Séminaire BOURBAKI Avril 2022
74e année, 2021–2022, no 1193

BINARY ADDITIVE PROBLEMS FOR POLYNOMIALS OVER FINITE FIELDS
[according to W. Sawin and M. Shusterman]

by **Emmanuel Kowalski**

1. INTRODUCTION

In the study of prime numbers, robust methods have been discovered during the late 19th century and over the course of the 20th century and the early 21st to solve and understand many of the most natural "additive" questions that curiosity had suggested to mathematicians. These methods include (to name a few): L-functions (from Dirichlet and Riemann to Artin and Langlands); sieve methods (Brun, Selberg, Iwaniec); combinations of these (Bombieri, Vinogradov, Maynard, Tao), bilinear forms methods (Vinogradov, Linnik); and ideas from ergodic theory and additive combinatorics (Green, Tao). We refer, for instance, to the surveys [1, 2, 3, 4, 5, 6, 7] in this seminar for some accounts of these methods and their achievements.

Among the remaining outstanding open questions, we have quite a few belonging to the class of *binary additive problems*, which include what are probably the two most popular among them: the twin prime conjecture, and Goldbach's conjecture for sums of two primes. There are a number of intrinsic limitations to the currently known methods which have blocked all attempts at solving these problems.

We will report in this survey on recent groundbreaking work of W. Sawin and M. Shusterman [36, 37] where, for the first time, problems of this kind are solved in the case of polynomials over a *fixed* finite field, in a very strong quantitative form. Furthermore, we will present results of Sawin [32] where similar ideas are used to prove extremely strong results concerning the level of distribution of arithmetic functions in arithmetic progressions, again in the case of polynomials over *fixed* finite fields.

Here are simple examples of the main achievements of Sawin and Shusterman. For a polynomial f with coefficients in a finite field k with $|k|$ elements, we denote

$$|f| = |k[\mathrm{T}]/fk[\mathrm{T}]| = |k|^{\deg(f)}$$

(the second formula when $f \neq 0$). Given a finite field k, we also denote by k_0 its prime subfield, so $|k_0|$ is the characteristic of k.[(1)]

[(1)] We will reserve the letter p to denote *either* prime numbers or irreducible polynomials, hence we do not want to waste it simply for the characteristic of the fixed field k.

THEOREM 1.1 (Sawin–Shusterman). — *Let k be a finite field such that $|k| > 10^6|k_0|^4$. Let $a \in k[\mathrm{T}]$ be a fixed non-zero polynomial.*

(1) (Twin prime conjecture) *We have*

$$|\{p \in k[\mathrm{T}] \mid \deg(p) = d,\ p \text{ and } p+a \text{ are monic irreducible polynomials}\}| \sim \mathfrak{L}_a \frac{|k|^d}{d^2}$$

as $d \to +\infty$, where

$$\mathfrak{L}_a = \prod_{p|a}\Big(1-\frac{1}{|p|}\Big)^{-1} \prod_{p\nmid a}\Big(1-\frac{2}{|p|}\Big)\Big(1-\frac{1}{|p|}\Big)^{-2},$$

which is a strictly positive absolutely convergent product over all monic irreducible polynomials $p \in k[\mathrm{T}]$.

(2) (Quadratic Bateman–Horn conjecture) *Assume that $|k|$ is odd. We have*

$$|\{f \in k[\mathrm{T}] \mid \deg(f) = d,\ f^2+a \text{ is a monic irreducible polynomial}\}| \sim \frac{\mathfrak{Q}_a}{2}\frac{|k|^d}{d}$$

as $d \to +\infty$, where

$$\mathfrak{Q}_a = \prod_{p}\Big(1-\frac{\varrho_a(p)}{|p|}\Big)\Big(1-\frac{1}{|p|}\Big)^{-2}, \quad \varrho_a(p) = |\{x \in k[\mathrm{T}]/pk[\mathrm{T}] \mid x^2+a(x) = 0\}|$$

which is a strictly positive conditionally *convergent product over all monic irreducible polynomials $p \in k[\mathrm{T}]$, taken as the limit as $d \to +\infty$ of partial products over polynomials of degree $\leqslant d$.*

Remark 1.2. — (1) For integers, one often states the twin prime conjecture simply as the question of existence of infinitely many examples, without emphasizing the quantitative version. In the case of polynomials over finite fields, it is interesting to note that one can prove quite easily that there are infinitely many irreducible polynomials p in $k[\mathrm{T}]$ such that $p+1$ is also irreducible. For instance, if $|k^\times|$ is divisible by an odd prime ℓ, one can look for binomials $p = \mathrm{T}^{\ell^m} - a$ where $m \geqslant 1$, in which case the question is to find some $a \in k^\times$ such that neither a nor $a+1$ is an ℓ-th power in k; there must be some of them, simply because the group of ℓ-th powers has index $\geqslant 2$ among the non-zero elements of k. (This observation can be found, with applications to ranks of twisted Legendre curves, in a paper of C. Hall, see [19, Cor. 14].)

(2) Observe the restriction on the size of k: although we haven't stated the sharpest forms of the results of Sawin and Shusterman (see [36, Th. 1.1] and [37, Th. 1.2], respectively), they require that $|k|$ be large enough compared to the characteristic of k. We will explain the source of this condition, and observe for the moment only that new ideas seem to be necessary to handle the case when $k = k_0$.

And here is a sample result from [32]:

THEOREM 1.3 (Sawin). — *Let k be a finite field such that $|k| \geqslant 23173$. For polynomials a and q in $k[\mathrm{T}]$, and for integers $d \geqslant 0$, let*

$$\pi(d; q, a) = \sum_{\substack{\deg(p)=d \\ p \equiv a \bmod q}} 1,$$

where the sum is over monic irreducible polynomials. Furthermore, let $\pi(d)$ be the number of monic irreducible polynomials of degree d.[(2)]

There exists $\mathrm{C} \geqslant 0$ and $\delta > 0$ such that, for all $d \geqslant 1$ and for $q \in k[\mathrm{T}]$ squarefree with $\deg(q) \leqslant 3d/4$ and $a \in k[\mathrm{T}]$ coprime with q, we have

$$\left|\pi(d; q, a) - \frac{\pi(d)}{\varphi(q)}\right| \leqslant \mathrm{C}|k|^{(1-\delta)(d-\deg(q))},$$

where $\varphi(q) = |(k[\mathrm{T}]/qk[\mathrm{T}])^\times|$.

Remark 1.4. — (1) Again, there is a more precise version (where, for $|k|$ large enough, the constant $3/4$ may be replaced by any number < 1) in [32, Th. 1.2], but note the absolute bound for the size of k, independent of k_0, which highlights a difference with the previous results.

(2) The analogue of this theorem over number fields would be the fact that the primes have level of distribution $3/4$ in *individual* arithmetic progressions, which is currently unknown even under the assumption of the Generalized Riemann Hypothesis (and would have enormous implications in analytic number theory). In fact, there is (to this writer's knowledge) currently no non-trivial example of a sequence of integers with individual level of distribution $\geqslant 3/4$ (see the paper of Nunes [27] for one of the best results currently known, for squarefree integers, with level of distribution $25/36$; a notable non-trivial sequence with level of distribution arbitrarily close to 1 on average over the modulus is the Thue–Morse sequence, by work of Spiegelhofer [30, Th. 1.1]).

The outline of the remaining of this survey is the following:

1. We recall the analogy between integers and polynomials over finite fields, and state in parallel the Bateman–Horn conjecture in both cases.
2. We will explain *why* the case of polynomials over finite fields may be more accessible; in particular, we will explain briefly the simpler setting of the conjecture where the finite field k is allowed to change while the degree of the polynomials p is fixed.
3. We then present the strategy of Sawin and Shusterman – this combines beautifully arguments from algebraic as well as analytic number theory, and algebraic geometry. We attempt especially to focus on the points where the case of polynomials presents new phenomena and methods.
4. We sketch briefly some of the key arguments, chosen to emphasize both the similarities with integers, and some of the new ingredients.

(2) As we will recall below, we have $\pi(d) \sim q^d/d$ (which may be checked elementarily by looking at elements of the extension of degree d of k which generate it; there are $\sim q^d$ such elements, each has minimal polynomial irreducible of degree d, and only d elements have the same minimal polynomial).

Finally, we wish to point out that the papers we discuss contain a wealth of other results, many of which have considerable independent interest (non only more general, precise and uniform versions of the statements above, but also, e.g., proof of existence of cancellation in sums of the Möbius function evaluated at polynomials, which includes Chowla's conjecture for polynomials over finite fields). We invite the interested reader to go back to the source for more details.

Notation.

If X is a set and f, g are complex-valued functions on X, with $g \geqslant 0$, we write equivalently $f \ll g$ or $f = \mathrm{O}(g)$ if there exists a constant $\mathrm{C} \geqslant 0$ such that $|f(x)| \leqslant \mathrm{C}g(x)$ for all $x \in \mathrm{X}$. We then say that C is an *implied constant*. On the other hand, if X is a topological space and x_0 is in X (or is "at infinity"), we write $f(x) \sim g(x)$ as $x \to x_0$ to mean that g is non-zero close to x_0 and f/g tends to 1 as $x \to x_0$.

Acknowledgments.

Many thanks to C. Dartyge, É. Fouvry, J. Fresán, Ph. Michel and Z. Rudnick for comments and corrections on a draft of this text.

2. THE POLYNOMIAL–INTEGER ANALOGY

We will present the classical analogy between integers and polynomials over finite fields, choosing notation so that the parallel is as literal as possible. In particular, we can then present the general Bateman–Horn conjecture (and "standard" level of distribution conjectures) in a uniform manner.

The analogy goes back at least to a famous paper of Dedekind and Weber [16], and the basic dictionary is well-established:

$\mathbf{Z}$	$k[\mathrm{T}]$ where k is a finite field
$n \geqslant 1$	$f \in k[\mathrm{T}]$ monic
p prime	p monic irreducible polynomial
$\lvert n\rvert$ for $n \in \mathbf{Z}$	$\lvert f\rvert = \lvert k\rvert^{\deg(f)}$ for a polynomial f.

The analogy is reinforced by the fact that both $\mathbf{Z}$ and $k[\mathrm{T}]$ are principal ideal domains, and that prime numbers and monic irreducible polynomials, respectively, are in bijection with the set of non-zero prime ideals in $\mathbf{Z}$ and $k[\mathrm{T}]$. Moreover, it is crucial to the arithmetic part of this analogy that for non-zero integer n or polynomial f, the quotient ring $\mathbf{Z}/n\mathbf{Z}$ or $k[\mathrm{T}]/fk[\mathrm{T}]$ is finite, and is a finite field if n is prime or f irreducible.

So for instance, the analogue of the Riemann zeta function for $k[\mathrm{T}]$ is

$$\zeta_{k[\mathrm{T}]}(s) = \prod_p (1 - |p|^{-s})^{-1} = \sum_f |f|^{-s},$$

where the product ranges over all monic irreducible polynomials in $k[\mathrm{T}]$, and the sum over all $f \in k[\mathrm{T}]$ monic. The Prime Number Theorem states that the number $\pi(x)$ of prime numbers $p \leqslant x$ satisfies

$$\pi(x) \sim \frac{x}{\log x}, \qquad x \to +\infty,$$

and the analogue for irreducible polynomials is that

$$\pi(d) \sim \frac{q^d}{d}, \qquad d \to +\infty.$$

If we note that q^d is the number of monic polynomials of degree d, then the two asymptotic are clearly comparable. Intuitively, the second states that a monic polynomial of degree $d \geqslant 1$ has probability about $1/d$ of being irreducible.

We will use the following notation to have completely uniform statements. We write $\mathscr{O} = \mathbf{Z}$ or $k[\mathrm{T}]$ for some finite field; we denote by n (resp. p) a positive integer or a monic polynomial (resp. a prime number or a monic irreducible polynomial). We call p a *prime* in all cases. We sometimes denote by $\mathscr{O}_+$ either the set of positive integers or the set of monic polynomials.

Certain arithmetic functions have definitions which are identical in both cases. For instance, the function τ maps n to the number of divisors d of n, where divisors are either positive integers or monic (i.e., $d \in \mathscr{O}_+$). By convention, this meaning of divisors of n will be used below implicitly; similarly, the notation $d \mid n$ will later on always contain this restriction on d, unless otherwise stated. Another crucial function is the Möbius function μ, with $\mu(n) = 0$ unless n is squarefree, in which case $\mu(n) = (-1)^k$ if n is the product of $k \geqslant 0$ primes.

For given non-zero $n \in \mathscr{O}$, we denote by $|n|$ the "norm" $|\mathscr{O}/n\mathscr{O}|$, recovering the usual absolute value of integers, or the previous definition for polynomials.

In some cases, separate definitions are needed to have a uniform presentation. Given a real number $x \geqslant 1$, we will write

$$n \sim x$$

to indicate that either $x < n \leqslant 2x$, or that $|n| = x$ for polynomials (in other words, $x = |k|^{\deg(f)}$, so x can only be a power of $|k|$). We write also

$$\mathbf{log}(x) = \begin{cases} \log(x) & \text{if } \mathscr{O} = \mathbf{Z} \\ \log(x)/\log(|k|) & \text{if } \mathscr{O} = k[\mathrm{T}]. \end{cases}$$

Furthermore, for non-zero $n \in \mathscr{O}_+$, we define

$$\mathbf{log}(n) = \begin{cases} \log(n) & \text{if } \mathscr{O} = \mathbf{Z} \\ \deg(n) & \text{if } \mathscr{O} = k[\mathrm{T}]. \end{cases}$$

Then, for instance, the von Mangoldt function Λ for $\mathscr{O}$ is defined by

$$\Lambda(n) = \begin{cases} \mathbf{log}(n) & \text{if } n = p^m \text{ for some prime } p \text{ and integer } m \geqslant 1 \\ 0 & \text{otherwise,} \end{cases}$$

and satisfies

$$\mathbf{log}(n) = \sum_{d|n} \Lambda(d),$$

with the sum over divisors of n (either positive integers, or monic divisors of a monic polynomial). By Möbius inversion, we obtain a formula for $\Lambda(n)$, namely

$$\Lambda(n) = \sum_{d|n} \mu(d)\mathbf{log}(n/d) = -\sum_{d|n} \mu(d)\mathbf{log}(d) \tag{1}$$

This equation is a fundamental "detector" of primes (although it also detects the powers of primes, these are often quite easily handled separately, as they are typically much sparser in the applications).

Let $q \in \mathscr{O}_+$ and $a \in \mathscr{O}$. For $x \geqslant 1$, we now define

$$\boldsymbol{\pi}(x; q, a) = \sum_{\substack{p \sim x \\ p \equiv a \bmod q}} 1, \qquad \boldsymbol{\pi}(x) = \sum_{p \sim x} 1.$$

The Prime Number Theorem then has the form

$$\boldsymbol{\pi}(x) \sim \frac{x}{\mathbf{log}(x)}, \qquad x \to +\infty,$$

in all cases.[3]

The key conjecture generalizing the main additive problems for primes is the *Bateman–Horn* conjecture. Given a finite family $\mathbf{F} = (\mathrm{F}_i)_{i \in \mathrm{I}}$ of distinct non-constant irreducible and primitive[4] polynomials in $\mathscr{O}[\mathrm{X}]$, this predicts how often an element n with $n \sim x$ will be such that $\mathrm{F}_i(n)$ is prime for all $i \in \mathrm{I}$. (So that, for instance, we recover the twin prime conjecture by taking $\mathrm{F}_1 = \mathrm{X}$, $\mathrm{F}_2 = \mathrm{X} + a$.)

To state it, we first define for p prime the quantity

$$\varrho_{\mathbf{F}}(p) = |\{\alpha \in \mathscr{O}/p\mathscr{O} \mid \mathrm{F}_i(\alpha) = 0 \text{ for some } i \in \mathrm{I}\}|$$

(where $\mathrm{F}_i(\alpha)$ denotes the value in the finite field $\mathscr{O}/p\mathscr{O}$ of the reduction of F_i modulo p, a polynomial in $(\mathscr{O}/p\mathscr{O})[\mathrm{X}]$). One can then show that the infinite product

$$\mathfrak{S}_{\mathbf{F}} = \prod_p \Big(1 - \frac{\varrho_{\mathbf{F}}(p)}{|p|}\Big)\Big(1 - \frac{1}{|p|}\Big)^{-|\mathrm{I}|},$$

converges, in the sense of the limit as $d \to +\infty$ of the finite products over p with $|p| \leqslant d$, and moreover that the value of the product is zero if and only if there exists some prime p such that $\varrho_{\mathbf{F}}(p) = |p|$.

There is a probabilistic interpretation for these products, which can basically motivate the Bateman–Horn conjecture below (see [8]). We present another well-known heuristic derivation which is closer to a proof (in the sense of being indeed the starting point of rigorous proofs, both in a number of classical results, as illustrated for instance

[3] In the polynomial case, x is taken to vary among powers of $|k|$; if one wishes to avoid this interpretation, define $n \sim x$ to mean $\lfloor x \rfloor = |k|^{\deg(f)}$ instead.

[4] By which we mean that the coefficients of F_i do not have a non-trivial common factor.

in [23, Ch. 19], and in the works of Sawin and Shusterman). For simplicity, consider the case $|\mathrm{I}| = 1$, so that we have a single polynomial F. Then using (1), we have

$$\begin{aligned}\sum_{n\sim x}\Lambda(\mathrm{F}(n)) &= -\sum_{n\sim x}\sum_{d\mid \mathrm{F}(n)}\mu(d)\mathbf{log}(d)\\ &= -\sum_{d}\mathbf{log}(d)\mu(d)\sum_{\substack{\alpha\in\mathscr{O}/d\mathscr{O}\\ \mathrm{F}(\alpha)=0}}\sum_{\substack{n\sim x\\ n\equiv\alpha \bmod d}}1\end{aligned}$$

(having split the sum over $n\sim x$ such that $d\mid \mathrm{F}(n)$ according to the value $\alpha\in\mathscr{O}/d\mathscr{O}$ of $\mathrm{F}(n)$ modulo d, which must satisfy $\mathrm{F}(\alpha)=0$).

The inner sum counts elements in an arithmetic progression, and at least when x is much larger than $|d|$, it is asymptotic to $x/|d|$; if we disregard all error terms, and extend ϱ_{F} to squarefree d by multiplicativity, we obtain the series

$$-x\sum_{d}\mathbf{log}(d)\frac{\mu(d)\varrho_{\mathrm{F}}(d)}{|d|},$$

and the series itself is formally $xf'(1)$, where[(5)]

$$f(s)=\sum_{d}\frac{\mu(d)\varrho_{\mathrm{F}}(d)}{|d|^s}=\prod_{p}\Big(1-\frac{\varrho_{\mathbf{F}}(p)}{|p|^s}\Big)=\zeta_{\mathscr{O}}(s)^{-1}\prod_{p}\Big(1-\frac{\varrho_{\mathbf{F}}(p)}{|p|^s}\Big)\Big(1-\frac{1}{|p|^s}\Big)^{-|\mathrm{I}|}.$$

Since $\zeta_{\mathscr{O}}(s)$ has a simple pole at $s=1$, with residue equal to 1 for $\mathscr{O}=\mathbf{Z}$ and to $1/\log|k|$ if $\mathscr{O}=k[\mathrm{T}]$, this implies that $f'(1)=\mathfrak{S}_{\mathrm{F}}$.

We can now state the Bateman–Horn conjecture, which essentially claims that this heuristic derivation gives the right answer:

Conjecture 2.1. — *If $\mathfrak{S}_{\mathbf{F}}\neq 0$, and if all (F_i) are separable, then as $x\to+\infty$, we have*

$$\sum_{n\sim x}\prod_{i\in\mathrm{I}}\Lambda(\mathrm{F}_i(n))\sim\mathfrak{S}_{\mathbf{F}}x,$$

$$|\{n\sim x\mid \mathrm{F}_i(n)\text{ is prime for all }i\in\mathrm{I}\}|\sim\frac{\mathfrak{S}_{\mathbf{F}}}{\Delta_{\mathbf{F}}}\frac{x}{\mathbf{log}(x)^{|\mathrm{I}|}}.$$

where $\Delta_{\mathbf{F}}$ is the product of the degrees of F_i for $i\in\mathrm{I}$.

(Note that the second form of this conjecture can easily be deduced from the first one; the factor $\Delta_{\mathbf{F}}$ is present because $|\mathrm{F}(n)|$ is of size $n^{\deg(\mathrm{F})}$ for $\mathrm{F}\in\mathscr{O}[\mathrm{X}]$.)

Example 2.2. — (1) (Twin prime) Take $(\mathrm{F}_1,\mathrm{F}_2)=(\mathrm{X},\mathrm{X}+a)$, where $a\in\mathscr{O}$ is non-zero. Then $\varrho_{\mathrm{F}}(p)=2$ unless $\mathrm{F}(a)=0 \bmod p$, in which case $\varrho_{\mathrm{F}}(p)=1$. It follows that

$$\mathfrak{S}_{(\mathrm{F}_1,\mathrm{F}_2)}=\prod_{p\mid a}\Big(1-\frac{1}{|p|}\Big)^{-1}\prod_{p\nmid a}\Big(1-\frac{2}{|p|}\Big)\Big(1-\frac{1}{|p|}\Big)^{-2},$$

which coincides with the constant $\mathfrak{L}_a$ in Theorem 1.1, (1).

(5) When $\mathscr{O}=k[\mathrm{T}]$, this derivative should be interpreted as $(\log|k|)^{-1}$ times the usual derivative, so that $(s\mapsto|d|^{-s})'=-\mathbf{log}(d)|d|^{-s}$.

(2) (Quadratic Bateman–Horn Conjecture) Take $\mathrm{F} = \mathrm{X}^2 + a$ for some non-zero $a \in \mathscr{O}$. Then $\mathfrak{S}_\mathrm{F}$ coincides with the constant $\mathfrak{Q}_a$ of Theorem 1.1, (2).

Thus Theorem 1.1 confirms these two cases of the Bateman–Horn conjecture, and in fact in quantitative form, as soon as $|k|$ is large enough.

(3) We snuck in the assumption that the polynomials F_i are separable; this is not simply a "belt and suspenders" kind of assumption, but a necessary restriction in many cases. Indeed, Conrad, Conrad and Gross [13] have shown that the asymptotic formula in the conjecture is not always true without such an assumption, and have studied the issue in depth, leading to a corrected conjecture that they expect to hold in all cases (see [13, Conj. 6.2]). A simple example demonstrating the failure of the statement in general is the following (see also [12, Ex. 4.3]): let k be a finite field of odd characteristic and

$$\mathrm{F} = \mathrm{X}^{4|k|} + \mathrm{T}^{2|k|-1} \in \mathscr{O}[\mathrm{X}]$$

(e.g., $\mathrm{F} = \mathrm{X}^{20} + \mathrm{T}^9$ if $|k| = 5$); then it is elementary that F is irreducible, and that the corresponding constant $\mathfrak{S}_\mathrm{F}$ is non-zero, and yet $\mathrm{F}(n)$ is *never* irreducible for $n \in k[\mathrm{T}]$ of degree $\geqslant 1$ (see Example 5.3 below).

Crucially, the investigations of Conrad, Conrad and Gross involve the fluctuations of the sign of $\mu(\mathrm{F}(n))$ as n varies: the above heuristic can only be potentially correct if the values 1 and -1 exhibit the statistical behavior of a fair coin toss, and it is shown that this is simply not the case.

3. DIGRESSION: THE CASE OF LARGE FINITE FIELDS

This section is essentially independent of the remainder of the text, and may be skipped in a first reading. We discuss here the "other" analogue of the basic additive number theory conjectures over $k[\mathrm{T}]$, namely the "large finite field" situation. Precisely, consider the single polynomial case of the Bateman–Horn conjecture. We then only consider input polynomials n with a *given* degree d, and vary the base field. Thus, for integers $\nu \geqslant 1$, let k_ν be the extension of k of degree ν in an algebraic closure $\bar{k}$ of k. We are looking at counting $n \in k_\nu[\mathrm{T}]$ of degree d such that $\mathrm{F}(n)$ is prime in $k_\nu[\mathrm{T}]$. Note that there is no analogue of this question for integers!

This problem has a nice geometric interpretation, whose origins lie in work of Birch and Swinnerton-Dyer [10] and Cohen [11]. Indeed, a polynomial $n \in k_\nu[\mathrm{T}]$ of degree d is irreducible if and only if n has d distinct roots and if the Frobenius automorphism $\alpha \mapsto \alpha^{|k_\nu|}$ induces on the set of roots of n in $\bar{k}$ a permutation which is a cycle of length d. We view F as a polynomial in $k_\nu[\mathrm{T}, \mathrm{X}]$. Then the roots of $\mathrm{F}(n)$ in $\bar{k}$ are those α such that $\mathrm{F}(\alpha, n(\alpha)) = 0$, and are therefore in canonical bijection with the intersection points of the graph of n with the plane curve $\mathscr{C}_\mathrm{F}$ defined by $\mathrm{F}(\alpha, \beta) = 0$. We can parameterize polynomials of degree d by an algebraic variety $\mathscr{X}_d$, and then construct a covering $\pi\colon \mathscr{Y}_{\mathrm{F},d} \to \mathscr{X}_d$ where the fiber over $n \in \mathscr{X}_d$ is the intersection $\mathscr{C}_\mathrm{F} \cap (\text{graph of } n)$.

Thus we are looking for elements in $\mathscr{X}_d(k_\nu)$ where the Frobenius acts in a a specified way on the fiber $\pi^{-1}(n)$; this interpretation brings the problem squarely within the confines of Galois theory for function fields (or for finite-degree coverings of algebraic varieties). Combined with a suitable version of the Chebotarev Density Theorem, this leads quickly to an *a priori* asymptotic formula of the form

$$|\{n \in k_\nu[\mathrm{T}] \mid \deg(n) = d \text{ and } \mathrm{F}(n) \text{ is prime}\}| \sim c_{\mathrm{F}}|k|^{\nu d}$$

as $\nu \to +\infty$, for some constant c_{F}, namely the proportion of elements in the Galois group G of a Galois closure of π which have the cycle type of a single m-cycle.(6) The degree m of π is the "generic" degree of $\mathrm{F}(n)$; thus one would like to prove that G is the full symmetric group, in which case $c_{\mathrm{F}} = 1/m$, and the asymptotic formula is then what is expected from the Bateman–Horn conjecture.(7)

Among the (relatively expansive) literature on this question, we refer to the paper [17] of Entin for very general results confirming this prediction; we note that one of Entin's innovations for the computation of the group G is a beautiful idea of using characterization of permutation groups containing the alternating group by their multiple-transitivity properties.(8)

4. STRATEGY

We now present a rough sketch of the strategy followed by Sawin and Shusterman to prove Theorem 1.1.(9) As indicated previously, the starting point is the analytic heuristic presented to motivate the conjecture. In the case of the quadratic Bateman–Horn conjecture, we put $\mathrm{F} = \mathrm{X}^2 + a$ and consider the sum

$$\sum_{n\sim x} \Lambda(\mathrm{F}(n)) = -\sum_d \mathbf{log}(d)\mu(d) \sum_{\substack{\alpha\in\mathscr{O}/d\mathscr{O}\\ \mathrm{F}(\alpha)=0}} \sum_{\substack{n\sim x\\ n\equiv\alpha \bmod d}} 1.$$

Extracting the main term can be done in the inner sum if d has small enough degree, and it is not very difficult to deduce that

$$-\sum_{|d|\leqslant x} \mathbf{log}(d)\mu(d) \sum_{\substack{\alpha\in\mathscr{O}/d\mathscr{O}\\ \mathrm{F}(\alpha)=0}} \sum_{\substack{n\sim x\\ n\equiv\alpha \bmod d}} 1 \sim \mathfrak{S}_{\mathrm{F}} x.$$

(6) We overlook here the potential distinction between the arithmetic and geometric Galois groups of the covering.

(7) Note that if we compute the constant $\mathfrak{S}_{\mathrm{F}}$ when viewing F as an element of $k_\nu[\mathrm{T},\mathrm{X}]$, we obtain a quantity that tends to 1 as $\nu \to +\infty$.

(8) This method may be viewed as a permutation-group analogue of the Larsen Alternative, as developed by Katz [26] for the computation of monodromy groups.

(9) We only discuss very briefly some aspects of Theorem 1.3 in Section 8.

What remains to be done (and remains unknown for integers) is to treat d of larger size. This must involve the sign fluctuations of the Möbius function, so we rewrite the sum as

$$-\sum_{x\leqslant y\leqslant x^2}\mathbf{log}(y)\sum_{b\sim x^2/y}\sum_{\substack{\alpha\in\mathscr{O}/b\mathscr{O}\\ \mathrm{F}(\alpha)=0}}\sum_{\substack{n\sim x\\ n\equiv\alpha \bmod b}}\mu\Big(\frac{n^2+a}{b}\Big), \tag{2}$$

and the saving will come from the inner sum

$$\sum_{\substack{n\sim x\\ n\equiv\alpha \bmod b}}\mu\Big(\frac{n^2+a}{b}\Big) \tag{3}$$

involving the Möbius function.

In the case of the twin primes (with $\mathrm{F}_1=\mathrm{X}$ and $\mathrm{F}_2=\mathrm{X}+a$), the procedure is similar; since there are two polynomials involved, the argument is a bit different, and one reduces similarly to sums of the type

$$-\sum_{x^{1/(2-\varepsilon)}\leqslant y\leqslant x}\mathbf{log}(y)\sum_{g\sim x/y}\sum_{q\sim y}\mu(q)\Lambda(gq+a)$$

after extracting the expected main term of the asymptotic from small degree polynomials. Applying the convolution formula (1) again leads to

$$\sum_{x^{1/(2-\varepsilon)}\leqslant y\leqslant x}\mathbf{log}(y)\sum_{g\sim x/y}\sum_{q\sim y}\sum_{b\mid gq+a}\mathbf{log}(b)\mu(b)\mu(q).$$

The sum over b is then split according to whether $\mathbf{log}(b)\leqslant y^{1/2}$ or not, and in the second range the classical trick of switching to the complementary divisor is used as above to sum $\mathbf{log}(c)\mu((gq+a)/c)$ over divisors c of $gq+a$ instead of $\mathbf{log}(b)\mu(b)$. The savings will come from the sum over q, which are of one of the two forms

$$\sum_{q\sim y}\mu(q),\qquad\qquad\sum_{\substack{q\sim y\\ q\equiv -a\bar{g} \bmod b}}\mu(q)\mu\Big(\frac{gq+a}{b}\Big), \tag{4}$$

in the two ranges described previously, respectively.

The key transformative steps that now escape from the common features of integers and polynomials are explained in the next two sections; they are first the existence of an *algebraic formula* for the Möbius function for a polynomial over a finite field, and the use of Deligne's Riemann Hypothesis over finite fields.

5. THE MÖBIUS FUNCTION FOR POLYNOMIALS

In final analysis, the most fundamental input to the work of Sawin and Shusterman, and that which explains the difference with the case of integers, is the fact that the Möbius function, whose importance to the study of additive problems is clear from the previous sections, has an algebraic expression in terms of multiplicative characters in

the case of polynomials over finite fields. This observation goes back to Pellet in 1878 at least (see the enlightening discussions in [13] and [12]).

PROPOSITION 5.1. — *Let k be a finite field of odd characteristic. Denote by λ_2 the non-trivial multiplicative character of order 2 of $k^\times$, extended to k by putting $\lambda_2(0) = 0$. For any non-zero $f \in k[\mathrm{T}]$, we have*

$$\mu(f) = (-1)^{\deg(f)}\lambda_2(\mathrm{disc}(f)),$$

where $\mathrm{disc}(f)$ *is the discriminant of* f.

Proof. — Both sides of this formula vanish for f having a multiple root, so we can assume that f is squarefree.

The Möbius function is of course multiplicative; we next observe that the function $f \mapsto (-1)^{\deg(f)}\lambda_2(\mathrm{disc}(f))$ is also. Indeed, if $(\alpha_i)_{1\leqslant i\leqslant \deg(f)}$ (resp. $(\beta_j)_{1\leqslant j\leqslant \deg(g)|}$) are the roots of a squarefree polynomial f (resp. g) in some algebraic closure $\bar{k}$ of k, then the condition that f and g are coprime means that $\alpha_i \neq \beta_j$ for all (i, j), hence

$$\mathrm{disc}(fg) = \mathrm{disc}(f)\,\mathrm{disc}(g)\gamma^2, \qquad \gamma = \prod_{i,j}(\alpha_i - \beta_j) \in \bar{k}.$$

The Frobenius automorphism Fr_k of $\bar{k}$ (given by $x \mapsto x^{|k|}$) permutes the (α_i) and the (β_j), and it follows that $\mathrm{Fr}_k(\gamma) = \gamma$. So $\gamma \in k$, hence $\lambda_2(\gamma^2) = 1$, and the multiplicativity follows.

We are thus reduced to proving that

$$(-1)^{\deg(f)}\lambda_2(\mathrm{disc}(f)) = -1$$

if f is irreducible. Let again $(\alpha_i)_{1\leqslant i\leqslant \deg(f)}$ be the roots of f. We have $\lambda_2(\mathrm{disc}(f)) = 1$ if and only if the product

$$\gamma = \prod_{i<j}(\alpha_i - \alpha_j) \in \bar{k},$$

which satisfies $\gamma^2 = \mathrm{disc}(f)$, belongs to k (and not to the unique quadratic extension of k in $\bar{k}$). Computing $\mathrm{Fr}_k(\gamma)$ we see that $\mathrm{Fr}_k(\gamma) = \varepsilon(\sigma)\gamma$, where $\varepsilon(\sigma)$ is the signature of the permutation of the roots induced by Fr_k. However, the irreducibility of f means (by the elementary Galois theory of finite fields) that this permutation is a cycle of length $\deg(f)$, with signature $(-1)^{\deg(f)-1}$. So we have $\lambda_2(\mathrm{disc}(f)) = \varepsilon(\sigma) = (-1)^{\deg(f)-1}$. □

Remark 5.2. — Conrad, Conrad and Gross [13, Th. 2.4] also explain a variant for fields of characteristic 2 (that goes back to work of R. Swan); this involves Witt vectors.

Example 5.3. — Assume that $|k|$ is odd and let $\mathrm{F} = \mathrm{X}^{4|k|} + \mathrm{T}^{2|k|-1}$. Consider $f \in k[\mathrm{T}]$ with positive degree. If $f(0) = 0$, we have $\mu(\mathrm{F}(f)) = 0$. Otherwise, $g = \mathrm{F}(f)$ has even degree, so by the formula for the discriminant in terms of roots of the derivative (implicitly used below) we get

$$\mu(\mathrm{F}(f)) = \lambda_2(\mathrm{disc}(g)) = \lambda_2\Big(\prod_{g(\alpha)=0} g'(\alpha)\Big).$$

But $g' = -\mathrm{T}^{2|k|-2}$, so this is $\lambda_2(g(0)^{2|k|-2}) = 1$. In particular, the polynomial $\mathrm{F}(f)$ cannot be irreducible.

The discriminant in the proposition can be transformed further. Recall that $\mathrm{disc}(f)$ is also the resultant of f and f'. Furthermore, for any non-zero polynomial g, there is a Jacobi symbol $f \mapsto (\frac{f}{g})$ modulo g on $k[\mathrm{T}]$, namely

$$\Big(\frac{f}{g}\Big) = \prod_{p^m || g} \Big(\frac{f}{p}\Big)^m$$

where, for p irreducible, we have

$$\Big(\frac{f}{p}\Big) = \begin{cases} 1 & \text{if } f \text{ is a non-zero square in } k[\mathrm{T}]/p \\ 0 & \text{if } p \mid f \\ -1 & \text{if } f \text{ is not a square in } k[\mathrm{T}]/p. \end{cases}$$

Then we have:

Proposition 5.4. — *For f non-zero in $k[\mathrm{T}]$, we have*

$$\mu(f) = (-1)^{\deg(f)}\Big(\frac{f'}{f}\Big)$$

See, e.g., [36, Lemma 3.1] for a proof.

Here is the key corollary of these facts used in [36].

Corollary 5.5. — *Let q be a monic polynomial in $k[\mathrm{T}]$ and $a \in k[\mathrm{T}]$ coprime to q. Let $\delta \in k[\mathrm{T}]$. There exist $\varepsilon \in \{-1, 0, 1\}$, $s \in k[\mathrm{T}]$ and a real Dirichlet character χ on $k[\mathrm{T}]$ with conductor dividing $d = q^2(\delta + (a/q)') \in k[\mathrm{T}]$ such that*

$$\mu(gq + a) = \varepsilon\chi(g + s)$$

for any polynomial $g \in k[\mathrm{T}]$ with derivative $g' = \delta$, provided $\deg(a) \neq \deg(qg)$.

Proof. — This is (essentially) a consequence of the previous propositions combined with quadratic reciprocity in $k[\mathrm{T}]$: we have first

$$\mu(gq + a) = (-1)^{\deg(gq+a)}\lambda_2(\mathrm{disc}(gq + a)) = (-1)^{\deg(gq+a)}\Big(\frac{gq + a}{(gq + a)'}\Big),$$

and then

$$\begin{aligned}\Big(\frac{gq + a}{(gq + a)'}\Big) = \varepsilon_1\Big(\frac{\delta q + gq' + a'}{gq + a}\Big) &= \varepsilon_2\Big(\frac{\delta q^2 + gqq' + a'q}{gq + a}\Big) \\ &= \varepsilon_2\Big(\frac{\delta q^2 + a'q - aq'}{gq + a}\Big) = \varepsilon_2\Big(\frac{d}{gq + a}\Big)\end{aligned}$$

where ε_1 and ε_2 are in $\{-1, 0, 1\}$ and depend only on (q, a, δ). By quadratic reciprocity again, this is (up to some $\varepsilon \in \{-1, 0, 1\}$ again) a real Dirichlet character evaluated at $gq + a$, or a real Dirichlet character evaluated at $g + \bar{q}a$, where $\bar{q}$ is the inverse of q modulo the conductor. □

Remark 5.6. — In fact, we see that the conductor of the Dirichlet character can be specified more precisely, and this is important in further arguments.

What is the key lesson from this corollary? It is that *when summing the Möbius function over an arithmetic progression, we are reduced to summing a shifted Dirichlet character, as long as we sum over the $gq+a$ where the derivative of g is fixed.* Thus, we can split a sum like

$$\sum_{g\sim x} \mu(gq+a)$$

according to the value of the derivative, namely

$$\sum_g \mu(gq+a) = \sum_\delta \sum_{g'=\delta} \mu(gq+a) = \sum_\delta \varepsilon(\delta) \sum_{g'=\delta} \chi(g+s(\delta)).$$

Given a fixed g_0 with $g'=\delta$, to say that $g'=\delta$ means that $g = g_0 + h(\mathrm{X}^{|k_0|})$ (recall that $|k_0|$ is the characteristic of k) and h is an arbitrary polynomial. If the size k is large enough compared with the characteristic, this means that *the inner sum is still long enough to be usefully attacked using methods from algebraic geometry*, as we now describe. Moreover, note that this principle applies equally well to sums involving a product of values of the Möbius function, such as

$$\sum_{g\sim x} \mu(g)\mu(gq+a),$$

which are in fact the type that occurs in the twin prime problem (because of the two von Mangoldt functions). With an arbitrary finite number of factors (with distinct linear polynomials), the resulting sums are those in the so-called *Chowla conjecture*, and indeed Sawin and Shusterman prove the version for $k[\mathrm{T}]$ of this conjecture (see [36, Th. 1.3]).

6. ALGEBRAIC GEOMETRY AND THE RIEMANN HYPOTHESIS

After the input from the previous section, one is faced with a task of a fairly common kind in analytic number theory: get good bounds (cancellation among the terms) for sums over finite fields which are related to character sums. More precisely, the sums that arise are of the type

$$\sum_{f\in \mathrm{P}_d(k)} t(f)$$

where $d \geqslant 0$ is a fixed integer, the set $\mathrm{P}_d(k)$ is the set of polynomials $f \in k[\mathrm{T}]$ with $\deg(f) < d$, and t is a function on $\mathrm{P}_d(k)$ of "algebraic nature".

The set $\mathrm{P}_d(k)$ can be interpreted as the set of k-rational points on an algebraic variety, indeed simply on the affine space of dimension d, with coordinates $\boldsymbol{a} = (a_i)_{0\leqslant i\leqslant d-1}$ which

are the coefficients of the polynomial f. The function t should be a "trace function"[10] on this algebraic variety, e.g.. something like

$$t(f) = \chi(\mathrm{A}(\boldsymbol{a}))\psi(\mathrm{B}(\boldsymbol{a}))$$

where A and B are polynomials in the coefficients (a_i), and we denote by χ (resp. ψ) a multiplicative character of $k^\times$ (resp. a non-trivial additive character of k).

Example 6.1. — Let $q \in k[\mathrm{T}]$ be a non-constant polynomial. Consider a non-trivial multiplicative character $\chi\colon (k[\mathrm{T}]/qk[\mathrm{T}])^\times \to \mathbf{C}^\times$. For $m \geqslant 1$ and $h_1, \dots, h_m \in k[\mathrm{T}]$ fixed, it is not very difficult to check that the function

$$t(f) = \chi(f + h_1) \cdots \chi(f + h_m)$$

defined for $f \in \mathrm{P}_d(k)$, can be expressed in the desired form.

So, in effect, we have multi-variable character sums over finite fields, including more complicated variants of those of this simple form. From the fundamental work of Deligne (which, for one-variable sums, goes back to Weil) it is known that such sums can be interpreted using methods of algebraic geometry, and especially that the *general form* of the Riemann Hypothesis [14] can be an extremely powerful tool to prove very strong estimates, potentially best possible in many cases.

We review the mechanism behind this method. The formalism of ℓ-adic cohomology (especially the Grothendieck–Lefschetz trace formula) leads to a transformation of the form

$$\sum_{f \in \mathrm{P}_d(k)} t(f) = \sum_{j \in \mathbf{Z}} (-1)^j \operatorname{tr}(\mathrm{F}_k \mid \mathrm{H}_c^j)$$

where only finitely many terms (typically, those terms with $0 \leqslant j \leqslant 2d$) can be non-zero, and H_c^j is then a finite-dimensional vector space on which a certain incarnation F_k of the Frobenius automorphism of k acts by a linear transformation; these spaces depend on the summation set P_d as well as on the trace function t.[11]

Deligne's version of the Riemann Hypothesis[12] states that, under certain conditions on t,[13] any complex eigenvalue α of F_k on H_c^j is an algebraic number, and satisfies the bound $|\alpha| \leqslant |k|^{j/2}$. This means that we obtain typically an estimate

$$\left| \sum_{f \in \mathrm{P}_d(k)} t(f) \right| \leqslant \sum_{0 \leqslant j \leqslant 2d} |k|^{j/2} \dim(\mathrm{H}_c^j) \leqslant \mathrm{C}|k|^{\beta/2}$$

where

$$\beta = \max\{j \mid \mathrm{H}_c^j \neq 0\}, \qquad \mathrm{C} = \sum_j \dim(\mathrm{H}_c^j).$$

[10] Precisely, that of some ℓ-adic complex for a prime ℓ invertible in k.

[11] We are omitting some considerations involving the choice of an auxiliary prime number ℓ different from the characteristic of k, which are not essential in this sketch.

[12] Which has been called the most important result in number theory of the 20th century.

[13] More precisely, on the geometric object that gives rise to t; these conditions are "local" and usually fairly easily checked.

Hence, it is clear that a successful application of Deligne's work requires two extra ingredients to reach a non-trivial outcome (and these ingredients must be correspondingly refined to obtain sharper bounds):

1. One should, at the minimum, prove that $\beta < 2d$, since otherwise the bound is of size $|k|^d = |\mathrm{P}_d(k)|$, which is trivial (because the summands $t(f)$ are bounded in practice). Of course, the smaller β is, the better the result.
2. But one must also find a manageable upper-bound for C, which should be as much as possible independent of k, since otherwise it could swamp the gain from the first point (e.g., if $\beta = d - 1$ but $\mathrm{C} = |k|$, then the estimate is again trivial).

In the cases considered by Sawin and Shusterman, the base field k is *fixed* and the degree d of the polynomials, hence the dimension of the underlying summation variety, tend to infinity. This has a considerable impact on the ease of applicability of Deligne's work. In particular, note that getting non-trivial bounds now essentially requires that $2d - \beta$ tends to infinity with d (otherwise the gain from the trivial bound is at most $|k|^{\mathrm{A}}$, for some fixed $\mathrm{A} \geqslant 1$, and this is a constant). In practice, we would like to have a power-saving, which means a bound of the type

$$\sum_{f \in \mathrm{P}_d(k)} t(f) \ll |k|^{d(1-\delta)}$$

for some $\delta > 0$), which requires β to grow like a $(2-\delta)d$ for some $\delta > 0$ (with $\delta \to 1$ corresponding to square-root cancellation).

This is probably the more challenging of the two problems above, but the second is also far from simple, and in fact is closely related. Geometrically, bounding C in this situation corresponds to finding upper-bounds for sums of "Betti numbers" on varieties of increasing dimension. In fact, if we now write C_d for the constant C above, then the bound

$$\left| \sum_{f \in \mathrm{P}_d(k)} t(f) \right| \leqslant \mathrm{C}_d |k|^{\beta/2}$$

can only be interesting if C_d grows *at most exponentially* with d, say $\mathrm{C}_d \leqslant \mathrm{A}^d$ for some $\mathrm{A} \geqslant 1$. Moreover, if in fact $\mathrm{C}_d \geqslant \mathrm{A}^d$ with $\mathrm{A} > 1$, then the bound now requires, indeed, to prove that $\beta \leqslant 2(1-\delta)d$ for some $\delta > 0$ to be non-trivial, namely

$$\left| \sum_{f \in \mathrm{P}_d(k)} t(f) \right| \leqslant \mathrm{A}^d |k|^{(1-\delta)d}$$

will give cancellation for $|k|$ large enough (depending on A, roughly $|k| > \mathrm{A}^{1/\delta}$).

This point is important, because most of the known explicit upper-bounds for Betti numbers (some of which could in principle be applicable) give an estimate for C which is typically super-exponential (see for instance the bounds by Katz in [25], or the general recent development [35] of Sawin's "quantitative sheaf theory"). Thus Sawin and Shusterman cannot rely on off-the-shelf tools here either.[(14)]

[(14)] Note a difference with many previous applications of Deligne's Riemann Hypothesis in classical analytic number theory, where the underlying summation variety is typically fixed, or of fixed dimension,

How do Sawin and Shusterman handle these difficulties? Remarkably, the proofs of the two parts of Theorem 1.1, and that of Theorem 1.3, use three different approaches to bounding β and C.

1. In the proof of the twin prime result [36], the key tool are the so-called *vanishing cycles*, a fundamental part of the formalism of étale cohomology, which has its origins in the methods developed by Lefschetz to study algebraic surfaces, by comparing the desired invariants (cohomology groups) of a "generic" variety, and of a "specialization". The (very rough) idea is that a specialization (or "deformation") might become geometrically extremely simple, in such a way that the relevant cohomology groups are easily computable (e.g., by reduction to one-dimensional cases, which are well-understood); if one can control the difference between the generic and special invariants, then one gets information on the difficult case.
2. In the quadratic Bateman–Horn problem [37], the method is in some sense simpler. Indeed, a commonly-used approach to find good bounds on (the analogue of) β for multi-variable exponential sums, already used for instance by Deligne for bounds for additive character sums or hyper-Kloosterman sums [15, Th. 7.4], is the "comparison of cohomology with and without support". Again, very roughly described, this uses the fact that in addition to cohomology groups H_c^j (with compact support), one can define groups H^j (cohomology without support condition); it is a very general fact (Artin's vanishing theorem) that for a variety like an affine space, under some conditions, we have $\mathrm{H}_c^j = 0$ for $0 \leqslant j < d$ (which doesn't help for estimating β) whereas $\mathrm{H}^j = 0$ for $d < j \leqslant 2d$. Hence, if one can prove that $\mathrm{H}^j = \mathrm{H}_c^j$ for (say) $j \neq d$, it follows that $\beta \leqslant d$, and in fact that only H_c^d may be non-zero (and typically is).
3. Both of the previous methods belong to the toolkit of algebraic geometry since the 1960's. However, in Theorem 1.3, Sawin uses a much more recent ingredient, namely the theory of the *characteristic cycle* of Beĭlinson and Saito (see [9] and [29]). The author of this report is far from being able to say much about this topic, except that this provides a means to understand the so-called wild ramification phenomena on algebraic varieties of dimension $\geqslant 2$, somewhat similarly to the way older results like the Euler–Poincaré characteristic formula of Grothendieck–Ogg–Shafarevich describe certain global invariants for wildly ramified sheaves in terms of local data (see for instance [24, 2.3.1]).

In all three cases, what Sawin and Shusterman actually prove is (under suitable assumptions) that $\mathrm{H}_c^j = 0$ except (at worse) for $j \in \{d, d+1\}$. This does not lead to perfect square-root cancellation, but for a fixed finite field, it is essentially as good as that. Moreover, they are also able to obtain upper-bounds for the dimensions of the remaining spaces H_c^d and H_c^{d+1} (this is intuitively reasonable in the first approach at

in which case bounds like those of Katz (and earlier ones due to Bombieri) are quite sufficient, and the key difficulty is to get analogues of the bounds for β (but see also [18] for cases where the Betti number bounds also require some innovation).

least, because one can expect that a good choice of deformation or specialization will not only reveal vanishing properties, but also give some insight on the dimensions of the spaces when non-zero).

Sawin [31] has written a very insightful and intuitive survey of the applications of the first method (vanishing cycles), which we recommend very warmly (these applications include, e.g., the proof [34] of the function-field version of the Michel–Venkatesh mixing conjecture). Similarly, Sawin and Shusterman [37, § 1.4.1] explain on an intuitive geometric level the ideas used in the second method). For the sake of illustrating the type of arguments involved, we will conclude this section with just a hint of the latter (this method appears in [37, § 3]), noting that it also implies the vanishing statement used in [36]. We assume here familiarity with the formalism of étale cohomology, and use the corresponding standard notation (but this short discussion can be safely skipped).

Sketch of proof. The sums we consider are over polynomials in $k[\mathrm{T}]$ of degree at most d, viewed as points in $\mathrm{P}_d(k)$, where P_d is the affine d-space of coefficients. We have a fixed squarefree polynomial $g \in k[\mathrm{T}]$ of degree $\geqslant d$, with zero set $\mathrm{Z} \subset \bar{k}$, and a distinguished zero $z_0 \in \mathrm{Z}$. For any $z \in \mathrm{Z}$, we denote by ev_z the morphism $\mathrm{P}_d \to \mathbf{A}^1$ given by evaluation at z.

Fix a prime ℓ distinct from the characteristic of k. The trace functions involved in the sums of interest are associated to an ℓ-adic sheaf $\mathscr{G}$ on P_d such that the pullback of $\mathscr{G}$ to P_d over $\bar{k}$ is of the form

$$\mathscr{F} = \bigotimes_{z \in \mathrm{Z}} \mathrm{ev}_z^* \mathscr{F}_z,$$

where the factors satisfy the following conditions:

1. Each $\mathscr{F}_z$ is an ℓ-adic sheaf on $\mathbf{A}^1_{\bar{k}}$ without punctual sections and tamely ramified at ∞.
2. For the distinguished point z_0, there exists $w_0 \in \bar{k}$ such that $\mathscr{F}_{z_0}$ is the extension by zero from $\mathbf{A}^1 - \{w_0\}$ of $\mathscr{L}_{\chi(\mathrm{T}-w_0)}$, a shifted Kummer sheaf associated to a non-trivial multiplicative character χ.

Remark 6.2. — So, if we had $\mathrm{Z} \subset k$, and all the above data were defined over k, we would have the trace function

$$t(f) = \chi(f(z_0) - w_0) \prod_{z \in \mathrm{Z} - \{z_0\}} t_{\mathscr{F}_z}(f(z)).$$

Using the Chinese Remainder Theorem, it is not difficult to show that, for instance, the function $f \mapsto \lambda_2(f + h)$ is of this form.

Under these conditions, we have the following vanishing result (combining the statement of [37, Corollary 3.7, Lemma 3.13] with the remark after [37, Def. 3.8]):

Theorem 6.3. — *We have* $\mathrm{H}^j_c(\mathrm{P}_{d,\bar{k}}, \mathscr{F}) = \{0\}$ *unless* $j \in \{d, d+1\}$. *Moreover, the sum of the dimensions of* $\mathrm{H}^j_c(\mathrm{P}_{d,\bar{k}}, \mathscr{F})$ *is at most equal to the coefficient of* B^d *in the polynomial*

$$\prod_{z \in \mathrm{Z}} (\mathrm{rank}(\mathscr{F}_z)(1 + \mathrm{B}) + \mathrm{rank}(\widehat{\mathscr{F}_z})\mathrm{B}) \in \mathbf{Z}[\mathrm{B}]$$

where $\widehat{\mathscr{F}_z}$ *is the* ℓ*-Fourier transform of* $\mathscr{F}_z$.

The proof of the vanishing is intricate. It is based on the following steps:

Step 1. The shifted sheaf $\mathscr{F}[d]$ is a *perverse sheaf* ([37, Lemma 3.6]). This fundamental "regularity" property implies in particular that $\mathrm{H}^j_c(\mathrm{P}_{d,\bar{k}}, \mathscr{F}) = 0$ for $j < d$; it is derived from the local nature of perverse sheaves, which is used to find an étale-local model of $\mathscr{F}$, and ultimately from the assumption that the $\mathscr{F}_z$ have no punctual sections.

Step 2. Let $\bar{\mathrm{P}}_d$ be the natural compactification of P_d, isomorphic to the projective d-space. Let H_{z_0} in $\bar{\mathrm{P}}_d$ be the projective closure of the affine hyperplane defined by $f(z_0) = w_0$. Let $\widetilde{\mathscr{F}}$ be the extension by zero to $\bar{\mathrm{P}}_d - \mathrm{H}_{z_0}$ of the restriction of $\mathscr{F}$ to $\mathrm{P}_d - \mathrm{H}_{z_0}$. Let v be the open immersion of $\bar{\mathrm{P}}_d - \mathrm{H}_{z_0}$ in $\bar{\mathrm{P}}_d$. There is a general excision long exact sequence

$$\cdots \to \mathrm{H}^j_c(\bar{\mathrm{P}}_d - \mathrm{H}_{z_0}, \widetilde{\mathscr{F}}) \to \mathrm{H}^j(\bar{\mathrm{P}}_d - \mathrm{H}_{z_0}, \widetilde{\mathscr{F}}) \to \mathrm{H}^j(\mathrm{H}_{z_0}, v_*\widetilde{\mathscr{F}}) \to \cdots$$

which shows that it is enough to check that if $j > d$, we have

$$\mathrm{H}^j(\bar{\mathrm{P}}_d - \mathrm{H}_{z_0}, \widetilde{\mathscr{F}}) = \mathrm{H}^j(\mathrm{H}_{z_0}, v_*\widetilde{\mathscr{F}}) = 0.$$

The vanishing of the left-hand side holds by Artin's vanishing theorem for an ℓ-adic sheaf on an affine variety of dimension d.

Step 3. The trickiest part of the proof is the fact that $\mathrm{H}^j(\mathrm{H}_{z_0}, v_*\widetilde{\mathscr{F}}) = 0$ for $j > d$. Here, the point is that the restriction of $v_*\widetilde{\mathscr{F}}[d]$ to H_{z_0} is at least semi-perverse (by general principles), so its stalks are supported in degree $\leqslant d$. By delicate arguments (which use the full force of the assumptions on the sheaves $\mathscr{F}_z$; in particular, the "multiplicativity" of the Kummer sheaves is exploited, which explains the particular restriction on the special point z_0), Sawin and Shusterman prove ([37, Lemmas 3.4, 3.5]) that $v_*\widetilde{\mathscr{F}}$ is supported on *finitely many points*,[(15)] hence its cohomology is supported in the same degrees as its stalks.

7. CONCLUSION OF THE PROOF

We now discuss briefly the end of the proofs of both parts of Theorem 1.1. In the case of twin primes, there is not much left to be done: the cohomological estimates are sufficient to provide bounds for the sums of the type (4), e.g.,

$$\sum_{q \sim y} \mu(q),$$

which (given the ranges of values of y considered and the strength of the bounds) lead to the result.

In the case of the Bateman–Horn conjecture, the estimate for (3), namely

$$\sum_{\substack{n \sim x \\ n \equiv \alpha \bmod b}} \mu\Big(\frac{n^2 + a}{b}\Big)$$

[(15)] At most those $f \in \bar{\mathrm{P}}_d$ such that $f(z)$ is a singularity of $\mathscr{F}_z$ for more than d zeros z of g.

is only good enough to deal with the terms where $y \geqslant x^{1+\varepsilon}$ in (2), for some $\varepsilon > 0$ arbitrarily small. It turns out that handling the remaining small range $x \leqslant y \leqslant x^{1+\varepsilon}$ is very involved. Sawin and Shusterman proceed roughly by writing the sum as

$$-\sum_{x\leqslant y\leqslant x^{1+\varepsilon}} \mathbf{log}(y) \sum_{b\sim y} \mu(b) \sum_{\substack{n\sim x\\ n^2+a\equiv 0 \bmod b}} 1.$$

Summing over all solutions of $n^2 + a \equiv 0 \bmod b$ with $n \sim y$, they detect those with $n \sim x$ using additive characters modulo b. This leads to the goal of estimating now sums of the type

$$\sum_{b\sim y} \mu(b) \sum_{n^2+a\equiv 0 \bmod b} \psi(n),$$

where ψ is a non-trivial additive character of $\mathscr{O}/b\mathscr{O}$.

Thus the situation is reminiscent of the question of equidistribution of roots of quadratic congruences, and indeed the long section 7 of [37] handles this problem by developing, in this context, some tools which are close analogues of those involved in the classical studies of quadratic congruences for integers, such as in [21] and [22]. In particular, this includes the parameterization of roots of $n^2 + a = 0$ in terms of classes of binary quadratic forms of discriminant a, which goes back to Gauss in principle (including the distinction between definite and indefinite forms, with analogues of either Heegner points or closed geodesics in the upper half-plane). Ultimately, these parameterizations lead again to sums of trace functions of the type

$$\sum_{\substack{n,m\\ \mathrm{Q}(n,m)\sim x}} \mu(\mathrm{Q}(n,m))\psi(n^{-1}m)$$

where $\mathrm{Q} = \alpha \mathrm{X}^2 + \beta \mathrm{XY} + \gamma \mathrm{Y}^2$ is a quadratic form of discriminant $4\alpha\gamma - \beta^2 = 4a$. Sawin and Shusterman are able to handle such sums by fixing the variable m, and viewing the sum over n as a combination of one or two sums of trace functions over polynomials of a given degree.

Remark 7.1. — It is of course natural to ask how far the methods of Sawin and Shusterman can go; for instance, how much harder is the Bateman–Horn conjecture for polynomials F of degree 3? As explained again in [37], it does seem that significant new ideas are involved, despite the fact that Sawin and Shusterman do prove non-trivial bounds for sums of $\mu(\mathrm{F}(n))$ for any separable polynomial F. One issue, valid over the integers also, is that there is currently no satisfactory understanding of the roots of cubic congruences comparable to what is known in the quadratic case.

8. LEVEL OF DISTRIBUTION

We conclude with a quick discussion of the results of Sawin [32] concerning the level of distribution of arithmetic functions of polynomials over finite fields. These apply to an extensive class of functions, namely the so-called *factorization functions*, which are

roughly those functions f of polynomials $n \in k[\mathrm{T}]$ which can be expressed in terms of the factorization pattern of n, i.e., the number of irreducible factors of each degree.

More precisely, suppose given an integer d and a finite-dimensional representation

$$\varrho\colon \mathfrak{S}_d \to \mathrm{GL}_r(\mathbf{C}).$$

For $n \in k[\mathrm{T}]$ squarefree of degree d, one can define $f_\varrho(n)$ to be the value of the character $\operatorname{tr}\varrho$ at the permutation corresponding to the Frobenius acting on the roots of the polynomial n. But there is in fact a natural extension to *all* polynomials of degree d (see [32, § 1.6] and [33, § 3]), namely

$$f_\varrho(n) = \operatorname{tr}(\mathrm{Fr}_k \mid (\mathrm{V}_f \otimes \varrho)^{\mathfrak{S}_d}),$$

where V_f is the permutation representation of $\mathfrak{S}_d$ associated to the permutation action on the tuples $(a_i) \in \bar{k}^d$ such that $f = (\mathrm{T} - a_1)\cdots(\mathrm{T} - a_d)$. This leads to a fairly straightforward interpretation of these arithmetic functions as trace functions, the key point being that if π is the morphism from the affine space of dimension d to the space of monic polynomials of degree d mapping $(a_1, \dots, a_d)$ to $(\mathrm{T} - a_1)\cdots(\mathrm{T} - a_d)$, then the representation V_f (with its Frobenius action) can be identified as the stalk at f of the sheaf $\pi_*\bar{\mathbf{Q}}_\ell$.

Remark 8.1. — This construction had already been exploited by Sawin [33] to study sums of factorization functions over "short intervals" (i.e., sums of $f_\varrho(n+a)$ over n of degree $\leqslant d$ where a is a fixed polynomial with $\deg(a) > d$). Other related interesting works in the case of short intervals include those of Rodgers [28] and of Hast–Matei [20].

Example 8.2. — The following table indicates which representations give rise to some of the standard arithmetic functions (in the last line, we have a virtual representation, i.e., we take the corresponding linear combinations for the representations indicated):

ϱ	$f_\varrho(n)$
signature	$(-1)^{\deg(n)}\mu(n)$
$(\mathbf{C}^m)^{\otimes d}$	$\tau_m(n)$, the m-th divisor function
$\sum_{i=0}^{d-1}(-1)^i \bigwedge^i(\mathbf{C}^d)$	$\Lambda(n)$

See [32, § 8.1, 8.2, 8.4] as well as [33, § 3].

The main result of Sawin is an explicit upper bound for sums of functions of type f_ϱ on arithmetic progressions to squarefree moduli, which translates to level of distribution $> 1/2$ for individual progressions when $|k|$ is large enough, at least for certain representations ϱ, including those in the table above (see [32, Th. 1.7]):

THEOREM 8.3 (Sawin). — *Let $q \in k[\mathrm{T}]$ be squarefree and $a \in (k[\mathrm{T}]/qk[\mathrm{T}])^\times$. There exist explicit quantities $c_1(\varrho)$ and $c_2(\varrho)$ such that for any $d \geqslant \deg(q)$, we have*

$$\Big| \sum_{\substack{\deg(n)=d \\ n\equiv a \bmod q}} f_\varrho(n) - \frac{1}{\varphi(q)} \sum_{\substack{\deg(n)=d \\ (n,q)=1}} f_\varrho(n) \Big| \leqslant 2(c_1(\varrho) + |k|^{1/2} c_2(\varrho))|k|^{(d-\deg(q))/2}.$$

Remark 8.4. — (1) For the Möbius or von Mangoldt functions, this theorem improves (except in very few cases) on the corresponding ones from [36] and [37]. However, it is not applicable in the proof of Theorem 1.1, because a function like $\mu(\mathrm{F}(n))$ for F of degree at least 2 is typically *not* a factorization function.

(2) This "direct" translation into trace functions explains why there is no requirement on the size of k, except that it be large enough, in Theorem 1.3.

REFERENCES

[1] J-M. Deshouillers: *Progrès récents des petits cribles arithmétiques*, Séminaire Bourbaki, exposé 520, Lecture Notes in Math. 510 (1979), Springer.

[2] J-M. Deshouillers: *Théorème de Fermat: la contribution de Fouvry*, Séminaire Bourbaki, exposé 648, Astérisque 133–134 (1986), SMF.

[3] Ph. Michel: *Progrès récents du crible et applications*, Séminaire Bourbaki, exposé 842, Astérisque 252 (1998).

[4] B. Host: *Progressions arithmétiques dans les nombres premiers [d'après B. Green et T. Tao]*, Séminaire Bourbaki, exposé 944, Astérisque 307 (2006).

[5] E. Kowalski: *Écarts entre nombres premiers successifs*, Séminaire Bourbaki, exposé 959, Astérisque 311 (2007).

[6] J. Wolf: *Arithmetic and polynomial progressions in the primes [after Gowers, Green, Tao and Ziegler]* , Séminaire Bourbaki, exposé 1053, Astérisque 352 (2013).

[7] E. Kowalski: *Gaps between prime numbers and prime numbers in arithmetic progressions, after Y. Zhang and J. Maynard*, Séminaire Bourbaki, exposé 1084, Astérisque 367–368 (2015).

[8] P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Mathematics of Computation 16 (1962), 363–367.

[9] A.A. Beĭlinson: *Constructible sheaves are holonomic*, Selecta Math. (N.S.) 22 (2016), 1797–1819.

[10] B. Birch and H. Swinnerton-Dyer: *Note on a problem of Chowla*, Acta Arithmetica 5 (1959), 417–423.

[11] S.D. Cohen: *The distribution of polynomials over finite fields*, Acta Arith. 17 (1970), 255–271.

[12] K. Conrad: *Irreducible values of polynomials: a non-analogy*, in "Number fields and function fields – two parallel worlds", 71–85, Progr. Math., 239, Birkhäuser Boston, Boston, MA, 2005.

[13] B. Conrad, K. Conrad and R. Gross: *Prime specialization in genus* 0, Trans. Amer. Math. Soc. 360 (2008), 2867–2908.

[14] P. Deligne: *La conjecture de Weil, II*, Publ. Math. de l'IHÉS 52 (1980), 137–252.

[15] P. Deligne: *Applications de la formule des traces aux sommes trigonométriques*, in SGA $4\frac{1}{2}$, Lecture Notes Math. 569, Springer, 1977.

[16] R. Dedekind and H. Weber: *Theorie der algebraischen Funktionen einer Veränderlichen*, J. reine angew. Math. XCII (1882), 181–290; `http://gdz.sub.uni-goettingen.de/dms/resolveppn/?PPN=GDZPPN002158280`

[17] A. Entin: *On the Bateman-Horn conjecture for polynomials over large finite fields*, Compositio Math. 152 (2016), 2525–2544.

[18] É. Fouvry, E. Kowalski and Ph. Michel: *Algebraic twists of modular forms and Hecke orbits*, Geom. Funct. Anal. 25 (2015), 580–657.

[19] C. Hall: L-*functions of twisted Legendre curves*, J. Number Theory 119 (2006), 128–147.

[20] D.R. Hast and V. Matei: *Higher moments of arithmetic functions in short intervals: a geometric perspective*, IMRN 2018, `doi:10.1093/imrn/rnx310`.

[21] C. Hooley: *On the number of divisors of quadratic polynomial*s, Acta Math. 110 (1963), 97–114.

[22] H. Iwaniec: *Almost-primes represented by quadratic polynomials* , Invent. math. 47 (1978), 171-188.

[23] H. Iwaniec and E. Kowalski: *Analytic number theory*, AMS Colloquium Publ. 53, AMS (2004).

[24] N.M. Katz: *Gauss sums, Kloosterman sums and monodromy*, Annals of Math. Studies 116, Princeton University Press, 1988.

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

[26] N.M. Katz: *Larsen's alternative, moments and the monodromy of Lefschetz pencils*, in "Contributions to automorphic forms, geometry and number theory (collection in honor of J. Shalika's 60th birthday)", Johns Hopkins University Press 2004, 521–560.

[27] R. Nunes: *On the least squarefree number in an arithmetic progression*, Mathematika 63 (2017), 483–498.

[28] B. Rodgers: *Arithmetic functions in short intervals and the symmetric group*, Alg. Number Theory 12 (2018), 1243–1279.

[29] T. Saito: *The characteristic cycle and the singular support of a constructible sheaf*, Invent. math. 207 (2017), 597–695.

[30] L. Spiegelhofer: *The level of distribution of the Thue–Morse sequence*, Compositio Math. 156 (2020), 2560–2587.

[31] W. Sawin: *Singularities and vanishing cycles in number theory over function fields*, Res. Math. Sci. 7 (2020), paper no. 12; also `arXiv:2005.09693v2`.

[32] W. Sawin: *Square-root cancellation for sums of factorization functions over square-free progressions in* $\mathbf{F}_q[t]$, preprint (2021), arXiv:2102.09730.

[33] W. Sawin: *Square-root cancellation for sums of factorization functions over short intervals in function fields*, Duke Math. J. (to appear); arXiv:1809.05137.

[34] W. Sawin: *Bounds for the stalks of perverse sheaves in characteristic p and a conjecture of Shende and Tsimerman*, Invent. math. (2021), 1–32.

[35] W. Sawin, *mis en forme* by A. Forey, J. Fresán and E. Kowalski: *Quantitative sheaf theory*, preprint (2021), arXiv:2101.00635.

[36] W. Sawin and M. Shusterman: *On the Chowla and twin primes conjectures over* $\mathbf{F}_q[\mathrm{T}]$, preprint (2018), arXiv:1808.04001.

[37] W. Sawin and M. Shusterman: *Möbius cancellation on polynomial sequences and the quadratic Bateman-Horn conjecture over function fields*, preprint (2020), arXiv:2008.09905.

Emmanuel Kowalski

ETH Zürich, DMATH, Rämistrasse 101
8092 Zürich, Switzerland
E-mail : kowalski@math.ethz.ch