will also employ uniform estimates in Chebotarev's density theorem (e.g. [May13]), since E may depend on p.

1.4. **Organization of the paper.** In Section 2, we lay out the technical setup of trace functions of sheaves of $\mathcal{O}_\lambda$-modules over finite fields, define coherent families, and show that (1) and (3) arise from such families. Finally, we state a variant of the large sieve for Frobenius in compatible systems (Theorem 2.7).

In Section 3, we get results on the Gaussian sums mentioned above, which will be used to compute the local densities in the sieve.

In Section 4, we apply the large sieve of Section 2 to obtain bounds of the type (4), by using the estimates from Section 3 and uniform bounds in Chebotarev's density theorem.

In Section 5, we start by explaining how this leads to the results for Kloosterman sums given in Section 1.2.2 above. Then, we work towards obtaining similar zero-density estimates for general exponential sums of the form (2), showing that coherent families can still be obtained through the results of Larsen and Pink (in particular with Theorem 5.3).

Acknowledgements. The author would like to thank Emmanuel Kowalski and Richard Pink for helpful discussions, as well as the anonymous referees for very valuable comments. This work was partially supported by DFG-SNF lead agency program grant 200021L_153647 and by the National Science Foundation under Grant No. 1440140, while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring semester of 2017. Some of the results also appeared in the author's PhD thesis.

2. The large sieve for Frobenius in compatible systems

We start by recalling the technical setup of trace functions over finite fields, before stating a version of the large sieve for Frobenius adapted to our needs.

Throughout this section, a number field E with ring of integers $\mathcal{O}$ is fixed, as well as a finite field $\mathbb{F}_q$ of characteristic p.

2.1. Trace functions over finite fields.

2.1.1. *Definitions.* Let λ be an ℓ-adic valuation corresponding to a prime ideal $\mathfrak{l}$ of $\mathcal{O}$, E_λ and $\mathcal{O}_\lambda$ the completions, and $\mathbb{F}_\lambda \cong \mathcal{O}/\mathfrak{l}$ the residue field.

Let $A = \overline{\mathbb{Q}}_\ell$, $\mathcal{O}_\lambda$ or $\mathbb{F}_\lambda$. We recall that a constructible middle-extension sheaf of A-modules over $\mathbb{P}^1/\mathbb{F}_p$ (or *sheaf of A-modules over $\mathbb{F}_p$* for simplicity) corresponds to a continuous ℓ-adic Galois representation

$$\rho_{\mathcal{F}} : \pi_{1,p} := \mathrm{Gal}\left(\mathbb{F}_p(T)^{\mathrm{sep}}/\mathbb{F}_p(T)\right) \to \mathrm{GL}(\mathcal{F}_{\overline{\eta}}) \cong \mathrm{GL}_n(A),$$

for $\overline{\eta}$ a geometric generic point and $\mathbb{F}_p(T)^{\mathrm{sep}}$ the corresponding separable closure. The associated *trace functions* are, for every finite extension $\mathbb{F}_q/\mathbb{F}_p$,

$$\begin{aligned} t_{\mathcal{F}} = t_{\mathcal{F},q} : \mathbb{F}_q &\to A \\ x &\mapsto \mathrm{tr}\left(\rho_{\mathcal{F}}(\mathrm{Frob}_{x,q}) \mid \mathcal{F}_{\overline{\eta}}^{I_x}\right), \end{aligned}$$

where[2] $\mathrm{Frob}_{x,q} \in (D_x/I_x)^\sharp \cong \mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ is the geometric Frobenius at $x \in \mathbb{F}_q$, for $I_x \trianglelefteq D_x \leqslant \pi_{1,p}$ the inertia (resp. decomposition) group at x. We will denote by $U_{\mathcal{F}} \subset \mathbb{P}^1$ the maximal open of lissité of $\mathcal{F}$.

We refer the reader to [Kat88, Chapter 2] for more details and references.

2.1.2. *Monodromy groups.* If $\mathcal{F}$ is a sheaf of A-modules over $\mathbb{F}_p$ as above, the *arithmetic and geometric monodromy groups* of $\mathcal{F}$ are the groups

$$G_{\mathrm{geom}}(\mathcal{F}) = \rho_{\mathcal{F}}\big(\pi_{1,p}^{\mathrm{geom}}\big) \leqslant G_{\mathrm{arith}}(\mathcal{F}) = \rho_{\mathcal{F}}(\pi_{1,p}) \leqslant \mathrm{GL}_n(A)$$

if A is discrete, and

$$G_{\mathrm{geom}}(\mathcal{F}) = \overline{\rho_{\mathcal{F}}\big(\pi_{1,p}^{\mathrm{geom}}\big)} \leqslant G_{\mathrm{arith}}(\mathcal{F}) = \overline{\rho_{\mathcal{F}}(\pi_{1,p})} \leqslant \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$$

if $A = \overline{\mathbb{Q}}_\ell$, where $\bar{\cdot}$ denotes Zariski closure, for $\pi_{1,p}^{\mathrm{geom}} := \mathrm{Gal}(\mathbb{F}_p(T)^{\mathrm{sep}}/\overline{\mathbb{F}}_p(T))$.

The works of Katz (see e.g. [Kat88, Kat90, KS99]) contain the determination of the monodromy groups over $\overline{\mathbb{Q}}_\ell$ of many sheaves of interest, such as Kloosterman sheaves. An important input is the fact that, for pointwise pure of weight 0 sheaves, the connected component of the geometric monodromy group is a semisimple algebraic group by a result of Deligne.

The determination of discrete monodromy groups is usually more difficult, since they have far less structure.

2.2. **Coherent families.**

Definition 2.1. Let Λ be a set of valuations on $\mathcal{O}$ and let $U \subset \mathbb{P}^1/\mathbb{F}_p$ be an open affine subset. A family $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$, where $\mathcal{F}_\lambda$ is a sheaf of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$ with maximal open of lissité U, is *coherent* if:

(1) It forms a *compatible system*: if $\rho_\lambda : \pi_{1,p} \to \mathrm{GL}_n(\mathcal{O}_\lambda)$ is the representation corresponding to $\mathcal{F}_\lambda$, then for every $\lambda \in \Lambda$, every finite extension $\mathbb{F}_q/\mathbb{F}_p$ and every $x \in U(\mathbb{F}_q)$, the characteristic polynomial
$$\mathrm{charpol}\,\rho_\lambda(\mathrm{Frob}_{x,q}) \in \mathcal{O}_\lambda[T]$$
lies in $E[T]$ and does not depend on λ.

(2) There exists $G \in \{\mathrm{SL}_m, \mathrm{Sp}_{2m}\}$ such that for every $\lambda \in \Lambda$ corresponding to a prime ideal $\mathfrak{l} \trianglelefteq \mathcal{O}$, the arithmetic and geometric monodromy groups of $\widetilde{\mathcal{F}}_\lambda := \mathcal{F}_\lambda \pmod{\mathfrak{l}}$ coincide and are conjugate to $G(\mathbb{F}_\lambda)$. We call G the *monodromy group structure* of the family.

The *conductor* of the family is defined to be $\sup_{\lambda\in\Lambda} \mathrm{cond}(\widetilde{\mathcal{F}}_\lambda)$, where

$$\mathrm{cond}(\widetilde{\mathcal{F}}_\lambda) = n + |\,\mathrm{Sing}(\widetilde{\mathcal{F}}_\lambda)| + \sum_{x\in\mathrm{Sing}(\widetilde{\mathcal{F}}_\lambda)} \mathrm{Swan}_x(\widetilde{\mathcal{F}}_\lambda) \quad (\lambda \in \Lambda)$$

is the conductor defined by Fouvry–Kowalski–Michel (see e.g. [FKM15]).

[2] The set of conjugacy classes of a group G will be denoted by $G^\sharp$.

Remark 2.2. Here, the prime p is fixed, and the bounds of type (4) would concern the trace functions on $\mathbb{F}_q$ obtained for every power q of p. However, it may also make sense to vary p (e.g. for Kloosterman sums of fixed rank, exponential sums (2) coming from the reduction of integer polynomials, etc.), and the conductor will allow to control this dependency. See also Remark 1.3.

If $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ is a compatible system as above, then in particular the trace function $t = t_{\mathcal{F}_\lambda} : \mathbb{F}_q \to \mathcal{O}_\lambda$ (as the opposite of the coefficient of order $n-1$ in the characteristic polynomial) is independent from λ and takes values in E. More precisely,

$$t(\mathbb{F}_q) \subset \bigcap_{\lambda\in\Lambda} \mathcal{O}_\lambda \cap E = \bigcap_{\mathfrak{l}\in\Lambda} \mathcal{O}_\mathfrak{l} = (\mathrm{Spec}(\mathcal{O}) - \Lambda)^{-1}\,\mathcal{O} \subset E, \tag{6}$$

where $\mathcal{O}_\mathfrak{l}$ is the localization at the ideal $\mathfrak{l}$ corresponding to the valuation λ.

2.2.1. *Fourier transforms and coherent families.* The sheaves we will consider arise by ℓ-adic Fourier transforms, as developed by Deligne, Laumon and others (see [Kat90, Section 7.3], [Kat88, Chapter 5]), corresponding to the discrete Fourier transform on the level of trace functions.

This often results in sheaves with large classical monodromy groups, which is part of Condition (2) above.

Concerning Condition (1) and the conductor, we recall:

Lemma 2.3. *Let us assume that $\mathbb{Q}(\zeta_{4p}) \leqslant E$ and let $\psi : \mathbb{F}_p \to \mathbb{C}$ be a nontrivial additive character. If $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ is a compatible system of Fourier sheaves*[3] *of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$, then the family $(\mathrm{FT}_\psi(\mathcal{F}_\lambda))_{\lambda\in\Lambda}$ is compatible as well and $\mathrm{cond}(\widetilde{\mathrm{FT}_\psi}(\mathcal{F}_\lambda)) \ll \mathrm{cond}(\widetilde{\mathcal{F}}_\lambda)^2$, where FT_ψ denotes the normalized Fourier transform with respect to ψ.*

Proof. Let $\mathcal{F} = \mathcal{F}_\lambda$ and $\mathcal{G} = \mathrm{FT}_\psi(\mathcal{F})$. By construction, for every finite extension $\mathbb{F}_q/\mathbb{F}_p$ and every $a \in U_\mathcal{G}(\mathbb{F}_q)$, the reverse characteristic polynomial $\det(1 - \mathrm{Frob}_{a,q}\,T \mid \mathcal{G}_{\overline{\eta}})$ is equal to

$$\prod_{i=0}^{2} \det\left(1 - \mathrm{Frob}_q\,T \mid H^i_c(U_\mathcal{G} \times \overline{\mathbb{F}}_p, \mathcal{F} \otimes \mathcal{L}_{\psi(ax)})\right)^{(-1)^{i+1}}_,$$

where $\mathcal{L}_{\psi(ax)}$ denotes an Artin–Schreier sheaf and H^i_c the ith ℓ-adic cohomology group with compact support. By the Grothendieck–Lefschetz trace formula [Del77, Exposé 2], this is $\exp\left(\sum_{n\geqslant 1} S(a,n)T^n/n\right)$, where $S(a,n) = \sum_{x\in U_\mathcal{G}(\mathbb{F}_{q^n})} t_{\mathcal{F},q^n}(x)\psi(\mathrm{tr}(ax))$ has image in E and does not depend on λ by hypothesis, whence the conclusion.

The assertion on the conductors can be found in [FKM15, Proposition 8.2], along with [Kat88, Remark 1.10]. □

2.2.2. *Examples.* For the examples below, we let $E = \mathbb{Q}(\zeta_{4p})$, with ring of integers $\mathcal{O} = \mathbb{Z}[\zeta_{4p}]$.

[3] See [Kat90, 7.3.5] for the relevant definitions.

Proposition 2.4 (Kloosterman sheaves)**.** *Let $n \geqslant 2$ be a fixed integer coprime to p. For*

$$\Lambda_n = \{\lambda \text{ } \ell\text{-adic valuation on } \mathcal{O} : \ p \neq \ell \gg_n 1, \ \ell \equiv 1 \pmod 4, \ \mathbb{F}_\lambda = \mathbb{F}_\ell\},$$

there exists a coherent family $(\mathcal{K}l_{n,\lambda})_{\lambda\in\Lambda_n}$ of sheaves of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$, with monodromy group structure

$$\begin{cases} \mathrm{SL}_n & : n \text{ odd} \\ \mathrm{Sp}_n & : n \text{ even}, \end{cases}$$

conductor bounded by $n+3$, and such that the trace function $t_{\mathcal{K}l_{n,\lambda},q}$ is equal to the Kloosterman sum $\mathrm{Kl}_{n,q}$ on $\mathbb{F}_q^\times$.

Proof. The construction of the Kloosterman sheaves is due to Deligne (see [Kat88] for the construction via recursive Fourier transforms). As already mentioned, the assertion on the integral monodromy groups over $\mathbb{F}_\lambda$ can be found in [PG18]. They form a compatible system for n fixed by Lemma 2.3 applied recursively. □

Remark 2.5. As an illustration of (6), note that $\mathrm{Kl}_{n,q} : \mathbb{F}_q \to \mathbb{Z}[\zeta_{4p}]_{q^{(n-1)/2}}$.

The following example, when unnormalized (hence replacing $\mathcal{O}_\lambda$ by $\mathbb{Z}_\ell$), was treated in [Kow06a] and [Kow08]:

Proposition 2.6 (Point counting on families of hyperelliptic curves)**.** *Let $f \in \mathbb{Z}[X]$ be a squarefree polynomial of degree $2g \geqslant 2$, and let Λ be the set of ℓ-adic valuations of $\mathcal{O}$ with $\ell \geqslant 3$. For p large enough, there exists a coherent family $(\mathcal{F}_{f,\lambda})_{\lambda\in\Lambda}$ of ℓ-adic sheaves of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$, with monodromy group structure Sp_{2g}, conductor depending only on f, and such that $t_{\mathcal{F}_{f,\lambda},q}(z)$ is given by (3) when $f(z) \neq 0$.*

Proof. For the construction, see [KS99, Section 10.1], and normalize by a Tate twist. Because of this normalization, [KS99, Theorem 10.1.16] and [KS99, Lemma 10.1.9] show that the arithmetic and geometric monodromy group preserve the same symplectic pairing. Finally, [Hal08, Theorem 1.2] shows that the geometric monodromy group is Sp_{2g}. □

2.3. **The large sieve for Frobenius.**

Theorem 2.7. *Let Λ be a set of valuations (or equivalently prime ideals) on $\mathcal{O}$. Given $L \geqslant 1$, we write Λ_L for the set of valuations in Λ corresponding to ideals of norm at most L. Let $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ be a coherent family, with monodromy group structure G, where $\widetilde{\mathcal{F}}_\lambda$ corresponds to a representation*

$$\rho_\lambda : \pi_{1,p} \to \mathrm{GL}_n(\mathcal{O}_\lambda) \to \mathrm{GL}_n(\mathbb{F}_\lambda).$$

For every $\lambda \in \Lambda$, let $\Omega_\lambda \subset G(\mathbb{F}_\lambda)$ be a conjugacy-invariant subset. Then, for all $L \geqslant 1$,

$$\frac{|\{x \in U_{\mathcal{F}_\lambda}(\mathbb{F}_q) : \rho_\lambda(\mathrm{Frob}_{x,q}) \notin \Omega_\lambda \text{ for all } \lambda \in \Lambda_L\}|}{q} \ll \left(1 + \frac{L^B}{q^{1/2}}\right) \frac{1}{P(L)},$$

where the implied constant depends only on the conductor of the family, and

$$P(L) = \sum_{\lambda\in\Lambda_L} \frac{|\Omega_\lambda|}{|G(\mathbb{F}_\lambda)|},\ B = \begin{cases} \frac{2n^2+n-1}{2} & : G = \mathrm{SL}_n \\ \frac{2n^2+3n+4}{4} & : G = \mathrm{Sp}_n \ (n \text{ even}). \end{cases} \tag{7}$$

Proof. This is a variant of [Kow06a, Proposition 3.3] (see also [Kow08, Chapter 8]). For $\lambda, \lambda' \in \Lambda$ distinct, the product map $\pi_{1,p} \to G(\mathbb{F}_\lambda) \times G(\mathbb{F}_{\lambda'})$ is surjective by [Kow06a, Corollary 2.6] (a variant of Goursat's Lemma), which extends with no modification to the case where $\mathbb{F}_\lambda$ and $\mathbb{F}_{\lambda'}$ do not necessarily have prime order (see [MT11, Part III]). By [MT11, Corollary 24.6], $B = 1 + \dim(G) + \mathrm{rank}(G)/2$. □

Remark 2.8. Note that in the case $E = \mathbb{Q}(\zeta_d)$ of the examples of Section 1.1, the size of the residue field $\mathbb{F}_\lambda$ corresponding to a prime ideal $\mathfrak{l} \trianglelefteq \mathbb{Z}[\zeta_d]$ depends on the multiplicative order modulo d of the prime ℓ above which $\mathfrak{l}$ lies (see [Was97, Theorem 2.13]). In particular, if $d = 4p$, then $|\mathbb{F}_\lambda|$ depends on p. This is a new phenomenon compared to the degree 1 case (i.e. $\mathcal{O}_\lambda = \mathbb{Z}_\ell$) studied in [Kow06a] and [Kow08].

Remark 2.9. The case of orthogonal monodromy group structures (that would appear in some variants of the examples in Section 5) is excluded in the definition of a coherent family, because the argument above does not apply in general: see the remark after [Kow06a, Corollary 2.6]. A similar difficulty arises in Theorem 5.3 later on: see Remark 5.4(2).

3. Traces of random matrices and Gaussian sums

In the next section, we will apply Theorem 2.7 to $\Omega_\lambda = \{g \in G(\mathbb{F}_\lambda) : \mathrm{tr}(g) \notin A_\lambda\}$, for some $A_\lambda \subset \mathbb{F}_\lambda$. In this section, we get estimates on the densities

$$P\left(\mathrm{tr}(g) \notin A_\lambda\right) := \frac{|\Omega_\lambda|}{|G(\mathbb{F}_\lambda)|}.$$

By the orthogonality relations in $\mathbb{F}_\lambda$, we get the following:

Proposition 3.1. *Let $G \leqslant \mathrm{GL}_n(\mathbb{F}_\lambda)$ be a subgroup and $A \subset \mathbb{F}_\lambda$. Then*

$$\begin{aligned} P\left(\mathrm{tr}(g) \in A\right) &:= \frac{1}{|G|}\sum_{g\in G} 1_A(\mathrm{tr}(g)) \\ &= \frac{|A|}{|\mathbb{F}_\lambda|} + O\left(\max_{1\neq\psi\in\hat{\mathbb{F}}_\lambda} \left|\frac{1}{|G|}\sum_{g\in G}\psi(\mathrm{tr}(g))\right| \left|\sum_{x\in A}\psi(-x)\right|\right). \end{aligned}$$

We expect, for nontrivial $\psi \in \hat{\mathbb{F}}_\lambda$,

$$\frac{1}{|G|}\sum_{g\in G}\psi(\mathrm{tr}(g)) \ll |\mathbb{F}_\lambda|^{-\alpha(G)} \tag{8}$$

for some $\alpha(G) > 0$, and similarly, if A is "well-distributed" in $\mathbb{F}_\lambda$, we expect

$$\frac{1}{|A|}\sum_{x\in A}\psi(x) \ll |\mathbb{F}_\lambda|^{-\alpha(A)} \tag{9}$$

for some $\alpha(A) > 0$. In both cases, the bounds should be uniform with respect to all nontrivial $\psi \in \hat{\mathbb{F}}_\lambda$.

Under (8) and (9), Proposition 3.1 becomes

$$P\left(\mathrm{tr}(g) \in A\right) = \frac{|A|}{|\mathbb{F}_\lambda|}\left(1 + O\left(|\mathbb{F}_\lambda|^{-\alpha(G)-\alpha(A)+1}\right)\right). \tag{10}$$

3.1. **Gaussian sums in linear groups** (8)**.**

3.1.1. *General result.* We start by a result that applies more generally to algebraic varieties in GL_n.

Proposition 3.2. *Let $V = \boldsymbol{V}(\mathbb{F}_\lambda)$ for $\boldsymbol{V} \subset \mathrm{GL}_n$ an algebraic variety over $\mathbb{F}_\lambda$. The bound (8) holds with $\alpha(V) = 1/2$, uniformly for all nontrivial $\psi \in \hat{\mathbb{F}}_\lambda$, unless $\mathrm{tr} : V \to \mathbb{F}_\lambda$ is constant.*

Proof. Let $\ell' \neq \mathrm{char}(\mathbb{F}_\lambda)$ be an auxiliary prime and let us consider the restriction $\mathcal{L}$ of the Lang torsor $\mathcal{L}_{\psi\circ\mathrm{tr}}$ on $\mathbb{A}^{n^2}/\mathbb{F}_\lambda$ to $\boldsymbol{V}$ (see [KR15, Example 7.17]), as sheaf of $\mathbb{Q}_{\ell'}$-modules. By the Grothendieck–Lefschetz trace formula,

$$\sum_{g\in V}\psi(\mathrm{tr}(g)) = \sum_{i=0}^{2\dim \boldsymbol{V}}(-1)^i \,\mathrm{tr}\left(\mathrm{Frob}_{\mathbb{F}_\lambda} \mid H_c^i(\boldsymbol{V}\times\overline{\mathbb{F}}_\lambda, \mathcal{L})\right).$$

By Deligne's generalization of the Riemann hypothesis over finite fields [Del80],

$$\mathrm{tr}\left(\mathrm{Frob}_{\mathbb{F}_\lambda} \mid H_c^i(\boldsymbol{V}\times\overline{\mathbb{F}}_\lambda, \mathcal{L})\right) \leqslant |\mathbb{F}_\lambda|^{i/2}\dim H_c^i(\boldsymbol{V}\times\overline{\mathbb{F}}_\lambda, \mathcal{L})$$

for $0 \leqslant i \leqslant 2\dim \boldsymbol{V}$, and by the coinvariant formula,

$$\mathrm{tr}\left(\mathrm{Frob}_{\mathbb{F}_\lambda} \mid H_c^{2\dim \boldsymbol{V}}(\boldsymbol{V}\times\overline{\mathbb{F}}_\lambda, \mathcal{L})\right) = 0$$

unless $\mathcal{L}$ is geometrically trivial, in which case $\mathrm{tr} : V \to \mathbb{F}_\lambda$ would be constant. Hence

$$\left|\sum_{g\in V}\psi(\mathrm{tr}(g))\right| \leqslant |\mathbb{F}_\lambda|^{\dim \boldsymbol{V}-1/2}\sum_{i=0}^{2\dim \boldsymbol{V}-1}\dim H_c^i(\boldsymbol{V}\times\overline{\mathbb{F}}_\lambda, \mathcal{L}).$$

By [Kat01, Theorem 12], we find that

$$\left|\sum_{g\in V}\psi(\mathrm{tr}(g))\right| \leqslant 3|\mathbb{F}_\lambda|^{\dim \boldsymbol{V}-1/2}(2+d)^{n^2+r}$$

if $\boldsymbol{V}$ is defined by r polynomials of degree at most d. The conclusion follows by [MT11, Corollary 24.6]. □

3.1.2. *Classical finite groups of Lie type.* Using the Bruhat decomposition, D.S. Kim actually explicitly evaluated the Gaussian sums (8) for classical finite groups of Lie type (see e.g. [Kim97, Kim98]). The expressions involve hyper-Kloosterman sums, and applying Deligne's bound yields the following, which greatly improves Proposition 3.2, in particular as n grows:

Proposition 3.3. *For $n \geqslant 1$ and $G = \mathrm{GL}_n(\mathbb{F}_\lambda)$, $\mathrm{SL}_n(\mathbb{F}_\lambda)$, $\mathrm{Sp}_{2n}(\mathbb{F}_\lambda)$, $\mathrm{SO}^{\pm}_{2n}(\mathbb{F}_\lambda)$ and $\mathrm{SO}_{2n+1}(\mathbb{F}_\lambda)$, the bound (8) holds with $\alpha(G) \geqslant 1$ given in Table 1.*

G	$\alpha(G)$
GL_n	$\frac{n(n-1)}{2}$
SL_n	$\frac{n^2-1}{2}$
Sp_n, SO_n^- (n even)	$\frac{n(n+2)}{8}$
SO_n (n odd)	$\frac{n^2-1}{8}$
SO_n^+ (n even)	$\frac{n(n-2)}{8}$

TABLE 1. Cancellation for Gaussian sums over finite groups of Lie type.

Proof. See [PG17a, Proposition 6.28]. □

3.2. **Gaussian sums in $\mathbb{F}_\lambda$.** Let us now consider Bound (9) for various subsets $A \subset \mathbb{F}_\lambda$.

3.2.1. *Squares.* Let $A = \mathbb{F}_\ell^{\times 2}$ be the subgroup of squares in $\mathbb{F}_\ell^\times$ with $\ell > 2$. Using the Legendre symbol and the evaluation of quadratic Gauss sums, we get that (9) holds with $\alpha(A) = 1/2$, uniformly for all nontrivial $\psi \in \hat{\mathbb{F}}_\ell$, corresponding to square-root cancellation since $|A| = (\ell-1)/2$.

3.2.2. *Powers/Multiplicative subgroups.* More generally, we have:

Proposition 3.4. *For $\alpha \in (0, 1/2)$, Bound (9) holds for any subgroup $H \leqslant \mathbb{F}_\lambda^\times$ such that $|H| \geqslant |\mathbb{F}_\lambda|^{1/2+\alpha}$, uniformly for all nontrivial $\psi \in \hat{\mathbb{F}}_\lambda$.*

Proof. This follows for example from the bound $\sum_{x\in H} \psi(x) \ll |\mathbb{F}_\lambda|^{1/2}$ that is deduced from Deligne's extension of the Riemann hypothesis over finite fields (see [PG17a, Proposition 5.7]). □

Example 3.5. For $m \geqslant 2$ fixed and $H = \mathbb{F}_\lambda^{\times m}$ the subgroup of mth powers, the condition $|H| \geqslant |\mathbb{F}_\lambda|^{1/2+\alpha}$ holds as soon as $|\mathbb{F}_\lambda|$ is large enough, since $|H| = \frac{|\mathbb{F}_\lambda|-1}{(m,|\mathbb{F}_\lambda|-1)}$.

Remark 3.6. When $|H|$ is arbitrarily small (say $|H| \geqslant |\mathbb{F}_\lambda|^\delta$ for some $\delta > 0$), the works of Bourgain and others (see e.g. [BC06]) give (9) for some $\alpha = \alpha(\delta) > 0$, up to some necessary restrictions if $\delta \leqslant 1/2$ and $\mathbb{F}_\lambda \neq \mathbb{F}_\ell$.

3.2.3. *Definable subsets.* For R a ring and $\varphi(x)$ a first-order formula in one variable in the language of rings, we define $\varphi(R) = \{a \in R : \varphi(a) \text{ holds}\}$.

Example 3.7. For $\varphi(x) = (\exists y : x = y^2)$, the set $\varphi(R)$ is the subset of squares, as in the previous section. More generally, we can take $\varphi(x) = (\exists y : x = f(y))$ for any polynomial $f \in \mathbb{Z}[Y]$.

We recall:

Theorem 3.8 (Chatzidakis–van den Dries–Macintyre [CvdDM92])**.** *For every formula $\varphi(x)$ in one variable in the language of rings, there exists a finite*

set $C(\varphi) \subset (0,1] \cap \mathbb{Q}$ such that for every finite field $\mathbb{F}_\lambda$,

$$|\varphi(\mathbb{F}_\lambda)| = C(\lambda,\varphi)|\mathbb{F}_\lambda| + O_\varphi(|\mathbb{F}_\lambda|^{1/2}) \quad (11)$$
$$\text{with } C(\lambda,\varphi) \in C(\varphi), \text{ or}$$
$$|\varphi(\mathbb{F}_\lambda)| \ll_\varphi |\mathbb{F}_\lambda|^{-1/2}. \quad (12)$$

The implied constants depend only on φ.

The following combined with Theorem 3.8 shows that Gaussian sums over definable subsets exhibit square-root cancellation:

Theorem 3.9 ([Kow07, Theorem 1, Corollary 12, Remark 19])**.** *Let $\varphi(x)$ be a formula in one variable in the language of rings such that $|\varphi(\mathbb{F}_\lambda)|$ is not bounded as $|\mathbb{F}_\lambda| \to +\infty$. Then, if $\psi \in \hat{\mathbb{F}}_\lambda$ is nontrivial, the bound (9) for $A = \varphi(\mathbb{F}_\lambda)$ holds with $\alpha(A) = 1/2$, with an implied constant depending only on φ.*

3.2.4. *Images of polynomials.* When $\varphi(x) = (\exists y : x = f(y))$ for some polynomial $f \in \mathbb{Z}[X]$, Theorem 3.8 also appears in [BSD59] (using the Weil conjectures for curves).

Proposition 3.10 ([BSD59, Theorem 1, Lemma 1])**.** *Let $f \in \mathbb{Z}[X]$ be of degree $d \geqslant 2$ and such that the Galois group of $f(X) - y \in \mathbb{C}(y)[X]$ over $\mathbb{C}(y)$ is equal to $\mathfrak{S}_d$. Then (11) for $\varphi(x) = (\exists y : x = f(y))$ and a finite field $\mathbb{F}_\lambda$ of characteristic $\ell \gg_f 1$ holds with*

$$C(\lambda,\varphi) = \sum_{n=1}^{\deg(f)} \frac{(-1)^{n+1}}{n!} \in (0,1).$$

This is extended to $f \in \mathbb{F}_\lambda(X)$ in [Coh70].

Remarks 3.11. (1) See [BSD59, p. 422] for sufficient conditions to verify the hypothesis of Proposition 3.10.

(2) By [vdW34] or [Gal73], the hypothesis of Proposition 3.10 holds for almost all monic $f \in \mathbb{Z}[X]$ of degree $d \geqslant 2$, with respect to the terminology of Footnote 1, p. 3.

4. Zero-density estimates for trace functions in algebraic subsets

We continue to fix a number field E with ring of integers $\mathcal{O}$.

4.1. General result.

Proposition 4.1. *Let Λ be a set of valuations on $\mathcal{O}$ and let $t : \mathbb{F}_q \to E$ be the trace function over $\mathbb{F}_q$ associated to a coherent family $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ of sheaves of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$, with monodromy group structure G. For $A \subset E$ and $\lambda \in \Lambda$ corresponding to a prime ideal $\mathfrak{l}$ of $\mathcal{O}$, we denote by $A_\lambda \subset \mathbb{F}_\lambda$ the*

reduction of $A \cap \mathcal{O}_\mathfrak{l}$ modulo $\mathfrak{l}$. Assume that

$$\sup_{\lambda\in\Lambda} \frac{|A_\lambda|}{|\mathbb{F}_\lambda|} < 1. \tag{13}$$

Then

$$P\big(t(x) \in A\big) \ll \frac{1}{|\Lambda_L|} \text{ with } L = \left\lfloor q^{\frac{1}{2B}} \right\rfloor, \tag{14}$$

where $B > 0$ is as in Theorem 2.7, with an implied constant depending only on the conductor of the family and on the left-hand side of (13).

Proof. For every $\lambda \in \Lambda$, we may reduce $t : \mathbb{F}_q \to \mathcal{O}_\mathfrak{l} \leqslant \mathcal{O}_\lambda$ to $\tilde{t} : \mathbb{F}_q \to \mathbb{F}_\lambda$, so that

$$P(t(x) \in A) \leqslant \frac{|\{x \in \mathbb{F}_q : \tilde{t}(x) \in A_\lambda \text{ for all } \lambda \in \Lambda_L\}|}{q}.$$

By Theorem 2.7 with

$$\Omega_\lambda = \{g \in G(\mathbb{F}_\lambda) : \operatorname{tr} g \notin A_\lambda\} \quad (\lambda \in \Lambda),$$

which are clearly conjugacy-invariant, we get

$$P(t(x) \in A) \ll \left(1 + \frac{L^B}{q^{1/2}}\right) \frac{1}{P(L)},$$

where $P(L) = \sum_{\lambda\in\Lambda_L} P(\operatorname{tr}(g) \notin A_\lambda)$. By (10) (Proposition 3.1),

$$P(\operatorname{tr}(g) \in A_\lambda) = \frac{|A_\lambda|}{|\mathbb{F}_\lambda|}\left(1 + O\left(\frac{1}{|\mathbb{F}_\lambda|^{\alpha(G)+\alpha(A_\lambda)-1}}\right)\right) \ll \frac{|A_\lambda|}{|\mathbb{F}_\lambda|},$$

since $\alpha(G) \geqslant 1$ by Proposition 3.3. Therefore, we get that for any $L \geqslant 1$,

$$P\big(t(x) \in A\big) \ll \left(1 + \frac{L^B}{q^{1/2}}\right) |\Lambda_L|^{-1} \left(1 - \max_{\lambda\in\Lambda_L} \frac{|A_\lambda|}{|\mathbb{F}_\lambda|}\right)^{-1}.$$

□

Remark 4.2. If we assume more generally that the monodromy group of $\widetilde{\mathcal{F}}_\lambda$ is $\boldsymbol{G}(\mathbb{F}_\lambda)$ for $\boldsymbol{G} \leqslant \mathrm{GL}_n$ any linear group over $\mathbb{F}_\lambda$, the results hold if $\alpha(A_\lambda) \geqslant 1/2$ for all $\lambda \in \Lambda$, by Proposition 3.2. Interestingly, in the case of SL_n, Sp_{2n} and $\mathrm{SO}_n^\pm$, Proposition 3.3 gives much more cancellation, so that we do not need information about the $\alpha(A_\lambda)$.

To apply Proposition 4.1, we need the local densities assumption (13) and lower bounds on $|\Lambda_L|$. We treat these aspects in the following subsections.

4.2. **Lower bounds on $|\Lambda_L|$.** For our applications, we will mainly consider Λ to be either:

Examples 4.3.
(1) The full set $\Lambda_{0,p}$ of valuations on $\mathcal{O}$ not lying above the p-adic valuation.
(2) For $m \geqslant 2$ and $C \subset (\mathbb{Z}/m)^\times$, the set of valuations $\lambda \in \Lambda_{0,p}$ such that $|\mathbb{F}_\lambda| \in C$.
(3) The restriction of these to ideals having degree 1 over $\mathbb{Q}$.

More generally, let F/E be a fixed finite Galois extension of number fields with Galois group H, $C \subset H$ be a conjugacy-stable subset, and

$$\begin{aligned} \Lambda(C) &= \{\mathfrak{l} \trianglelefteq E \text{ prime, not ramified in } F : \mathrm{Frob}_{\mathfrak{l}} \in C\} \\ \Lambda_1(C) &= \{\mathfrak{l} \in \Lambda(C) \text{ of degree 1 over } \mathbb{Q}\}. \end{aligned} \tag{15}$$

Example 4.3 (1) then corresponds to $E = F$, while (2) corresponds to $F = E(\zeta_m)$ with $H \cong (\mathbb{Z}/m)^\times$.

By Chebotarev's density theorem, if E and F are fixed,

$$|\Lambda(C)_L| \geqslant |\Lambda_1(C)_L| \gg \frac{|C|}{|H|}\frac{L}{\log L} \quad (L \to +\infty) \tag{16}$$

with an absolute implied constant. Hence, if F and E do not depend on p, (14) is

$$P\big(t(x) \in A\big) \ll_{C,H} \frac{\log q}{Bq^{1/(2B)}} \to 0 \quad (q = p^e \to +\infty). \tag{17}$$

If E and/or F depend on p (e.g. for Kloosterman sums, where $E = \mathbb{Q}(\zeta_{4p})$), we must either fix p or deal with uniformity with respect to E and F. We discuss this situation in the following paragraphs.

4.2.1. *Uniformity in the prime ideal theorem.* By [Fri80] (extending Chebychev's method to number fields), if $E/\mathbb{Q}$ is normal[4], then

$$\pi_E(L) = |\{\mathfrak{l} \trianglelefteq E \text{ prime} : N(\mathfrak{l}) \leqslant L\}| \gg_\varepsilon \frac{L}{\log(2L)^{1+\varepsilon}\Delta_E^{1/2+\varepsilon}}$$

for $\Delta_E = |\operatorname{disc}_{\mathbb{Q}}(E)|$, and any $\varepsilon > 0$ if $n_E = [E : \mathbb{Q}] \gg_\varepsilon 1$. This is nontrivial only when $L \gg \Delta_E^{1/2+\varepsilon'}$ for some $\varepsilon' > 0$.

4.2.2. *Uniformity in Chebotarev's density theorem.* The unconditional results due to Lagarias–Odlyzko and Serre (see [Ser81, Section 2.2]) show that (16) holds with an absolute implied constant under the restriction $\log L \gg n_E(\log \Delta_E)^2$.

Assuming the generalized Riemann hypothesis (GRH) for the Dedekind zeta function of E, this range can be improved to $L \gg (\log \Delta_E)^{2+\varepsilon}$ for an arbitrary $\varepsilon > 0$ (see [Ser81, Section 2.4]).

4.2.3. *Cyclotomic fields.* If $E = \mathbb{Q}(\zeta_d)$, $F = E(\zeta_m)$ are cyclotomic fields, it is possible to improve the unconditional uniform range in Chebotarev's density theorem by relying on estimates for primes in arithmetic progressions.

Proposition 4.4. *For $d, m \geqslant 1$ coprime integers, let $E = \mathbb{Q}(\zeta_d)$ and $F = E(\zeta_m)$. For $C \subset \mathrm{Gal}(F/E) \cong (\mathbb{Z}/m)^\times$, we have*

$$|\Lambda(C)_L| \geqslant |\Lambda_1(C)_L| \gg \frac{|C|L}{(dm)^\varepsilon \varphi(m)\log L}$$

when either:

(1) $\varepsilon > 0$ and $L \geqslant (dm)^8$, or

(2) under GRH, $\varepsilon = 0$ and $L \geqslant (dm)^{2+\varepsilon'}$ for some $\varepsilon' > 0$.

[4] In Friedlander's paper, it is only assumed that E is in a tower of normal extensions. If $E/\mathbb{Q}$ is itself normal, we can improve the result by using more a precise version of Stark's estimates [Sta74] on the residue at 1 of the Dedekind zeta function of E.

Proof. Since every unramified rational prime ℓ of inertia/residual degree f_ℓ (equal to the order of ℓ in $(\mathbb{Z}/d)^\times$) gives rise to $\varphi(d)/f_\ell$ prime ideals with norm ℓ^{f_ℓ},

$$|\Lambda(C)_L| \quad = \quad \varphi(d) \sum_{f|\varphi(d)} \frac{|\{\ell \leqslant L^{1/f} \text{ prime } : \ell \nmid \Delta_E,\ f_\ell = f,\ \ell^f \in C\}|}{f}.$$

The summand with $f = 1$ gives:

$$|\Lambda_1(C)_L| \quad \geqslant \quad \varphi(d)|\{\ell \leqslant L \text{ prime } : \ell \nmid \Delta_E,\ \ell \equiv 1 \pmod d,\ \ell \in C\}|.$$

If $(d, m) = 1$, then by the Chinese remainder theorem

$$|\Lambda_1(C)_L| \geqslant \varphi(d) \left[\sum_{c \in C} \pi(c, dm, L) - \omega(d) \right],$$

where $\pi(a, d, L) = |\{\ell \leqslant L \text{ prime } : \ell \equiv a \pmod d\}|$ for $a \in (\mathbb{Z}/d)^\times$. Uniformly, one has

$$\pi(a, d, L) \gg \frac{L}{\varphi(d) d^\varepsilon \log L} \tag{18}$$

under (1) (by [May13, Theorem 3.3], using Linnik-type arguments) or (2) assuming GRH. □

Remark 4.5. Similarly, this shows that for a Galois extension $E/\mathbb{Q}$, the set of prime ideals with inertia degree 1 has natural density 1, so we cannot hope to substantially improve the lower bound by taking into account the $f > 1$ in the proof of Proposition 4.4.

Remarks 4.6.
(1) By the Bombieri–Vinogradov theorem, the range (2) in (18) holds unconditionally for all a on average over d.
(2) By a conjecture of Montgomery, one may be able to take $\varepsilon = 0$ and $L \gg (dm)^{1+\delta}$ for any $\delta > 0$. By Barban–Davenport–Halberstam, Montgomery, and Hooley, this holds true in (18) on average over d and a.

4.3. **Explicit zero-density estimates.** The results from the previous section along with Proposition 4.1 give:

Proposition 4.7. *Under the hypotheses of Proposition 4.1 and (13), with $E/\mathbb{Q}$ normal, F/E a finite Galois extension with Galois group H, a conjugacy-invariant subset $C \subset H$ and $\Lambda = \Lambda(C)$ or $\Lambda_1(C)$ as in (15), we have that for any $\varepsilon > 0$:*

(1) If $F = E$ is normal,

$$P\big(t(x) \in A\big) \ll_\varepsilon \frac{\Delta_E^{1/2+\varepsilon} (\log q)^{1+\varepsilon}}{B^{1+\varepsilon} q^{1/(2B)}},$$

which is nontrivial when $\Delta_E^{B+\varepsilon'} = o(q)$ for some $\varepsilon' > 0$.

(2) Under GRH, if $q \geqslant (\log \Delta_E)^{2B+\varepsilon}$,

$$P\big(t(x) \in A\big) \ll_\varepsilon \frac{m \log q}{|C| B q^{1/(2B)}} \ll_{m,C} \frac{\log q}{B q^{1/(2B)}}.$$

(3) Assume that $E = \mathbb{Q}(\zeta_d)$ and $F = E(\zeta_m)$ with $(d, m) = 1$. If $q \geqslant (dm)^{16B}$, then

$$P\big(t(x) \in A\big) \ll_\varepsilon \frac{m(dm)^\varepsilon \log q}{|C|Bq^{1/(2B)}} \ll_{m,C} \frac{d^\varepsilon \log q}{Bq^{1/(2B)}}.$$

The implied constants depend only on the conductor of the family and the quantities indicated.

4.3.1. *The case $E = \mathbb{Q}(\zeta_{4p})$.* For exponential sums, we are interested in the case $E = \mathbb{Q}(\zeta_{4p})$, where $n_E = 2(p-1)$ and $\Delta_E = 4^{2p-3}p^{2(p-2)}$.

The restrictions $q \gg g(E)$ (for some $g(E) = g(n_E, \Delta_E) \geqslant 1$) of Proposition 4.7 impose limitations on the range of e, p when $q = p^e \to +\infty$:

Corollary 4.8. *Under the hypotheses of Proposition 4.7 for $E = \mathbb{Q}(\zeta_{4p})$ and $F = E(\zeta_m)$ with $(m, 4p) = 1$, we have*

$$P\big(t(x) \in A\big) \ll_\varepsilon \frac{m(pm)^\varepsilon \log q}{|C|Bq^{1/(2B)}} \ll_{m,C} \frac{p^\varepsilon \log q}{Bq^{1/(2B)}} \to 0 \quad (q = p^e \to +\infty)$$

when either

(1) $\varepsilon > 0$ and $e \geqslant 16B$, or (2) under GRH, $\varepsilon = 0$ and $e > 4B$.

The implied constants depend only on the conductor of the family and the quantities indicated.

Remarks 4.9. (1) Had we not taken advantage of the fact that E is a cyclotomic field, the unconditional results mentioned in Section 4.2.2 would have forced to take $q = p^e \to +\infty$ with $e \gg p$.

(2) Under Montgomery's conjecture mentioned in Remarks 4.6, we may take $\varepsilon = 0$ and $e > 2B$. Without an improvement in the error term of the large sieve bound (14), $e = 2B + 1 \geqslant 10$ is the minimal value the method could handle.

4.4. **Local densities.** In this section, we finally give examples of sets $A \subset E$ for which the local densities assumption (13) holds.

4.4.1. *Powers/finite index subgroups.*

Proposition 4.10. *Let E, $\mathcal{O}$ be as in Proposition 4.1 and for $m \geqslant 2$, let*

$$\Lambda = \begin{cases} \{\lambda \in \Lambda_{0,p} : |\mathbb{F}_\lambda| \equiv 1 \pmod m\} & : m \text{ odd} \\ \{\lambda \in \Lambda_{0,p} : \text{not lying above } 2\} & : m \text{ even.} \end{cases}$$

Then (13) holds for $A = E^m \subset E$.

Proof. We have $A_\lambda = \mathbb{F}_\lambda^m$ and for $|\mathbb{F}_\lambda| \geqslant 3$,

$$\frac{|A_\lambda|}{|\mathbb{F}_\lambda|} = \left(1 - \frac{1}{|\mathbb{F}_\lambda|}\right) \frac{1}{(|\mathbb{F}_\lambda^\times|, m)} + \frac{1}{|\mathbb{F}_\lambda|} \ll \begin{cases} \frac{1}{m} + \frac{1}{|\mathbb{F}_\lambda|} & : m \text{ odd} \\ \frac{1}{2} + \frac{1}{|\mathbb{F}_\lambda|} & : m \text{ even.} \end{cases}$$

□

Note that the set Λ in Proposition 4.10 is of the form given in Example 4.3 (2).

4.4.2. *Definable subsets.*

Proposition 4.11. *Let E, $\mathcal{O}$ and Λ be as in Proposition 4.1 and let $\varphi(x)$ be a first order formula in one variable in the language of rings such that:*

(1) Neither $|\varphi(\mathbb{F}_\lambda)|$ nor $|\neg\varphi(\mathbb{F}_\lambda)|$ are bounded as $|\mathbb{F}_\lambda| \to +\infty$, where $\neg$ denotes negation.

(2) For every $\lambda \in \Lambda$ corresponding to an ideal $\mathfrak{l}$, $\varphi(E) \cap \mathcal{O}_\mathfrak{l} \pmod{\mathfrak{l}}$ is contained in $\varphi(\mathbb{F}_\lambda)$.

Then (13) holds with $A = \varphi(E) \subset E$.

Proof. Condition (2) implies that $A_\lambda \subset \varphi(\mathbb{F}_\lambda)$ for all $\lambda \in \Lambda$. Under condition (1), Theorem 3.8 shows that

$$\begin{aligned} |\varphi(|\mathbb{F}_\lambda|)| &= C_{\lambda,\varphi}|\mathbb{F}_\lambda|(1+o(1)) \\ |\neg\varphi(|\mathbb{F}_\lambda|)| &= C_{\lambda,\neg\varphi}|\mathbb{F}_\lambda|(1+o(1)) \\ &= (1-C_{\lambda,\varphi})|\mathbb{F}_\lambda|(1+o(1)) \end{aligned}$$

with $C_{\lambda,\varphi}, C_{\lambda,\neg\varphi} \in (0,1]$. Hence, $C_{\lambda,\varphi} \neq 0,1$ for $|\mathbb{F}_\lambda|$ large enough, and $\limsup_{|\mathbb{F}_\lambda|\to+\infty} \frac{|A_\lambda|}{|\mathbb{F}_\lambda|} \leqslant \limsup_{|\mathbb{F}_\lambda|\to+\infty} \frac{|\varphi(\mathbb{F}_\lambda)|}{|\mathbb{F}_\lambda|} \leqslant \max C(\varphi) < 1$, recalling that $C(\varphi)$ is finite. □

Remark 4.12. Condition (2) of Proposition 4.11 holds if both

$$\text{(a) } \varphi(E) \cap \mathcal{O}_\mathfrak{l} \subset \varphi(\mathcal{O}_\mathfrak{l}), \qquad \text{and} \qquad \text{(b) } \varphi(\mathcal{O}_\mathfrak{l}) \pmod{\mathfrak{l}} \subset \varphi(\mathbb{F}_\lambda)$$

hold. Note that:

- Condition (a) holds when $\mathrm{char}(\mathbb{F}_\lambda) \gg_f 1$ if $\varphi(x) = (\exists y : x = f(y))$ for some $f \in \mathbb{Z}[X]$. Indeed, for $x \in E$, we have $\lambda(f(x)) = \min(0, \deg(f)\lambda(x))$ if no coefficient of f is divisible by $\mathrm{char}(\mathbb{F}_\lambda)$.
- Condition (b) holds if φ contains no negations or implications. On the other hand, for $\varphi(x) = \neg(\exists y : x = y^2)$, the reduction of a nonsquare in $\mathcal{O}$ may be a square in $\mathbb{F}_\lambda$.

Example 4.13 (Images of polynomials). Consider the case $\varphi(x) = (\exists y : x = f(y))$ for $f \in \mathbb{Z}[X]$ of Section 3.2.4. Then Proposition 4.11 applies for

- almost all monic f of fixed degree $d \geqslant 2$ (with respect to the terminology of Footnote 1, p. 3), and
- all f satisfying the Galois group condition of Proposition 3.10,

up to restricting to a cofinite subset of Λ. Indeed:

- By Proposition 3.10 and Remarks 3.11, Condition (1) of Proposition 4.11 holds for almost all monic f of fixed degree.
- Condition (2) holds if $\mathrm{char}(\mathbb{F}_\lambda) \gg_f 1$ by Remark 4.12.

5. Examples

5.1. **Kloosterman sums.** Proposition 1.1, given in the introduction, now follows directly from Corollary 4.8 with Proposition 2.4 and the local densities estimates from Proposition 4.10.

Similarly, replacing the latter with Proposition 4.11, we obtain:

Proposition 5.1. *Let $\varphi(x)$ be a first-order formula in the language of rings as in Proposition 4.11. Then, for $n \geqslant 2$ and $\varepsilon > 0$,*

$$P\Big(\mathrm{Kl}_{n,q}(x) \in \varphi(\mathbb{Q}(\zeta_{4p}))\Big) \ll_{n,\varphi,\varepsilon} \frac{p^{\varepsilon}\log q}{B_n q^{1/(2B_n)}} \to 0 \tag{19}$$

when $q = p^e \to +\infty$ coprime to n with $e \geqslant 16B_n$, for B_n as in (5). The implied constant depends only on n, φ and ε.

Proposition 1.2 is a particular case of the latter, using Example 4.13.

5.1.1. *Results for unnormalized sums.* Replacing A by $q^{(n-1)/2}A$ in Proposition 4.1 and using uniformity shows that the above results also hold for unnormalized Kloosterman sums.

5.1.2. *Galois actions.* When considering densities of the form (19), it is interesting to take into account the following Galois actions:

(1) For all $\sigma \in \mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p) \cong \mathbb{Z}/e$ and $x \in \mathbb{F}_q^\times$,
$$\mathrm{Kl}_{n,q}(x) = \mathrm{Kl}_{n,q}(\sigma(x)).$$
The orbit of x has size $\deg(x) \in \{1,\dots,e\}$. Fisher [Fis92, Corollary 4.25] has actually shown that if $p > (2n^{2e}+1)^2$, the Kloosterman sums are distinct up to this action.

(2) For $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \cong \mathbb{F}_p^\times$ corresponding to $c \in \mathbb{F}_p^\times$ and $x \in \mathbb{F}_q^\times$, we have
$$\sigma(\mathrm{Kl}_{n,q}(x)) = \mathrm{Kl}_{n,q}(c^n x).$$
Moreover, orbits have size $\frac{p-1}{(p-1,n)} \in \{(p-1)/n,\dots,p-1\}$.

If φ is a first-order formula in the language of rings, let $A_p = \varphi(\mathbb{Q}(\zeta_{4p}))$. Since $\sigma(A_p) = A_p$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$, we can define an equivalence relation $\sim$ on $\{x \in \mathbb{F}_q^\times : \mathrm{Kl}_{n,q}(x) \in A_p\}$ generated by $x \sim c^n x$ for all $c \in \mathbb{F}_p^\times$, $x \in \mathbb{F}_q^\times$, and we have

$$\begin{aligned}|\{x \in \mathbb{F}_q^\times : \mathrm{Kl}_{n,q}(x) \in A_p\}/\sim| &= \frac{|\{x \in \mathbb{F}_q^\times : \mathrm{Kl}_{n,q}(x) \in A_p\}|(p-1,n)}{p-1} \\ &\ll_n \frac{|\{x \in \mathbb{F}_q^\times : \mathrm{Kl}_{n,q}(x) \in A_p\}|}{p-1}.\end{aligned}$$

If in addition the hypotheses of Proposition 5.1 are satisfied, this yields

$$|\{x \in \mathbb{F}_q^\times : \mathrm{Kl}_{n,q}(x) \in A_p\}/\sim| \ll_{n,\varepsilon} \frac{q^{1-1/(2B_n)}\log q}{p^{1-\varepsilon}}.$$

Remark 5.2. The right-hand side can tend to 0 with $p \to +\infty$ only when $e < \frac{2B_n}{2B_n-1}$. Since $\frac{2B_n}{2B_n-1} \in (1,2)$, this is the case only for $e = 1$. Unfortunately, our estimate on the number of prime ideals of bounded norm in $\mathbb{Q}(\zeta_{4p})$ requires to take $e \gg 1$. If it could be extended to $e = 1$ (but see Remarks 4.9 (2)), the above would show that for p large enough, there is no $x \in \mathbb{F}_p^\times$ such that $\mathrm{Kl}_{n,p}(x) \in \varphi(\mathbb{Q}(\zeta_{4p}))$.

5.2. **Exploiting monodromy over $\mathbb{C}$.** As we mentioned in the previous section, determining integral monodromy groups (as required by Definition 2.1 (2)), say for a subset of valuations of density 1, is usually difficult.

By using some deep results of Larsen and Pink (relying in particular on the classification of finite simple groups in [Lar95]), the following result allows to obtain coherent families from the knowledge of the monodromy groups over $\overline{\mathbb{Q}}_\ell$, up to passing to a subfamily of density 1 depending on p.

Theorem 5.3. *Let $E \subset \mathbb{C}$ be a Galois number field with ring of integers $\mathcal{O}$ and let Λ be a set of valuations on $\mathcal{O}$ of natural density 1. Let $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ be a compatible system with $\mathcal{F}_\lambda$ a sheaf of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$. We assume that:*

(2') There exists $G \in \{\mathrm{SL}_n, \mathrm{Sp}_{2n}\}$ such that for every $\lambda \in \Lambda$, the arithmetic monodromy group of $\mathcal{F}_\lambda \otimes \overline{\mathbb{Q}}_\ell$ is conjugate to $G(\overline{\mathbb{Q}}_\ell)$.

Then there exists a subset $\Lambda_p \subset \Lambda$ of natural density 1, depending on p and on the family, such that $(\mathcal{F}_\lambda)_{\lambda\in\Lambda_p}$ is coherent, with monodromy group structure G.

After using Theorem 5.3, we may apply Proposition 4.1 with the coherent subfamily $(\mathcal{F}_\lambda)_{\lambda\in\Lambda_p}$ to get

$$P\big(t(x) \in A\big) \ll \frac{1}{|(\Lambda_p)_L|} \ll_p \frac{1}{|\Lambda_L|}, \tag{20}$$

when $L = \lfloor q^{1/(2B)} \rfloor \to +\infty$, with the implied constant depending on p and on the original family.

5.2.1. *Proof of Theorem 5.3.* The idea of the argument, based on [LP92] and [Lar95], is due to Katz and appears partly in [Kow06a, p. 29], [Kow06b, p. 7], [Kow08, pp. 188–189] (however see Remark 5.5 below), and [Kat12, Section 7].

To reduce as much as possible to the situation of [LP92] and [Lar95], we consider the subset $\Lambda_1 \subset \Lambda$ corresponding to ideals of degree 1 over $\mathbb{Q}$, so that $E_\lambda = \mathbb{Q}_\ell$, $\mathcal{O}_\lambda = \mathbb{Z}_\ell$ and $\mathbb{F}_\lambda = \mathbb{F}_\ell$ if $\lambda \in \Lambda_1$ is an ℓ-adic valuation. By [Jan05, 4.7.1], for any $S \subset \mathrm{Spec}(\mathcal{O})$, the Dirichlet density of S is equal to the Dirichlet density of the elements of S having degree 1 over $\mathbb{Q}$. In particular, Λ_1 has Dirichlet density 1, and actually natural density 1 by [Nar04, Corollary 2, p. 248] (for cyclotomic fields, see also the proof of Proposition 4.4).

In the notations of [Lar95, Section 3] and definitions of [LP92, Section 6], we have the compact F-group $\pi_{1,p}$ with compatible system of representations

$$\big(\rho_\lambda = \rho_{\mathcal{F}_\lambda} : \pi_{1,p} \to \mathrm{GL}_n(\mathcal{O}_\lambda) = \mathrm{GL}_n(\mathbb{Z}_\ell)\big)_{\lambda\in\Lambda_1}$$

and Frobenius Frob_α for $\alpha \in \mathcal{A} = \{(x, p^n) : n \geqslant 1, x \in \mathbb{F}_{p^n}\}$. Note that G is a simply connected reductive group scheme over $\mathbb{Z}$, and by hypothesis ρ_λ is semisimple.

For every $\lambda \in \Lambda$, we let $G_\lambda = G_{E_\lambda}$, $\Gamma_\lambda = \rho_\lambda(\pi_{1,p}) \leqslant G(\mathcal{O}_\lambda)$ the integral monodromy group, $\widetilde{\Gamma}_\lambda := \Gamma_\lambda \pmod{\lambda}$ its reduction, and

$$B = \{\lambda \in \Lambda : \widetilde{\Gamma}_\lambda \lneq G(\mathbb{F}_\lambda)\} \subset \Lambda$$

the set of valuations where the monodromy group is smaller than expected. We let moreover:

- For every $\alpha \in \mathcal{A}$,
$$\xi(\alpha) \in \mathcal{O}_\lambda[T] \leqslant E[T]$$
the characteristic polynomial of $\rho_\lambda(\mathrm{Frob}_\alpha)$ (which does not depend on $\lambda \in \Lambda$ by hypothesis).
- $K \subset \xi(G_\mathbb{Q})$ the $\mathbb{Q}$-rational closed subvariety of codimension $\geqslant 1$ given by [Lar95, (3.8)]. There exists a constant $C_\alpha > 0$ such that $\xi(\alpha) \pmod{\ell} \notin K \pmod{\ell}$ if $\ell > C_\alpha$.
- $\mathcal{A}' \subset \mathcal{A}$ the set of the $\alpha \in \mathcal{A}$ such that:
  (1) $\rho_\lambda(\mathrm{Frob}_\alpha)$ is *regular with respect to* GL_n (see [Lar95, (3.4)], [LP92, (4.5)]) for every $\lambda \in \Lambda$.
  (2) $\xi(\alpha) \notin K$.
  By [Lar95, (3.11)], $\{\mathrm{Frob}_\alpha : \alpha \in \mathcal{A}'\} \subset \pi_{1,p}$ is still dense and by [LP92, (4.7)]:
  (1) $\rho_\lambda(\mathrm{Frob}_\alpha)$ lies in a unique maximal torus of $T_{\lambda,\alpha}$ of G_{E_λ}.
  (2) $\xi(\alpha)$ is associated to a torus T_α in $\mathrm{GL}_{n,E}$, unique up to $\mathrm{GL}_n(E)$-conjugacy, such that $T_\alpha \times_E E_\lambda$ is conjugate to $T_{\lambda,\alpha}$.
  (3) The splitting field of these tori is equal to the splitting field L_α of $\xi(\alpha)$ over E [LP92, (4.4)].
- $C'_\alpha \geqslant 1$ such that $L_\alpha/\mathbb{Q}$ is unramified at any $\ell > C'_\alpha$.
- L the intersection of the L_α for $\alpha \in \mathcal{A}'$, so that $\mathbb{Q} \subset E \subset L \subset L_\alpha$.

We decompose

$$B = (B \cap (\Lambda \backslash \Lambda_1)) \bigcup \bigcup_{x \in \mathrm{Gal}(L/E)^\sharp} B_x,$$

where $B_x = \{\lambda \in \Lambda_1 \cap B : [\lambda, L/E] = x\}$.

The upper natural density of B is

$$\begin{aligned} \overline{\delta}(B) &= \limsup_{S \to +\infty} \frac{|\{\lambda \in B : N(\lambda) \leqslant S\}|}{|\{\lambda \in \Lambda : N(\lambda) \leqslant S\}|} \\ &\leqslant \overline{\delta}(\Lambda \backslash \Lambda_1) + \sum_{x \in \mathrm{Gal}(L/E)^\sharp} \overline{\delta}\left(B_x\right) = \sum_{x \in \mathrm{Gal}(L/E)^\sharp} \overline{\delta}\left(B_x\right). \end{aligned}$$

Let us fix a class $x \in \mathrm{Gal}(L/E)^\sharp$ and an ℓ'-adic valuation $\lambda' \in \Lambda_1$ with Frobenius $[\lambda', L/E] = x$.

If $\lambda \in B_x$, then Γ_λ is a proper subgroup of $G(\mathbb{F}_\lambda) = G(\mathbb{F}_\ell)$. By [Lar95, (1.1), (1.19)], when $\ell \gg_G 1$, every maximal subgroup of $G(\mathbb{F}_\ell)$ is of the form $H(\mathbb{F}_\ell)$, for $H \subset G_{\mathbb{Z}_\ell}$ a smooth $\mathbb{Z}_\ell$-subgroup scheme. By [Lar95, (3.17)] (see also [Lar95, (3.8)]), it follows that there exists a maximal proper reductive $\mathbb{Q}_\ell$-subgroup N of G_λ (containing a Levi component of $H_{\mathbb{Q}_\ell}$) such that

$$\mathrm{FM}(\lambda, \alpha) \in \mathrm{FM}_{N^\circ} \subsetneq \mathrm{FM}_{G_\lambda}$$

for every $\alpha \in \mathcal{A}'$ such that $\ell > D_\alpha = \max(C_\alpha, C'_\alpha)$, where:

- N° is the identity component of N.
- $\mathrm{FM}(\lambda, \alpha)$ is the isomorphism class of the Frobenius module (i.e. free $\mathbb{Z}$-module of finite rank with an endomorphism of finite order) arising from the character group of the maximal torus $T_{\lambda,\alpha} \leqslant G_\lambda$ containing $\rho_\lambda(F_\alpha)$, with the action of $\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) = \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/E_\lambda)$. By [Lar95, (3.14)], this depends only on $[\ell, L_\alpha/\mathbb{Q}] = [\lambda, L_\alpha/E]$ up to isomorphism.
- FM_{G_λ} and FM_{N° are the set of isomorphism classes of Frobenius modules arising from unramified tori of G_λ, resp. N°.

Let $M \in \mathrm{FM}_{G_\lambda} \setminus \mathrm{FM}_{N^\circ}$. As in [Lar95, (3.15)], and [LP92, (8.2)], we will show that

($\star$) For every $R \geqslant 1$, there exist $\alpha_1, \dots, \alpha_R \in \mathcal{A}'$ such that $[M] = \mathrm{FM}(\lambda', \alpha_i)$ with L_{α_i} linearly disjoint[5].

Assuming this, it follows that if $\ell > \max_{1\leqslant i\leqslant R} D_{\alpha_i}$, then for $1 \leqslant i \leqslant R$,

$$[\lambda, L_{\alpha_i}/E] = [\ell, L_{\alpha_i}/\mathbb{Q}] \neq [\ell', L_{\alpha_i}/\mathbb{Q}] = [\lambda', L_{\alpha_i}/E]$$

in $\mathrm{Gal}(L_{\alpha_i}/\mathbb{Q}) \geqslant \mathrm{Gal}(L_{\alpha_i}/E)$, since $M \neq \mathrm{FM}(\lambda, \alpha_i)$. Therefore, by Chebotarev's theorem,

$$\begin{aligned} \overline{\delta}(B_x) &\leqslant \overline{\delta}\left(\{\lambda \in \Lambda : [\lambda, L_{\alpha_i}/E] \neq [\lambda', L_{\alpha_i}/E] \text{ for } 1 \leqslant i \leqslant R\}\right) \\ &= \left(1 - \frac{1}{n!}\right)^R \frac{|x|}{|\mathrm{Gal}(L/E)|} \end{aligned}$$

since $[L_{\alpha_i} : E] \leqslant n!$ and by linear disjointedness. Hence $\overline{\delta}(B) \leqslant (1 - 1/n!)^R$ for every $R \geqslant 1$, so that B has natural density 0 by taking $R \to +\infty$.

We now prove ($\star$). It suffices to show that for any finite Galois extension F/L, there exists $\alpha \in \mathcal{A}'$ such that $[M] = \mathrm{FM}(\lambda', \alpha)$ with L_α and F linearly disjoint over L. We proceed as in [LP92, (8.2)] (where $E = \mathbb{Q}$).

For $K_1, \dots, K_m$ the intermediate fields of F/L normal over L and minimal with respect to inclusion with this property, we have that L_α is linearly disjoint with F over L if and only if $K_i \not\subset L_\alpha$ for all $1 \leqslant i \leqslant m$. This holds in particular if for every i there exists $\lambda_i \in \Lambda_1$ corresponding to a prime that splits in L_α, but not in K_i.

For every $1 \leqslant i \leqslant m$, let $\beta_i \in \mathcal{A}'$ be such that $K_i \not\subset E_{\beta_i}$. By minimality of K_i, we have $E_{\beta_i} \cap K_i = L$, so that $\mathrm{Gal}(L_{\beta_i}/L) \times \mathrm{Gal}(K_i/L)$ is contained in

$$\{(\sigma_1, \sigma_2) \in \mathrm{Gal}(L_{\beta_i}/E) \times \mathrm{Gal}(K_i/E) : \sigma_1 \mid_L = \sigma_2 \mid_L\} \cong \mathrm{Gal}(L_{\beta_i}K_i/E).$$

By Chebotarev's theorem, the set of $\lambda \in \mathrm{Spec}(\mathcal{O})$ that split in L_{β_i} but does not split in K_i has positive Dirichlet density, so the same holds for the $\lambda \in \Lambda_1$ with this property, since Λ_1 has Dirichlet density 1. Hence, there exists $\lambda_i \in \Lambda_1 \setminus \{\lambda'\}$ that splits in L_α but not in K_i, and we may suppose all the λ_i distinct.

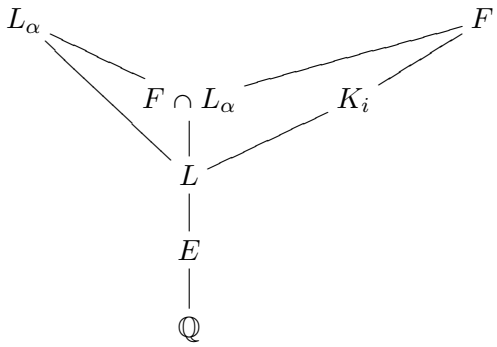
By [LP92, (7.5.3)], there exists $\alpha \in \mathcal{A}'$ such that:

(1) $T_{\lambda',\alpha}$ is conjugate in $\mathrm{GL}_n(E_{\lambda'})$ to the unramified maximal torus of $G_{\lambda'}$ corresponding to M, so $[M] = \mathrm{FM}(\lambda', \alpha)$.

[5]Here, this means that for any $2 \leqslant i \leqslant R$, $L_{\alpha_1} \dots L_{\alpha_{i-1}}$ and L_{α_i} are linearly disjoint over L, i.e. their intersection is equal to L.

(2) $T_{\lambda_i,\alpha}$ is conjugate in $\mathrm{GL}_n(E_{\lambda_i})$ to T_{λ_i,β_i}. Since λ_i splits in L_{β_i}, this torus is split, so that λ_i also splits in L_α.

This concludes the argument.



Finally, concerning the geometric integral monodromy group $\Gamma_\lambda^{\mathrm{geom}} = \rho_\lambda(\pi_{1,p}^{\mathrm{geom}}) \trianglelefteq \Gamma_\lambda$, note that:

(1) $\widetilde{\Gamma}_\lambda/\widetilde{\Gamma}_\lambda^{\mathrm{geom}}$ is a finite quotient of $\pi_{1,p}/\pi_{1,p}^{\mathrm{geom}} \cong \widehat{\mathbb{Z}}$, hence a finite cyclic group.

(2) If $|\mathbb{F}_\lambda| \gg_G 1$, the group $G'(\mathbb{F}_\lambda) := G(\mathbb{F}_\lambda)/Z(G(\mathbb{F}_\lambda))$ is simple non-abelian (see e.g. [MT11, Theorem 24.17]).

Hence, by (2), if $\widetilde{\Gamma}_\lambda^{\mathrm{geom}} \lneq G(\mathbb{F}_\lambda)$, then it is contained in $Z(G(\mathbb{F}_\lambda))$, so that

$$G'(\mathbb{F}_\lambda) \cong \frac{\widetilde{\Gamma}_\lambda/\widetilde{\Gamma}_\lambda^{\mathrm{geom}}}{Z(G(\mathbb{F}_\lambda))/\widetilde{\Gamma}_\lambda^{\mathrm{geom}}}$$

would be cyclic by (1), a contradiction. □

Remarks 5.4. (1) We consider compatible systems of representations $\rho_\lambda : \pi \to \mathrm{GL}_n(\mathcal{O}_\lambda)$, where λ is a valuation on the ring of integers $\mathcal{O}$ of a number field $E/\mathbb{Q}$, while the results in [LP92, Part II] and [Lar95] are stated for the case $E = \mathbb{Q}$. One needs to be cautious before stating the natural generalizations of the results of Larsen and Pink. For example, under the notations of the theorem, the maximal subgroups of $G(\mathbb{F}_\lambda)$ are not all of the form $H(\mathbb{F}_\lambda)$ for $H \subset G_{\mathcal{O}_\lambda}$ a smooth $\mathcal{O}_\lambda$-subgroup scheme, unless $\mathbb{F}_\lambda = \mathbb{F}_\ell$ as in [Lar95, (1.1), (1.19)]: for instance, one has subfield subgroups.

(2) Theorem 5.3 cannot be used when $G = \mathrm{SO}_n$, since it is not simply connected, and this assumption is required for [Lar95, (1.19)]. In even dimension, note that one would need additional input to determine the type (+ or −) of the monodromy groups over $\mathbb{F}_\lambda$.

5.2.2. *Arithmetic and geometric monodromy groups.* Often, only the geometric monodromy group is determined, while Theorem 5.3 and Definition 2.1 require knowledge of the arithmetic monodromy. By twisting a sheaf $\mathcal{F}_\lambda$ by a constant or a Tate twist, it is often possible to get a sheaf $\mathcal{F}'_\lambda$ with

$$G_{\mathrm{geom}}(\mathcal{F}_\lambda) = G_{\mathrm{geom}}(\mathcal{F}'_\lambda) \leqslant G_{\mathrm{arith}}(\mathcal{F}'_\lambda) \leqslant G_{\mathrm{geom}}(\mathcal{F}'_\lambda),$$

so that $G_{\mathrm{geom}}(\mathcal{F}'_\lambda) = G_{\mathrm{arith}}(\mathcal{F}'_\lambda) = G_{\mathrm{geom}}(\mathcal{F}_\lambda)$. Examples will be given in the next sections.

Remark 5.5. In [Kow06a, Kow06b, Kow08], the results of Larsen–Pink are applied to deduce the geometric monodromy group over $\mathbb{F}_\ell$ from the geometric group over $\overline{\mathbb{Q}}_\ell$. However, this is incorrect since the geometric group does not contain a dense subset of the Frobenius. Moreover, note that the arithmetic monodromy group is not contained in $\mathrm{Sp}_{2g}(\overline{\mathbb{Q}}_\ell)$ (but in $\mathrm{GSp}_{2g}(\overline{\mathbb{Q}}_\ell)$).

For the unnormalized family of first cohomology groups of hyperelliptic curves, this is not an issue because the results of J.-K. Yu and C. Hall also apply to give the geometric monodromy groups. Alternatively, one may normalize by a Tate twist as in Proposition 2.6 and apply Theorem 5.3 to the normalized sheaf (see above).

For the characteristic 2 example of [Kow06b, Proposition 3.3], the result of Hall can also be applied because the local monodromy at 0 is a unipotent pseudoreflection. Again, one could also apply Theorem 5.3 after normalizing.

On the other hand, the statement [Kow06a, Theorem 6.1] must be modified to assume for example that the *arithmetic* monodromy group is Sp, or that the geometric monodromy groups over $\mathbb{F}_\ell$ are known for all $\ell \gg 1$.

5.3. **General exponential sums.** Finally, we use the previous section to give examples of coherent families of the form (2).

5.3.1. *Construction of coherent families.*

Proposition 5.6 (Exponential sums (2), $h = 0$, $\chi = 1$)**.** *Let $f \in \mathbb{Q}(X)$ and let $Z_{f'}$ be the set of zeros of f' in $\mathbb{C}$, having cardinality k_f. We assume that the zeros of f' are simple, that $|f(Z_{f'})| = |Z_{f'}|$ (i.e. f is* supermorse*), and that either:*

- *(H_1): k_f is even, $\sum_{z\in Z_{f'}} f(z) = 0$, and if $s_1 - s_2 = s_3 - s_4$ with $s_i \in f(Z_{f'})$, then $s_1 = s_3, s_2 = s_4$ or $s_1 = s_2, s_3 = s_4$.*
- *(H_2): f is odd, and if $s_1 - s_2 = s_3 - s_4$ with $s_i \in f(Z_{f'})$, then $s_1 = s_3, s_2 = s_4$ or $s_1 = s_2, s_3 = s_4$ or $s_1 = -s_4, s_2 = -s_3$.*

If p is large enough, for $E = \mathbb{Q}(\zeta_{4p})$ and $\Lambda_{0,p}$ as in Example 4.3(1), there exists a family $(\mathcal{G}_{f,\lambda})_{\lambda\in\Lambda_{0,p}}$ of sheaves of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$, with trace function

$$x \mapsto \frac{-1}{\sqrt{q}} \sum_{\substack{y\in\mathbb{F}_q \\ f(y)\neq\infty}} e\left(\frac{\mathrm{tr}(xf(y))}{p}\right) \quad (x \in \mathbb{F}_q),$$

and conductor depending only on f.

Moreover, there exists $\alpha_p \in \overline{\mathbb{Q}}$ and a set of valuations $\Lambda' = \Lambda'_{f,p}$ of density 1 on $E' = E(\alpha_p)$, depending only on f and p, such that

$$\left(\mathcal{G}_{f,\lambda} \otimes \alpha_p \mathcal{O}'_\lambda\right)_{\lambda\in\Lambda'}$$

is a coherent family of sheaves of $\mathcal{O}'_\lambda$-modules over $\mathbb{F}_p$, for $\mathcal{O}'$ the ring of integers of E', with monodromy group structure

- *$G = \mathrm{SL}_{k_f}$ if (H_1) holds.*
- *$G = \mathrm{Sp}_{k_f}$ if (H_2) holds, and one may take $\alpha_p = 1$.*

Proof. See [Kat90, Theorem 7.9.4, Lemmas 7.10.2.1, 7.10.2.3] for the construction and [Kat90, 7.9.6–7, 7.10] for the determination of $G^\circ_{\mathrm{geom}}(\mathcal{G}_{f,\lambda})$

over $\mathbb{C}$. The family forms a compatible system by Lemma 2.3. The definition over $\mathcal{O}_\lambda$ comes from the definition of the ℓ-adic Fourier transform on the level of sheaves of $\mathcal{O}_\lambda$-modules (see [Kat88, Chapter 5]). Under our hypotheses, $G_{\mathrm{geom}}(\mathcal{G}_{f,\lambda})$ contains $\mathrm{SL}_{k_f}(\overline{\mathbb{Q}}_\ell)$, resp. $\mathrm{Sp}_{k_f}(\overline{\mathbb{Q}}_\ell)$. Moreover:

- In the (H_2) case, $G_{\mathrm{arith}}(\mathcal{G}_{f,\lambda}) \leqslant \mathrm{Sp}_{k_f}(\mathbb{F}_\lambda)$ by [Kat90, 7.10.4 (3)], and we can apply Theorem 5.3.
- In the (H_1) case, since $\pi_{1,p}/\pi_{1,p}^{\mathrm{geom}} \cong \hat{\mathbb{Z}}$ is abelian, there exists by Clifford theory an element $\beta_p \in \mathcal{O}_\lambda^\times \cap \overline{\mathbb{Q}}$ not depending on λ (since we have a compatible system) such that the determinant is isomorphic to $\beta_p \otimes \mathcal{O}_\lambda$.

  As in Section 5.2.2, we obtain that with $\alpha_p = \beta_p^{-1/k_f} \in \mathcal{O}'_{\lambda'}$ for any valuation λ' of $\mathcal{O}'$ extending λ, the arithmetic and geometric monodromy groups of $\mathcal{G}_{f,\lambda'} \otimes \alpha_p \mathcal{O}'_{\lambda'}$ coincide and are conjugate to $\mathrm{SL}_{k_f}(\overline{\mathbb{Q}}_\ell)$, so that we can apply Theorem 5.3.

□

Example 5.7. The hypotheses hold for the rational function $f = aX^{r+1} + bX$, where $a, b \in \mathbb{Z}$, $r \in \mathbb{Z} - \{1\}$, $rab \neq 0$, with (H_1) if r is odd and (H_2) otherwise (see [FM03, p. 7]), or for the polynomial $f = X^n - naX$, where $a \in \mathbb{Z}\backslash\{0\}$ and $n \geqslant 3$, with (H_1) if n is even, (H_2) otherwise.

The following include for example Birch sums (with $h = X^3$):

Proposition 5.8 (Exponential sums (2), $f = X$, $\chi = 1$, h polynomial)**.** *Let $h = \sum_{i=1}^n a_i X^i \in \mathbb{Z}[X]$ be a polynomial of degree $n \geqslant 3$ with $n \neq 7, 9$ and $a_{n-1} = 0$. If p is large enough, for $E = \mathbb{Q}(\zeta_{4p})$ and $\Lambda_{0,p}$ as in Example 4.3(1), there exists a family $(\mathcal{G}_{h,\lambda})_{\lambda \in \Lambda_{0,p}}$ of sheaves of $\mathcal{O}_\lambda$-modules over $\mathbb{F}_p$ with trace function*

$$x \mapsto \frac{-1}{\sqrt{q}} \sum_{y \in \mathbb{F}_q} e\left(\frac{\mathrm{tr}(xy + h(y))}{p}\right) \quad (x \in \mathbb{F}_q),$$

and conductor depending only on h.

Moreover, there exists $\alpha_p \in \overline{\mathbb{Q}}$ and a set of valuations $\Lambda' = \Lambda'_{h,p}$ of density 1 on $E' = E(\alpha_p)$, depending only on h and p, such that

$$\left(\mathcal{G}_{h,\lambda} \otimes \alpha_p \mathcal{O}'_\lambda\right)_{\lambda \in \Lambda'}$$

is a coherent family of sheaves of $\mathcal{O}'_\lambda$-modules over $\mathbb{F}_p$, for $\mathcal{O}'$ the ring of integers of E', with monodromy group structure:

(1) $G = \mathrm{Sp}_{n-1}$ if n is odd and h has no monomial of even positive degree; one may take $\alpha_p = 1$.
(2) $G = \mathrm{SL}_{n-1}$ otherwise.

Proof. This is similar to the proof of Proposition 5.6. See [Kat90, 7.12] for the construction of the sheaves and the determination of the monodromy groups over $\mathbb{C}$. In the symplectic case, ibidem shows that the arithmetic monodromy group is itself contained in Sp_{n-1}. □

Proposition 5.9 (Exponential sums (2), f polynomial, $\chi \neq 1$)**.** *Let*

- $h \in \mathbb{Q}(X)$ *odd with a pole of order* $n \geqslant 1$ *at* ∞.
- $f \in \mathbb{Z}[X]$ *odd nonzero of degree* d *with* $(d, n) = 1$.
- χ *a character of* $\mathbb{F}_p^\times$ *of order* $r \geqslant 2$.
- $g \in \mathbb{Q}(X)$ *nonzero, with the order of any zero or pole not divisible by* r.

For p *large enough, for* $E = \mathbb{Q}(\zeta_{4p})$ *and* $\Lambda_{0,p}$ *as in Example 4.3(1), there exists a family* $(\mathcal{G}_{h,f,\chi,g,\lambda})_{\lambda\in\Lambda_{0,p}}$ *of sheaves of* $\mathcal{O}_\lambda$*-modules over* $\mathbb{F}_p$ *with trace function (2) and conductor depending only on* f, g, h, r.

Moreover, if we assume that there exists $L \in \mathbb{Q}(X)$ *even with* $L(x)^r = g(x)g(-x)$ *and either* $N = \operatorname{rank}(\mathcal{G}_{h,f,\chi,g,\lambda}) \neq 8$ *or* $|n - d| \neq 6$, *then there exists a set of valuations* $\Lambda_p \subset \Lambda_{0,p}$ *of density* 1, *depending only on* h, f, g, χ *and* p, *such that*

$$(\mathcal{G}_{h,f,\chi,g,\lambda})_{\lambda\in\Lambda_p}$$

is a coherent family, with monodromy group structure $G = \mathrm{Sp}_N$.

Proof. This is again similar to the proof of Proposition 5.6. See [Kat90, 7.7, 7.13 (Sp-example(2))] for the construction of the sheaves and the determination of the monodromy groups over $\mathbb{C}$; [Kat90, 7.13] shows that the arithmetic monodromy group is itself contained in Sp_N. □

Remark 5.10. If L as in the statement of Proposition 5.9 is odd, there exists $\alpha_p \in \{\pm 1\}$ such that the arithmetic and geometric monodromy groups over $\mathbb{C}$ of $\alpha_p \otimes \mathcal{G}_{h,f,\chi,g,\lambda}$ coincide and are conjugate to $\mathrm{SO}_N(\mathbb{C})$ (see [Kat90, 7.14 (O-example(2))]). However, Theorem 5.3 does not apply in that case (see Remarks 5.4 (2)).

5.3.2. *Zero-density estimates.* Hence, for the three families above, we get by Corollary 4.8 with Propositions 4.10 and 4.11:

Proposition 5.11. *We fix a prime* p *and we set* $q = p^e$. *Let* $t : \mathbb{F}_q \to \mathbb{Q}(\zeta_{4p})$ *be the trace function associated with one of the families from Propositions 5.6, 5.8 or 5.9, and let* B *be as in (7).*

For $\varphi(x)$ *a first-order formula in the language of rings as in Proposition 4.11,*

$$P\big(t(x) \in \varphi(\mathbb{Q}(\zeta_{4p}))\big) \ll_{p,f,\varphi} \frac{\log q}{Bq^{\frac{1}{2B}}} \to 0 \ (e \to +\infty).$$

In particular, for almost all monic $f \in \mathbb{Z}[X]$ *of fixed degree* $\geqslant 2$ *(such as* $f(X) = X^m$ *for* $m \geqslant 2$ *coprime to* p*),*

$$P\big(t(x) \in f(\mathbb{Q}(\zeta_{4p}))\big) \ll_{p,f} \frac{\log q}{Bq^{\frac{1}{2B}}} \to 0 \ (e \to +\infty).$$

Proof. In the symplectic case, this is immediate. In the special linear cases, we get the result for the twisted trace function $t' : \mathbb{F}_q \to \mathcal{O}'_\lambda$, $t'(x) = \alpha_p^e t(x)$. The result for the unnormalized function is obtained as in Section 5.1.1, replacing A by $\alpha_p^{-e}A$ in Proposition 4.1 and using uniformity. □

Remark 5.12. In the special linear case, the implied constant depends on p both because of the use of Theorem 5.3, and because of the twisting factor α_p.

5.3.3. *Galois actions.* Note that for the sums

$$\frac{-1}{\sqrt{q}}\sum_{y\in\mathbb{F}_q} e\left(\frac{\mathrm{tr}(xf(y)+h(y))}{p}\right)\chi(g(y)) \quad (x\in\mathbb{F}_q^\times)$$

with $h(Y)=Y^m$ and $f(Y)=Y^n$ $(m,n\in\mathbb{Z})$, we have $\sigma_{c^m}(t(x)))=t(c^{m-n}x)$, where $\sigma_{c^m}\in\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})\cong\mathbb{F}_p^\times$ corresponds to c^m for some $c\in\mathbb{F}_p^\times$. Hence, as in Section 5.1.2, it makes sense to study the *integer*

$$\frac{|\{x\in\mathbb{F}_q^\times : t(x)\in\varphi(\mathbb{Q}(\zeta_{4p}))\}|(p-1,m-n)}{p-1} \ll_{m,n} \frac{|\{x\in\mathbb{F}_q^\times : t(x)\in\varphi(\mathbb{Q}(\zeta_{4p}))\}|}{p-1}$$

when $\varphi(x)$ is a first-order formula in the language of rings. However, doing so requires an estimate of the form (4) uniform in p, for example through a more precise knowledge of the integral monodromy instead of relying on Theorem 5.3.

5.4. **Hypergeometric sums.** The same methods also apply to the hypergeometric sums defined by Katz [Kat90, Chapter 8], generalizing Kloosterman sums: under some conditions, the arithmetic and geometric monodromy groups over $\overline{\mathbb{Q}}_\ell$ coincide and are conjugate to SL_n, without needing to twist (see the references to [Kat90] in [PG17b, Proposition 7.7]).

References

[BC06] Jean Bourgain and Mei-Chu Chang. A Gauss sum estimate in arbitrary finite fields. *C. R. Math. Acad. Sci. Paris*, 342(9):643–646, 2006.

[BSD59] Bryan Birch and Peter Swinnerton-Dyer. Note on a problem of Chowla. *Acta Arith.*, 5(4):417–423, 1959.

[Cha97] Nick Chavdarov. The generic irreducibility of the numerator of the zeta function in a family of curves with large monodromy. *Duke Math. J.*, 87(1):151–180, 1997.

[Coh70] Stephen Cohen. The distribution of polynomials over finite fields. *Acta Arith.*, 17(3):255–271, 1970.

[CvdDM92] Zoé Chatzidakis, Lou van den Dries, and Angus Macintyre. Definable sets over finite fields. *J. Reine Angew. Math.*, 427:107–136, 1992.

[Del77] Pierre Deligne. *Cohomologie étale, séminaire de géométrie algébrique du Bois-Marie SGA* $4\frac{1}{2}$, volume 569 of *Lecture notes in Math.* Springer, 1977.

[Del80] Pierre Deligne. La conjecture de Weil. II. *Inst. Hautes Études Sci. Publ. Math.*, 52(1):137–252, 1980.

[Fis92] Benji Fisher. Distinctness of Kloosterman sums. In *p-adic methods in number theory and algebraic geometry*, volume 133 of *Contemp. Math.*, pages 81–102. Amer. Math. Soc., 1992.

[Fis95] Benji Fisher. Kloosterman sums as algebraic integers. *Math. Ann.*, 301(1):485–505, 1995.

[FKM15] Étienne Fouvry, Emmanuel Kowalski, and Philippe Michel. Algebraic twists of modular forms and Hecke orbits. *Geom. Funct. Anal.*, 25(2):580–657, 2015.

[FM03] Étienne Fouvry and Philippe Michel. Sommes de modules de sommes d'exponentielles. *Pacific J. Math.*, 209(2):261–288, 2003.

[Fri80] John B. Friedlander. Estimates for prime ideals. *J. Number Theory*, 12(1):101–105, 1980.

[Gal73] Patrick X. Gallagher. The large sieve and probabilistic Galois theory. In *Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972)*, pages 91–101. Amer. Math. Soc., 1973.

[Hal08] Chris Hall. Big symplectic or orthogonal monodromy modulo ℓ. *Duke Math. J.*, 141(1):179–203, 2008.

[Jan05] Gerald J. Janusz. *Algebraic number fields*, volume 7 of *Grad. Stud. Math.* Amer. Math. Soc., second edition, 2005.

[Kat88] Nicholas M. Katz. *Gauss sums, Kloosterman sums, and monodromy groups*, volume 116 of *Annals of Math. Studies*. Princeton University Press, 1988.

[Kat90] Nicholas M. Katz. *Exponential sums and differential equations*, volume 124 of *Annals of Math. Studies*. Princeton University Press, 1990.

[Kat01] Nicholas M. Katz. Sums of Betti numbers in arbitrary characteristic. *Finite Fields Appl.*, 7(1):29–44, 2001.

[Kat12] Nicholas M. Katz. Report on the irreducibility of L-functions. In Dorian Goldfeld, Jay Jorgenson, Peter Jones, Dinakar Ramakrishnan, Kenneth Ribet, and John Tate, editors, *Number Theory, Analysis and Geometry*, pages 321–353. Springer, 2012.

[Kim97] Dae San Kim. Gauss sums for general and special linear groups over a finite field. *Arch. Math. (Basel)*, 69(4):297–304, 1997.

[Kim98] Dae San Kim. Gauss sums for symplectic groups over a finite field. *Monatsh. Math.*, 126(1):55–71, 1998.

[Kow06a] Emmanuel Kowalski. The large sieve, monodromy and zeta functions of curves. *J. Reine Angew. Math.*, 601:29–69, 2006.

[Kow06b] Emmanuel Kowalski. Weil numbers generated by other Weil numbers and torsion field of abelian varieties. *J. Lond. Math. Soc. (2)*, 74(2):273–288, 2006.

[Kow07] Emmanuel Kowalski. Exponential sums over definable subsets of finite fields. *Israel J. Math.*, 160(1):219–251, 2007.

[Kow08] Emmanuel Kowalski. *The large sieve and its applications: Arithmetic geometry, random walks and discrete groups*, volume 175 of *Cambridge*

Tracts in Math. Cambridge University Press, 2008.

[KR15] Lars Kindler and Kay Rülling. *Introductory course on ℓ-adic sheaves and their ramification theory on curves.* September 2015. https://arxiv.org/abs/1409.6899.

[KS99] Nicholas M. Katz and Peter Sarnak. *Random matrices, Frobenius eigenvalues and monodromy*, volume 45 of *Amer. Math. Soc. Colloq. Publ.* Amer. Math. Soc., 1999.

[Lar95] Michael Larsen. Maximality of Galois actions for compatible systems. *Duke Math. J.*, 80(3):601–630, 1995.

[Liv87] Ron Livné. The average distribution of cubic exponential sums. *J. Reine Angew. Math.*, 375/376:362–379, 1987.

[LP92] Michael Larsen and Richard Pink. On ℓ-independence of algebraic monodromy groups in compatible systems of representations. *Invent. Math.*, 107(1):603–636, December 1992.

[May13] James Maynard. On the Brun-Titchmarsh theorem. *Acta Arith.*, 157(3):249–296, 2013.

[MT11] Gunther Malle and Donna Testerman. *Linear algebraic groups and finite groups of Lie type*, volume 133 of *Cambridge Stud. Adv. Math.* Cambridge University Press, 2011.

[Nar04] Władysł aw Narkiewicz. *Elementary and analytic theory of algebraic numbers.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, third edition, 2004.

[PG17a] Corentin Perret-Gentil. Distribution questions for trace functions with values in the cyclotomic integers and their reductions. *Trans. Amer. Math. Soc.*, 2017. To appear.

[PG17b] Corentin Perret-Gentil. Gaussian distribution of short sums of trace functions over finite fields. *Math. Proc. Cambridge Philos. Soc.*, 163(3):385–422, 2017.

[PG18] Corentin Perret-Gentil. Integral monodromy groups of Kloosterman sheaves. *Mathematika*, 64(3):652–678, June 2018.

[Ser81] Jean-Pierre Serre. Quelques applications du théorème de densité de Chebotarev. *Inst. Hautes Études Sci. Publ. Math.*, 54:123–201, 1981.

[Sta74] Harold M. Stark. Some effective cases of the Brauer-Siegel theorem. *Invent. Math.*, 23(2):135–152, 1974.

[vdW34] Bartel Leendert van der Waerden. Die Seltenheit der Gleichungen mit Affekt. *Math. Ann.*, 109:13–16, 1934.

[Was97] Lawrence C. Washington. *Introduction to cyclotomic fields*, volume 83 of *Grad. Texts in Math.* Springer, 1997.

Centre de recherches mathématiques, Montréal, Canada
Email address: corentin.perretgentil@gmail.com