

# Indifferentiable Deterministic Hashing to Elliptic and Hyperelliptic Curves

Reza R. Farashahi<sup>1</sup>, Pierre-Alain Fouque<sup>2</sup>, Igor E. Shparlinski<sup>1</sup>,  
Mehdi Tibouchi<sup>2</sup>, and J. Felipe Voloch<sup>3</sup>

<sup>1</sup> Macquarie University  
Department of Computing  
Sydney, NSW 2109, Australia  
{reza.farashahi,igor.shparlinski}@mq.edu.au

<sup>2</sup> École normale supérieure  
Département d'informatique, Équipe de cryptographie  
45 rue d'Ulm, F-75230 Paris CEDEX 05, France  
{pierre-alain.fouque,mehdi.tibouchi}@ens.fr

<sup>3</sup> University of Texas  
Department of Mathematics  
Austin, TX 78712, USA  
voloch@math.utexas.edu

**Abstract.** At Crypto 2010, Brier *et al.* proposed the first construction of a hash function into ordinary elliptic curves that was indifferentiable from a random oracle, based on Icart's deterministic encoding from Crypto 2009. Such a hash function can be plugged into any cryptosystem that requires hashing into elliptic curves, while not compromising proofs of security in the random oracle model. However, the proof relied on relatively involved tools from algebraic geometry, and only applied to Icart's deterministic encoding from Crypto 2009.

In this paper, we present a new, simpler technique based on bounds of character sums to prove the indifferentiability of similar hash function constructions based on essentially any deterministic encoding to elliptic curves or curves of higher genus, such as the algorithms by Shallue, van de Woestijne and Ulas, or the Icart-like encodings recently presented by Kammerer, Lercier and Renault. In particular, we get the first constructions of well-behaved hash functions to Jacobians of hyperelliptic curves.

Our technique also provides more precise estimates on the statistical behavior of those deterministic encodings and the hash function constructions based on them. Additionally, we can derive pseudorandomness results for partial bit patterns of such encodings.

**Keywords:** Elliptic Curve Cryptography, Hashing, Random Oracle Model, Exponential Sums, Pseudorandomness.

## 1 Introduction

### 1.1 Hashing into elliptic curves

Many elliptic curve (especially pairing-based) cryptosystems require hashing into the group of points of an elliptic curve. For example in the Boneh-Franklin IBE scheme [4], the public-key for identity  $id \in \{0, 1\}^*$  is a point  $Q_{id} = H_1(id)$  on the curve. This is also the case in many other pairing-based cryptosystems including IBE and HIBE schemes [1,14,15], signature and identity-based signature schemes [2,5,6,10,25] and identity-based signcryption schemes [8,19].

Those cryptosystems are proved to be secure when the hash function is modeled as a random oracle into the curve, but it is not obvious how to instantiate such a hash function so that the security proof

can go through. As discussed in by Brier *et al.* [9], simple constructions that are easily distinguished from a random oracle are sufficient in some cases, owing to random self-reducibility properties of the underlying problems, but it is generally desirable to have proper hash functions that can be plugged into any cryptosystem that requires hashing into elliptic curves while not compromising proofs of security in the random oracle model.

The first example of such a hash function construction is due to Boneh and Franklin [4]. They use a particular supersingular elliptic curve  $E$  endowed with a one-to-one mapping  $f$  from the base field  $\mathbb{F}_p$  to  $E(\mathbb{F}_p)$ , and define their hash function as  $H(m) = f(h(m))$ , where  $h$  is a classical hash function from  $\{0, 1\}^*$  to  $\mathbb{F}_p$ . They are able to prove that their IBE scheme remains secure when  $h$  is seen as a random oracle into  $\mathbb{F}_p$  (they don't have to assume that  $H$  itself is a random oracle into  $E(\mathbb{F}_p)$ ).

In situations where shorter key sizes or asymmetric pairings are preferred, however, one wants to use ordinary elliptic curves, so the Boneh-Franklin construction does not apply. Two constructions for that case have been given by Brier *et al.* [9]. The main construction is valid for any ordinary elliptic curve  $E$  over a field  $\mathbb{F}_q$  such that  $q \equiv 2 \pmod{3}$ , and takes the form:

$$H(m) = f(h_1(m)) + f(h_2(m)),$$

where  $h_1, h_2$  are regarded as independent random oracles with values in  $\mathbb{F}_q$ , and  $f$  is Icart's encoding into  $E$ , described in [16]. This construction is quite efficient, but the proof requires rather technical tools from algebraic geometry, and uses many particular properties of Icart's function, which makes it difficult to adapt to other encodings or different settings. The alternate construction:

$$H(m) = f(h_1(m)) + h_2(m)G$$

with  $G$  a generator of the group of points, can use a wider range of encoding functions  $f$  instead of just Icart's encoding, but is significantly less efficient (typically five times slower). Furthermore, neither construction generalizes in a natural way to hyperelliptic curves.

## 1.2 Deterministic encodings

For hashing into an ordinary elliptic curve, the classical approach is inherently probabilistic: one can first compute an integer hash value  $h(m)$  and add a short counter to get  $x = 0^{\log k} \| h(m)$ . If  $x$  is the abscissa of a point on the elliptic curve  $y^2 = x^3 + ax + b$ , this gives the desired point; otherwise, one increments the counter and tries again. Each step succeeds with probability about  $1/2$ , so if  $k$  is the security parameter,  $k$  steps are heuristically enough to construct a point except with negligible probability.

However, the length of the hash computation depends on the message  $m$ , which can lead to side-channel attacks [7], unless all  $k$  steps are run for all messages, and Legendre symbols and square roots are computed in constant time, in which case computational cost becomes prohibitive. More importantly for pairing-based cryptography, it is difficult to assess the security of a scheme in which such a "probabilistic" hash function is used, even when the underlying integer hash function  $h$  is considered ideal.

Therefore, it has been desirable to devise point construction algorithms on elliptic and hyperelliptic curves that are more robust and easier to analyze.

The first algorithm to generate elliptic curve points in *deterministic* polynomial time has been given by Shallue and van de Woestijne [22], and later simplified and extended to hyperelliptic curves by Ulas in [23]. Icart's [16] sparked renewed interest in such algorithms, and several new ones have

been presented recently [9,17] both for elliptic and hyperelliptic curves. See §2.1 for a run-down of published encodings.

Techniques are known to establish a number of properties of these encodings, such as the size of their image into the curve [12,13]. If  $f$  is such an encoding function to an elliptic curve, the methods of [9] can also establish that

$$H(m) = f(h_1(m)) + h_2(m)G$$

is a well-behaved hash function (in the sense alluded to in the previous section). However, the more efficient construction from [9]:

$$H(m) = f(h_1(m)) + f(h_2(m))$$

is only shown to be well-behaved when  $f$  is Icart's function.

### 1.3 Our contributions

We introduce a new approach to deal with hash function constructions of the more general form:

$$H(m) = f(h_1(m)) + \cdots + f(h_s(m))$$

when  $f$  is any of the known deterministic encodings. We can show in particular that this construction is well-behaved (indifferentiable from a random oracle, in the random oracle model for the  $\mathbb{F}_q$ -valued functions  $h_i$ ) as soon as  $s$  is greater than the genus of the target curve (that is,  $s \geq 2$  for elliptic curves,  $s \geq 3$  for genus 2 curves, etc.). We recover the results from Brier *et al.* [9] about Icart's function as a special case with sharper bounds than in the original paper, and extend them to all known deterministic encodings, including new encodings to hyperelliptic curves.

For that purpose, we introduce the notion of *well-distributed encoding*, based on a new type of character sums associated with characters of the groups of points of the Jacobians of the target curves. We show that these sums can be estimated using classical results of Weil [24] and Bombieri [3], and combine these estimates with standard number theoretic technique in order to get explicit regularity results for functions of the form  $(u_1, \dots, u_s) \mapsto f(u_1) + \cdots + f(u_s)$ .

As a side contribution, we also investigate the pseudorandomness properties of sequences of bits extracted from these encoding functions. For example, while it is easy to distinguish the  $x$ -coordinate of a point constructed using Icart's function from the  $x$ -coordinate of a random point on the same curve over  $\mathbb{F}_p$ , it is not possible to construct a distinguisher when we are only given the top  $(1/2 - \varepsilon) \log p$  bits of  $x$ , where throughout this paper  $\log z$  means the base 2 logarithm of  $z$ .

### 1.4 Organization of the paper

The paper is organized as follows:

- section 2 is a summary of material from previous works, including a run-down of currently known deterministic encodings, and a review of relevant notions and results from Brier *et al.* [9];
- in section 3, we introduce the notion of *well-distributed encoding*, and show how it can be used to derive regularity results formally (Theorem 1 and Theorem 2);
- section 4 is more technical in nature and the details are not essential for applications: some machinery is introduced with the purpose of establishing Theorem 3, a convenient tool for proving well-distributedness;

- in section 5, we pick three illuminating examples of deterministic encodings to elliptic and hyperelliptic curves, prove that they are well-distributed, and deduce from our general results that they give rise to well-behaved hash functions. In particular, we give the first hash function constructions for curves of genus 2;
- finally, section 6 is devoted to the separate problem of studying the uniformity of bit substrings from a deterministic encoding.

## 2 Previous Work

### 2.1 Deterministic encodings: a roundup

Table 1 lists known deterministic encodings to ordinary elliptic curves and hyperelliptic curves. They fit in two families:

- SWU-like encodings, similar to those proposed by Shallue and van de Woestijne in [22]. They are based on the construction of explicit rational curves on a surface associated to the target curve.
- Icart-like encodings, similar to Icart’s function [16]. They are obtained by writing down a root of the curve equation using radicals of degrees prime to the order of the multiplicative group. This is only possible if the curve equation is solvable.

The techniques presented in this paper make it possible to construct well-behaved hash functions from any of these encodings. We work out some examples in detail in §5.

	char.	curve equation	genus	encoding	condition
SWU-like	$\neq 2, 3$	$y^2 = x^3 + ax + b$	1	SW [22]	—
		$y^2 = x^{2g+1} + ax + b$	$g$	Ulas [23]	—
	2	$y^2 + xy = x^3 + ax^2 + b$	1	SW [22]	—
	3	$y^2 + xy = x^3 + ax^2 + b$		Brier et al. [9, §8]	—
Icart-like	$\neq 2, 3$	$y^2 = x^3 + ax + b$	1	Icart [16, §2]	$q \equiv 2 \pmod{3}$
		$x^3 = (1 - b^2 - ay - y^2)(3x + 2)$		KLR [17, §3.1]	
		$x^3 + (y + c)(3x + 2a + 2b/y) = 0$	2	KLR [17, §3.2]	
		$y^2 = x^{2d} + x^d + a$	$d - 1$	KLR [17, §4.1]	$(d, q - 1) = 1$
		$y^2 = p_{a,b}^{(d)}(x)$	$d - 1$	KLR [17, §4.2]	$(d, q - 1) = 1$ $q \equiv 2 \pmod{3}$
		$y^2 + y = p_{a,b}^{(d)}(x)$	2		
2	$y^2 + xy = x^3 + ax^2 + b$	1	Icart [16, App. A]	$q \equiv 2 \pmod{3}$	

**Table 1.** Known deterministic encodings to ordinary elliptic curves and hyperelliptic curves. Some minor variants are omitted.

### 2.2 Admissible encodings and indifferenciability

Brier *et al.* [9] use Maurer’s indifferenciability framework [20] to analyze the conditions under which their hash function constructions can be plugged into a scheme that is proved secure in the random

oracle model in such a way that the proof of security goes through. As shown by Maurer, it suffices that the hash function construction be indifferentiable from a random oracle.

Then, Brier *et al.* [9] establish a sufficient condition for a hash function construction into an elliptic curve  $E$  to be indifferentiable from a random oracle. It applies to hash functions of the form:

$$H(m) = F(h(m)),$$

where  $F : S \rightarrow E(\mathbb{F}_q)$  is a deterministic encoding, and  $h$  is seen as a random oracle to  $S$ . Assuming that  $h$  is a random oracle, the construction is indifferentiable whenever  $F$  is an *admissible encoding* into  $E(\mathbb{F}_q)$ , in the following sense.

A function  $F : S \rightarrow R$  between finite sets is an  $\varepsilon$ -admissible encoding if it satisfies the following properties:

1. Computable:  $F$  is computable in deterministic polynomial time.
  2. Regular: for  $s$  uniformly distributed in  $S$ , the distribution of  $F(s)$  is  $\varepsilon$ -statistically indistinguishable from the uniform distribution in  $R$ .
  3. Samplable: there is an efficient randomized algorithm  $\mathcal{I}$  such that for any  $r \in R$ ,  $\mathcal{I}(r)$  induces a distribution that is  $\varepsilon$ -statistically indistinguishable from the uniform distribution in  $F^{-1}(r)$ .
- $F$  is an admissible encoding if  $\varepsilon$  is a negligible function of the security parameter.

This definition is motivated by the result of [9, Theorem 1] which asserts that if  $F : S \rightarrow R$  be an admissible encoding then the construction  $H(m) = F(h(m))$  is indifferentiable from a random oracle, in the random oracle model for  $h : \{0, 1\}^* \rightarrow S$ .

### 3 Well-Distributed Encodings

#### 3.1 Character sums

Consider an encoding  $f$  into a curve  $X$ , and let  $J$  denote the Jacobian of  $X$ . Assume that  $X$  has an  $\mathbb{F}_q$ -rational point  $O$ , so that we can fix an embedding  $X \rightarrow J$  (sending a point  $P$  to the degree 0 divisor  $(P) - (O)$ ). Regularity properties of functions  $f^{\otimes s}$  of the form:

$$\begin{aligned} f^{\otimes s} : (\mathbb{F}_q)^s &\rightarrow J(\mathbb{F}_q) \\ (u_1, \dots, u_s) &\mapsto f(u_1) + \dots + f(u_s) \end{aligned}$$

can be derived formally from the behavior of  $f$  with respect to characters of  $J(\mathbb{F}_q)$ . More precisely, introduce the character sums

$$S_f(\chi) = \sum_{u \in \mathbb{F}_q} \chi(f(u)), \tag{1}$$

where  $\chi$  is any character of the abelian group  $J(\mathbb{F}_q)$ . We say that  $f$  is *well-distributed* if we have good bounds on the magnitude of  $S_f(\chi)$  for nontrivial characters  $\chi$ .

**Definition 1.** *Let  $X$  be a smooth projective curve over a finite field  $\mathbb{F}_q$ ,  $J$  its Jacobian,  $f$  a function  $\mathbb{F}_q \rightarrow X(\mathbb{F}_q)$  and  $B$  a positive constant. We say that  $f$  is  $B$ -well-distributed if for any nontrivial character  $\chi$  of  $J(\mathbb{F}_q)$ , the following holds:*

$$|S_f(\chi)| \leq B\sqrt{q}. \tag{2}$$

*We say that  $f$  is well-distributed if it is  $B$ -well-distributed for some  $B$  bounded independently of the security parameter.*<sup>4</sup>

<sup>4</sup> Which only makes sense if we are implicitly looking at a family of functions  $f_k : \mathbb{F}_{q_k} \rightarrow X_k(\mathbb{F}_{q_k})$  instead of just one function, of course.

As we show in §5, essentially all known deterministic encoding functions into elliptic and hyper-elliptic curves satisfy (2) and we can thus establish results on the regularity of  $f^{\otimes s}$  for any such encoding similar to those that Brier *et al.* [9] have obtained for  $f^{\otimes 2}$  when  $f$  is Icart's function.

### 3.2 Collision probability

Besides character sums our estimates also depend on the number of solution  $W_f$  to the equation  $f(u) = f(v)$  where  $u, v \in \mathbb{F}_q$ . We note that

$$\rho_f = W_f/q^2$$

is the probability of a collision. For all functions considered in this paper we have abounded of the type

$$\rho_f \leq \frac{A_0}{q} + \frac{B_0}{q^2} \quad (3)$$

with some explicit constants  $A_0$  and  $B_0$  (different for each function  $f$ ).

### 3.3 Distribution of image sums

Hence we immediately obtain the following result.

**Theorem 1.** *If  $f : \mathbb{F}_q \rightarrow X(\mathbb{F}_q)$  is a  $B$ -well-distributed encoding into a curve  $X$ , then for all  $D \in J(\mathbb{F}_q)$ , we have:*

$$\left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right| \leq \frac{B^{s-2}}{q^{s/2-1}} \left( \rho_f - \frac{1}{\#J(\mathbb{F}_q)} \right).$$

*Proof.* Fix a positive integer  $s$ , and consider for any element  $D \in J(\mathbb{F}_q)$  the number of tuples  $(u_1, \dots, u_s)$  such that  $D = f(u_1) + \dots + f(u_s)$ :

$$N_s(D) = \#\{(u_1, \dots, u_s) \in (\mathbb{F}_q)^s \mid D = f(u_1) + \dots + f(u_s)\}$$

$N_s(D)$  can be expressed in terms of character sums:

$$\begin{aligned} N_s(D) &= \sum_{u_1, \dots, u_s \in \mathbb{F}_q} \frac{1}{\#J(\mathbb{F}_q)} \sum_{\chi} \chi(f(u_1) + \dots + f(u_s) - D) \\ &= \sum_{\chi} \frac{\chi(-D)}{\#J(\mathbb{F}_q)} \sum_{u_1, \dots, u_s \in \mathbb{F}_q} \chi(f(u_1) + \dots + f(u_s)) = \sum_{\chi} \frac{\chi(-D)}{\#J(\mathbb{F}_q)} (S_f(\chi))^s, \end{aligned}$$

where the summation is over all characters of  $J(\mathbb{F}_q)$ . Putting aside the contribution of the trivial character  $\chi_0$ , we get:

$$N_s(D) - \frac{q^s}{\#J(\mathbb{F}_q)} = \frac{1}{\#J(\mathbb{F}_q)} \sum_{\chi \neq \chi_0} \chi(-D) (S_f(\chi))^s.$$

Similarly, for  $W_f$ , defined in Section 3.2, we get

$$W_f - \frac{q^2}{\#J(\mathbb{F}_q)} = \frac{1}{\#J(\mathbb{F}_q)} \sum_{\chi \neq \chi_0} S_f(\chi) S_f(\bar{\chi}) = \frac{1}{\#J(\mathbb{F}_q)} \sum_{\chi \neq \chi_0} |S_f(\chi)|^2, \quad (4)$$

from which we the result follows immediately.

We see from (4) that

$$\rho_f - \frac{1}{\#J(\mathbb{F}_q)} \leq \frac{B^2}{q}. \quad (5)$$

Therefore we have:

**Corollary 1.** *If  $f : \mathbb{F}_q \rightarrow X(\mathbb{F}_q)$  is a  $B$ -well-distributed encoding into a curve  $X$ , then for all  $D \in J(\mathbb{F}_q)$ , we have:*

$$\left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right| < \frac{B^s}{q^{s/2}}.$$

Suppose that  $X$  is of genus  $g_X$ . Then  $\#J(\mathbb{F}_q) = q^{g_X} + O(q^{g_X-1})$ , so the bound of Corollary 1 is negligible compared to  $1/\#J(\mathbb{F}_q)$  provided that  $s/2 > g_X$ . In other words, if  $f$  is a well-distributed encoding, then for  $s > 2g_X$ , all elements  $D \in J(\mathbb{F}_q)$  have the same number of preimages by  $f^{\otimes s}$  up to negligible deviation.

When  $f$  is Icart's function, this says that all the points of the target elliptic curve have almost the same number of preimages by  $f^{\otimes s}$  for  $s \geq 3$ . This cannot be improved to  $s = 2$ , as the analysis in [9] shows that there is in fact a bounded number of points which have several times more preimages by  $f^{\otimes 2}$  than the others.

Nevertheless, Brier *et al.* [9] could obtain their indifferentiability result by bounding the statistical distance between the distribution defined by  $f^{\otimes 2}$  and the uniform distribution. We can establish a general result of this type for well-distributed encodings.

**Theorem 2.** *If  $f : \mathbb{F}_q \rightarrow X(\mathbb{F}_q)$  is a  $B$ -well-distributed encoding into a curve  $X$ , then the statistical distance between the distribution defined by  $f^{\otimes s}$  on  $J(\mathbb{F}_q)$  and the uniform distribution is bounded as:*

$$\sum_{D \in J(\mathbb{F}_q)} \left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right| \leq \frac{B^{s-1}}{q^{(s-1)/2}} \sqrt{\rho_f \#J(\mathbb{F}_q) - 1}.$$

*Proof.* We can first write the sum of squared deviations. Let

$$V_s = \sum_{D \in J(\mathbb{F}_q)} \left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right|^2.$$

Then, as in the proof of Theorem 1, using (4), we have:

$$\begin{aligned} V_s &= \sum_{D \in J(\mathbb{F}_q)} \frac{1}{q^{2s} \#J(\mathbb{F}_q)^2} \sum_{\chi, \eta \neq \chi_0} \chi(-D) \bar{\eta}(-D) \left( \sum_{u, v \in \mathbb{F}_q} \chi(f(u)) \bar{\eta}(f(v)) \right)^s \\ &= \frac{1}{q^{2s} \#J(\mathbb{F}_q)^2} \sum_{\chi, \eta \neq \chi_0} \left( \sum_{D \in J(\mathbb{F}_q)} \chi(D) \bar{\eta}(D) \right) \left( \sum_{u, v \in \mathbb{F}_q} \chi(f(u)) \bar{\eta}(f(v)) \right)^s \\ &= \frac{1}{q^{2s} \#J(\mathbb{F}_q)} \sum_{\chi \neq \chi_0} |S_f(\chi)|^{2s} \leq \frac{B^{2s-2}}{q^{s-1}} \left( \rho_f - \frac{1}{\#J(\mathbb{F}_q)} \right). \end{aligned}$$

Now applying the Cauchy-Schwarz inequality we conclude the proof.  $\square$

Recalling (5) we derive

**Corollary 2.** *If  $f : \mathbb{F}_q \rightarrow X(\mathbb{F}_q)$  is a  $B$ -well-distributed encoding into a curve  $X$ , then the statistical distance between the distribution defined by  $f^{\otimes s}$  on  $J(\mathbb{F}_q)$  and the uniform distribution is bounded as:*

$$\sum_{D \in J(\mathbb{F}_q)} \left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right| \leq \frac{B^s}{q^{s/2}} \sqrt{\#J(\mathbb{F}_q)}.$$

In particular, we see from Corollary 2 that if  $f$  is a well-distributed encoding, then for  $s > g_X$ , the distribution defined by  $f^{\otimes s}$  on  $J(\mathbb{F}_q)$  is statistically indistinguishable from the uniform distribution. If  $f$  is also computable and samplable (which is easily verified to be the case when  $f$  is any of the known deterministic encodings), then it is admissible. In particular, the following hash function construction:

$$m \mapsto f(h_1(m)) + \cdots + f(h_s(m)) \quad (s = g_X + 1)$$

is indifferentiable from a random oracle if  $h_1, \dots, h_s$  are seen as independent random oracles into  $\mathbb{F}_q$ .

## 4 Character Sums on Curves

Let  $Y \rightarrow X$  be a morphism of curves<sup>5</sup> over  $\mathbb{F}_q$  which is an abelian covering of group  $G$  (that is, a non-constant morphism such that the corresponding extension of function fields  $\mathbb{F}_q(Y)/\mathbb{F}_q(X)$  is abelian of group  $G$ ).

Any character of  $G$  determines, via the Artin map, a corresponding character on the group of  $\mathbb{F}_q$ -divisors on  $X$  prime to the ramification locus  $S$  of  $Y \rightarrow X$ , which extends to a multiplicative map  $\chi : \text{Div}_{\mathbb{F}_q}(X) \rightarrow \mathbb{C}$  vanishing on divisors not prime to  $S$ . Let us call such a map  $\chi$  an *Artin character* of  $X$ . One associates to  $\chi$  a distinguished effective divisor  $\mathfrak{f}(\chi)$  of support  $S$  called the conductor (in particular, if  $Y \rightarrow X$  is unramified,  $\mathfrak{f}(\chi) = 0$ ; the character itself is then said to be unramified).

*Example 1.* We consider the following examples of Artin characters.

- Let  $E$  be an elliptic curve over  $\mathbb{F}_q$ . Then any character of the abelian group  $E(\mathbb{F}_q)$  extends to an unramified Artin character of  $E$ . Indeed, if  $F$  denotes the Frobenius endomorphism of  $E$ ,  $1 - F : E \rightarrow E$  is an unramified abelian covering of group  $G = E(\mathbb{F}_q)$ , and characters of  $G$  determine Artin characters of  $E$  whose restriction to  $E(\mathbb{F}_q)$  is as expected.
- More generally, let  $X$  be any curve over  $\mathbb{F}_q$  with an  $\mathbb{F}_q$ -point, and  $J$  its Jacobian. As usual, we can embed  $X$  in  $J$  using this rational point. Then any character of the group  $J(\mathbb{F}_q)$  extends to an Artin character of  $X$ . It is constructed similarly;  $1 - F : J \rightarrow J$  is again an unramified abelian covering of group  $J(\mathbb{F}_q)$  which can be pulled back to an abelian covering  $Y \rightarrow X$  of group  $J(\mathbb{F}_q)$  along the embedding  $X \rightarrow J$ .
- If  $\chi$  is an Artin character on  $X$ , and  $h : \tilde{X} \rightarrow X$  is a non constant morphism of curves, there is a natural Artin character  $\tilde{\chi} = h^*\chi$  on  $\tilde{X}$  obtained by pulling back the abelian covering of  $X$  along  $h$ . On divisors,  $\tilde{\chi}$  can be defined as  $\tilde{\chi}(D) = \chi(h_*D)$ . Clearly, if  $\chi$  is unramified, then  $\tilde{\chi}$  is too, and more generally,  $\mathfrak{f}(\tilde{\chi}) = h^*\mathfrak{f}(\chi)$ .
- Assume that  $q$  is odd. The Legendre symbol  $\left(\frac{\cdot}{q}\right)$  on  $\mathbb{P}^1$ , which sends the point of abscissa  $x$  to 1,  $-1$ , or 0 according as whether  $x$  is a quadratic residue, a quadratic nonresidue or 0,  $\infty$ , extends to the nontrivial Artin character  $\chi_2$  defined by the ramified quadratic covering  $\mathbb{P}^1 \rightarrow \mathbb{P}^1 : x \mapsto x^2$ . One has  $\mathfrak{f}(\chi_2) = (0) + (\infty)$ .

<sup>5</sup> In this section, “curve over  $k$ ” means smooth, projective, geometrically connected curve over  $k$ .



- More generally, if  $X$  is a curve over  $\mathbb{F}_q$  and  $\varphi$  a non constant, rational function on  $X$ , the Legendre symbol of  $\varphi$  extends to a Artin character on  $X$ , namely  $\varphi^*\chi_2$ . Its conductor is the sum of the divisor of zeroes of  $\varphi$  and its divisor of poles. In particular,  $\deg \mathfrak{f}(\varphi^*\chi_2) = 2 \deg \varphi$ .

When  $\chi$  is nontrivial, Weil [24] has proved the following estimate for sums related to  $\chi$ , as a consequence of the Riemann hypothesis for curves (see, for example, [18, §2], [21, Chapter 9]). For any Artin character  $\chi$  of  $X$ , let:

$$S_X(\chi) = \sum_{P \in X(\mathbb{F}_q)} \chi(P).$$

**Lemma 1.** *If  $\chi$  is nontrivial and  $X$  is of genus  $g$ , one has*

$$|S_X(\chi)| \leq (2g - 2 + \deg \mathfrak{f}(\chi)) \sqrt{q}.$$

We can now easily deduce the following result, which forms the basis of the proofs in the next section.

**Theorem 3.** *Let  $h : \tilde{X} \rightarrow X$  be a non constant morphism of curves, and  $\chi$  be any nontrivial character of  $J(\mathbb{F}_q)$ , where  $J$  is the Jacobian of  $X$ . Assume that  $h$  does not factor through a nontrivial unramified morphism  $Z \rightarrow X$ . Then:*

$$\left| \sum_{P \in \tilde{X}(\mathbb{F}_q)} \chi(h(P)) \right| \leq (2\tilde{g} - 2) \sqrt{q} \quad (6)$$

where  $\tilde{g}$  is the genus of  $\tilde{X}$ . Furthermore, if  $q$  is odd and  $\varphi$  is a non constant rational function on  $\tilde{X}$ :

$$\left| \sum_{P \in \tilde{X}(\mathbb{F}_q)} \chi(h(P)) \left( \frac{\varphi(P)}{q} \right) \right| \leq (2\tilde{g} - 2 + 2 \deg \varphi) \sqrt{q}. \quad (7)$$

*Proof.* Denote also by  $\chi$  the Artin character of  $X$  extending the given character of  $J(\mathbb{F}_q)$ . The left-hand side of (6) is then just  $|S(h^*\chi)|$ , and we know that  $h^*\chi$  is an unramified Artin character of  $\tilde{X}$ , so the inequality follows from Lemma 1 provided that we can prove that  $h^*\chi$  is nontrivial. But if it is a trivial character,  $h_*$  maps all divisors of  $\tilde{X}$  to the kernel of  $\chi$ : this means that  $h$  factors through the unramified covering  $Z \rightarrow X$  defined by the kernel of  $\chi$ , which is impossible by hypothesis. Hence inequality (6).

Similarly, the left-hand side of (7) is  $|S(\tilde{\chi})|$  for  $\tilde{\chi}$  the Artin character of  $\tilde{X}$  defined as the product of  $h^*\chi$  and  $\varphi^*\chi_2$ . This character cannot be trivial: otherwise,  $\varphi^*\chi_2$  would be the inverse of  $h^*\chi$ , and hence unramified, which it is not. Thus, Lemma 1 gives  $|S(\tilde{\chi})| \leq (2\tilde{g} - 2 + \deg \mathfrak{f}(\tilde{\chi})) \sqrt{q}$ . Since  $h^*\chi$  is unramified, we have  $\deg \mathfrak{f}(\tilde{\chi}) = \deg \mathfrak{f}(\varphi^*\chi_2) = 2 \deg \varphi$ , which concludes the proof.  $\square$

## 5 Examples of Well-Distributed Encodings

### 5.1 Icart's function

In [16], Icart defined a deterministic functions to elliptic curves  $E: y^2 = x^3 + ax + b$  over finite fields  $\mathbb{F}_q$  such that  $q \equiv 2 \pmod{3}$ , essentially by solving for  $x$  using Cardano's formula. His map

$f : \mathbb{F}_q \rightarrow E(\mathbb{F}_q)$  is given by  $u \mapsto (x, y)$  with

$$x = \left( v^2 - b - \frac{u^2}{27} \right)^{1/3} + \frac{u^2}{3} \quad \text{and} \quad y = ux + v, \quad (8)$$

where  $v = (3a - u^4)/(6u)$ . The image of 0 is chosen as the neutral element of the elliptic curve.

Icart showed that a point  $(x, y)$  is the image of  $u$  if and only if

$$u^4 - 6xu^2 + 6yu - 3a = 0. \quad (9)$$

This makes it possible to give a simple geometric interpretation of Icart's function (and other encodings of the same type, as listed in Table 1), as has been done in [12,13,9].

Indeed, let  $K = \mathbb{F}_q(E)$  be the function field of  $E$ , and introduce the smooth projective curve  $C$  whose function field is the quartic extension  $L = K[u]/(P)$  of  $K$ , where  $P = u^4 - 6xu^2 + 6yu - 3a$ . The inclusions  $\mathbb{F}_q(u) \subset L$  and  $K \subset L$  give rise to birational maps  $g : C \rightarrow \mathbb{P}^1$  and  $h : C \rightarrow E$  which are in fact morphisms, since these curves are smooth and complete.

Then, Icart's function can be described as  $f(u) = h(g^{-1}(u))$ . This is well-defined when  $q \equiv 2 \pmod{3}$  because in that case,  $g$  induces a bijection from the set of points in  $C(\mathbb{F}_q)$  which are not poles of  $u$  to  $\mathbb{A}^1(\mathbb{F}_q) = \mathbb{F}_q$ .

This geometric point of view makes it possible to express the character sum  $S_f(\chi)$ , for any character  $\chi$  of  $E(\mathbb{F}_q)$ , in terms of the Artin character sum  $S_C(h^*\chi)$  (they are the same up to a few "bad points"), and then to use Theorem 3 to show that Icart's function is well-distributed.

**Theorem 4.** *Let  $f$  be Icart's function (8). For any nontrivial character  $\chi$  of  $E(\mathbb{F}_q)$ , the character sum  $S_f(\chi)$  given by (1) satisfies:*

$$|S_f(\chi)| \leq 12\sqrt{q} + 3.$$

*Proof.* The map  $h : C \rightarrow E$  defined above is a non constant morphism of curves. Moreover, we know from the analysis of [12] that if  $a \neq 0$ ,  $C$  is of genus 7, and that the Galois closure of the quartic extension  $\mathbb{F}_q(C)/\mathbb{F}_q(E)$  has Galois group  $S_4$  (for  $a = 0$ , the discussion is analogous). It follows that  $\mathbb{F}_q(C)/\mathbb{F}_q(E)$  has no nontrivial intermediate extension, and it is clearly ramified (because an unramified covering of an elliptic curve must be of genus 1). Thus,  $h$  fulfills the hypotheses of Theorem 3, and we get

$$\left| \sum_{P \in C(\mathbb{F}_q)} \chi(h(P)) \right| \leq (2 \cdot 7 - 2)\sqrt{q} = 12\sqrt{q}.$$

Now recall that  $g$  induces a bijection from  $C(\mathbb{F}_q) \setminus \{\text{poles of } u\}$  to  $\mathbb{F}_q$ . Thus:

$$\sum_{P \in C(\mathbb{F}_q)} \chi(h(P)) = S_f(\chi) + \sum_{\substack{P \in C(\mathbb{F}_q) \\ u(P) = \infty}} \chi(h(P)).$$

It is shown in the proof of [9, Lemma 7] that  $u$  has exactly 3 poles on  $C$ , hence the result.  $\square$

In other words, Theorem 4 says that  $f$  is a  $(12 + 3q^{-1/2})$ -well-distributed encoding.

In particular, as a well-distributed encoding to a curve of genus 1, Icart's function  $f$  satisfies that  $f^{\otimes s}$  is regular for  $s \geq 2$ . Since  $f^{\otimes 2}$  is clearly computable and samplable, it is admissible. This, we get a new, simpler and conceptually different proof of the main result of [9], that

$$m \mapsto f(h_1(m)) + f(h_2(m))$$

is indifferentiable from a random oracle when  $h_1, h_2$  are seen as random oracles to  $\mathbb{F}_q$ .

We also get more general results, since we have information about  $f^{\otimes s}$  for  $s > 2$ , and the bounds we obtain are also sharper than those in [9]. We note that for the collision probability for  $f$  by  $\rho_f$ , we have:

$$\rho_f = q^{-2} \sum_{P \in E(\mathbb{F}_q)} \#f^{-1}(P)^2 \leq 4q^{-2} \sum_{P \in E(\mathbb{F}_q)} \#f^{-1}(P) = 4q^{-1},$$

because points on  $E$  have at most 4 preimages.

More precisely, Theorem 1 gives:

**Corollary 3.** *For all  $P \in E(\mathbb{F}_q)$  and all  $s$ , we have*

$$\left| \frac{N_s(P)}{q^s} - \frac{1}{\#E(\mathbb{F}_q)} \right| \leq \frac{(12 + 3q^{-1/2})^{s-2} (4\#E(\mathbb{F}_q)q^{-1} - 1)}{q^{s/2-1}\#E(\mathbb{F}_q)}.$$

Similarly, Theorem 2 gives:

**Corollary 4.** *For all  $s$ , the statistical distance between the distribution given by  $f^{\otimes s}$  and the uniform distribution on  $E(\mathbb{F}_q)$  is bounded as*

$$\sum_{P \in E(\mathbb{F}_q)} \left| \frac{N_s(P)}{q^s} - \frac{1}{\#E(\mathbb{F}_q)} \right| \leq \frac{(12 + 3q^{-1/2})^{s-1} \sqrt{4\#E(\mathbb{F}_q)q^{-1} - 1}}{q^{(s-1)/2}}.$$

Since  $\#E(\mathbb{F}_q) = q + O(q^{1/2})$ , the bounds of Corollaries 3 and 4 simplify as

$$\left| \frac{N_s(P)}{q^s} - \frac{1}{\#E(\mathbb{F}_q)} \right| \leq \frac{3 \cdot 12^{s-2} + O(q^{-1/2})}{q^{s/2}} \quad (10)$$

and

$$\sum_{P \in E(\mathbb{F}_q)} \left| \frac{N_s(P)}{q^s} - \frac{1}{\#E(\mathbb{F}_q)} \right| \leq \frac{\sqrt{3} \cdot 12^{s-1} + O(q^{-1/2})}{q^{(s-1)/2}}. \quad (11)$$

For  $s = 2$ , this gives a bound of the statistical distance of the form  $12\sqrt{3}q^{-1/2} + O(q^{-1})$ , which is a significant improvement over the estimate from [9, §4.1]. We can refine this further by computing the collision probability precisely:

*Remark 1.* The Chebotarev density theorem gives the prevalence of points with any given number of preimages: the number of permutations in  $\text{Gal}(\mathbb{F}_q(C)/\mathbb{F}_q(E)) = S_4$  with exactly 1 (respectively 2, 4) fixed point(s) is 8 (respectively 6, 1), so the density of points with 1 (respectively 2, 4) preimages is  $8/24$  (respectively  $6/24$ ,  $1/24$ ). Using an effective version of the Chebotarev density theorem, this gives:

$$q^2 \rho_f = 1^2 \cdot \left( \frac{8q}{24} + O(\sqrt{q}) \right) + 2^2 \cdot \left( \frac{6q}{24} + O(\sqrt{q}) \right) + 4^2 \cdot \left( \frac{q}{24} + O(\sqrt{q}) \right).$$

Thus  $\rho_f = 2q^{-1} + O(q^{-3/2})$ , which allows us to drop the factors 3 and  $\sqrt{3}$  from (10) and (11), respectively.

## 5.2 Kammerer-Lercier-Renault

In a recent paper [17], Kammerer, Lercier and Renault have introduced a series of new encodings based on the same principles as Icart's function, namely solving curve equations in radicals. One such example is an encoding  $f$  to hyperelliptic curves of genus 2 over fields  $\mathbb{F}_q$  such that  $q \equiv 2 \pmod{3}$  of the form:

$$H: x^3 + (y + c) \left( 3x + 2a + \frac{2b}{y} \right) = 0.$$

The precise description of the encoding is rather complicated, and we refer the reader to [17, §3.2.2] for details, but the geometric picture is the same as for Icart's function. The parameter  $u$  defining the encoding satisfies a relation of the form:

$$4u^2(u^2 - 3y - a^2 - c)^3 + u^8 + \alpha u^4 + \beta u^2 + \gamma = 0,$$

where  $\alpha, \beta, \gamma$  are constants in  $\mathbb{F}_q$  defined in terms of  $a, b, c$ , which we assume are nonzero.

In particular, if  $K = \mathbb{F}_q(x, y)$  is the function field of  $H$ , one can consider the extension  $L = K[u]/(P)$  where  $P$  is the polynomial of degree 8 given by the equation above. This defines a covering  $h: C \rightarrow H$  of degree 8 by a certain smooth projective curve  $C$ , and the rational function  $u$  on  $C$  provides a morphism  $g: C \rightarrow \mathbb{P}^1$  of degree 9 which induces a bijection on  $\mathbb{F}_q$ -rational points over  $\mathbb{A}^1(\mathbb{F}_q) \setminus S$  where  $S$  is a finite set of points shown in [17] to be of size at most 75. The encoding  $f: \mathbb{F}_q \rightarrow H(\mathbb{F}_q)$  is then defined as  $u \mapsto h(g^{-1}(u))$  for  $u \in \mathbb{F}_q \setminus S$ , and is extended in some way to all of  $\mathbb{F}_q$ .

Our machinery applies again to show that  $f$  is a well-distributed encoding. Denote by  $J$  the Jacobian of  $H$ , and fix an embedding  $H \rightarrow J$ .

**Theorem 5.** *Let  $f$  be the encoding function described above. For any nontrivial character  $\chi$  of  $J(\mathbb{F}_q)$ , the character sum  $S_f(\chi)$  given by (1) satisfies:*

$$|S_f(\chi)| \leq 96\sqrt{q} + 759.$$

*Proof.* The map  $h: C \rightarrow H$  defined above is a non constant morphism of curves. Let us compute the genus of  $C$ .

Note first that  $P(u)$  can be written as  $Q(t)$  where  $t = u^2$ , which defines a factorization  $C \rightarrow D \rightarrow H$ , with  $[D : H] = 4$ . The discriminant of  $Q$  is a polynomial of degree 12 in  $y$ , and each of its 12 roots corresponds to 3 ramified points on  $H$ , since each value of  $y$  corresponds to 3 values of  $x$ . Moreover, when regarded as a polynomial over  $\mathbb{F}_q((1/y))$ , the quartic  $Q$  has a Newton polygon with integer slopes ( $-3$  with length 1 and 1 with length 3). Thus,  $D$  is unramified over points with  $y = \infty$ . All in all, the Riemann-Hurwitz formula gives  $2g_D - 2 = 4 \cdot (2g_H - 2) + 3 \cdot 12 = 44$ :  $D$  is of genus 23.

Then, the quadratic covering  $C \rightarrow D$  is ramified exactly at points such that  $t = 0$  or  $t = \infty$ . At finite distance, this gives  $\gamma = 0$ , which is excluded, so all the ramification is at infinity. By the previous Newton polygon argument, over each point of  $H$  with  $y = \infty$  lie 4 points of  $D$ , one with  $t = 0$  and three with  $t = \infty$ . And there are 2 such points of  $H$ , by another Newton polygon argument. Hence  $2g_C - 2 = 2 \cdot (2g_D - 2) + 2 \cdot 4 = 96$ , and  $C$  is of genus 49.

Let us now show that  $h: C \rightarrow H$  does not factor nontrivially through an unramified covering. To see this, consider  $D_0 \rightarrow \mathbb{P}^1$ , the ramified covering of degree 4 defined by  $Q$  (which pulls back to  $D \rightarrow H$  along  $x: H \rightarrow \mathbb{P}^1$ ). We see like before that all points of  $D_0$  ramified over  $\mathbb{P}^1$  have ramification index 2. Thus, the monodromy group of this covering is a transitive subgroup of  $S_4$  generated by transpositions, hence all of  $S_4$ .

By inspection of the ramification of  $x: H \rightarrow \mathbb{P}^1$ , it follows that  $D \rightarrow H$  also has  $S_4$  as its monodromy group. In particular, it has no quadratic subcovering. Now suppose  $h: C \rightarrow H$  factors through some abelian unramified extension  $Z \rightarrow H$ , which we can assume is quadratic. Then the function fields  $\mathbb{F}_q(Z)$  and  $\mathbb{F}_q(D)$  are everywhere linearly disjoint over  $\mathbb{F}_q(H)$  (i.e. all of their embeddings in some algebraic closure of  $\mathbb{F}_q(H)$  are linearly disjoint). Thus  $\mathbb{F}_q(C) = \mathbb{F}_q(D) \otimes_{\mathbb{F}_q(H)} \mathbb{F}_q(D)$ , and in particular,  $C \rightarrow D$  is the pullback of  $Z \rightarrow H$  along  $D \rightarrow H$ . This implies that  $C \rightarrow D$  is unramified, which we know is not the case.

Thus,  $h$  does not factor nontrivially through an abelian unramified covering, and hence fulfills the hypotheses of Theorem 3. We get

$$\left| \sum_{P \in C(\mathbb{F}_q)} \chi(h(P)) \right| \leq (2g_C - 2)\sqrt{q} = 96\sqrt{q}.$$

Now recall that  $g$  induces a bijection from  $C(\mathbb{F}_q) \setminus g^{-1}(S \cup \{\infty\})$  to  $\mathbb{F}_q \setminus S$ . Thus:

$$\begin{aligned} \sum_{P \in C(\mathbb{F}_q)} \chi(h(P)) &= \sum_{u \in \mathbb{F}_q \setminus S} \chi(f(u)) + \sum_{P \in g^{-1}(S \cup \{\infty\})} \chi(h(P)) \\ &= S_f(\chi) - \sum_{u \in S} \chi(f(u)) + \sum_{P \in g^{-1}(S \cup \{\infty\})} \chi(h(P)). \end{aligned}$$

Since  $\#S \leq 75$  and  $g$  is of degree 9, we get

$$|S_f(\chi)| \leq 96\sqrt{q} + 9 \cdot 76 + 75 = 96\sqrt{q} + 759$$

as required.  $\square$

In other words,  $f$  is a  $(96 + 759q^{-1/2})$ -well-distributed encoding to  $H$ . In particular, as a well-distributed encoding to a curve of genus 2, it satisfies that  $f^{\otimes s}$  is regular for any  $s \geq 3$ . Thus,  $f^{\otimes 3}$  is regular and clearly also computable and samplable, so the following construction:

$$m \mapsto f(h_1(m)) + f(h_2(m)) + f(h_3(m)) \in J(\mathbb{F}_q)$$

is indifferentiable from a random oracle when  $h_1, h_2, h_3$  are seen as random oracles to  $\mathbb{F}_q$ . This is the first example of an efficient, well-behaved hash function to the Jacobians of a large family of curves of genus 2.

We now estimate (quite coarsely) the collision probability for  $f$ : since  $h$  is of degree 8, any given point on  $H$  has at most 8 preimages by  $f$ . Thus,

$$\rho_f = q^{-2} \sum_{D \in J(\mathbb{F}_q)} \#f^{-1}(D)^2 \leq 8q^{-1} \sum_{D \in J(\mathbb{F}_q)} \#f^{-1}(D) = 8q^{-1}$$

and so  $q\rho_f \leq 8$  as required. So, now Theorem 1, gives the following estimate:

**Corollary 5.** *For all  $D \in J(\mathbb{F}_q)$  and all  $s$ , we have*

$$\left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right| \leq \frac{8(96 + 759q^{-1/2})^{s-2}}{q^{s/2}}.$$

Furthermore, from Theorem 2 we derive:

**Corollary 6.** *For all  $s$ , the statistical distance between the distribution given by  $f^{\otimes s}$  and the uniform distribution on  $J(\mathbb{F}_q)$  is bounded as*

$$\sum_{D \in J(\mathbb{F}_q)} \left| \frac{N_s(D)}{q^s} - \frac{1}{\#J(\mathbb{F}_q)} \right| \leq \frac{\sqrt{8} \cdot (96 + 759q^{-1/2})^{s-1} \sqrt{\#J(\mathbb{F}_q)}}{q^{s/2}}.$$

Using that  $\#J(\mathbb{F}_q) = q^2 + O(q^{3/2})$  one can also obtain simplified versions of Corollaries 5 and 6 similar to the bounds (10) and (11).

Of course, we have considered only one of the encodings from [17], but the same technique applies to all of them. In some cases, it is even much easier to apply. For example, in the case of the encoding to hyperelliptic curves  $H_a: y^2 = x^{2d} + x^d + a$ , the covering curve is  $H_a$  itself (the parameter  $u$  is in fact just  $y - x^d \in \mathbb{F}_q(x, y)$ ), so the hypotheses of Theorem 3 are trivially verified.

### 5.3 Shallue-Woestijne-Ulas

The first published type of encoding functions to ordinary elliptic curves, the functions of Shallue and van de Woestijne [22], have a rather different geometric interpretation than Icart-like encodings, but they can nevertheless be proved to be well-distributed by essentially the same techniques.

Here, we consider the simplified Shallue-Woestijne-Ulas encoding over fields  $\mathbb{F}_q$  with  $q \equiv 3 \pmod{4}$ , as introduced in [9] with the sign tweak from [13]. The function is based on the following observation. If we let  $g(x) = x^3 + ax + b \in \mathbb{F}_q[x]$  with  $ab \neq 0$ , and define:

$$X_2(u) = -\frac{b}{a} \left( 1 + \frac{1}{u^4 - u^2} \right) \quad X_3(u) = -u^2 X_2(u) \quad Z(u) = u^3 g(X_2(u)).$$

then we have  $Z(u)^2 = -g(X_2(u)) \cdot g(X_3(u))$ .

Let  $S = \{0, 1, -1\} \cup \{\text{roots of } g(X_j(u)) = 0, j = 2, 3\}$ . For any  $u \notin S$ ,  $X_2(u)$  and  $X_3(u)$  are well-defined and non zero. Since  $-1$  is a quadratic nonresidue in  $\mathbb{F}_q$ , this implies that for any  $u \in \mathbb{F}_q \setminus S$ , exactly one of  $g(X_2(u))$  or  $g(X_3(u))$  is a square. Therefore, we can define an encoding function  $f$  to the elliptic curve  $E: y^2 = x^3 + ax + b$  by

$$f(u) = \left( X_j(u) ; (-1)^j \sqrt{g(X_j(u))} \right),$$

where  $j = 2$  or  $3$  is the index such that  $g(X_j(u))$  is a square, and  $\sqrt{\cdot}$  is the standard square root in  $\mathbb{F}_q$ , given by exponentiation by  $(q+1)/4$  (thus, the  $y$ -coordinate is a quadratic residue if  $j = 2$  and a quadratic nonresidue if  $j = 3$ ).

As discussed in [13], this encoding function admits the following geometric interpretation. It is easy to see that for all  $u \in \mathbb{F}_q \setminus \{-1, 0, 1\}$ ,

$$\begin{aligned} x = X_2(u) &\iff u^4 - u^2 + \frac{1}{\omega} = 0 \\ x = X_3(u) &\iff u^4 - \omega u^2 + \omega = 0 \end{aligned}$$

where  $\omega = \frac{a}{b}x + 1$ . Thus, we can introduce coverings  $h_j: C_j \rightarrow E$ ,  $j = 2, 3$ , by the smooth projective curves whose function fields are the extensions of  $\mathbb{F}_q(x, y)$  defined respectively by  $u^4 - u^2 + 1/\omega = 0$

and  $u^4 - \omega u^2 + \omega = 0$ . The parameter  $u$  is a rational function on each of the  $C_j$  giving rise to morphisms  $g_j: C_j \rightarrow \mathbb{P}^1$ , such that any point in  $\mathbb{A}^1(\mathbb{F}_q) \setminus S$  has exactly two preimages in  $C_j(\mathbb{F}_q)$  for one of  $j = 2, 3$ , and none in the other.

If  $u \in \mathbb{F}_q \setminus S$  has its preimages in  $C_j(\mathbb{F}_q)$ , those preimages are conjugate under  $y \mapsto -y$ , so that exactly one of them satisfies  $\left(\frac{y}{q}\right) = (-1)^q$ . Let  $P \in C_j(\mathbb{F}_q)$  be that preimage. Then,  $f(u) = h_j(P)$ . This geometric interpretation is enough to apply Theorem 3 and establish that  $f$  is well-distributed.

**Theorem 6.** *Let  $f$  be the encoding described above. For any nontrivial character  $\chi$  of  $E(\mathbb{F}_q)$ , the character sum  $S_f(\chi)$  given by (1) satisfies:*

$$|S_f(\chi)| \leq 52\sqrt{q} + 151.$$

*Proof.* In light of the previous discussion, the character sum  $S_f(\chi)$ , when restricted to parameters  $u$  in  $\mathbb{F}_q \setminus S$ , can be written as:

$$\sum_{u \notin S} \chi(f(u)) = \sum_{\substack{P \in C_2(\mathbb{F}_q) \setminus S_2 \\ \left(\frac{y}{q}\right) = +1}} \chi(h_2(P)) + \sum_{\substack{P \in C_3(\mathbb{F}_q) \setminus S_3 \\ \left(\frac{y}{q}\right) = -1}} \chi(h_3(P)), \quad (12)$$

where  $S_j = g_j^{-1}(S \cup \{\infty\})$ . To estimate such sums, observe that:

$$\sum_{P \in C_j(\mathbb{F}_q)} \chi(h_j(P)) \cdot \left( \frac{1 + (-1)^j \left(\frac{y}{q}\right)}{2} \right) = \sum_{\substack{P \in C_j(\mathbb{F}_q) \\ \left(\frac{y}{q}\right) = (-1)^j}} \chi(h_j(P)) + \frac{1}{2} \sum_{\substack{P \in C_j(\mathbb{F}_q) \\ \left(\frac{y}{q}\right) = 0}} \chi(h_j(P)). \quad (13)$$

The first sum on the right-hand side of (13) is almost what we want (up to a bounded number of “bad terms”) and the second sum on the right-hand side contains at most  $3 \cdot 4 = 12$  terms.

Furthermore, the left-hand side of (13) is directly estimated using Theorem 3. Indeed, by the Eisenstein criterion,  $h_2$  and  $h_3$  are totally ramified over points in  $H$  such that  $\omega = 0$  (that is,  $x = -b/a$ ), so they cannot factor through any unramified covering of  $E$ . Hence, Theorem 3 applies and gives:

$$\left| \sum_{P \in C_j(\mathbb{F}_q)} \chi(h_j(P)) \cdot \left( \frac{1 + (-1)^j \left(\frac{y}{q}\right)}{2} \right) \right| \leq (2g_{C_j} - 2 + \deg y)\sqrt{q},$$

where  $g_{C_j}$  is the genus of  $C_j$ , and  $\deg y$  is the degree of  $y$  as a rational function on  $C_j$ . Clearly,  $\deg y = [\mathbb{F}_q(x, y, u) : \mathbb{F}_q(x, y)] \cdot [\mathbb{F}_q(x, y) : \mathbb{F}_q(y)] = 4 \cdot 3 = 12$ . Furthermore, to compute  $g_{C_j}$ , note that in addition to the totally ramified points mentioned previously,  $C_j \rightarrow E$  has ramification type  $(2, 2)$  at points with  $\omega = 4$  and at infinity, and is unramified elsewhere. Thus, the Riemann-Hurwitz formula gives  $2g_{C_j} - 2 = 0 + 2 \cdot 3 + 2 \cdot (1 + 1) + 2 \cdot (1 + 1) = 14$ : the curves  $C_j$  are of genus 8. Thus:

$$\left| \sum_{P \in C_j(\mathbb{F}_q)} \chi(h_j(P)) \cdot \left( \frac{1 + (-1)^j \left(\frac{y}{q}\right)}{2} \right) \right| \leq 26\sqrt{q}.$$

Plugging this estimate back into (12) using (13), we get:

$$\left| \sum_{u \notin S} \chi(f(u)) \right| \leq (26\sqrt{q} + \#S_2 + 6) + (26\sqrt{q} + \#S_3 + 6) = 52\sqrt{q} + 12 + \#S_2 + \#S_3.$$

Thus  $|S_f(\chi)| \leq 52\sqrt{q} + 12 + \#S_2 + \#S_3 + \#S$ . Now  $\#S \leq 3 + 2 \cdot 12 = 27$ , and since  $g_j$  is a map of degree 2,  $\#S_j \leq 2(\#S + 1) \leq 56$ , which concludes the proof.  $\square$

Thus, the simplified Shallue-Woestijne-Ulas encoding is  $(52 + 151q^{-1/2})$ -well-distributed. As in Sections 5.1 and 5.2, using Theorem 1 and Theorem 2, we can deduce precise bounds on the maximum deviation of functions of the form  $f^{\otimes s}$  and on the statistical distance of the distribution they define on  $E(\mathbb{F}_q)$  and the uniform distribution.

In particular, we get that  $f^{\otimes s}$  is regular for  $s \geq 2$ . Since  $f^{\otimes 2}$  is clearly computable and samplable, it is admissible, and we obtain that

$$m \mapsto f(h_1(m)) + f(h_2(m))$$

is indifferentially random from a random oracle when  $h_1, h_2$  are seen as random oracles to  $\mathbb{F}_q$ .

Once again, the method generalizes to other SWU-like encodings, such as the Ulas encoding to hyperelliptic curves of the form  $y^2 = x^5 + ax + b$ .

## 6 Pseudorandomness of Bits

The previous sections have focused on establishing regularity results for functions of the form  $f^{\otimes s}$  when  $f$  is a deterministic encoding to an elliptic or hyperelliptic curve and  $s$  is large enough (at least greater than the genus of the target curve). We now turn to a different problem: we would like to investigate the uniformity of  $f$  itself.

To fix ideas, let  $f$  be Icart's function (the arguments adapt to other encodings easily). It is proved in [12,13] that  $f(\mathbb{F}_q)$  consists of about 5/8 of the points of the target elliptic curve, so it is easy to construct a distinguisher between points constructed with  $f$  and points picked at random, by checking whether the point is in  $f(\mathbb{F}_q)$  or not (which can be done in polynomial time by computing the roots of (9)). If we have  $k$  points to check, the distinguisher succeeds with probability roughly  $1 - (5/8)^k$ . The same can be done if we only have the  $x$ -coordinate of the given points.

Suppose however that a device leaks only a fraction of the bits of  $x$ . Is it still possible to distinguish between points coming from  $f$  and random points? What we can show is that it is in fact impossible if we are given less than half of the bits.

More precisely, let  $q = p$  be a prime number with  $p \equiv 2 \pmod{3}$ . In particular, we identify the elements of  $\mathbb{F}_p$  with the integers from the set  $\{0, \dots, p-1\}$ . Given a binary string  $\Sigma$ , we denote by  $R_f(\Sigma)$  the number of  $u \in \mathbb{F}_p^*$  for which the least significant bits of the binary representation of the  $x$ -coordinate of  $f(u)$  coincide with  $\Sigma$ .

**Theorem 7.** *Let  $f$  be Icart's function (8). For any bit string  $\Sigma$  of length  $k$ , we have:*

$$R_f(\Sigma) = p \cdot 2^{-k} + O\left(p^{1/2} \log p\right)$$

and the implied constant in the big-O is universal.

*Proof.* Let us consider the curve

$$C : \left(x - \frac{u^2}{3}\right)^3 = \left(\frac{3a - u^4}{6u}\right)^2 - b - \frac{u^2}{27}.$$



Clearly for every  $u \in \mathbb{F}_p^*$  there is a unique point  $(u, x) \in C(\mathbb{F}_p)$  and the corresponding  $x$  is exactly the  $x$ -coordinate  $f(u)$ . We consider the character sums

$$T_f(\psi) = \sum_{(u,x) \in C(\mathbb{F}_p)} \psi(x),$$

where  $\psi$  is an additive character of  $\mathbb{F}_p$ . By the Bombieri bound of character sums along a curve [3], we see that

$$T_f(\psi) = O(p^{1/2}). \quad (14)$$

Let  $s$  be the integer formed by  $\Sigma$ . We now note that the least significant bits of the binary representation of the  $x \in \mathbb{F}_p$  coincide with  $\Sigma$  if and only if we have

$$2^{-k}(x - s) \equiv z \pmod{p} \quad (15)$$

for some integer  $z$  with  $0 \leq z < p2^{-k}$ . We now combine (14) with the classical Erdős-Turán inequality (see [11, Theorem 1.21]) that relates the uniformity of distribution to character sums. This immediately implies that the deviation of the number of solutions to (15) from the expected number of  $p2^{-k}$  is  $O(p^{1/2} \log p)$ .  $\square$

In particular, if  $k \leq (1/2 - \varepsilon) \log p$ , then for large enough  $p$ , the quantity  $R_f(\Sigma)$  is independent of  $\Sigma$  up to negligible deviations. This implies that the top  $(1/2 - \varepsilon) \log p$  bits of the  $x$ -coordinate are indistinguishable from a random bit string of the same length.

## Acknowledgments

This work was partly supported by the French ANR-07-TCOM-013-04 PACE Project and by the European Commission through the IST Program under Contract ICT-2007-216646 ECRYPT II.

## References

1. J. Baek and Y. Zheng. Identity-based threshold decryption. In F. Bao, R. H. Deng, and J. Zhou, editors. *PKC 2004*, volume 2947 of *Lecture Notes in Computer Science*, pages 262–276. Springer, 2004.
2. A. Boldyreva. Threshold signatures, multisignatures and blind signatures based on the gap-Diffie-Hellman-group signature scheme. Y. Desmedt, editor. *PKC 2003*, volume 2567 of *Lecture Notes in Computer Science*, pages 31–46. Springer, 2002.
3. E. Bombieri. On exponential sums in finite fields. In *Les Tendances Géom. en Algèbre et Théorie des Nombres*, pages 37–41. Éditions du Centre National de la Recherche Scientifique, Paris, 1966.
4. D. Boneh and M. K. Franklin. Identity-based encryption from the Weil pairing. In J. Kilian, editor, *CRYPTO*, volume 2139 of *Lecture Notes in Computer Science*, pages 213–229. Springer, 2001.
5. D. Boneh, C. Gentry, B. Lynn, and H. Shacham. Aggregate and verifiably encrypted signatures from bilinear maps. In E. Biham, editor, *EUROCRYPT*, volume 2656 of *Lecture Notes in Computer Science*, pages 416–432. Springer, 2003.
6. D. Boneh, B. Lynn, and H. Shacham. Short signatures from the Weil pairing. In C. Boyd, editor, *ASIACRYPT*, volume 2248 of *Lecture Notes in Computer Science*, pages 514–532. Springer, 2001.
7. C. Boyd, P. Montague and K.Q. Nguyen. Elliptic Curve Based Password Authenticated Key Exchange Protocols. In *ACISP 2001*, volume 2119 of *Lecture Notes in Computer Science*, pages 487–501. Springer, 2001.
8. X. Boyen. Multipurpose identity-based signcryption (a swiss army knife for identity-based cryptography). In D. Boneh, editor, *CRYPTO*, volume 2729 of *Lecture Notes in Computer Science*, pages 383–399. Springer, 2003.
9. 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*, Lecture Notes in Computer Science. Springer, 2010. To appear.

10. J. C. Cha and J. H. Cheon. An identity-based signature from gap Diffie-Hellman groups. Y. Desmedt, editor. *PKC 2003*, volume 2567 of *Lecture Notes in Computer Science*, pages 18–30. Springer, 2002.
11. M. Drmota and R. Tichy. Sequences, discrepancies and applications. Springer-Verlag, Berlin, 1997.
12. R.R. Farashahi, I. Shparlinski and F. Voloch. On hashing into elliptic curves. In *J. Math. Cryptology*, volume 3, pages 353–360. Springer, 2009.
13. P.-A. Fouque and M. Tibouchi. Estimating the size of the image of deterministic hash functions to elliptic curves. In M. Abdalla and P. Baretto, editors, *LATINCRYPT*, *Lecture Notes in Computer Science*. Springer, 2010. To appear.
14. C. Gentry and A. Silverberg. Hierarchical ID-based cryptography. In Y. Zheng, editor. *Advances in Cryptology - ASIACRYPT 2002*, volume 2501 of *Lecture Notes in Computer Science*, pages 548–566. Springer, 2002.
15. J. Horwitz and B. Lynn. Toward hierarchical identity-based encryption. In L. R. Knudsen, editor, *EUROCRYPT*, volume 2332 of *Lecture Notes in Computer Science*, pages 466–481. Springer, 2002.
16. T. Icart. How to hash into elliptic curves. In S. Halevi, editor, *CRYPTO*, volume 5677 of *Lecture Notes in Computer Science*, pages 303–316. Springer, 2009.
17. J.-G. Kammerer, R. Lercier, and G. Renault. Encoding points on hyperelliptic curves over finite fields in deterministic polynomial time. *Preprint*, 2010, available from <http://arxiv.org/abs/1005.3758>.
18. D. R. Kohel and I. Shparlinski. On exponential sums and group generators for elliptic curves over finite fields. In W. Bosma, editor, *ANTS*, volume 1838 of *Lecture Notes in Computer Science*, pages 395–404. Springer, 2000.
19. B. Libert and J.-J. Quisquater. Efficient signcryption with key privacy from gap Diffie-Hellman groups. In F. Bao, R. H. Deng, and J. Zhou, editors. *PKC 2004*, volume 2947 of *Lecture Notes in Computer Science*, pages 187–200. Springer, 2004.
20. U. M. Maurer, R. Renner, and C. Holenstein. Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In M. Naor, editor, *TCC*, volume 2951 of *Lecture Notes in Computer Science*, pages 21–39. Springer, 2004.
21. M. Rosen. *Number Theory in Function Fields*, volume 210 of *Graduate Texts in Mathematics*. Springer, 2002.
22. 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 *Lecture Notes in Computer Science*, pages 510–524. Springer, 2006.
23. M. Ulas. Rational points on certain hyperelliptic curves over finite fields. *Bull. Polish Acad. Sci. Math.*, 55(2):97–104, 2007.
24. A. Weil. *Basic number theory*. Classics in Mathematics. Springer-Verlag, Berlin, 1995.
25. F. Zhang and K. Kim. ID-based blind signature and ring signature from pairings. In Y. Zheng, editor. *Advances in Cryptology - ASIACRYPT 2002*, volume 2501 of *Lecture Notes in Computer Science*, pages 533–547. Springer, 2002.