Improved Elliptic Curve Hashing and Point Representation

Taechan Kim, Mehdi Tibouchi

Improved Elliptic Curve Hashing and Point Representation

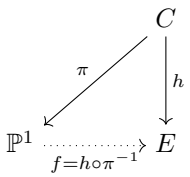
Taechan Kim and Mehdi Tibouchi

NTT Secure Platform Laboratories
{taechan.kim,tibouchi.mehdi}@lab.ntt.co.jp

Abstract. For a large class of functions $f\colon \mathbb{F}_q \to E(\mathbb{F}_q)$ to the group of points of an elliptic curve $E/\mathbb{F}_q$, Farashahi et al. (Math. Comp. 2013) established that the map $(u,v) \mapsto f(u) + f(v)$ is regular, in the sense that for a uniformly random choice of $(u,v) \in \mathbb{F}_q^2$, the elliptic curve point $f(u) + f(v)$ is close to uniformly distributed in $E(\mathbb{F}_q)$. This result has several applications in cryptography, mainly to the construction of elliptic curve-valued hash functions and to the "Elligator Squared" technique for representating uniform points on elliptic curves as close to uniform bitstrings. In this paper, we improve upon Farashahi et al.'s character sum estimates in two ways: we show that regularity can also be obtained for a function of the form $(u,v) \mapsto f(u) + g(v)$ where g has a much smaller domain than $\mathbb{F}_q$, and we prove that the functions f considered by Farashahi et al. also satisfy requisite bounds when restricted to large intervals inside $\mathbb{F}_q$. These improved estimates can be used to obtain more efficient hash function constructions, as well as much shorter "Elligator Squared" bitstring representations.

1 Introduction

Mapping to elliptic curves. Many elliptic curve cryptosystems involve representing a base field element $u \in \mathbb{F}_q$ as a rational point $f(u) \in E(\mathbb{F}_q)$ of the elliptic curve $E/\mathbb{F}_q$ where the computations are carried out. Moreover, it is often desirable for the corresponding "encoding function" $f\colon \mathbb{F}_q \to E(\mathbb{F}_q)$ to be efficiently computable in constant time (i.e. independently of the value of u) rather than in an iterative, probabilistic manner. Many methods have been proposed to construct such functions f, starting with the technique used in Boneh and Franklin's identity-base encryption scheme [3], and especially with Shallue and van de Woestijne's construction [12], which applies to essentially all isomorphism classes of elliptic curves. All constructions proposed so far are of a geometric nature. For some of them (notably Icart's function [9] and its variants), there exists a diagram:



(1)

where $h\colon C \to E$ is a covering of E over $\mathbb{F}_q$, and $\pi\colon C \to \mathbb{P}^1$ induces a *bijection on points* (it is an *exceptional cover* of $\mathbb{P}^1$ in the terminology of Fried [8]). The map $f\colon \mathbb{F}_q \subset \mathbb{P}^1(\mathbb{F}_q) \to E(\mathbb{F}_q)$ can then be defined on points as $h \circ \pi^{-1}$. The other constructions (including Shallue and van de Woestijne's [12] and their variants, say SW-type encodings) arise from several coverings and admit a similar geometric description [14].

Hashing to elliptic curves. The first application of the above constructions to cryptography was hashing to curve points. It is common, especially in pairing-based cryptography, that a certain value (a message, an identity, etc.) has to be hashed to an element of the group of points of an elliptic curve, in such a way that the hash function can be reasonably modeled as a random oracle. One approach that has been considered is to take a function $f\colon \mathbb{F}_q \to E(\mathbb{F}_q)$ as above, a hash function $\mathfrak{h}\colon \{0,1\}^* \to \mathbb{F}_q$, and combine the two together by hashing a message m as $\mathfrak{H}(m) = f\big(\mathfrak{h}(m)\big)$. This is actually sufficient for certain cryptographic schemes, in the sense that they can be proved secure in the random oracle model for $\mathfrak{h}$, but this is not the case in general. Indeed, $\mathfrak{H}$ is typically easy to distinguish from a random oracle to $E(\mathbb{F}_q)$, because the image of f consists of only a fraction of all points on the curve and image membership can be tested for efficiently.

Formal conditions under which a hash function construction can securely replace a random oracle in essentially any cryptographic protocol are given by Maurer et al.'s indifferentiability framework, and Brier et al. [4] have applied them to the elliptic curve setting, establishing that a hash function construction $\mathfrak{H}(m) = F\big(\mathfrak{h}(m)\big)$ is indifferentiable (and hence can be used securely in the random oracle model for $\mathfrak{h}$) for any map F to $E(\mathbb{F}_q)$ which is efficiently computable, efficiently samplable (one can compute a close to uniform preimage of any point efficiently) and regular (the image of a uniformly random element is close to uniformly distributed in $E(\mathbb{F}_q)$). They call such a function F admissible, and prove the admissibility of:

$$F_1\colon \mathbb{F}_q \times \mathbb{Z}/N\mathbb{Z} \to E(\mathbb{F}_q);\ (u,v) \mapsto f(u) + vG \tag{2}$$

for any curve such that $E(\mathbb{F}_q)$ is a cyclic group of order N generated by G and f is a function $f\colon \mathbb{F}_q \to E(\mathbb{F}_q)$ verifying mild conditions. That construction is rather slow, however. They also prove the admissibility of the much more efficient construction:

$$F_2\colon \mathbb{F}_q \times \mathbb{F}_q \to E(\mathbb{F}_q);\ (u,v) \mapsto f(u) + f(v) \tag{3}$$

but only when f is Icart's function, and the proof involves rather painful geometric arguments. A much simpler approach was later proposed by Farashahi et al. [6], who prove that the function F_2 is regular whenever f satisfies certain bounds expressed in terms of character sums on $E(\mathbb{F}_q)$, and they show how such bounds can be obtained for any map f of Icart type or SW type as a consequence of the Riemann hypothesis for function fields. This yields a relatively efficient hash function construction to elliptic curves from any function f of one of those forms.

Representing elliptic curve points as uniform bitstrings. For certain applications related to anonymity and privacy, elliptic curve cryptography presents a weakness: points on a given elliptic curve, when represented in a usual way, are easy to distinguish from random bit strings. This makes it relatively easy for an attacker to distinguish ECC traffic (the transcripts of multiple ECDH key exchanges, say) from random traffic, and then proceed to intercept, block or otherwise tamper with such traffic.

An efficient approach to solve this problem was proposed by Bernstein et al. [2]. Their idea is to leverage an efficiently computable, efficiently invertible algebraic function ι that maps the integer interval $S = \{0, \dots, (p-1)/2\}$ *injectively* to $E(\mathbb{F}_p)$. Since ι is injective, a uniformly random point P in $\iota(S) \subset E(\mathbb{F}_p)$ has a uniformly random preimage $\iota^{-1}(P)$ in S, so that P can be represented as the binary expansion of the integer $\iota^{-1}(P)$ if it exists. If p is close to a power of 2, a uniform point in $\iota(S)$ will have a close to uniform bit string representation. This approach is simple and efficient, but limited to special elliptic curves such as Edwards and Montgomery curves [7,2] for which ι exists.

A variant of that approach, "Elligator Squared", was recently suggested by Tibouchi [13], eliminating most of the limitations of Bernstein et al.'s method. The idea is to represent $P \in E(\mathbb{F}_q)$ by a randomly sampled preimage under an admissible encoding F_2 of the form (3). By Farashahi et al.'s results, such encodings can be obtained for all known point encodings, and in particular for all elliptic curves. Moreover, the representation of a uniformly random point is close to uniformly distributed in $(\mathbb{F}_q)^2$ by the regularity of F_2. Since F_2 is essentially surjective, no rejection sampling is necessary contrary to Bernstein et al.'s method, yielding record performance [1]. Its main drawback, however, is that points are represented as elements of $(\mathbb{F}_q)^2$ (or rather, as bitstring representations thereof), which take up at least twice as much space as Bernstein et al.'s representations.

Our contributions. In this paper, we revisit Farashahi et al.'s character sum estimates with the goal of improving the efficiency of hash function constructions and Tibouchi's "Elligator Squared" method for representing points on elliptic curves as uniform bitstrings.

Our improvements are twofold. Firstly, in §2, we establish that for any function f subject to the same conditions as introduced by Farashahi et al. (and verified by constructions of the form (1) and SW-type encodings), the following map is regular:

$$F_3 \colon \mathbb{F}_q \times V \to E(\mathbb{F}_q);\ (u, v) \mapsto f(u) + g(v) \tag{4}$$

for any map $g \colon V \to E(\mathbb{F}_q)$ from a set of cardinality $\#V \gg q^\varepsilon$ and with small collision probability. This implies that the following variants of (2), (3) are also regular:

$$\begin{aligned} F_1' \colon \mathbb{F}_q \times [0, q^\varepsilon) &\to E(\mathbb{F}_q) & \qquad F_2' \colon \mathbb{F}_q \times V_\varepsilon &\to E(\mathbb{F}_q) \\ (u, v) &\mapsto f(u) + vG & (u, v) &\mapsto f(u) + f(v) \end{aligned} \tag{5}$$

where $G \in E(\mathbb{F}_q)$ is any point of order $\geq q^\varepsilon$ and $V_\varepsilon \subset \mathbb{F}_q$ is any subset of cardinality $\gg q^\varepsilon$. This result is technically very simple, but has valuable consequences.

It is especially interesting for point representation, as it provides a way to obtain Elligator Squared-like representations of length $(1+\varepsilon)\log_2 q$ instead of $2\log_2 q$ (by sampling preimages under F_2' instead of F_2), which makes the Elligator Squared construction almost as space efficient as Bernstein et al.'s. For hash function constructions, it says that indifferentiable hashing can be obtained from shorter random oracles, and also settles the long-standing open question of whether admissibility can be obtained for all elliptic curves at a cost of less than two base field exponentiations: indeed, for ε small enough, the scalar multiplication in F_1' becomes cheaper than a base field exponentiation!

Secondly, in §3, we show that for all nontrivial characters χ of $E(\mathbb{F}_q)$ and all *intervals*[1] $I \subset \mathbb{F}_q$, we obtain a bound of the form $\left|\sum_{u\in I} \chi\big(f(u)\big)\right| \ll \sqrt{q}\log p$, where p is the characteristic of $\mathbb{F}_q$, which is an extension of the result by Farashahi et al.. As a consequence, we get admissibility for the maps F_1', F_2' in which the variable u is taken from any large interval (of length $q/O(1)$) within $\mathbb{F}_q$. This is of practical relevance for hashing or point representation on elliptic curves defined over prime fields $\mathbb{F}_p$ when p is not pseudo-Mersenne (such as most pairing-friendly elliptic curves, and many other standardized curves). Indeed, when hashing to a 256-bit curve of that type, for example, one traditionally needs to obtain a hash value in $\mathbb{F}_p$, which typically involves reducing a digest of at least 384 bits modulo p, since usual hash functions return bitstrings. A similar problem arises when representing an Elligator Squared value $(u, v) \in (\mathbb{F}_p)^2$ (or $\mathbb{F}_p \times V_\varepsilon$ when using F_2') as a bitstring: to get uniform bitstrings, elements of $\mathbb{F}_p$ have to be greatly enlarged. Our result solves this problem completely by allowing u to be chosen from an interval of length a power of 2 (say $[0, 2^{255})$ in the 256-bit case), making it possible to use the output of a standard hash function directly, and to directly obtain representation as bitstrings instead of base field elements.

2 Stronger regularity bounds for encodings

In this section, we show how we can improve upon the regularity bounds from [6] for encodings to elliptic curves. We first formulate and prove a simple generalization of [6, Th. 3] in §2.1, and then discuss applications to elliptic curves in §2.2. We refer to the full version of this paper [10] for standard definitions regarding probability distributions on finite sets and regularity.

2.1 A general regularity bound

Let A be any finite abelian group (denoted additively), and $f\colon U \to A$, $g\colon V \to A$ arbitrary functions from finite sets U, V to A. We consider, for some $s \geq 1$, the following mapping:

$$F\colon U^s \times V \to A;\ (u_1, \ldots, u_s, v) \mapsto f(u_1) + \cdots + f(u_s) + g(v).$$

[1] An *interval* in a not necessarily prime finite field $\mathbb{F}_q$ is any subset of the form $H + x[m, \ldots, m+k]$ where H is an additive subgroup of $\mathbb{F}_q$, x an element of $\mathbb{F}_q$, and m, k non negative integers.

We can obtain bounds on the regularity of F from bounds on the character sums $S_f(\chi)$ defined by $S_f(\chi) = \sum_{u \in U} \chi\big(f(u)\big)$ for nontrivial characters χ of A on the one hand, and on the collision probability of g on the other hand. Indeed, the following theorem is a simple generalization of [6, Th. 3], proved in the full version of this paper [10].

Theorem 1. *Assume that for all nontrivial characters χ of A, the inequality $|S_f(\chi)| \leq S$ holds, and denote by ρ the collision probability of g. Then, the mapping F defined above is α-regular with $\alpha = (S/\#U)^s \sqrt{\rho \# A}$.*

Corollary 1. *Assume that for all nontrivial characters χ of A, the inequality $|S_f(\chi)| \leq S$ holds, and that ρ has preimage size bounded by some constant d (i.e. $\#\rho^{-1}(\{a\}) \leq d$ for all $a \in A$). Then, the mapping F defined above is α-regular with $\alpha = (S/\#U)^s \sqrt{d \# A / \# V}$.*

2.2 Application to elliptic curve encodings

Consider now an encoding $f \colon \mathbb{F}_q \to E(\mathbb{F}_q)$ to the group of points of an elliptic curve over $\mathbb{F}_q$, and recall the following definition proposed by Farashahi et al. [6].

Definition 1. *The encoding f is said to be B-well-distributed for some positive constant B if for all nontrivial characters χ of $E(\mathbb{F}_q)$, the character sum $S_f(\chi) = \sum_{u \in \mathbb{F}_q} \chi\big(f(u)\big)$ is bounded as $\big|S_f(\chi)\big| \leq B\sqrt{q}$.*

Farashahi et al. have shown that known encoding functions to elliptic curves (of the form (1) or SW-type encodings, say) are indeed B-well-distributed for some small B depending on the encoding construction but not on q.

Then, fix $g \colon V \to E(\mathbb{F}_q)$ any function from a set V of cardinality $\#V \geq q^\varepsilon$ and with preimage size bounded by d. Assuming that f is B-well-distributed, Corollary 1 shows that the map $F_3 \colon \mathbb{F}_q \times V \to E(\mathbb{F}_q)$ from (4), given by $F_3(u, v) = f(u) + g(v)$, is α-regular for $\alpha = \frac{B\sqrt{q}}{q} \sqrt{\frac{d \cdot (q + 2\sqrt{q} + 1)}{\#V}} \leq \frac{B\sqrt{d}}{q^{\varepsilon/2}} \cdot (1 + q^{-1/2})$, which is negligible for bounded B and d. As a result, the maps F_1', F_2' defined in (5) are indeed regular. If we assume furthermore that f is efficiently invertible and has preimage size bounded by d (which usually follows trivially for a suitable d from the fact that it is algebraic), then F_3 (and hence F_1', F_2') are also efficiently and uniformly samplable using Algorithm 1, and thus admissible [4].

Thus, if $\mathfrak{h}_1 \colon \{0,1\}^* \to \mathbb{F}_q, \mathfrak{h}_2 \colon \{0,1\}^* \to [0, q^\varepsilon)$ are hash functions modeled as independent random oracles, then $m \mapsto f\big(\mathfrak{h}_1(m)\big) + \mathfrak{h}_2(m)G$ is indifferentiable from a random oracle to $E(\mathbb{F}_q)$ for any element $G \in E(\mathbb{F}_q)$ of order $\geq q^\varepsilon$. This provides indifferentiable hashing to $E(\mathbb{F}_q)$ from as few as $(1 + \varepsilon) \log_2 q$ random oracle bits (and for ε small enough, it gives indifferentiable hashing for a smaller computational cost than 2 base fields exponentiations). Similarly, if $\mathfrak{h}_1 \colon \{0,1\}^* \to \mathbb{F}_q, \mathfrak{h}_2 \colon \{0,1\}^* \to V_\varepsilon$ are hash functions modeled as independent random oracles with $V_\varepsilon \subset \mathbb{F}_q$ a subset of cardinality $\geq q^\varepsilon$, then $m \mapsto f\big(\mathfrak{h}_1(m)\big) + f\big(\mathfrak{h}_2(m)\big)$ is indifferentiable from a random oracle to $E(\mathbb{F}_q)$. And in the spirit of [13], if $P \in E(\mathbb{F}_q)$ is a uniformly random point, then a uniformly random preimage

Algorithm 1 Preimage sampling algorithm for F_3 assuming f has preimage size bounded by d.

1: **function** SAMPLEPREIMAGE(P)
2: **repeat**
3: $v \xleftarrow{\$} V$
4: $Q \leftarrow P - g(v)$
5: $t \leftarrow \#f^{-1}(Q)$
6: $j \xleftarrow{\$} \{1, \ldots, d\}$
7: **until** $j \leq t$
8: $\{u_1, \ldots, u_t\} \leftarrow f^{-1}(Q)$
9: **return** (u_j, v)
10: **end function**

of P under F'_1 (resp. F'_2) is statistically close to uniform in $\mathbb{F}_q \times [0, q^\varepsilon)$ (resp. $\mathbb{F}_q \times V_\varepsilon$), and can be efficiently sampled using Algorithm 1, which provides a close-to-uniform point representation technique from a set of cardinality as small as $q^{1+\varepsilon}$.

As mentioned in the introduction, we can also extend those results to the restriction of f to a large enough interval of $\mathbb{F}_q$. Indeed, we introduce the following definition.

Definition 2. *The encoding f is said to be B-strongly well-distributed for some positive constant B if for all nontrivial characters χ of $E(\mathbb{F}_q)$ and all intervals $I \subset \mathbb{F}_q$, the restricted character sum $S_f(\chi; I) = \sum_{u \in I} \chi\big(f(u)\big)$ is bounded as $\big|S_f(\chi; I)\big| \leq B(1 + \log p)\sqrt{q}$, where p is the characteristic of $\mathbb{F}_q$.*

We will show in §3 that the same techniques as used by Farashahi et al. to show that encodings are well-distributed can be adapted to prove that they are also strongly well-distributed.

Now, consider a B-strongly well-distributed f to $E(\mathbb{F}_q)$ and a mapping $g\colon V \to E(\mathbb{F}_q)$ with preimage size bounded by d from a set of cardinality $\#V \geq q^\varepsilon$, and fix an interval $I \subset \mathbb{F}_q$ of cardinality $\#I \geq q/c$ for some constant c. Corollary 1 shows that the map $F_{3,I}\colon I \times V \to E(\mathbb{F}_q)$ given by $F_{3,I}(u, v) = f(u) + g(v)$ is α-regular for $\alpha = \frac{B(1+\log p)\sqrt{q}}{q/c}\sqrt{\frac{d\cdot(q+2\sqrt{q}+1)}{\#V}} \leq \frac{cB(1+\log p)\sqrt{d}}{q^{\varepsilon/2}} \cdot (1 + q^{-1/2})$, which is again negligible for bounded B, c, d. This is especially interesting in the case when $q = p$ is prime. Then, for some $\varepsilon > 0$, let $k_1 = \lfloor \log_2 p \rfloor$ and $k_2 = \lceil (1+\varepsilon) \log_2 p \rceil - k_1$, and identify bitstrings in $\{0,1\}^k$ with integers in $[0, 2^k)$. We can then introduce:

$$F'_{1,I}\colon \{0,1\}^{k_1+k_2} \to E(\mathbb{F}_p) \qquad F'_{2,I}\colon \{0,1\}^{k_1+k_2} \to E(\mathbb{F}_p)$$
$$(u, v) \mapsto f(u) + vG \qquad (u, v) \mapsto f(u) + f(v).$$

The previous bound says that both of $F'_{1,I}$ and $F'_{2,I}$ are regular. If f has bounded preimage size, they are thus both admissible encodings (using the variant of Algorithm 1 where only preimages in $f^{-1}(Q) \cap I$ are considered in Steps 5, 8 for preimage sampling) from *bitstrings* of length $k_1 + k_2 \sim (1 + \varepsilon) \log_2 p$ to $E(\mathbb{F}_p)$.

As a result, we get an efficient indifferentiable hash functions to $E(\mathbb{F}_p)$ from random oracles to the set $\{0,1\}^{k_1+k_2}$ of bitstrings of length $k_1+k_2 \sim (1+\varepsilon)\log_2 p$, as well as an efficient representation of uniform points in $E(\mathbb{F}_p)$ as close to uniform bitstrings of length $k_1+k_2 \sim (1+\varepsilon)\log_2 p$. This is a major improvement over the approach described in [4,13] for hashing and point representation, which requires strings of length $\sim (5/2)\log_2 p$ when p is not close to a power of 2 (not a pseudo-Mersenne prime, say).

3 Character sums on intervals of curves

Throughout this paper, a "curve" means a smooth, projective, geometrically integral curve over a finite field (the field $\mathbb{F}_q$ unless otherwise specified).

Let $h\colon X \to Y$ be a branched covering (i.e. a finite separable morphism) of curves over $\mathbb{F}_q$, and $\xi, \pi\colon X \to \mathbb{P}^1$ be rational functions. We also assume that Y has an $\mathbb{F}_q$ rational point, and fix the embedding $Y \hookrightarrow J$ of Y into its Jacobian variety J defined by that rational point. The goal of this section is to obtain bounds on character sums:

$$S_{h,\xi,\pi}(\chi,\omega;I) = \sum_{\substack{P\in X(\mathbb{F}_q)\\ \pi(P)\in I, \xi(P)\neq\infty}} \chi\big(h(P)\big)\omega\big(\xi(P)\big), \tag{6}$$

where χ is a nontrivial character of $J(\mathbb{F}_q)$, ω is a multiplicative character of $\mathbb{F}_q$, and $I \subset \mathbb{F}_q$ is any interval. Under mild conditions on h, ξ, π, we will obtain a bound of the form $\big|S_{h,\xi,\pi}(\chi,\omega;I)\big| \ll q^{1/2}\log p$, where p is the characteristic of $\mathbb{F}_q$. This extends the results of Farashahi et al. [6, §4] giving similar bounds (without the $\log p$ factor) in the case when $I = \mathbb{F}_q$, and makes it possible to prove that encoding functions $\mathbb{F}_q \to Y(\mathbb{F}_q)$ constructed as in (1) or SW-type encodings are *strongly well-distributed* in the sense of Definition 2. The idea is to express $S_{h,\xi,\pi}(\chi,\omega;I)$ in terms of the sums:

$$S_{h,\xi,\pi}(\chi,\omega,\psi) = \sum_{\substack{P\in X(\mathbb{F}_q)\\ \pi(P),\xi(P)\neq\infty}} \chi\big(h(P)\big)\omega\big(\xi(P)\big)\psi\big(\pi(P)\big) \tag{7}$$

for all additive characters ψ of $\mathbb{F}_q$, which are Artin character sums along X and can therefore be bounded using a theorem by Weil.

3.1 Background on Artin characters

Consider an abelian covering $\widetilde{Y} \to Y$ of curves over $\mathbb{F}_q$ with Galois group G (i.e. a finite morphism such that $\mathbb{F}_q(\widetilde{Y})/\mathbb{F}_q(Y)$ is abelian with Galois group G). Any character of G determines, via the Artin map, a corresponding character on the group of $\mathbb{F}_q$-divisors on Y prime to the ramification locus S of $\widetilde{Y} \to Y$, which extends to a multiplicative map $\chi\colon \mathrm{Div}_{\mathbb{F}_q}(Y) \to \mathbb{C}$ vanishing on divisors not prime to S. Let us call such a map χ an *Artin character* of Y. One associates

to χ a distinguished effective divisor $\mathfrak{f}(\chi)$ of support S called the conductor (in particular, if $\widetilde{Y} \to Y$ is unramified, $\mathfrak{f}(\chi) = 0$; the character itself is then said to be unramified).

We mainly consider three types of Artin characters: those arising from Artin–Schreier coverings, those arising from Kummer coverings, and those attached to characters of the Jacobian. They are discussed in more details in the full version of this paper [10], but suffice it to say that their actions on rational points $y \in Y(\mathbb{F}_q)$ are respectively given by $y \mapsto \psi(\pi(y))$ for some additive character ψ of $\mathbb{F}_q$ and some rational function π of Y not of the form $u^p - u$, $u \in \bar{\mathbb{F}}_q(Y)$ (Artin–Schreier), by $y \mapsto \omega(\xi(y))$ for some multiplicative character ω of $\mathbb{F}_q$ and some rational function ξ which is not a perfect power in $\bar{\mathbb{F}}_q(Y)$ (Kummer), and simply $\chi(y)$ some some character χ of the Jacobian group $J(\mathbb{F}_q)$, where we implicitly identify $Y(\mathbb{F}_q)$ as a subset of $J(\mathbb{F}_q)$ using the embedding $Y \hookrightarrow J$ given by some rational point.

The product $\chi_1\chi_2$ of two Artin characters χ_1, χ_2 is an Artin character, and if χ_1,χ_2 have disjoint ramification loci (which is in particular the case when one of them is unramified), the conductor of the product is given as $\mathfrak{f}(\chi_1\chi_2) = \mathfrak{f}(\chi_1) + \mathfrak{f}(\chi_2)$. Furthermore, one can pull back Artin characters along morphisms: if χ is an Artin character on Y and $h\colon X \to Y$ any non constant morphism of curves, one can define an Artin character $h^*\chi$ on X by pulling back the Galois covering. It is given on divisors by $h^*\chi(D) = \chi(h_*D)$, and is unramified if χ is unramified.

The main tool for estimating sums of Artin character is the following theorem by Weil, which gives a bound on sums of the form $S_Y(\chi) = \sum_{P \in Y(\mathbb{F}_q)} \chi(P)$ where χ is a nontrivial Artin character on Y.

Lemma 1. *If χ is a nontrivial Artin character on the curve Y is of genus g_Y, the following bound holds: $|S_Y(\chi)| \leq (2g_Y - 2 + \deg \mathfrak{f}(\chi))\sqrt{q}$.*

3.2 A bound for $S_{h,\xi,\pi}(\chi, \omega; I)$

Now consider the situation described at the beginning of this section: we have a branched covering $h\colon X \to Y$, rational functions $\xi, \pi\colon X \to \mathbb{P}^1$ (which are not constant, say), a nontrivial character χ of $J(\mathbb{F}_q)$ where J is the Jacobian of Y, an additive character ψ of $\mathbb{F}_q$ and a multiplicative character ω of $\mathbb{F}_q$. We want to estimate the sum $S_{h,\xi,\pi}(\chi, \omega, \psi)$ defined by (7).

Suppose for simplicity that ξ is not a perfect power in $\bar{\mathbb{F}}_q(X)$ and π is not of the form $u^p - u$ in $\bar{\mathbb{F}}_q(X)$. Then, we have Artin characters $\omega(\xi)$ and $\psi(\pi)$ on X. We denote their product by λ. This character has been studied in details by Perel'muter [11] and more recently by Castro and Moreno [5]. In particular, they show [5, Th. 13]:

Lemma 2. *Suppose that ω and ψ are not both trivial characters. Then λ is a ramified Artin character, and its conductor satisfies $\deg \mathfrak{f}(\lambda) \leq \deg(\pi)_\infty + l + s - r - a$ (with equality when ω and ψ are both nontrivial), where $(\pi)_\infty$ is the divisor of poles of π (counted positively), and l the number of poles of π, s the number*

of points in the support of (ξ)*,* r *the number of points common to the supports of* $(\pi)_\infty$ *and* (ξ)*, and* a *the number of points in the union of the supports of* $(\pi)_\infty$ *and* (ξ) *where* λ *is unramified.*

Furthermore, χ also defines an unramified Artin character on Y which can be pulled back to an unramified Artin character $h^*\chi$ of X. Let $\widetilde{\chi} = \lambda \cdot h^*\chi$. Then by definition, for any point $P \in X(\mathbb{F}_q)$ such that $\pi(P), \xi(P) \neq \infty$, we have: $\widetilde{\chi}(P) = \chi\big(h(P)\big)\omega\big(\xi(P)\big)\psi\big(\pi(P)\big)$. As a result, the sum $S_{h,\xi,\pi}(\chi,\omega,\psi)$ is almost the same as $S_X(\widetilde{\chi})$: they differ at most by the number of points $P \in X(\mathbb{F}_q)$ which are poles of π or ξ but where λ is nonzero (hence unramified), and there are at most a such points, using the notations of Lemma 2. The following extension of [6, Th. 7] follows (the proof details are provided in the full version of this paper [10]).

Theorem 2. *Let* $h\colon X \to Y$ *be a branched covering of curves,* $\xi, \pi\colon X \to \mathbb{P}^1$ *non constant rational functions,* χ *a nontrivial character of* $J(\mathbb{F}_q)$ *where* J *is the Jacobian of* Y*,* ψ *an arbitrary additive character of* $\mathbb{F}_q$ *and* ω *an arbitrary multiplicative character* ω *of* $\mathbb{F}_q$*. Assume that* h *does not factor through a nontrivial unramified covering of* Y*, and that* ξ *is not a perfect power in* $\bar{\mathbb{F}}_q(X)$ *and* ψ *not of the form* $u^p - u$ *in* $\bar{\mathbb{F}}_q(X)$*. Then, we have:*

$$\big|S_{h,\xi,\pi}(\chi,\omega,\psi)\big| \leq \big(2g_X - 2 + 2\deg\xi + 2\deg\pi\big)q^{1/2}.$$

Remark 1. It is easy to verify that the result still holds even when the condition on ξ and π isn't verified, as those cases essentially reduce to the case of trivial characters. Similarly, the theorem remains true when ξ or π is constant, with the convention that the degree is then zero.

Let $I \subset \mathbb{F}_q$ an arbitrary interval of $\mathbb{F}_q$. We can obtain the following estimate $S_{h,\xi,\pi}(\chi,\omega;I)$ by combining Theorem 2 with a standard bound on sums of additive characters on intervals (see the full version of this paper [10]).

Theorem 3. *Let the notations as described earlier. Assume that the branded covering* $h\colon X \to Y$ *does not factor through a nontrivial unramified covering of* Y*. Then, we have (denoting by* p *the characteristic of* q*):*

$$\big|S_{h,\xi,\pi}(\chi,\omega;I)\big| \leq \big(2g_X - 2 + 2\deg\xi + 2\deg\pi\big)q^{1/2}(1+\log p).$$

3.3 Application to encodings

Let us now succinctly discuss how Theorem 3 enables us to prove that encodings are strongly well-distributed. Take Icart's function [9] as an example: let $E\colon y^2 = x^3 + ax + b$ $(a \neq 0)$ an elliptic curve over a field $\mathbb{F}_q$ such that $q \equiv 2 \pmod 3$. It admits a geometric description of the form (1) with $h\colon C \to E$ such that $\mathbb{F}_q(C) = \mathbb{F}_q(E)[u]/(u^4 - 6xu^2 + 6yu - 3a)$, and a morphism $\pi\colon C \to \mathbb{P}^1$ induced by u. Icart's function $f\colon \mathbb{F}_q \to E(\mathbb{F}_q)$ can then be defined as $h \circ \pi^{-1}$ on $\mathbb{F}_q \subset \mathbb{P}^1(\mathbb{F}_q)$. Therefore, for any interval $I \subset \mathbb{F}_q$ and any nontrivial character χ of $E(\mathbb{F}_q)$:

$$S_f(\chi;I) = \sum_{u\in I}\chi\big(f(u)\big) = \sum_{\substack{P\in C(\mathbb{F}_q)\\ \pi(P)\in I}}\chi\big(h(P)\big) = S_{h,1,\pi}(\chi,\omega_0;I)$$

for ω_0 the trivial multiplicative character of $\mathbb{F}_q$. Moreover, C is of genus $g_C = 7$ and it is easy to see (by eliminating y between $y^2 - x^3 - ax - b$ and $u^4 - 6xu^2 + 6yu - 3a$, say) the rational function π is of degree $\deg \pi = 3$. As a result, we obtain $|S_f(\chi; I)| \leq (2 \cdot 7 - 2 + 2 \cdot 3)q^{1/2}(1 + \log p) = 18q^{1/2}(1 + \log p)$. In other words:

Theorem 4. *Icart's function f is 18-strongly well-distributed.*

In particular, the results of §2.2 apply to Icart's function. Similarly, the same approach as in [6] shows that all other known types of encodings, such as the Shallue–van de Woestijne–Ulas encodings [12,4] are also strongly well-distributed.

References

1. D. F. Aranha, P. Fouque, C. Qian, M. Tibouchi, and J. Zapalowicz. Binary Elligator Squared. In A. Joux and A. M. Youssef, editors, *SAC*, volume 8781 of *LNCS*, pages 20–37. Springer, 2014.
2. D. J. Bernstein, M. Hamburg, A. Krasnova, and T. Lange. Elligator: Elliptic-curve points indistinguishable from uniform random strings. In V. Gligor and M. Yung, editors, *ACM CCS*, 2013.
3. D. Boneh and M. K. Franklin. Identity-based encryption from the Weil pairing. In J. Kilian, editor, *CRYPTO*, volume 2139 of *LNCS*, pages 213–229. Springer, 2001.
4. E. Brier, J.-S. Coron, T. Icart, D. Madore, H. Randriam, and M. Tibouchi. Efficient indifferentiable hashing into ordinary elliptic curves. In T. Rabin, editor, *CRYPTO*, volume 6223 of *LNCS*, pages 237–254. Springer, 2010.
5. F. N. Castro and C. J. Moreno. Mixed exponential sums over finite fields. *Proc. Amer. Math. Soc.*, 128(9):2529–2537, 2000.
6. R. R. Farashahi, P.-A. Fouque, I. Shparlinski, M. Tibouchi, and J. F. Voloch. Indifferentiable deterministic hashing to elliptic and hyperelliptic curves. *Math. Comp.*, 82(281):491–512, 2013.
7. P.-A. Fouque, A. Joux, and M. Tibouchi. Injective encodings to elliptic curves. In C. Boyd and L. Simpson, editors, *ACISP*, volume 7959 of *LNCS*, pages 203–218. Springer, 2013.
8. M. D. Fried. Global construction of general exceptional covers. In G. L. Mullen and P. J. Shiue, editors, *Finite Fields: Theory, Applications, and Algorithms*, number 168 in Contemporary Mathematics, pages 69–100. American Mathematical Socity, 1994.
9. T. Icart. How to hash into elliptic curves. In S. Halevi, editor, *CRYPTO*, volume 5677 of *LNCS*, pages 303–316. Springer, 2009.
10. T. Kim and M. Tibouchi. Improved elliptic curve hashing and point representation. *IACR Cryptology ePrint Archive*, 2015. Full version of this paper.
11. G. I. Perel'muter. Estimation of a sum along an algebraic curve. *Mat. Zametki*, 5:373–380, 1969.
12. A. Shallue and C. van de Woestijne. Construction of rational points on elliptic curves over finite fields. In F. Hess, S. Pauli, and M. E. Pohst, editors, *ANTS*, volume 4076 of *LNCS*, pages 510–524. Springer, 2006.
13. M. Tibouchi. Elligator Squared: Uniform points on elliptic curves of prime order as uniform random strings. In N. Christin and R. Safavi-Naini, editors, *Financial Cryptography*, volume 8437 of *LNCS*, pages 139–156. Springer, 2014.
14. M. Tibouchi. Impossibility of surjective icart-like encodings. In S. S. M. Chow, J. K. Liu, L. C. K. Hui, and S. Yiu, editors, *ProvSec*, volume 8782 of *LNCS*, pages 29–39. Springer, 2014.