Advances in asymmetric cryptographic algorithms
Benjamin Smith

To cite this version:

HAL Id: tel-04238166
https://inria.hal.science/tel-04238166
Submitted on 12 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Synthèse d’Activité Scientifique
Advances in asymmetric cryptographic algorithms

Benjamin Smith

July 2023

Abstract

This document gives a synthetic overview of a large part of my research since 2006, for the degree of Habilitation à Diriger des Recherches at the Institut Polytechnique de Paris.

We focus on algorithms for asymmetric (or public-key) cryptography—especially, though not exclusively, Elliptic Curve Cryptography; its generalizations including Hyperelliptic Curve Cryptography; and its derivatives including Isogeny-Based Cryptography, one of the newer branches of post-quantum cryptography. We also consider the implementation of post-quantum cryptosystems in a broader sense. The contributions range from theoretical and mathematical foundations through to low-level implementation techniques.

Habilitation à Diriger des Recherches
Defence: 6th of October, Campus Cyber, La Défense Paris, France

Emmanuel Thomé, Inria Nancy & LORIA, Président
Pierre-Alain Fouque, Université Rennes 1, Rapporteur
Florian Hess, Carl von Ossietzky Universität Oldenburg, Rapporteur
Alfred Menezes, University of Waterloo, Rapporteur
Catuscia Palamidessi, Inria Saclay & LIX, Examinatrice
Nigel Smart, Katholieke Universiteit Leuven, Examinateur

Contents

1 Introduction and highlights 1
  1.1 Context .......................................................... 1
  1.2 Organisation, highlights, and collaborations .................. 2
  1.3 Other topics .................................................... 3
## CONTENTS

### 2 Classical elliptic-curve cryptography 4
- 2.1 Elliptic curves for cryptography 4
- 2.2 Isogenies and endomorphisms 6
- 2.3 Endomorphisms and scalar decomposition 6
- 2.4 The $\mathbb{Q}$-curve construction 8
- 2.5 High-speed key exchange software 9
- 2.6 Accelerating point counting 10

### 3 Classical hyperelliptic cryptography 12
- 3.1 Hyperelliptic curves and cryptosystems 12
- 3.2 Realizing higher-genus isogenies with correspondences 13
- 3.3 Hyperelliptic endomorphisms and scalar multiplication 13
- 3.4 Point counting with efficient real multiplication 15
- 3.5 Isogenies and the DLP in genus 3 16
- 3.6 Trustless unknown-order groups 18

### 4 Isogeny-based cryptography 19
- 4.1 Theoretical foundations 19
- 4.2 From theory to practice 20
- 4.3 Supersingularity 21
- 4.4 The Velusqrt algorithm 22
- 4.5 Constant-time CSIDH and CTIDH 24
- 4.6 The genus-2 superspecial graph 24

### 5 Compact public-key cryptography 26
- 5.1 qDSA and muKummer 26
- 5.2 Post-quantum software updates for low-end IoT devices 28
- 5.3 Wavelet: adventures in code-based cryptography 30

### 6 Looking forward 32

#### Publication list
- International journal articles 33
- Articles in reviewed international conference proceedings 34
- Book chapters 37
- Preprints 37
- Ph.D. thesis 38

#### Bibliography 39
1 Introduction and highlights

This document gives a synthetic overview of my research since 2006. It is focused on cryptographic algorithms: especially, though not exclusively, Elliptic Curve Cryptography (ECC), its generalizations, and its derivatives.

1.1 Context

My research exploits algebraic geometry and number theory to attack and improve cryptosystems, especially those built around elliptic curves and abelian varieties (higher-dimensional generalizations of elliptic curves, notably Jacobians of hyperelliptic curves). The fundamental tools are explicit \textit{isogenies}, which are algebraic transformations of abelian varieties, and \textit{endomorphisms}, which are transformations of an abelian variety into itself.

ECC, introduced by Miller and Koblitz in 1985 \cite{Mil85;Kob87} is now crucial to internet security: it is the state-of-the-art for \textit{key exchange} (establishing shared cryptographic keys), and offers the most compact \textit{digital signatures}. For industrial applications we target a minimum \textit{security level} of 128 bits: adversaries require roughly $2^{128}$ operations (far beyond the limits of feasibility) to have any non-negligible probability of breaking the scheme. At the 128-bit security level, ECC public keys are only 256 bits long—twelve times shorter than the equivalent 3072-bit RSA keys. Working with smaller keys always reduces bandwidth, storage, and infrastructure costs for key distribution. Often, it even saves time and energy: ECC is substantially faster than RSA for key establishment and public-key encryption.

Conventional ECC is based on the supposed hardness of the elliptic-curve DLP. But powerful general-purpose quantum computers would render ECC (and indeed, virtually all public-key cryptosystems currently deployed on the internet) completely insecure: Shor’s famous quantum algorithm \cite{Sho97} solves DLP instances in polynomial time. Cryptographic research is therefore turning towards \textit{post-quantum} cryptosystems, which run on conventional computers yet resist quantum-equipped adversaries. ECC is finding a post-quantum future in \textit{isogeny-based cryptography} (IBC). ECC deals with points on a single curve $\mathcal{E}$: given $P$ and $Q = [m]P$ on $\mathcal{E}$, the hard problem is to compute the algebraic relationship between them, which is the discrete logarithm $m$. In IBC, points are irrelevant: instead, we deal with multiple curves in a single class, and the hard problem is to compute the algebraic relationship between curves $\mathcal{E}$ and $\mathcal{E}'$: that is, an isogeny.

In real-world applications, reducing the computational cost of cryptography is critical. Faster cryptography means reduced latency and server costs, and improved battery lifetime for mobile computing. Reducing memory footprints becomes crucial when we move to low-end IoT devices operated by low-cost microcontrollers where very little RAM is available (even less once the enveloping ap-
plication is taken into account). Like ECC, IBC generally promises small keys and low bandwidth, but at the cost of intensive algebraic calculations. Compactness makes ECC and IBC interesting in constrained devices, where reducing the intensive computations, and providing even the most basic protections against side-channel attacks, is a fascinating scientific challenge. My work here has spilled over into compact implementations of non-curve-based public-key systems.

1.2 Organisation, highlights, and collaborations

Chapter 2 describes some of my results in classical ECC, setting the scene for later chapters. In [Smi16] and [Smi13], I introduced a new family of elliptic curves with efficient endomorphisms, which can be used to accelerate scalar multiplication following the Gallant–Lambert–Vanstone method (§2.4). François Morain, Charlotte Scribot, and I used these endomorphisms to accelerate point-counting algorithms [MSS16] (§2.6). Craig Costello, Huseyin Hisil, and I used one family from [Smi16;Smi13] to develop free, high-speed, and side-channel protected software for Diffie–Hellman key exchange [CHS14] (§2.5).

Chapter 3 surveys my work in hyperelliptic-curve cryptography. My doctoral thesis with David Kohel [Smi06] developed algorithms for hyperelliptic isogenies and endomorphisms (§3.2), and exploited these to speed up scalar multiplication [KS06] (§3.3). David Kohel, Pierrick Gaudry, and I developed faster point-counting algorithms for families of curves with efficient endomorphisms in [GKS11], computing cryptographic parameters (and finding secure curves) far more rapidly in practice (§3.4). Over a decade later, [GKS11] still holds the world record in genus-2 point counting; this paper won the Best Paper award at Asiacrypt 2011. Hyperelliptic point counting without efficient endomorphisms remains extremely slow. Samuel Dobson, Steven Galbraith, and I have proposed higher-dimensional Jacobians without efficient endomorphisms as a source of cryptographic unknown-order groups with trustless setup [DGS22] (§3.6).

Isogenies also have destructive applications: my genus-3 attack [Smi08] was a totally new approach to DLP transfer, translating a technical and abstract result from algebraic geometry (Recillas’ trigonal construction) into explicit formulas and algorithms for cryptographers (§3.5). This result essentially killed genus 3 in DLP-based cryptography; [Smi08] won the best paper award at Eurocrypt 2008, and [Smi09] was an invited submission to Journal of Cryptology.

Chapter 4 covers commutative isogeny-based cryptography, which is based on actions of ideal class groups on isogeny classes of elliptic curves. My contributions to the foundations of group-action-based IBC (§4.2) include the survey article [Smi18], an invited contribution to WAIFI 2018, and a proof of the quantum equivalence of the group-action analogues of the Discrete Logarithm and Computational Diffie–Hellman Problems [Gal+21] with Steven Galbraith, Lorenz Panny, and Frederik Vercauteren. Luca De Feo, Jean Kieffer, and I tested the practical limits of group-action cryptography with ordinary curves in [DKS18], exposing serious parameterization problems at practical security levels. This led Castryck, Lange, Martindale, Panny, and Renes to switch from ordinary to supersingular curves with CSIDH [Cas+18]. Mathilde Chenu and I showed that the families of

¹[MSS16] was the result of Charlotte Scribot’s masters internship with us.
²Commutative IBC appears immune to the devastating recent attacks on SIDH and SIKE.
³[DKS18] was the subject of Jean Kieffer’s masters internship with us.
curves from [Smi16] admitted class group actions in [CS22], extending and generalizing CSIDH and its derivatives (§4.2). Gustavo Banegas, Valerie Gilchrist, and I explored fast supersingularity testing for CSIDH key-validation [BGS22] (§4.3).

Daniel J. Bernstein, Luca De Feo, Antonin Leroux, and I created the Vélusqrt algorithm [Ber+20] to compute isogenies in square-root instead of linear time (§4.4). This is the first improvement in complexity for prime isogeny computation since the 1970s; it immediately improves any IBC scheme using large-prime-degree isogenies, including CSIDH. The CTIDH project [Ban+21a] set new speed records for CSIDH through Vélusqrt and a fundamental reconception of the private keyspace, while improving resistance to side-channel attacks (§4.5).

Moving to higher dimensions, Wouter Castryck, Thomas Decru, and I designed a genus-2 isogeny-based hash function [CDS20], before Craig Costello and I substantially reduced its security with a new high-dimensional isogeny-finding algorithm [CS20] (with serious implications for the (in)security of higher-dimensional abelian varieties in IBC). Enric Florit and I investigated the structure and random-walk properties of the genus-2 superspecial isogeny graph [FS21a; FS21b] (§4.6).

Chapter 5 concerns implementations of public-key cryptosystems for constrained environments. We begin with an example from hyperelliptic cryptography: Joost Renes, Peter Schwabe, Lejla Batina, and I used Kummer surfaces (the genus-2 analogue of the $x$-coordinate on elliptic curves) to produce compact, high-speed key exchange and signature software for microcontroller platforms with very low memory requirements [Ren+16; RS17] (§5.1).

The Inria Défi RIOT-FP [RFP] aims to provide “future-proof” security for RIOT [Bac+18: RIO], a free and open-source operating system for low-end IoT devices. In this context, Gustavo Banegas, Adrian Herrmann, Koen Zandberg, Emmanuel Baccelli, and I evaluated postquantum signature schemes in the context of secure software updates with the SUIT protocol in [Ban+22] (§5.2). This work was selected for presentation at Real-World Cryptography 2022.

Most recently, Gustavo Banegas, Thomas Debris-Alazard, Milena Nedeljković, and I developed the first implementation of Wave, a post-quantum code-based signature scheme with fast verification on microcontrollers [Ban+21b]. I have continued this work in the Wave submission [Ban+23a] to the NIST signature on-ramp for post-quantum cryptography standardization [NIST-sigs].

### 1.3 Other topics

Some of my cryptographic research is not covered in this document: notably, pairings in genus 2 [Gal+09], Weil descent attacks on binary elliptic curves with special endomorphisms [CRS21], eliminating semi-abelian varieties as candidate cryptographic groups [GS06], classical improvements to quantum factoring algorithms [Gro+15], and related algorithms for deterministic factoring with number-theoretic oracles [MRS21]. We also set aside some results from explicit number theory including low-degree isogenies in genus 2 [Smi12], sporadic isogeny families in higher genus [Smi10; Smi11], and arithmetic statistics in genus 3 [Ler+19].
2 Classical elliptic-curve cryptography

We begin with a quick review of elliptic curves for cryptography in §§2.1-2.3. Then we summarize some contributions: §2.4 describes an original family of curves equipped with efficient endomorphisms, §2.5 describes high-speed key-exchange software based on one of these curves, and §2.6 describes a special point-counting algorithm for curves in the family. (For further background, see [Sil09; Cas91; Gal12; CS18].)

2.1 Elliptic curves for cryptography

An elliptic curve is, by definition, a curve of genus 1 with a distinguished rational point. More concretely, elliptic curves are realised by nonsingular plane cubics: for example, in characteristic $p \neq 2$ we have the short Weierstrass model

\[
\mathcal{E} : y^2 = x^3 + ax + b
\]

(2.1)

where $4a^3 + 27b^2 \neq 0$, or the Montgomery model

\[
\mathcal{E} : By^2 = x(x^2 + Ax + 1)
\]

(2.2)

where $A^2 \neq 4$ and $B \neq 0$. (The inequalities ensure that the curve is nonsingular.)

For cryptography, the coefficients $a$ and $b$ (or $A$ and $B$) are in a finite field $\mathbb{F}_q$, where $q = p^d$ is a prime power; typically $p$ is large and $d = 1$ or 2, though the "binary" case where $p = 2$ and $d$ is prime has also been of interest in the past, especially where dedicated hardware for finite field arithmetic is involved.

Points on $\mathcal{E}$ are solutions $(x, y) = (\alpha, \beta)$ to the equation of $\mathcal{E}$, with $\alpha$ and $\beta$ in $\mathbb{F}_q$, together with the unique "point at infinity" $\mathcal{O}$. The points form a commutative group $\mathcal{E}(\mathbb{F}_q) \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ for some $m$ and $n$ with $m \mid \gcd(n, q - 1)$ and $|\#\mathcal{E}(\mathbb{F}_q) - (q + 1)| \leq 2\sqrt{q}$. The group law is geometric, and therefore algebraic: it can be efficiently computed by a series of polynomial operations over $\mathbb{F}_q$.

To make the group law explicit for the Montgomery model in (2.2) above: $\mathcal{O}$ is the zero element; negation is $\ominus : (x : y : 1) \mapsto (x : -y : 1)$; and for addition, if $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ are points on $\mathcal{E}$, then $P \oplus Q = (x_\oplus, y_\oplus)$ where

\[
\begin{align*}
  x_\oplus &= B\lambda^2 - (x_P + x_Q) - A \\
  y_\oplus &= \lambda(x_P - x_\ominus) - y_P
\end{align*}
\]

where $\lambda = \begin{cases} \frac{y_Q - y_P}{x_Q - x_P} & \text{if } P \neq Q \text{ or } \ominus Q, \\
\frac{(3x_P^2 + 2Ax_P + 1)}{2By_P} & \text{if } P = Q \end{cases}$

if $P = \ominus Q$, then $P \oplus Q = \mathcal{O}$. We observe that $\lambda$ is the slope of the secant through $P$ and $Q$ (or the tangent to $\mathcal{E}$ at $P$, in the case $P = Q$).
In cryptographic applications, the critical operation is scalar multiplication: Given $P \in \mathcal{E}(\mathbb{F}_p)$ and $m \in \mathbb{Z}$, compute $[m]P := P + \cdots + P$ ($m$ copies of $P$).

For each positive integer $m$, the $m$-torsion subgroup is $\mathcal{E}[m] := \{P \in \mathcal{E} : [m]P = \mathcal{O}\}$. We say $\mathcal{E}$ is supersingular if $\mathcal{E}[p](\mathbb{F}_q) = 0$; otherwise, $\mathcal{E}$ is ordinary.

In theory, scalar multiplication is efficient: computing $[m]P$ takes $O(\log m)$ group operations in $\mathcal{E}(\mathbb{F}_p)$. Optimizing scalar multiplication in practice has long been a subject of intense and competitive research. The simplest scalar multiplication algorithm is double-and-add. Start with $R = P$; then, for each bit in the binary expansion of $m$ (from the most significant bit down), double $R$, and add $P$ if the bit is 1. Generalizing from the binary expansion to $2^w$-adic expansions yields a family of fixed-window scalar multiplications.

The Discrete Logarithm Problem (DLP) is the inverse of scalar multiplication: Given $P$ and $Q \in \mathcal{E}(\mathbb{F}_p)$ with $Q = [m]P$ for some $m$, compute $m$.

For general $\mathcal{E}(\mathbb{F}_p)$, the fastest DLP algorithms known run in time $O(\sqrt{N})$, where $N$ is the largest prime divisor of $\#\mathcal{E}(\mathbb{F}_p)$; so for prime-order curves, these algorithms solve DLPs in $O(\sqrt{p})$ by Hasse’s theorem, and the DLP is exponentially hard (in the input size, which is in $O(\log p)$).

Once we get beyond the algebraic geometry, elliptic curve cryptosystems are generally quite simple. The simplest is Diffie–Hellman key exchange, which will serve to illustrate all the relevant ideas.

Alice and Bob want to establish a common encryption key over a public channel, without prior contact. We have a public, fixed elliptic curve $\mathcal{E}$ and a point $P$ in $\mathcal{E}(\mathbb{F}_p)$. Alice samples a secret integer $a$, computes $A = [a]P$, and publishes $A$; Bob samples a secret integer $b$, computes $B = [b]P$, and publishes $B$. Then, Alice computes $S = [a]B$ and Bob computes $S = [b]A$; they have computed the same $S$ because $[a][b]P = [b][a]P = [ab]P$.

The public keys $A = [a]P$ and $B = [b]P$ represent instances of the DLP in $\mathcal{E}(\mathbb{F}_p)$: to recover a private key, we must solve a DLP instance, so $\mathcal{E}$ and $p$ must be chosen such that this is computationally infeasible. Recovering $S$ from $(P, A, B)$ is the Computational Diffie–Hellman Problem (CDHP). The CDHP is not obviously equivalent to the DLP but there exist polynomial-time reductions under strong number-theoretic hypotheses, and subexponential-time reductions under more plausible ones [Mau94; MW99; MSV04; Ben05]. In practice, all known CDHP algorithms are DLP algorithms. So: if the DLP is hard in $\mathcal{E}(\mathbb{F}_p)$ then $S$ is regarded as a shared secret, and Alice and Bob may derive a common secret encryption key from it.

The hard work for Alice and Bob is all in scalar multiplication. The scalars are secret keys, so the scalar multiplication must be done in (cryptographic) constant time: this means that the execution path, including the sequence of instructions and memory access locations, must be independent of the scalars so as to reduce leakage of secret values to side channels. For this we use the Montgomery ladder [Mon87; CS18], which computes scalar multiplication “up to sign”: that is,

$$(x(P), m) \mapsto x([m]P).$$
The ladder does not use $y$-coordinates, or the full elliptic-curve group law: instead, we use (projective versions of) the relations
\[
x_{P \oplus Q} x_{P \ominus Q} (x_P - x_Q)^2 = (x_P x_Q - 1)^2 \quad \text{if} \ P \neq Q, \quad (2.3)
4x_{[2]P} x_P (x_P^2 + Ax_P + 1) = (x_P^2 - 1)^2. \quad (2.4)
\]
If $y([m]P)$ is required, then it is easily recovered from $y(P)$, $x([m]P)$, and $x([m+1]P)$, which is in fact an auxiliary output of the ladder.

### 2.2 Isogenies and endomorphisms

We do not speak of vector spaces without speaking of linear transformations and matrices, or of groups without speaking of homomorphisms. Similarly, we cannot go far with elliptic curves without introducing the morphisms between them.

A **homomorphism** of elliptic curves over $\mathbb{F}_q$ is an algebraic map $\phi : \mathcal{E}_1 \rightarrow \mathcal{E}_2$ (defined by coherent systems of tuples of polynomials, or rational functions, with coefficients in $\mathbb{F}_q$) mapping $\mathcal{O}_{\mathcal{E}_1}$ to $\mathcal{O}_{\mathcal{E}_2}$. Such a mapping is necessarily a homomorphism in the usual sense: it respects the group laws on $\mathcal{E}_1$ and $\mathcal{E}_2$, and induces a homomorphism $\mathcal{E}_1(\mathbb{F}_q) \rightarrow \mathcal{E}_2(\mathbb{F}_q)$ (see for example [Sil09, Theorem III.4.8]).

An **isogeny** is a homomorphism $\phi : \mathcal{E}_1 \rightarrow \mathcal{E}_2$ with finite kernel that is (geometrically) surjective. For example, if $\mathcal{E} : y^2 = x^3 + a_2 x^2 + a_4 x + a_6$ is a curve over $\mathbb{F}_q$, then there is a $p$-power *Frobenius isogeny*
\[
\pi_p : (x, y) \mapsto (x^p, y^p)
\]
from $\mathcal{E}$ to the Galois-conjugate curve
\[
\mathcal{E}^{(p)} : y^2 = x^3 + a_2^p x^2 + a_4^p x + a_6^p.
\]
An isogeny is **separable** if it does not factor through $\pi_p$. If $S$ is a finite subgroup of $\mathcal{E}$, then there exists a curve $\mathcal{E}/S$ and a separable isogeny $\phi : \mathcal{E} \rightarrow \mathcal{E}/S$ with kernel $S$; Vélu’s formulæ [Vél71] compute $\mathcal{E}/S$ and $\phi$ (see [Koh96, §2.4]).

An **endomorphism** is a homomorphism from a curve to itself. The simplest examples are the scalar multiplications $[m]$ and the *Frobenius endomorphism*
\[
\pi_q : (x, y) \mapsto (x^q, y^q).
\]

The endomorphisms of $\mathcal{E}$ form a ring, $\text{End}(\mathcal{E})$: composing with $\oplus$ on $\mathcal{E}$ induces addition on endomorphisms, and multiplication is composition of endomorphisms. The structure of $\text{End}(\mathcal{E})$ is a fundamental invariant of $\mathcal{E}$; it is used, implicitly or explicitly, in many of the algorithms discussed here. Any isogeny $\phi : \mathcal{E}_1 \rightarrow \mathcal{E}_2$ induces a close relationship between $\text{End}(\mathcal{E}_1)$ and $\text{End}(\mathcal{E}_2)$: in particular, the endomorphism algebras $\text{End}(\mathcal{E}_1) \otimes \mathbb{Q}$ and $\text{End}(\mathcal{E}_2) \otimes \mathbb{Q}$ are isomorphic.

### 2.3 Endomorphisms and scalar decomposition

Since elliptic curve scalar multiplication is analogous to exponentiation in finite fields, many algorithms originally developed for multiplicative groups transfer directly to scalar multiplication: finite-field square-and-multiply becomes elliptic-curve double-and-add, for example. But the geometry of elliptic curves can offer

---

1Equivalently, we can define elliptic isogenies to be nonzero homomorphisms; but when we move to higher dimensions, there are nonzero homomorphisms that are not isogenies.
us new algorithms with no true finite field analogues. A spectacular (and easy) example of this phenomenon is scalar multiplication with endomorphism decompositions, originally proposed by Gallant, Lambert, and Vanstone [GLV01].

Let $E/F_q$ be an elliptic curve with a cyclic subgroup $G \subset E(F_q)$ of order $N$. Given an integer $0 \leq m < N$ our goal is to evaluate $[m]$ on $G$ as quickly as possible.

Let $\psi$ be an endomorphism of $E$ such that $\psi(G) \subseteq G$ (this is typical in cryptographic applications, where $N$ is so close to $\#E(F_q)$ that there is no room for the image of $G$ to be anything but $G$ itself.) Then $\psi$ restricts to an endomorphism of $G$; since $G$ is cyclic, its endomorphism ring is isomorphic to $\mathbb{Z}/N\mathbb{Z}$; and so $\psi$ acts on $G$ as multiplication by an integer eigenvalue $-N/2 < \lambda \leq N/2$ such that

$$\psi|_G = [\lambda]|_G.$$  

A decomposition of $m$ is any pair $(a, b)$ such that

$$m \equiv a + b\lambda \pmod{N}.$$  

Given a decomposition $(a, b)$ of $m$, we can compute $[m]P$ for any $P$ in $G$ using a multiexponentiation algorithm\(^2\) on $(P, \psi(P))$ to compute $[m]P$ as

$$[m]P = [a]P \oplus [b]\psi(P).$$  

For example: given $(P, \psi(P))$, Straus’ algorithm [Str64] computes $[a]P \oplus [b]\psi(P)$ with $\log_2 \max(|a|, |b|)$ doubles and (on average) $\frac{3}{4} \log_2 \max(|a|, |b|)$ adds, while traditional 1-dimensional double-and-add computes $[m]P$ with $\log_2 |m|$ doubles and (on average) $\frac{1}{2} \log_2 |m|$ adds.

For scalar decomposition to offer an advantage over basic scalar multiplication, the endomorphism $\psi$ and its eigenvalue $\lambda$ must satisfy two conditions:

1. $\psi$ must be efficiently computable—that is, $P \rightarrow \psi(P)$ should cost no more than a few group operations—to ensure the cost of initialising the multiexponentiation does not overwhelm its benefits.

2. $\lambda$ must be large, to ensure $\max(|a|, |b|) \ll |m| \approx N$ so multiexponentiation is faster than basic scalar multiplication. If $|\lambda| > \sqrt{N}$, then we can achieve $\max(|a|, |b|) \in O(\sqrt{N})$ (and this is optimal); finding such $(a, b)$ is easy, and almost trivial for the $\psi$ and $\lambda$ in our applications [Smi15].

Achieving C1 and C2 simultaneously is hard: for general $E/F_q$, the only endomorphisms we know are linear combinations of $[1]$ and $\pi_q$. While $[1]$ and $\pi_q$ are obviously efficient, both have eigenvalue 1. Any combination meeting C1 will have tiny coefficients, and thus a tiny eigenvalue, violating C2; and conversely, any combination meeting C2 violates C1.

We therefore need a supply of elliptic curves over $F_q$ equipped with efficiently-computable endomorphisms that are not small linear combinations of $[1]$ and $\pi_q$. Gallant, Lambert, and Vanstone proposed using what are essentially reductions modulo $p$ of CM curves over number fields.

\(^2\)The literature on this topic is vast, and we will not attempt to summarize it here; for an introduction, we recommend [Knu97, §4.6.3], [Gal12, §2.8, §11.2] and [Ava+06, Chapter 9]).
Example 1 (GLV). Suppose \( p \equiv 1 \pmod{4} \), so \(-1\) is a square in \( \mathbb{F}_p \). The curve

\[
\mathcal{E} : y^2 = x^3 + x
\]

has a highly efficient endomorphism

\[
\psi : (x, y) \mapsto (-x, \sqrt{-1}y)
\]

and \( \psi^2 = [-1] \). Clearly \( \psi \) meets C1. It also meets C2: given a subgroup \( \mathcal{G} \subset \mathcal{E}(\mathbb{F}_p) \) as above, the eigenvalue \( \lambda \) is a square root of \(-1\) modulo \( N \), so \( |\lambda| \geq \sqrt{N-1} \).

More generally, we can use any \( \mathcal{E}/\mathbb{F}_q \) equipped with a very low-degree non-integer endomorphism (see [GLV01] for examples). The problem is that there are very few isomorphism classes of curves with such endomorphisms. If \( p \) is fixed—for example, if \( p \) is a special prime facilitating high-speed field arithmetic—then perhaps none of these curves have secure subgroups.

Galbraith, Lin, and Scott showed in [GLS11] that we can overcome the scarcity of secure GLV curves over \( \mathbb{F}_p \) for fixed \( p \) by using special curves over \( \mathbb{F}_p^2 \).

Example 2 (GLS). Fix a nonsquare \( \delta \) in \( \mathbb{F}_p^2 \) and let \( \mathcal{E}/\mathbb{F}_p^2 \) be any curve in the form

\[
\mathcal{E}^0/\mathbb{F}_p^2 : y^2 = x^3 + \delta^2 a_0 x + \delta^3 b_0
\]

with \( a_0, b_0 \in \mathbb{F}_p \).

Now \( \tau : (x, y) \mapsto (u, v) = (\delta^{-1} x, \delta^{-3/2} y) \) is an isomorphism of \( \mathcal{E} \) onto \( \mathcal{E}^0/\mathbb{F}_p^2 : v^2 = u^3 + \delta a_0 u + b_0 \), which is the base-extension to \( \mathbb{F}_p^2 \) of a curve over \( \mathbb{F}_p \), and as such has a \( p \)-power Frobenius endomorphism \( \pi_p : \mathcal{E}_0 \to \mathcal{E}_0 \). The twisted endomorphism

\[
\psi := \tau^{-1} \circ \pi_p \circ \tau : (x, y) \mapsto (\delta^{1-p} x^p, \delta^{3(1-p)/2} y^p)
\]

in \( \text{End}(\mathcal{E}) \) is defined over \( \mathbb{F}_p^2 \) because \( 2 \mid (p-1) \), and meets C1 (especially with a good choice of \( \delta \)): if \( p \equiv 5 \pmod{8} \), then \( \delta = \sqrt{2} \) yields \( \psi : (x, y) \mapsto (-x^p, \sqrt{-1} y^p) \). It also meets C2: \( \pi_p \tau = -\tau \pi_p^2 \), so

\[
\psi^2 = -\pi_p^2 
\]

so \( \lambda \) is a square root of \(-1\) on any \( \mathcal{G} \subset \mathcal{E}(\mathbb{F}_p^2) \) fixed by \( \psi \), hence \( |\lambda| \geq \sqrt{N-1} \).

This construction gives us \( O(p) \) isomorphism classes of curves over any \( \mathbb{F}_p^2 \), so we can find secure GLS curves for any given \( p \). However, we cannot find GLS curves \( \mathcal{E}/\mathbb{F}_p^2 \) such that \( \mathcal{E} \) and its quadratic twist \( \mathcal{E}_0 \) simultaneously have secure subgroups, because \( \mathcal{E}_0 \) is a subfield curve. Twist-security is an important property for efficient elliptic-curve Diffie–Hellman constructions (see [Ber06a], [Fou+08]).

2.4 The \( \mathbb{Q} \)-curve construction

The \( \mathbb{Q} \)-curve construction of [Smi13] and [Smi16] is a generalization of GLV and GLS allowing a large choice of twist-secure curves when \( p \) is fixed.

Fix a nonsquare \( \Delta \) in \( \mathbb{F}_p \), so \( \mathbb{F}_p^2 = \mathbb{F}_p(\sqrt{\Delta}) \). As we noted above, every

\[
\mathcal{E}/\mathbb{F}_p^2 : y^2 = x^3 + a_2 x^2 + a_4 x + a_6
\]

has Frobenius isogenies \( \pi_p : (x, y) \mapsto (x^p, y^p) \) to and from its Galois conjugate

\[
(\pi \mathcal{E}) : y^2 = x^3 + a_2^p x^2 + a_4^p x + a_6.
\]
Suppose there exists an isogeny $\phi : \mathcal{E} \to \mathcal{E}$ of very small degree $d$. Composing with $\pi_p : \mathcal{E}(\mathbb{F}_p) \to \mathcal{E}$ yields an endomorphism $\psi : \mathcal{E} \to \mathcal{E}$ of degree $dp$. Now $\psi$ meets $C_1$ because $d = \deg(\phi)$ is tiny and $\pi_p$ can be evaluated almost for free on points of $\mathcal{E}(\mathbb{F}_p)$ because $p$-th powering amounts to changing the sign on $\sqrt{\Delta}$.

Curves over $\mathbb{F}_p$ with $d$-isogenies to their conjugates are parametrized by the $\mathbb{F}_p$-points of an Atkin–Lehner quotient of the modular curve $X_0(d)$; in particular, for $d = 2, 3, 5$, and 7 this quotient is rational, and we get one-parameter families of curves with efficiently computable endomorphisms. (The case $d = 1$, where $\phi$ is an isomorphism, is isomorphic to GLS.) In [Smi13] and [Smi16], this is expressed in terms of reductions of quadratic $\mathbb{Q}$-curves—that is, elliptic curves defined over quadratic number fields with isogenies to their conjugates—modulo inert $p$.

The curves of Example 3 were constructed in this way from Hasegawa’s universal family of degree-2 quadratic $\mathbb{Q}$-curves [Has97, Theorem 2.2].

Example 3 ([Smi16, §5]). Fix a nonsquare $\Delta$ in $\mathbb{F}_p$ and consider the family $\mathcal{E}_{2,s}/\mathbb{F}_p^2 : y^2 = x^3 + 2(C_2(s) - 24)x - 8(C_2(s) - 16)$ where $C_2(s) := 9(1 + s\sqrt{\Delta})$ for $s$ in $\mathbb{F}_p$. There is a 2-isogeny $\phi_{2,s} : \mathcal{E}_{2,s} \to \mathcal{E}_{2,s}$ with kernel $((4,0))$, defined by $\phi_{2,s} : (x, y) \mapsto \left(\frac{-x}{2} - \frac{C_2(s)}{x - 4}, \frac{y}{\sqrt{-2}} \left(\frac{-1}{2} + \frac{C_2(s)}{x - 4}\right)\right)$.

The composition $\psi_{2,s} := \pi_p \circ \phi_{2,s}$ satisfies $C_1$—and also $C_2$, because

$$\psi_{2,s}^2 = \epsilon_p |_{\mathcal{E}_{2,s}}, \quad \text{where} \quad \epsilon_p := -(-2/p) = \begin{cases} -1 & \text{if } p \equiv 1, 3 \pmod{8}, \\ 1 & \text{if } p \equiv 5, 7 \pmod{8}. \end{cases}$$

### 2.5 High-speed key exchange software

In [CHS14], we present high-speed Diffie–Hellman key exchange software based on the Montgomery model of a curve from the family in Example 3 over $\mathbb{F}_p^2 = \mathbb{F}_p(\sqrt{-1})$ where $p$ is the Mersenne prime $2^{127} - 1$ (this allows very fast field arithmetic). The curve parameters and their selection are detailed in [CHS14, §2].

The chief novelty here is the combination of the new endomorphism with $x$-only Montgomery arithmetic. This means multiexponentiation using a 2D differential addition chain: we tried 2D chains from Montgomery [Mon87], Azares-Akhsh and Karabina [AK14], and Bernstein [Ber06b]. To initialise these chains, we need to efficiently compute the $x$-line action not only of $\psi$, but also of $\psi - 1$ (see [CHS14, §3]). We also tested 1D scalar multiplication on $\mathcal{E}$ using the Montgomery Ladder, to isolate the impact of scalar decomposition.

Table 2.1 presents our results. Compiled and ran on the same platform, our curve with the 1D Ladder (and no endomorphism) is $\approx 1.14 \times$ faster than Bernstein’s ladder-based Curve25519 software [Ber06a], which is the state of the art for constant-time key exchange at the 128-bit security level. This speedup is mostly due to the faster field arithmetic afforded by working with a Mersenne prime of half the bitlength. Using the endomorphism with the 2D Bernstein chain, we are $\approx 1.23 \times$ faster than Curve25519. This small improvement relative to the $\approx 1.14 \times$ speedup already achieved for the 1D case is disappointing: in fact, while we halve...
Table 2.1: Scalar multiplication performance, on $10^6$ random inputs, for the curve of §2.5. C/assembly implementation on an Intel Core i7-3520M processor (2.89GHz, hyperthreading and Turbo Boost disabled), compiled with gcc v4.6.3 -O2. Cycles counted using the SUPERCOP toolkit [BL]. CT denotes constant-time. xDBLADD denotes combined xDBL and xADD with shared operands.

<table>
<thead>
<tr>
<th>Multiexponentiation</th>
<th>Addition chain</th>
<th>Dim.</th>
<th>CT</th>
<th>xDBL av. s.d.</th>
<th>xADD av. s.d.</th>
<th>xDBLADD av. s.d.</th>
<th>kCycles av.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ladder [Mon87]</td>
<td>1 ✓</td>
<td>1</td>
<td></td>
<td>253</td>
<td>159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ber06b]</td>
<td>2 ✓</td>
<td>1</td>
<td>128</td>
<td></td>
<td>127</td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>[AK14]</td>
<td>2 x</td>
<td>1</td>
<td>1</td>
<td>179.6 6.7</td>
<td></td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>PRAC [Mon87]</td>
<td>2 x</td>
<td>0.2</td>
<td>0.4</td>
<td>73.4 11.1</td>
<td></td>
<td>109</td>
<td></td>
</tr>
</tbody>
</table>

the number of xDBLs, most of our savings are negated by awkward field exponentiations needed to evaluate $\psi - 1$ and normalize the images of $\psi$ and $\psi - 1$.

The $\mathbb{Q}$-curve construction was subsequently used to develop Microsoft’s FourQ software, currently the world’s fastest Diffie–Hellman key exchange. The FourQ curve is in the intersection of the $d = 2$ and $d = 5 \mathbb{Q}$-curve families over $\mathbb{F}_{2^{127}}$ with the Mersenne $p = 2^{127} - 1$; of the handful of intersections over this fast $\mathbb{F}_{2^{127}}$, this curve luckily has a cryptographically strong subgroup! This allows 4D scalar decompositions, which speeds up scalar multiplication dramatically when used with Edwards coordinates and a fixed-window multiexponentiation.

### 2.6 Accelerating point counting

To apply scalar decomposition in real cryptosystems, we must find secure elliptic curves from the families above—which means computing their orders. It would be wrong to do this without exploiting their special endomorphism structures.

The task of computing $\#\mathcal{E}(\mathbb{F}_q)$ is called point counting. Hasse’s theorem tells us that

$$\#\mathcal{E}(\mathbb{F}_q) = q + 1 - t$$

where $t^2 \leq 4q$.

Here, $t$ is the trace of the Frobenius endomorphism $\pi_q^2$: all efficient point-counting algorithms are based on computing invariants of the action of Frobenius on some module associated with $\mathcal{E}$. For large $p$, which is the case most relevant here, the classic algorithm is the Schoof–Elkies–Atkin (SEA) algorithm [Sch85; Sch95].

GLV curves (as in Example 1) are reductions of CM curves over number fields, so we can derive the curve order from $p$ and the CM discriminant, as in [Sch95, §4]. For GLS curves (as in Example 2), we can use the fact that

$$\#\mathcal{E}_0(\mathbb{F}_p) = p + 1 - t_0 \implies \#\mathcal{E}(\mathbb{F}_{p^2}) = (p + 1 - t_0)(p + 1 + t_0)$$

to reduce point-counting on GLS curves over $\mathbb{F}_{p^2}$ to point-counting on general elliptic curves over $\mathbb{F}_p$, where we apply SEA as usual. The $\mathbb{Q}$-curve reductions of §2.4 do not have fixed CM discriminants or subfield structures, so none of those tricks apply. But [MSS16] shows that we can still improve on SEA.

Recall that in the SEA algorithm, we detect small primes $\ell$ splitting in $\mathbb{Z}[\pi_{p^2}]$ by computing roots of $\Phi_j(\ell, X)$ where $\Phi_j$ is a modular polynomial; for split $\ell$, we construct the polynomials $f_\ell(X)$ defining the corresponding $\ell$-isogeny kernels
CHAPTER 2. CLASSICAL ELLIPTIC-CURVE CRYPTOGRAPHY

Table 2.2: Average time to compute $\#\mathcal{E}_{2, s}(\mathbb{F}_{p^2})$ for $1 \leq s \leq 100$, with $\mathcal{E}_{2, s}$ as in §2.4. Algorithms implemented in C++ using NTL 9.6.4, compiled with gcc 4.9.2, and ran on an Intel Xeon platform (E5520 CPU at 2.27GHz). “Powering” denotes time spent computing $X^p$ or $X^{p^2}$ modulo kernel polynomials.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>128-bit $p = 314159...459$</th>
<th>255-bit $p = 314159...963$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>av. max $\ell$</td>
<td>Powering</td>
</tr>
<tr>
<td>SEA [Sch95]</td>
<td>164</td>
<td>9.11s</td>
</tr>
<tr>
<td>New [MSS16]</td>
<td>62</td>
<td>2.62s</td>
</tr>
</tbody>
</table>

using Elkies’ algorithm, and then compute the eigenvalues of $\pi_{p^2}$ on these kernels using explicit symbolic computations (dominated by computing $X^{p^2}$ mod $f_\ell(X)$). The eigenvalues determine the trace $t_\ell$ of $\pi_{p^2}$ modulo $\ell$, and after sufficiently many $\ell$ we can recover $t_\ell$, and hence $\#\mathcal{E}(\mathbb{F}_{p^2}) = p^2 + 1 - t_\ell$, using the CRT.

In [MSS16] we use $\psi$ instead of $\pi_{p^2}$. The eigenvalues of $\psi$ determine the eigenvalues of $\pi_{p^2}$ (and hence $t_\ell$, and $\#\mathcal{E}(\mathbb{F}_{p^2})$) because $\psi^2 = \pm [d]\pi_{p^2}$; and

$$\psi^2 - [r] [d] \psi + [dp] = 0 \quad \text{in } \text{End}(\mathcal{E}), \quad \text{where } r^2 < 4p/d.$$

Compared with $t_\ell = c_p(dr^2 - 2p)$, the integer $r$ is determined by its value modulo half as many primes $\ell$, and of about half the size. As a bonus, $\psi$ is easier to evaluate than $\pi_{p^2}$ on symbolic points (we only need $X^p$, rather than $X^{p^2}$, modulo $f_\ell(X)$).

Globally, the algorithm of [MSS16] has the same asymptotic complexity as SEA, but with better constants: halving the number and size of the primes $\ell$ implies a $\approx 4 \times$ speedup, and our experiments confirm this (see Table 2.2). This speedup is welcome when searching for twist-secure curves, which requires point counting on thousands of candidate curves; when doing this, we use an early-abort variant which stops when the order mod $\ell$ is incompatible with a secure curve order.
3 Classical hyperelliptic cryptography

What is so special about elliptic curves? Formally, we can replace elliptic curves in cryptosystems with more general algebraic groups over $\mathbb{F}_p$; *principally polarized abelian varieties* (PPAVs). A $g$-dimensional PPAV is a projective algebraic variety—hence, with points corresponding to tuples of field elements—and with a group law defined by polynomial mappings. Elliptic curves are the case $g = 1$.

This chapter surveys some contributions to *hyperelliptic* curve-based cryptography. Hyperelliptic curves of genus $g$ are the most accessible source of efficient $g$-dimensional PPAVs. After quickly reviewing some background in §3.1, we consider the problem of realising isogenies and endomorphisms of hyperelliptic Jacobians in §3.2, before applying this to scalar multiplication in §3.3 and point counting in §3.4. These results are all specialized to the case $g = 2$; in §3.5, we explain how to use explicit isogenies to attack discrete logarithm problem instances in genus $g = 3$. Finally, we describe a surprising new possibility for constructing trustless unknown-order groups from genus-3 curves in §3.6.

3.1 Hyperelliptic curves and cryptosystems

Jacobians of hyperelliptic curves are the most accessible examples of PPAVs. A hyperelliptic curve of genus $g$ is defined by an affine equation

$$\mathcal{X} : y^2 = f(x)$$

with $f$ squarefree and $\deg(f) \in \{2g + 1, 2g + 2\}$.

Divisors on $\mathcal{X}$ are formal sums of points on $\mathcal{X}$; the divisors on $\mathcal{X}$ form a group $\text{Div}(\mathcal{X})$ under addition. The degree of a divisor $\sum_{P \in \mathcal{X}} n_P P$ is the sum of its coefficients, $\sum_{P \in \mathcal{X}} n_P$; the divisors of degree $0$ form a subgroup $\text{Div}^0(\mathcal{X}) \subset \text{Div}(\mathcal{X})$. Each nonzero rational function $f$ on $\mathcal{X}$ has an associated *principal divisor* $\text{div}(f)$, which is the formal sum of its zeroes minus its poles (all counted with multiplicity). Now every principal divisor has degree $0$ and $\text{div}(fg) = \text{div}(f) + \text{div}(g)$, so $\text{Prin}(\mathcal{X})$ is a subgroup of $\text{Div}^0(\mathcal{X})$. The quotient $\text{Div}(\mathcal{X})/\text{Prin}(\mathcal{X})$ is the *Picard group* $\text{Pic}(\mathcal{X})$, and the quotient $\text{Div}^0(\mathcal{X})/\text{Prin}(\mathcal{X})$ is the *degree-0 Picard group* $\text{Pic}^0(\mathcal{X})$.

In fact $\text{Pic}(\mathcal{X}) \cong \text{Pic}^0(\mathcal{X}) \times \mathbb{Z}$, and $\text{Pic}^0(\mathcal{X})$ is parametrized by a $g$-dimensional PPAV $\text{Jac}(\mathcal{X})$, the *Jacobian* of $\mathcal{X}$. That is: points on $\text{Jac}(\mathcal{X})$ correspond to degree-$0$ divisor classes on $\mathcal{X}$. We usually compute with $\text{Jac}(\mathcal{X})$ using the *Mumford representation* [Mum84], where points on $\text{Jac}(\mathcal{X})$ are encoded as pairs of polynomials $(a(x), b(x))$ such that $a(x)$ is monic, $\deg(b) \leq \deg(a) \leq g$, and $b^2 \equiv f \pmod{a}$. The group law is worked out in detail by Cantor in [Can87].

Koblitz proposed Jacobians of genus-$g$ hyperelliptic curves as a basis for cryptosystems in 1989 [Kob89]. If $\mathcal{X}$ has genus $g \ll q$ then $\#\text{Jac}(\mathcal{X}) \approx q^g$, and Pollard
rho solves the DLP in time $O(q^{g/2})$; so we can trade higher genera $g$ against proportionally smaller field primes $p$, potentially reducing the overall cost of cryptographic operations. But index calculus algorithms such as [Gau07], which runs in time $O(q^{2g})$ for $g > 2$, have reduced the values of $g$ for which the resulting cryptosystems offer competitive efficiency and key sizes.

Nowadays, especially after the results described in §3.5, we are mostly confined to $g = 2$ for (constructive) cryptographic applications. In that case, the Kummer surface $Kum(K) := Jac(K)/\langle \pm 1 \rangle$ is a particularly convenient object to compute with: Kummer arithmetic in genus 2 corresponds to $x$-only arithmetic for elliptic curves [Hud05; CF96]. Gaudry [Gau07] initiated the use of Kummer surfaces in high-speed cryptographic implementations, where they have proven competitive with elliptic curve implementations (see Chapter 5).

### 3.2 Realizing higher-genus isogenies with correspondences

When computing isogenies of higher-dimensional abelian varieties, the first question is how the isogenies should be represented. In my Ph.D. thesis [Smi06], I argued that (divisorial) correspondences are a convenient representation for isogenies and endomorphisms of Jacobians of curves.

Let $X$ and $Y$ be algebraic curves, and consider the product surface $X \times Y$; we write $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ for the projections onto each factor. A correspondence is a divisor on $X \times Y$: that is, a formal sum of curves $R$ on $X \times Y$.

If we suppose $R$ is a prime divisor and not supported on the fibres of $\pi_1$ or $\pi_2$, then we can compose the pullback $(\pi_1|_R)^* : \text{Div}(X) \to \text{Div}(R)$ with the pushforward $(\pi_2|_R)_* : \text{Div}(R) \to \text{Div}(Y)$ to get an induced homomorphism $\phi_R : \text{Div}(X) \to \text{Div}(Y)$. This homomorphism maps principal divisors onto principal divisors, and degree-0 divisor classes onto degree-0 divisor classes, so we have induced homomorphisms $\phi_R : \text{Pic}(X) \to \text{Pic}(Y)$ and $\phi_R : \text{Jac}(X) \to \text{Jac}(Y)$. In fact, every homomorphism $\phi : \text{Jac}(X) \to \text{Jac}(Y)$ can be represented by a correspondence on $X \times Y$, but this representation is never unique: we can add and subtract principal divisors and fibres of $\pi_1$ and $\pi_2$ without changing the induced homomorphism.

Suppose we want to evaluate $\phi_R$ at a divisor class $[D]$ in $\text{Pic}(X)$. We can always choose a representative $\sum_i P_i$ for $[D]$ such that the fibres $\pi_1^{-1}(P_i)$ intersect transversally with $R$, taking that intersection yields a divisor on $R$, which we then map onto $Y$ via $\pi_2$. Transforming this geometric process into a fast algorithm depends on the geometry of $X$ and $Y$, and on the choice of correspondence.

### 3.3 Hyperelliptic endomorphisms and scalar multiplication

In [Smi06, Chapter 7] and [KS06], we extend endomorphism-accelerated scalar multiplication to families of hyperelliptic Jacobians over large prime fields. As in §2.4, the advantage is that we can construct one- and two-parameter families of curves whose Jacobians have efficiently computable endomorphisms over a fixed $\mathbb{F}_p$; we can certainly find cryptographically strong Jacobians in these families.

¹Going further: the map $R \to \phi_R$ extends $\mathbb{Z}$-linearly to give a homomorphism $\text{Div}(X \times Y) \to \text{Hom}(\text{Jac}(X), \text{Jac}(Y))$, which factors through the Picard group of $X \times Y$ to give an isomorphism $\text{Pic}(X \times Y) \cong \text{Pic}(X) \times \text{Pic}(Y) \times \text{Hom}(\text{Jac}(X), \text{Jac}(Y))$.
briefly recalls a simple one-parameter construction due to Tautz, Top, and Verberkmoes [TTV91]—itself a degeneration of a more complicated two-parameter construction due to Mestre [Mes91].

Example 4. Let $p$ be a prime such that $5$ is a square in $\mathbb{F}_p$, and let $\eta_5 = -(1 + \sqrt{5})/2$ and $\overline{\eta}_5 = -(1 - \sqrt{5})/2$ in $\mathbb{F}_p$ (note that $\eta_5 = \rho_5 + 1/\rho_5$ for some primitive $5$-th root of unity in $\mathbb{F}_p^*$). Consider the one-parameter family of genus-2 curves

$$\mathcal{X}_1/\mathbb{F}_p: y^2 = D_3(x) + t$$

where $D_3(x) = x^5 - 5x^3 + 5x$ is the degree-5 Dickson polynomial of the first kind (with parameter $1$). Now

$$D_3(x_1) - D_3(x_2) = (x_1 - x_2)A_3(x_1, x_2)\overline{A}_3(x_1, x_2)$$

where

$$\begin{cases} A_3(x_1, x_2) = x_1^2 + x_2^2 + \eta_5 x_1 x_2 + \eta_5^2 - 4, \\
\overline{A}_3(x_1, x_2) = x_1^2 + x_2^2 + \overline{\eta}_5 x_1 x_2 + \overline{\eta}_5^2 - 4,
\end{cases}$$

so the hypersurface $V(y_1 - y_2) \subset \mathcal{X}_1 \times \mathcal{X}_1$ decomposes into three prime components: the diagonal $\Delta = (x_1 - x_2, y_1 - y_2)$, and correspondences

$$\mathcal{R}_5 := V(y_1 - y_2, A_3(x_1, x_2)) \quad \text{and} \quad \overline{\mathcal{R}}_5 := V(y_1 - y_2, \overline{A}_3(x_1, x_2)).$$

The diagonal $\Delta$ induces the identity map. There is an injection $\mathbb{Z}[(1 + \sqrt{5})/2] \to \text{End}(\text{Jac}(\mathcal{X}_1))$ mapping $(1 + \sqrt{5})/2$ to $\phi_{\mathcal{R}_5}$ and $(1 - \sqrt{5})/2$ to $\phi_{\overline{\mathcal{R}}_5}$.

The endomorphisms $\phi_{\mathcal{R}_5}$ can be used to accelerate scalar multiplication on the Jacobians $\text{Jac}(\mathcal{X}_1)$: their eigenvalues are very large (and the resulting scalar decompositions are suitably short) because the minimal polynomial of $(1 + \sqrt{5})/2$ has very small coefficients. As with GLS and the $\mathbb{Q}$-curve construction of §2.4, the fact that $\mathcal{X}_1$ is a one-parameter family means that we can hope to find secure Jacobians over $\mathbb{F}_p$ with special $p$ affording fast field arithmetic. It remains to show that $\phi_{\mathcal{R}_5}$ can be evaluated efficiently on points of $\text{Jac}(\mathcal{X}_1)$. In [Smi06] and [KS06], we give a general method for “compiling” correspondences in the form of $\mathcal{R}_5$ to efficient programs evaluating $\phi_{\mathcal{R}_5}$ on the Mumford representation.\footnote{Similar results appeared independently in [Tak06].}

The results of [KS06] have not yet been included in a “competitive” scalar multiplication implementation. Initially, this was due to a lack of fast point counting in genus 2, which meant computing secure instances was hard (this problem is solved in §3.4). Another problem is the explicit arithmetic of genus-2 Jacobians, which lags far behind elliptic curves when it comes to basic side-channel protections: we still do not have satisfactory constant-time group operations.

Remark 1. Example 4 is the case $n = 5$ of a more general construction in [TTV91] which, for each odd prime $n$, constructs a one-parameter family of curves of genus $(n - 1)/2$ whose Jacobians have explicit RM by the maximal real subring of the cyclotomic field $\mathbb{Q}(\zeta_n)$. (The case $n = 3$ yields a one-parameter family of elliptic curves, but the explicit endomorphism constructed is just $[-1]$.) See [Smi06, Chapter 7] and [KS06] for more discussion and algorithms for $n \neq 5$ and the two-parameter family of [Mes91]. The cryptographic interest of $n > 5$ is limited by the impact of index calculus [Gau+07] and the results of §3.5.
Chapter 3. Classical Hyperelliptic Cryptography

3.4 Point counting with efficient real multiplication

As we saw in §2.6, efficient endomorphisms are not only useful for accelerating scalar multiplication: they can also yield significant speed-ups in point counting. The results are even more spectacular for genus-2 curves, because point-counting on these curves is already much harder than it is on elliptic curves.

Schoof’s polynomial-time elliptic point-counting algorithm [Sch85] was generalized to PPAVs—including Jacobians of higher-genus curves—by Pila [Pil90]. But Pila’s algorithm is only polynomial-time for PPAVs of fixed dimension: the implicit dependence on the dimension is badly exponential, to the point that this algorithm has never been implemented (not even in genus 2).³ Instead, we have Schoof-type algorithms designed specifically for genus-2 Jacobians by Gaudry and Schost [GS04b; GS04a] and Pitcher [Pit09].

The record for point counting for general genus-2 curves over prime fields was set by Gaudry and Schost while searching for twist-secure Jacobians at the 128-bit security level [GS12]. Working over the Mersenne prime field \( \mathbb{F}_p \) where \( p = 2^{127} - 1 \), they found a Jacobian whose order is 16 times a 250-bit prime and whose quadratic twist also has order 16 times a 250-bit prime. (The cofactor of 16 is required for compatibility with the high-speed Kummer arithmetic of [Gau07].) Computing one Jacobian order over this field cost around 1000 core-hours (on an Intel Xeon L5640 processor running at 2.27GHz); finding the twist-secure curve cost around \( 10^6 \) core-hours on the same processor. Looking at Table 2.2, where computing the order of a similar-sized elliptic curve on a similar processor running at the same speed cost around 20 core-seconds using the SEA algorithm, we can see that the transition from genus 1 to genus 2 takes point-counting from a routine to an extremely challenging problem.

In [GKS11], we accelerate point counting for genus-2 curves with efficient real multiplication endomorphisms, focusing on the family of Example 4. The result is a point-counting algorithm that, surprisingly enough, has the same asymptotic complexity as Schoof’s original elliptic point-counting algorithm. This was used to set a new point-counting record in genus 2 (which still stands twelve years later): [GKS11] reports a kilobit Jacobian order (over \( \mathbb{F}_p \) with \( p = 2^{312} + 1273 \)) computed in 80 core-days on an Intel Core 2 running at 2.83GHz. Computing the order of one Jacobian from the family over a 128-bit field took approximately 3 core-hours.

There are two key ideas here. The first is to compute the trace of Frobenius with respect to the real multiplication subring \( \mathbb{Z}[(-1 + \sqrt{5})/2] \) instead of \( \mathbb{Z} \). Thus, instead of the quartic characteristic polynomial of Frobenius, we compute a smaller quadratic polynomial—just as in Schoof’s elliptic-curve algorithm. Second, we can focus on small primes \( \ell \) that split in \( \mathbb{Z}[(-1 + \sqrt{5})/2] \). This gives a speed-up analogous to Elkies’ speed-up for elliptic curves—but crucially, unlike SEA,

- We do not need to solve modular equations to detect \( \ell \) splitting in the unknown endomorphism ring (corresponding to rational \( \ell \)-isogeny kernels).
- We do not need an analogue of Elkies’ algorithm⁴ to reconstruct \( \ell \)-isogeny kernels from moduli points: instead, we derive the kernel from the corre-

³If we fix \( p \) and consider the point-counting problem for curves over extension fields \( \mathbb{F}_{p^n} \), then there is an array of efficient point counting algorithms based on explicit \( p \)-adic cohomology, derived from Kedlaya’s algorithm [Ked04], which run in polynomial time with respect to \( n \) and the dimension (but exponential time with respect to \( \log p \)).

⁴The missing genus-2 “modular” techniques have only recently been developed by Jean Kief-
sponding ideal above \( \ell \) in \( \mathbb{Z}[\phi_\mathcal{R}_t] \cong \mathbb{Z}[(-1 + \sqrt{5})/2] \), using the equation of \( \mathcal{R}_t \) to compute the ideal of relations on the coefficients of the Mumford representation of kernel elements.

The first appearances of modular polynomials in genus 2 in [Mil15; Mar18] inspired [Bal+17], which gives Atkin-style results for curves with real multiplication by \( \mathbb{Z}[\sqrt{-5}] \) (including the family of Example 4). Atkin’s contributions to the SEA algorithm deduce information on the trace of Frobenius from factorizations of modular polynomials. Unlike Elkies, this does not change the fundamental complexity of Schoof’s algorithm, but it can make an important difference in practice. Our results could speed up computations like those of [GKS11] in the future, but they have not yet been put into practice: in any case, the complexity of the associated modular systems is so high even for small \( \ell \) that it is unclear if or when these theoretical improvements might make a useful practical difference.

3.5 Isogenies and the DLP in genus 3

The genus-3 isogeny attack of [Smi08; Smi09] was a totally new approach to the DLP. It translates a technical and abstract result from algebraic geometry (Recillas’ trigonal construction) into explicit formulas and algorithms for curves over finite fields, suitable for use by cryptographers. The impact of [Smi08] and [Smi09] is clear: these works ended all serious cryptographic research in genus 3.

Let \( \mathcal{H}/\mathbb{F}_q \) be a hyperelliptic curve of genus 3; we want to solve discrete logarithms in \( \text{Jac}(\mathcal{H})(\mathbb{F}_q) \). The state-of-the-art solution was the index calculus of Gaudry, Thomé, Thériault, and Diem [Gau+07] running in time \( \widetilde{O}(q^{4/3}) \). We want to construct an explicit isogeny \( \phi : \text{Jac}(\mathcal{H}) \to \text{Jac}(\mathcal{X}) \) over \( \mathbb{F}_q \), where \( \mathcal{X} \) is a non-hyperelliptic genus-3 curve—that is, a smooth plane quartic—and use \( \phi \) to map discrete log instances from \( \text{Jac}(\mathcal{H})(\mathbb{F}_q) \) into \( \text{Jac}(\mathcal{X})(\mathbb{F}_q) \), where they can be solved in time \( \widetilde{O}(q) \) using Diem’s index calculus for low-degree plane curves [Die06].

The key theoretical ingredient is Recillas’ theorem [Rec74]: if \( g_1 : \mathcal{E} \to \mathbb{P}^1 \) is a tritrigonal map—that is, a map of degree 3—and \( \pi : \mathcal{E} \to \mathcal{C} \) is an unramified double cover, then there is a curve \( \mathcal{X} \) of genus \( g(\mathcal{C}) - 1 \) such that \( \text{Jac}(\mathcal{X}) \) is isomorphic to the Prym variety \( \text{Prym}(\mathcal{E}/\mathcal{C}) \). This \( \mathcal{X} \) comes equipped with a tetragonal (degree-4) map \( f : \mathcal{X} \to \mathbb{P}^1 \), and conversely if we start from a tetragonal map \( f : \mathcal{X} \to \mathbb{P}^1 \), we can recover an unramified double cover and tritrigonal map \( \mathcal{E} \to \mathcal{C} \to \mathbb{P}^1 \) such that \( \text{Prym}(\mathcal{E}/\mathcal{C}) \cong \text{Jac}(\mathcal{X}) \). The isomorphism is induced by a (3, 2)-correspondence \( \mathcal{R} \subset \mathcal{E} \times \mathcal{X} \), which is one of the two components of \( \mathcal{E} \times \mathbb{P}^1 \mathcal{X} \) (the other component induces the negative).

Donagi and Livné [DL99] show that this also applies when \( \mathcal{C} \) is singular, provided the double cover \( \mathcal{E} \to \mathcal{C} \) has certain properties over the singularities—and then, if we apply Recillas’ theorem to the normalization \( \mathcal{N} \) of \( \mathcal{E}/\mathcal{C} \), then the normalization maps induce an isogeny \( \text{Jac}(\mathcal{X}) \cong \text{Prym}(\mathcal{N}/\mathcal{N}) \to \text{Prym}(\mathcal{E}/\mathcal{C}) \). The interesting case for us is when \( \mathcal{N} \) is a hyperelliptic genus-3 curve \( \mathcal{H} \), and \( \mathcal{N} \cong \mathbb{P}^1 \) is its \( x \)-line; this happens when \( v : \mathcal{H} \to \mathcal{E} \) is a singularization of \( \mathcal{H} \) identifying disjoint pairs of Weierstrass points (and \( v : \mathbb{P}^1 \to \mathcal{C} \) identifies their images).
CHAPTER 3. CLASSICAL HYPERELLPTIC CRYPTOGRAPHY

Figure 3.1: Recillas’ trigonal construction in the genus-3 hyperelliptic case. The correspondence $\mathcal{R}$ induces a $(2,2,2)$-isogeny $\text{Jac}(\mathcal{H}) \to \text{Jac}(\mathcal{X})$, which is the $(2,2,2)$-isogeny $\text{Jac}(\mathcal{H}) = \text{Prym}(\mathcal{H}/\mathbb{P}^1) \to \text{Prym}(\mathcal{E}/\mathcal{E})$ induced by $\tilde{v}$ composed with Recillas’ isomorphism $\text{Prym}(\mathcal{E}/\mathcal{E}) \cong \text{Jac}(\mathcal{X})$.

Since the hyperelliptic cover $\pi : \mathcal{H} \to \mathbb{P}^1$ (i.e., the $x$-coordinate) ramifies precisely at the Weierstrass points, the induced cover $\tilde{\pi}$ is “of Beauville type” and amenable to the Donagi–Livné generalization of Recillas’ theorem, as in Figure 3.1.

Now, let us apply this to construct a $(2,2,2)$-isogeny in genus 3, as in [Smi08]. We begin with a genus-3 hyperelliptic curve $\mathcal{H}/\mathbb{F}_q$. We want to construct an explicit isogeny from $\text{Jac}(\mathcal{H})$ to the Jacobian of a non-hyperelliptic genus-3 curve $\mathcal{X}$: that is, a smooth plane quartic curve.

Choose a maximal 2-Weil-isotropic subgroup $S \subset \text{Jac}(\mathcal{H})[2]$ generated by differences of Weierstrass points: $S = \langle [W_i - W'_i] \rangle_{i=1}^{4}$. (Not every maximal 2-Weil isotropic subgroup is generated in this way, but such subgroups will be enough for our application.) In [Smi08] we compute the composition $g = g_1 \circ \nu$ directly via linear algebra after solving a single quadratic equation, before writing down explicit equations for the codomain curve $\mathcal{X}$ in the Mumford coordinates of $\text{Jac}(\mathcal{H})$. Equations for the correspondence $\mathcal{R}$ appear as a by-product of this representation of $\mathcal{X}$. We can easily compute the smooth plane quartic model of $\mathcal{X}$ required for Diem’s index calculus using explicit Riemann–Roch [Hes02].

For this to be useful in reducing the complexity of DLP computations, we need all of the above to be done over $\mathbb{F}_q$. In fact, all of this can be done over the field of definition of $S$, except for two critical square roots: one to separate the two candidates for $g$, and one to separate the two components of $\mathcal{H} \times_{g_1} \mathcal{X}$ (one of which is $\mathcal{R}$). Heuristically, we expect the first square root to be in $\mathbb{F}_q$ with probability 1/2, and if so, then the second should also be in $\mathbb{F}_q$ with (independent) probability 1/2; our experiments support this heuristic.

There are 105 choices for $S$ over $\mathbb{F}_q$, but we need an $S$ defined over $\mathbb{F}_q$; the number available depends on the factorization of the hyperelliptic polynomial of $\mathcal{H}$, but there is always at least one defined over $\mathbb{F}_q$. For each candidate $S$, the attack succeeds with probability 1/4. Summing over the possible factorization types and their probabilities, we find that the success probability of this attack on a uniformly randomly sampled hyperelliptic genus-3 curve $\mathcal{H}/\mathbb{F}_q$ is $\approx 0.1857$.

This attack is only a proof of concept; it can be strengthened by using $(2,2,2)$-isogenies whose kernels are not generated by differences of Weierstrass points, or by $(\ell, \ell, \ell)$-isogenies where $\ell$ is an odd prime, as in [LR12] and [Tia21].
As we remarked above, point-counting on hyperelliptic Jacobians of fixed genus over prime fields is polynomial-time in theory (thanks to Pila [Pil90]), but extremely hard in practice. We have already seen that for 256-bit group orders, elliptic curves take around twenty CPU-seconds with SEA, while genus-2 Jacobians take 1000 CPU-hours—five orders of magnitude greater! I am not aware of any serious attempt at large-characteristic point counting in genus 3 to date.

The concrete difficulty of point counting in genus 3 and beyond is not just a practical challenge: in [DGS22], we argue that for generic genus-3 curves over prime fields, point-counting can be made cryptographically infeasible. (While the idea that problems with polynomial-time solutions can be used as the basis for secure cryptosystems is obviously controversial, it goes back to the dawn of public-key cryptography [Mer78].) If this hypothesis is admitted, then we get an interesting scenario for so-called trustless unknown-order groups.

Many advanced cryptographic protocols run in groups of order unknown to the participants. The simplest examples of unknown-order groups are RSA groups $(\mathbb{Z}/N\mathbb{Z})^*$ (with $N = pq$ where $p$ and $q$ are large primes), but establishing such a group requires a trusted third party to generate the modulus $N$ (and to erase, or at least promise to not disclose, the prime factors $p$ and $q$, from which the order of $(\mathbb{Z}/N\mathbb{Z})^*$ can be derived). In the trustless setting, where no such trusted third party exists, we generally turn to ideal class groups of quadratic imaginary fields: if the participants choose a sufficiently large prime $p$, then the order of $\text{Cl}(-p)$ is not known; the best known algorithms for computing the order are subexponential in $\log p$, so $p$ can therefore be chosen in public without compromising security.

In [DGS22] we re-evaluate parameter sizes for secure trustless unknown-order groups in the light of Sutherland’s generic order-finding algorithm [Sut07], and argue that random hyperelliptic genus-3 Jacobians might provide a competitive alternative (in terms of key sizes; it was hard to compare speeds given the lack of competitive implementations of genus-3 hyperelliptic curve arithmetic). Independent of the genus-3 proposal, [DGS22] contains a very useful result for class-group based cryptography: inspired by Bleichenbacher’s Rabin signature compression algorithm, I introduced a simple method to efficiently compress class group elements to $3/4$ of their previous encoded size. This improvement is particularly welcome given our new, higher estimates of parameter sizes required for secure class groups.
4 Isogeny-based cryptography

Isogeny-based cryptography (IBC) is one of the youngest areas of post-quantum cryptography. The best-known isogeny-based protocol is Jao and De Feo’s key exchange [JD11; DJP14], named SIDH in [CLN16]. The key-encapsulation mechanism SIKE [Aza+17], based on SIDH, passed to the fourth round of the NIST Postquantum Standardization Project [NIST] before being spectacularly broken, along with SIDH, by Castryck and Decru [CD23], Maino, Martindale, Panny, Pope, and Wesolowski [Mai+23], and Robert [Rob23]. SIDH was itself a reaction to earlier key-exchange constructions by Couveignes [Cou06] and Rostovtsev and Stolbunov [RS06; Sto09; Sto10], which still resist these new attacks. These systems were widely considered impractical, but their more recent descendents have been more successful, and we focus on those in this chapter.

We begin with a quick tour of the theory in §4.1, before summarizing some theoretical and early practical contributions in §4.2. Going into a little more depth, we discuss the Vélusqrt algorithm in §4.4, supersingularity testing in §4.3, CSIDH implementations in §4.5, and hyperelliptic IBC in §4.6.

4.1 Theoretical foundations

An action of a commutative group \( G \) (written multiplicatively) on a set \( X \) is a mapping \( * : G \times X \to X \) such that

\[
1 \ast x = x \quad \text{and} \quad g \ast (h \ast x) = (gh) \ast x \quad \text{for all } g, h \in G \text{ and } x \in X.
\]

Each choice of \( x \) in \( X \) defines a map \( \varphi_x : g \mapsto g \ast x \) of \( G \) into \( X \). Any commutative group action \( G \times X \to X \) can be used to construct a simple key exchange analogous to Diffie–Hellman, but based on the hardness of inverting \( \varphi_x \). Alice and Bob have keypairs \((g_A, x_A = g_A \ast x_0)\) and \((g_B, x_B = g_B \ast x_0)\), respectively (where \( x_0 \) is some fixed base in \( X \)), and they derive a shared secret

\[
g_A \ast x_B = g_A \ast (g_B \ast x_0) = g_B \ast (g_A \ast x_0) = g_B \ast x_A.
\]

For example, if \( X = \langle x \rangle \) is cyclic of order \( p \) and \( G = (\mathbb{Z}/p\mathbb{Z})^* \) acts on \( X \setminus \{1\} \) by \( g \ast x = x^g \), then inverting \( \varphi_x \) is the discrete logarithm problem (DLP) in \( X \). But inverting \( \varphi_x \) for other actions may not be related to any DLP, and might resist attacks based on Shor’s quantum algorithm. Indeed, inverting \( \varphi_x \) can be recognised as the abelian hidden shift problem, for which the best quantum attacks are subexponential algorithms due to Kuperberg [Kup05; Kup13] and Regev [Reg04].

Couveignes [Cou06] and—independently—Rostovtsev and Stolbunov [RS06; Sto09; Sto10] proposed using a classic group action based on the theory of Complex Multiplication [Cox13]. Let \( X \) be the set of elliptic curves \( \mathcal{E}/\mathbb{F}_p \) with \( \text{End}(\mathcal{E}) \cong \mathbb{Z}/n\mathbb{Z} \)
CHAPTER 4. ISOGENY-BASED CRYPTOGRAPHY

𝒪 for some quadratic order 𝒪. If 𝒔 is an ideal in 𝒪, then the finite subgroup 𝒔[𝒪] := ∩ₜ∈𝒪kerφ is the kernel of an isogeny 𝒔 → 𝒔 := 𝒔[𝒪]. Principal ideals correspond to endomorphisms, so this action induces an action of the ideal class group of 𝒪 on 𝒳. In a sense, the resulting group-action Diffie–Hellman obfuscates Buchmann and Williams’ classical class-group Diffie–Hellman [BW88; BW89].

Rostovtsev and Stolbunov advertised their cryptosystems as potential post-quantum candidates; Childs, Jao, and Soukharev led the quantum analysis using Kuperberg’s algorithm with subexponential quantum isogeny evaluation [CJS14]. (See [Ber+19] for further analysis.) Isogeny-based group-actions provide the only known post-quantum non-interactive key exchange (NIKE), most notably with CSIDH [Cas+18]. This group action is also a fruitful source of new schemes, including CSI-FiSh [BKV19] and its many derivatives.

4.2 From theory to practice

The survey article [Smi18] investigates the theoretical foundations of commutative IBC, focusing on the obvious analogies (and the limits of the analogies) between classical and group-action-based Diffie–Hellman. For example, the classical Pollard and Baby-step giant-step attacks on the DLP in generic groups apply almost without modification to the group-action analogue of the DLP; however, Pohlig–Hellman does not extend, and neither does Shor’s quantum attack.

The equivalence of the DLP and CDHP in the group setting is a long-standing open problem. We prove a remarkably simple quantum polynomial-time equivalence for their group-action analogues in [Gal+21] (assuming an error-free, idealistic version of Shor’s algorithm; this result has been extended to more “realistic” quantum scenarios by Montgomery and Zhandry [MZ22]).

Couveignes’ original key-exchange scheme was never implemented. Rostovtsev implemented his scheme only at very low security levels, and when we reimplemented it in [DKS18] at the 128-bit security level we found that each action took about 2000 seconds, which is clearly impractical.

Working with with Luca De Feo and Jean Kieffer, in [DKS18] we revisited Stolbunov’s scheme in the hope of improving its efficiency. Our main contribution was to show that great improvements could be made for primes ℓ where the isogeny kernels are generated by rational points: in this case, computations with modular polynomials can be replaced with a simple application of Vélu’s formulae. The problem, then, is to construct elliptic curves with orders divisible by many small primes. It is tempting to apply the CM method, but this can only produce curves whose endomorphism rings have very small class groups (which means easily brute-forceable private keyspaces here). Instead, we were reduced to exhaustive search of ordinary curves, looking for suitable group orders with point counting. Such weak curves are extremely rare, and even after many CPU-months, the best curve we could find only improved over Stolbunov by a factor of 4.

However, this work directly inspired the authors of CSIDH (Commutative Supersingular Isogeny Diffie–Hellman) [Cas+18] to switch from ordinary to supersingular curves over ℱₚ, where the curve order can be made as smooth as desired by a careful choice of the field prime 𝒫. The resulting action of the class group of ℤ[√₋ℙ] on the set of supersingular curves over ℱₚ has the Vélu trick of [DKS18] enabled for every ℓ, reducing total computation time to milliseconds. CSIDH has since become the reference group action for post-quantum cryptography.
CSIDH and its derivatives work in the $\mathbb{F}_p$-subgraph of the full supersingular graph: that is, the subgraph supported on vertices defined over $\mathbb{F}_p$ (or with $j$-invariants in $\mathbb{F}_p$). In [CS22] we investigate a family of generalizations of the $\mathbb{F}_p$-subgraph, one for each squarefree integer $d$. The key is to recognise that a curve $\mathcal{E}/\mathbb{F}_p$ has its $j$-invariant in $\mathbb{F}_p$ precisely when $\mathcal{E}$ is isomorphic to its Galois conjugate. In [CS22] we relax the isomorphisms to $d$-isogenies to form the graph whose vertices are curves over $\mathbb{F}_p$ with a $d$-isogeny to their conjugate and whose edges are $\ell$-isogenies commuting with the $d$-isogenies (all up to the obvious notion of isomorphism). The reader may recognise that the curves we saw in §2.4 (in the context of scalar decomposition) are a perfect source of examples. Narrowing our focus to supersingular curves, we use the theory of orientations to set up an action of the ideal class group of $\mathbb{Z}[\sqrt{-dp}]$. We thus have a natural setting for generalized variants of CSIDH (which is the case $d = 1$). We give arguments for the security of such cryptosystems, and define a broad generalization of the Delfs–Galbraith isogeny-finding algorithm [DG16] reflecting ideas from [CLG09, §7], [Eis+20], and [Arp+21].

### 4.3 Supersingularity

Many historic isogeny-based cryptosystems including the Charles–Goren–Lauter hash function [CLG09] and SIDH were based on the difficulty of finding paths through supersingular isogeny graphs. Now, the emergence of CSIDH and its derivatives has maintained the emphasis on supersingular curves in IBC. Efficiently identifying and generating supersingular curves is therefore critical.

Many public-key cryptographic protocols require incoming public keys to be validated to mitigate some adaptive attacks. A CSIDH public key is deemed valid if the given Montgomery coefficient specifies a supersingular elliptic curve over the prime field. In [BGS22], we survey the current supersingularity tests used for CSIDH key validation, and implement and measure two new alternative algorithms. Our implementation shows that we can determine supersingularity substantially faster, and using less memory, than the state-of-the-art.

**Table 4.1: Supersingularity test comparison over $\mathbb{F}_p$.** Here, $n$ denotes the number of primes $\ell | (p + 1)$. Experiments ran on 500 random valid (supersingular) and 500 random invalid (ordinary) curves for CSIDH-512 on an Intel i7-10610U processor running at 4.90 GHz with TurboBoost and SpeedStep disabled, compiled with gcc 12.1.0. Our C code used the $\mathbb{F}_p$-arithmetic library from the CTIDH package [Ban+21a]. The product tree algorithm is CTIDH’s key validation.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Asymptotic complexity</th>
<th>Supersingular input</th>
<th>Ordinary input</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>M.Cycles</td>
<td>Stack</td>
</tr>
<tr>
<td>Random point</td>
<td>$O(n \log p)$</td>
<td>63.4</td>
<td>2890</td>
</tr>
<tr>
<td>Product tree</td>
<td>$O((\log n) \log p)$</td>
<td>6.7</td>
<td>4344</td>
</tr>
<tr>
<td>Sutherland</td>
<td>$O(\log^p p)$</td>
<td>35.4</td>
<td>2696</td>
</tr>
<tr>
<td>Doliskani</td>
<td>$O(\log p)$</td>
<td>4.5</td>
<td>3280</td>
</tr>
</tbody>
</table>

Existing CSIDH implementations check supersingularity of a curve $\mathcal{E}/\mathbb{F}_p$ by determining the order of a random point in $\mathcal{E}(\mathbb{F}_p)$, using the fact that $\mathcal{E}$ is supersingular if and only if $\#\mathcal{E}(\mathbb{F}_p) = p + 1$ (this requires the prime factorization of $p + 1$, ...
which is known in the context of CSIDH). We can save a lot of redundant scalar multiplication, at the cost of a little more memory, by traversing a product tree for \( p + 1 \). Sutherland’s supersingularity test \([Sut12]\) operates on curves over \( \mathbb{F}_{p^2} \). This test is based on distinguishing between the 2-isogeny graph structures of supersingular and ordinary elliptic curves over \( \mathbb{F}_{p^2} \). The asymptotic complexity of this algorithm is in \( O(\log^2 p) \), but we find that it performs surprisingly well in practice; our adaptation to CSIDH is extremely easy to implement, and requires very little memory. It is also the only one of the four tests considered here that always returns a proven positive or negative result. We also developed a new variant of Doliskani’s test \([Dol18]\), which uses Polynomial Identity Testing to distinguish between the \( p \)-th division polynomials of supersingular and ordinary curves over \( \mathbb{F}_{p^2} \). We adapted and optimized this algorithm for Montgomery models over \( \mathbb{F}_{p^2} \), replacing division polynomial evaluation with a single scalar multiplication over \( \mathbb{F}_{p^2} \) followed by an easy field exponentiation.

Table 4.1 presents our experimental results. We see that Sutherland’s algorithm is easily the fastest when the input is ordinary/invalid, but it lags well behind the product-tree algorithm on supersingular/valid input. Our new version of Doliskani’s test is the fastest on valid input, and only twice as slow as the product-tree algorithm on invalid input.

While distinguishing supersingular curves from ordinary curves is easy, sampling random supersingular curves with unknown endomorphism ring\(^1\) is curiously hard. In \([Boo+22]\), we focus on the special problem of constructing any supersingular curve with unknown endomorphism ring. We try a variety of approaches: “numerically” iterating to a root of the supersingular polynomial, roots of semi-evaluated modular polynomials (where the curves of §2.4 make yet another appearance), an “inverse Schoof” algorithm, the geometry of superspecial genus-2 Jacobians and Kummer surfaces, and quantum walks on isogeny graphs. While these results have some methodological interest, ultimately this project was an epic failure: none of these approaches made the slightest impact on what increasingly appears to be a fundamentally hard problem.

### 4.4 The Velusqrt algorithm

Daniel J. Bernstein, Luca De Feo, Antonin Leroux, and I defined the Vélusqrt algorithm in \([Ber+20]\), radically improving isogeny evaluation complexity from \( O(\ell) \) to \( \tilde{O}(\sqrt{\ell}) \) operations. This is the first genuine advance in elliptic isogeny evaluation in the half-century since Vélu’s algorithm appeared. Vélusqrt improves the efficiency of all IBC systems that use larger-degree isogenies, most notably CSIDH.

The central idea is that Vélu’s formulae suggest a linear sweep of the points in the kernel, and we can improve on this using a space-time tradeoff inspired by Shanks’ baby-step giant-step algorithm.

Let \( \mathcal{E}/\mathbb{F}_q \) be an elliptic curve with a cyclic subgroup \( \mathcal{G} = \langle P \rangle \) of odd order \( \ell \); for convenience, we write \( x_s \) for \( x([s]P) \). We want to compute the quotient isogeny \( \phi : \mathcal{E} \to \mathcal{E}/\mathcal{G} \). The key task is to evaluate the kernel polynomial

\[
\Psi_{\mathcal{G}}(X) = h_S(X) := \prod_{s \in S} (X - x_s) \quad \text{where} \quad S = \{1, 3, \ldots, \ell - 4, \ell - 2\}.
\]

---

\(^1\)Unknown endomorphism rings are critical to the security of some isogeny-based cryptosystems: the KLPT algorithm \([Koh+14]\) uses the endomorphism ring to efficiently solve many supersingular isogeny problems.
The classic approach to evaluating \( h_S(X) \) is to apply Horner's rule, computing the sequence of \( x_i \) as \( x_1 = x(P), x_2 = xDBL(x_1), \) and then \( x_{i+1} = xADDD(x_i, x_j, x_{i-1}) \) for \( i \in \{2, \ldots, (\ell - 3)/2\} \) for a total of \( (\ell - 3)/2 \) curve operations and hence \( O(\ell) \mathbb{F}_q \)-operations. (We explore a new, faster approach using mostly \( xDBL \)s in \cite{Ban+23b}, along with techniques exploiting the action of Galois when \( P \) is defined over an extension field, but these methods seem basically incompatible with Vélusqrt.)

But now, observe that if we define \( b = \lfloor \sqrt{\ell - 1}/2 \rfloor \) and \( b' = \lfloor (\ell - 1)/(4b) \rfloor \), and then \( I = \{2b(2i + 1) \mid 0 \leq i < b'\} \), \( J = \{2j + 1 \mid 0 \leq j < b\} \), \( K = \{4bb' + 1, \ldots, \ell - 4, \ell - 2\} \), then \#I, \#J, and \#K are in \( O(\sqrt{\ell}) \), and 

\[
S = (I + J) \cup (I - J) \cup K.
\]

Writing \( I \pm J \) for \((I + J) \cup (I - J)\), we have a corresponding factorization:

\[
h_S(X) = h_{I \pm J}(X) h_K(X)
\]

where \( h_{I \pm J}(X) := \prod_{s \in I \pm J} (X - x_s) \) and \( h_K(X) := \prod_{s \in K} (X - x_s) \). While \( h_{I \pm J} \) has \( O(\ell) \) terms, we can evaluate it (and hence \( h_S \)) using only \( O(\sqrt{\ell}) \mathbb{F}_q \)-operations. The trick is to use the bijections \( I \times J \rightarrow I + J \) and \( I \times J \rightarrow I - J \) to express \( h_{I \pm J} \) as a multivariate resultant of polynomials corresponding to \( I \) and \( J \) with degrees in \( O(\sqrt{\ell}) \). Concretely, let

\[
E_s(X, Z) := F_0(Z, x_s)X^2 + F_1(Z, x_s)X + F_2(Z, x_s)
\]

for \( s \in \mathbb{Z} \),

where \( F_0(X, Z), F_1(X, Z), \) and \( F_2(X, Z) \) are the biquadratic forms such that

\[
(X - x(P \oplus Q))(X - x(P \ominus Q)) = X^2 + \frac{F_1(x(P), x(Q))}{F_0(x(P), x(Q))} X + \frac{F_2(x(P), x(Q))}{F_0(x(P), x(Q))}
\]

for all \( P \) and \( Q \) on \( \mathcal{E} \) such that \( 0 \notin \{P, Q, P \oplus Q, P \ominus Q\} \). Observe that if \( s \equiv i \pm j \pmod{\ell} \) with \( i \equiv \pm j \pmod{\ell} \), then \( x_s \) is a root of \( E_s(X, x_s) \). Hence

\[
h_{I \pm J}(X) = \Delta_{I \pm J}^{-1} \text{Res}_Z(h_I(Z), E_j(X, Z)) \quad \text{where} \quad E_j(X, Z) := \prod_{j \in J} E_j(X, Z),
\]

which we can use to evaluate \( h_{I \pm J}(X) \) with multipoint multievaluation in only \( \tilde{O}(\sqrt{\ell}) \mathbb{F}_q \)-operations once we have computed

\[
\Delta_{I \pm J} := \text{Res}_Z(h_I(Z), D_j(Z)) \quad \text{where} \quad D_j(Z) := \prod_{j \in J} F_0(Z, x_j).
\]

Our proof-of-concept Vélusqrt implementation already beats the classic algorithm for \( \ell \geq 113 \); see \cite[Appendix A]{Ber+20} for more concrete analysis and crossover points with optimized implementations. Vélusqrt is already useful for the larger primes used in the smallest CSIDH parameter set, and will be even more useful for larger parameters for higher security levels).
4.5 Constant-time CSIDH and CTIDH

Making group-action key exchanges practical is one thing; making implementations practical and secure is another. Some first steps towards efficient, constant-time² class-group actions were made by Meyer, Campos, and Reith [MCR19] and Onuki, Aikawa, Yamazaki, and Takagi [Onu+20]. At first, it looks like the actions by all of the small primes need to be made indistinguishable. For the CSIDH-512 parameter set, this would mean making 3-isogenies look like 587-isogenies, which is obviously wasteful. Instead, one of the key techniques common to all constant-time CSIDH algorithms is grouping the primes \( \ell \) into similar-sized batches; the actions by primes in the same batch are then made indistinguishable, typically by filling in smaller \( \ell \)-isogeny computations with dummy operations. We give some faster constant-time CSIDH actions in [Cer+19], including some without dummy operations, to provide some protection against simple fault attacks.

The fastest and most sophisticated constant-time CSIDH implementation is CTIDH [Ban+21a], which is actually faster than the original (unprotected) CSIDH algorithm. The key idea is a fundamental reimagining of the private key space. Zooming in on a single batch of primes, of norms \( \ell_1 < \cdots < \ell_n \), say, most constant-time versions of CSIDH would set an exponent bound \( B \) for the batch, then sample private keys from \( [-B, B]^n \) for this batch. There are \((2B + 1)^n\) keys, and computing the action of any one of them costs the same as \( nB \ell_n \)-isogenies. In CTIDH, we instead sample exponent vectors \((e_1, \ldots, e_n)\) of bounded 1-norm from \( \mathbb{Z}^n \); that is, \( \sum_{i=1}^n |e_i| \leq N \) for some \( N \). The keyspace has size \( \Phi(n, N) := \sum_{\ell \in [B]} (\ell^n)^n \), and computing the action of any one key costs \( N \ell_n \)-isogenies. We can generally make \( \Phi(n, N) \geq (2B + 1)^n \) with \( N \ll nB \), which means we can save a substantial amount of work while maintaining a large keyspace (and hence a high security level).

For example, on a batch of \( n = 4 \) primes, taking \( B = 2 \) in the classic approach yields \( 5^4 = 625 \) keys, each costing \( 4 \times 2 = 8 \) isogenies to apply. But with the new approach, choosing arbitrary exponent vectors of 1-norm 5 yields \( \Phi(4, 5) = 681 \) keys (a slight increase), each costing only 5 isogenies to apply—a 27% saving.

4.6 The genus-2 superspecial graph

As in Chapter 3, the natural question is: why stop at elliptic curves? In [CDS20], Wouter Castryck, Thomas Decru, and I defined a genus-2 analogue of the Charles–Goren–Lauter (CGL) hash function [CLG09] in the \((2, 2)\)-isogeny graph of superspecial principally polarized abelian surfaces (PPASes), which are isomorphic as unpolarized abelian varieties to products of supersingular elliptic curves.

The CGL hash uses its binary input string to drive a non-backtracking walk through the elliptic supersingular \( 2\)-isogeny graph, with the hash value being derived from the \( j \)-invariant of the end-point of the walk. Computing preimages amounts to solving the isogeny problem for this graph: given supersingular curves \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \), efficiently compute a \( 2 \)-power isogeny \( \mathcal{E}_1 \to \mathcal{E}_2 \).

Our superspecial hash function takes the input three bits at a time, using each three-bit chunk to choose the next step in a restricted walk in the \((2, 2)\)-isogeny graph of superspecial PPASes. The graph navigation is based on the modern treat-

²All of these algorithms are fundamentally probabilistic so we need a more subtle definition of "constant-time": the distribution of the run-time behaviours must be independent of any secret input.
Table 4.2: Constant-time CSIDH-512 and CSIDH-1024 implementation performance with 256-bit private keyspaces. keygen is key generation (one CSIDH action), DH is computing a shared secret (key validation and one CSIDH action). M, S, and a respectively denote multiplications, squarings, and additions in $\mathbb{F}_p$. Average number of millions of cycles on an Intel Xeon E3-1220 v5 (Skylake) CPU at 3GHz (except [Hut+20], ran on an Intel Core i7-7500k) with Turbo Boost disabled. For CTIDH [Ban+21a], the average was over 65 actions for each of $2^{14} - 1$ private keys. Cycle counts omitted for [Onu+19] (unknown Turbo Boost) and [ACR20] (Python implementation). See [Ban+21a, §8] for more detailed measurements.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Operation</th>
<th>MCycles</th>
<th>M</th>
<th>S</th>
<th>a</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSIDH-512</td>
<td>keygen</td>
<td>125.53</td>
<td>321207</td>
<td>116798</td>
<td>482311</td>
<td>[Ban+21a]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>keygen</td>
<td>—</td>
<td>624000</td>
<td>165000</td>
<td>893000</td>
<td>[ACR20]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>DH</td>
<td>129.64</td>
<td>330966</td>
<td>121787</td>
<td>497476</td>
<td>[Ban+21a]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>DH</td>
<td>218.42</td>
<td>665876</td>
<td>189377</td>
<td>691231</td>
<td>[CR22]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>DH</td>
<td>238.51</td>
<td>632444</td>
<td>209310</td>
<td>704576</td>
<td>[Hut+20]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>DH</td>
<td>239.00</td>
<td>657000</td>
<td>210000</td>
<td>691000</td>
<td>[Cer+19]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>DH</td>
<td>—</td>
<td>732966</td>
<td>243838</td>
<td>680801</td>
<td>[Onu+19]</td>
</tr>
<tr>
<td>CSIDH-512</td>
<td>DH</td>
<td>395.00</td>
<td>1054000</td>
<td>410000</td>
<td>1053000</td>
<td>[MCR19]</td>
</tr>
<tr>
<td>CSIDH-1024</td>
<td>keygen</td>
<td>469.52</td>
<td>287739</td>
<td>87944</td>
<td>486764</td>
<td>[Ban+21a]</td>
</tr>
<tr>
<td>CSIDH-1024</td>
<td>keygen</td>
<td>—</td>
<td>552000</td>
<td>133000</td>
<td>924000</td>
<td>[ACR20]</td>
</tr>
<tr>
<td>CSIDH-1024</td>
<td>DH</td>
<td>511.19</td>
<td>310154</td>
<td>99371</td>
<td>521400</td>
<td>[Ban+21a]</td>
</tr>
</tbody>
</table>

The real value of [CDS20] is in its crystallisation of problems on the superspecial graph for future research. The fundamental problem is the difficulty of solving the isogeny problem in the superspecial $(2,2)$-isogeny graph, but there are also more subtle problems on the expansion properties of the graph and the limiting behaviour of random walks, which Enric Florit and I addressed in [FS21a] and [FS21b].

In [CS20], Craig Costello and I attack the isogeny problem for superspecial PPAs. Consider the superspecial PPAEs over $\mathbb{F}_{p^2}$: there are $O(p^3)$ of them, almost all Jacobians of genus-2 curves, but there are also $O(p^2)$ elliptic products. In [CDS20], the elliptic products are an inconvenience for implementation, which we essentially ignore (because the $O(1/p)$ chance that a random walk will hit one is vanishingly small). In [CS20], random walks to elliptic products become the target: we expect to find them in $O(p)$ field operations, and then we can complete the isogeny using elliptic isogenies (in only $O(\sqrt{p})$ field operations). The resulting $O(p)$ algorithm is exponentially faster than the usual generic random-walk-based algorithms, which would run in $O(p^{3/2})$ time on the same graph. In higher dimensions, walking to PPAEs with elliptic factors lets us recursively solve in lower dimensions with an overall complexity growing linearly in the exponent (with respect to the dimension) rather than quadratically. This result suggests that higher-dimensional PPAs are actually a surprisingly poor choice for IBC.
5 Compact public-key cryptography

We now turn to cryptographic implementations targeting constrained devices—in particular, microcontroller platforms with little processing power and memory. In §5.1 we see that hyperelliptic genus-2 curves can give interesting practical improvements for key exchange and signatures in this domain. In §5.2 we estimate the real-world cost of the transition from curve-based to post-quantum signatures on these devices. Finally, in §5.3 we highlight some implementation results for a new code-based post-quantum signature scheme.

5.1 qDSA and muKummer

Gaudry first proposed Kummer surfaces as a setting for DLP-based cryptography in [Gau07], and Gaudry and Schost [GS12] computed parameters for a fast Kummer surface Kum(ℰ) over \( \mathbb{F}_{2^{127} - 1} \) (targeting the 128-bit security level). In the sequel we write ℰ for the Gaudry–Schost curve.

Bernstein, Chuengsatiansup, Lange, and Schwabe’s work on high-speed key exchange in [Ber+14] showed that switching from elliptic 𝑥-lines to Kummer surfaces could improve efficiency on PC architectures. However, it was not at all obvious that Kummers could give faster key exchange on microcontroller architectures while simultaneously reducing the RAM footprint.

Joost Renes, Peter Schwabe, Lejla Batina, and I developed \( \mu \text{Kummer} \) [Ren+16], a software package for high-speed, high-security key exchange and signatures, with a minimal memory footprint for use in low-end 8- and 32-bit microcontrollers. \( \mu \text{Kummer} \) includes Diffie–Hellman key exchange on Kum(ℰ) and Schnorr signatures [Sch89] on Jac(ℰ), but with scalar multiplications computed on Kum(ℰ). This project was based on theoretical ideas from [CCS16], which defined a genus-2 analogue of elliptic 𝑦-coordinate recovery formulæ for use with the Montgomery ladder: given a point \( P \) on Jac(ℰ) and the Kummer images \( x([m]P) \) and \( x([m + 1]P) \) in Kum(ℰ) for some integer \( m \), we can recover the full point \([m]P \) on Jac(ℰ).

Our implementations targeted the ARM Cortex M0 (a typical low-end 32-bit architecture) and the AVR ATmega (an 8-bit architecture). \( \mu \text{Kummer} \) was the first genus-2 hyperelliptic curve implementation on either of these architectures, so we could only compare with elliptic curve-based alternatives at the same security level: notably [Zhe14], [HS13], [WUW13], and [Dül+15]. This comparison is not superficial: all use efficient 𝑥-only arithmetic, which is the exact elliptic-curve analogue of Kummer surface arithmetic. To provide full scalar multiplication in a group, [WUW13] uses the 𝑦-coordinate recovery of [BJ02]; again, this is the elliptic-curve analogue of our methods from [CCS16].

Table 5.1 shows great results for genus-2 scalar multiplication on the Cortex M0. Compared with the then-fastest Diffie–Hellman implementation [Dül+15],
we reduce clock cycles by about 27%, while roughly halving code and stack foot-
prints. For signature-compatible scalar multiplication on group elements, the
state-of-the-art is [WUW13]; $\mu$Kummer reduces the cycle count by a very impres-
sive 75%, at the cost of an increase in code size and stack.

On the AVR ATmega we reduce the cycle count for Diffie–Hellman by about
32% compared with [Dül+15], again roughly halving the code size, and reducing
the stack by about 80%. Jacobian scalar multiplication (for signatures) uses 71% fewer cycles than [WUW13], while increasing the stack by 25%.

Table 5.1: Scalar multiplication at the 128-bit security level on the ARM Cor-
tex M0 platform (STM32F051R8 MCU on the STMF0Discovery board), compiled
with clang v3.5.0, and an Arduino MEGA development board with an AVR AT-
mega2560 MCU, compiled with gcc v4.8.1. Only [WUW13] and [Ren+16] provide
full signature-compatible SM; the others provide $x$-only SM for Diffie–Hellman.
Code size for [Zhe14] includes a fixed-basepoint SM routine not used here; [HS13]
do not report code size for stand-alone SM.

<table>
<thead>
<tr>
<th>Source</th>
<th>ARM Cortex M0</th>
<th>AVR ATmega</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cycles</td>
<td>Stack</td>
</tr>
<tr>
<td>[HS13] (Curve25519)</td>
<td>3589850</td>
<td>548 B</td>
</tr>
<tr>
<td>[Dül+15] (Curve25519)</td>
<td>2633662</td>
<td>248 B</td>
</tr>
<tr>
<td>[Ren+16] (Kum($\mathcal{X}$))</td>
<td>2709401</td>
<td>968 B</td>
</tr>
</tbody>
</table>

Looking more closely at the figures for [Ren+16] in Table 5.1, we can compare
the costs of pure Kummer scalar multiplication (for Diffie–Hellman) and Jacobian
scalar multiplication (for signatures) using our point recovery technique. The cost
of using full group elements in the Jacobian is small in terms of speed, but large in
terms of stack: this is almost entirely due to the point-recovery technique, which
requires many simultaneously-live temporary variables.

This prompted Joost Renes and me to develop qDSA [RS17], a Schnorr signa-
ture scheme designed for Kummer varieties (including Kummer surfaces and el-
liptic $x$-lines) that does not require the full group structure of a Jacobian or an
elliptic curve. We achieve this by modifying the usual verification equation in the
Schnorr scheme (or, in our case, EdDSA [Ber+12]) to use relations on level-2 theta
functions instead of group operations and group element equality. As a result, we
retain all of the speed of [Ren+16] while losing the substantial memory overhead.

From a formal point of view the signature protocols of $\mu$Kummer and qDSA,
while closely related, are not the same: $\mu$Kummer is an instantiation of the clas-

<table>
<thead>
<tr>
<th>Source</th>
<th>ARM Cortex M0</th>
<th>AVR ATmega</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cycles</td>
<td>Stack</td>
</tr>
<tr>
<td>[WUW13] (NIST P-256)</td>
<td>$\approx$ 10730000</td>
<td>540 B</td>
</tr>
<tr>
<td>[Ren+16] (Jac($\mathcal{X}$))</td>
<td>2709401</td>
<td>968 B</td>
</tr>
</tbody>
</table>
On Cortex M0, genus-2 qDSA is significantly faster than elliptic qDSA (as should be expected), with a similar memory footprint. Comparing genus-2 qDSA with its predecessor $\mu Kummer$, signing and verification are slightly slower (corresponding to the additional work in an EdDSA-based scheme over a classic Schnorr scheme), but we get a serious improvement on the memory front: the stack for $\text{sign}$ resp. $\text{verify}$ is reduced by about 57% resp. 43%, while code size is reduced by about 8%. (We did not compare with other signature schemes on Cortex M0, because $\mu Kummer$ appeared to be the first.)

Comparing $\mu Kummer$ with its predecessor $\mu Kummer$, signing and verification are slightly slower (corresponding to the additional work in an EdDSA-based scheme over a classic Schnorr scheme), but we get a serious improvement on the memory front: the stack for $\text{sign}$ resp. $\text{verify}$ is reduced by about 57% resp. 43%, while code size is reduced by about 8%. (We did not compare with other signature schemes on Cortex M0, because $\mu Kummer$ appeared to be the first.)

Table 5.2: Signatures on the same ARM Cortex M0 and AVR ATmega platforms as Table 5.1. The SHAKE128 implementation for Cortex M0 forms 8 448 B of the code size in [RS17] and 6 938 B in [Ren+16]; there is room for some optimization here.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Function</th>
<th>ARM Cortex M0</th>
<th></th>
<th></th>
<th>AVR ATmega</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cycles</td>
<td>Stack</td>
<td>Code</td>
<td>Cycles</td>
<td>Stack</td>
<td>Code</td>
</tr>
<tr>
<td>Ed25519 [NLD15]</td>
<td>sign</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>19 047 706</td>
<td>1 473 B</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>verify</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>30 776 942</td>
<td>1 226 B</td>
<td>—</td>
</tr>
<tr>
<td>FourQ [Liu+20]</td>
<td>sign</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>5 174 800</td>
<td>1 572 B</td>
<td>25 354 B</td>
</tr>
<tr>
<td></td>
<td>verify</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>11 003 800</td>
<td>4 957 B</td>
<td>33 372 B</td>
</tr>
<tr>
<td></td>
<td>verify</td>
<td>6 793 695</td>
<td>788 B</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\mu Kummer/[\text{Jac}(\mathbb{F})]$ [Ren+16]</td>
<td>sign</td>
<td>2 865 351</td>
<td>1 360 B</td>
<td>19 606 B</td>
<td>10 404 033</td>
<td>926 B</td>
<td>29 242 B</td>
</tr>
<tr>
<td></td>
<td>verify</td>
<td>4 453 978</td>
<td>1 432 B</td>
<td>—</td>
<td>16 240 510</td>
<td>992 B</td>
<td>—</td>
</tr>
<tr>
<td>qDSA/Kum($\mathbb{F}$) [RS17]</td>
<td>sign</td>
<td>2 908 215</td>
<td>580 B</td>
<td>18 064 B</td>
<td>10 477 347</td>
<td>417 B</td>
<td>17 880 B</td>
</tr>
<tr>
<td></td>
<td>verify</td>
<td>5 694 414</td>
<td>808 B</td>
<td>—</td>
<td>20 423 937</td>
<td>609 B</td>
<td>—</td>
</tr>
</tbody>
</table>

We found two schemes to compare with on AVR ATmega: [NLD15], based on Ed25519, and [Liu+20], based on FourQ (which uses endomorphisms to accelerate scalar multiplication). Our Curve25519-base qDSA implementation outperforms [NLD15]: sign resp. verify are more than 26% resp. 17% faster, while using 65% resp. 47% less stack. Comparing against [Liu+20], we see a clear trade-off between speed and size: FourQ is clearly faster, but qDSA on Curve25519 requires only a fraction of the stack space.

Perhaps surprisingly, the advantages of genus 2 are more pronounced on the 8-bit ATmega platform. $\mu Kummer$ is almost twice as fast as [NLD15] for signing and verification, while significantly reducing the stack; qDSA is roughly twice as fast for signing with only 28% of the stack, and one-and-a-half times faster for verification with just under half the stack. Comparing to [Liu+20], again we see a clear trade-off between speed and size, but this time the loss of speed is less pronounced.

5.2 Post-quantum software updates for low-end IoT devices

My IoT-focused research currently forms part of the Inria Défi RIOT-FP [RFP], which aims to provide “future-proof” high-assurance pre- and post-quantum security for RIOT OS [Bac+18; RIO], a free and open-source operating system for low-end IoT devices where Linux cannot run. In this context, I have been working not only on isogeny-based systems like [Ban+21a],¹ but also on implementation aspects of signature schemes across the whole post-quantum spectrum [Ban+22].

In [Ban+22], we make a transverse evaluation of candidate postquantum signature schemes across several embedded platforms. The novelty in [Ban+22] is in

¹While CTIDH [Ban+21a] (see §4.5) was implemented on a PC, it was developed with a view towards microcontroller applications.
its methodology. Cryptographers typically optimize a single scheme in a stand-alone application targeting a single specific architecture. Instead, we evaluate several competing schemes in the context of a real-world application—secure software updates for RIOT-OS using the SUIT protocol [Mor+21b; Mor+21a; ZS20]—and, emphasising portability, across three different platforms (Cortex-M4, RISC-V, and ESP32).

We took ECC signatures as a pre-quantum baseline. For post-quantum alternatives we took Falcon [Fou+] and Dilithium [Ava+], both based on structured lattices, together with hash-based Leighton–Micali signatures (LMS) [MCF19]. Falcon and Dilithium are both NIST standardization candidates [NIST]. Being stateful, LMS does not fit the requirements of the NIST call; but it is accepted as a high-security post-quantum scheme, and statefulness is actually quite natural in the context of certifying software package updates.

Table 5.3 compares keylengths and signature sizes for each of the schemes. The post-quantum algorithms have larger public key and signature sizes, generally by well over an order of magnitude. Compared with ECC signatures, Falcon’s public keys are $28 \times$ larger and its signatures are $10.4 \times$ larger; Dilithium’s public keys are $41 \times$ larger than elliptic-curve keys, and its signatures are $38 \times$ larger. LMS avoids this spectacular growth in public key sizes, with keys only $1.875 \times$ larger than elliptic-curve public keys; but its signatures are a massive $74.3 \times$ larger than elliptic-curve signatures.

To compare running times and memory requirements we ran the schemes on popular off-the-shelf IoT hardware provided by the open-access IoT-Lab testbed [Adj+15; IoT-LAB]. We chose platforms representing the landscape of modern 32-bit microcontroller architectures, including ARM Cortex-M, Espressif ESP32, and RISC-V. Table 5.4 presents our Cortex-M4 benchmarks (see [Ban+22] for similar results on ESP32 and RISC-V platforms).

Looking at Table 5.4, we see that post-quantum signing is generally slower than ECC: the fastest post-quantum algorithm is $7.94 \times$ slower than Ed25519 (Monocypher). But the tables are turned when we compare signature verification: the fastest pre-quantum algorithm is $2.65 \times$ slower than post-quantum Falcon. Efficient verification is valuable, but here it comes at the price of an increase in memory: post-quantum flash requirements can grow to over $11 \times$ the smallest pre-quantum flash sizes, and there is also a considerable increase in stack memory.

As a concrete example: suppose we want to add quantum resistance to a RIOT firmware update with SUIT for the nRF52840dk board used for our tests. Table 5.5 estimates the relative impact on the entire SUIT software update process, including the crucial aspect of network transfer costs and the memory resources required to actually apply the firmware update on the device.
Chapter 5. Compact Public-Key Cryptography

Table 5.4: Pre- and post-quantum signature performance on a Nordic nRF52840 Development Kit: ARM Cortex-M4 MCU with 256 kB RAM and 1 MB flash. **Static Dilithium** is the implementation from [Ava+], modified to use (fixed) public keys pre-loaded in Flash.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Flash (B)</th>
<th>Sign Time (ms)</th>
<th>Sign Stack (kB)</th>
<th>Verify Time (ms)</th>
<th>Verify Stack (kB)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ed25519</td>
<td>5106</td>
<td>845</td>
<td>5411</td>
<td>1953</td>
<td>125012</td>
<td>1300</td>
</tr>
<tr>
<td>Ed25519</td>
<td>13852</td>
<td>17</td>
<td>1136</td>
<td>40</td>
<td>2599</td>
<td>1936</td>
</tr>
<tr>
<td>ECDSA</td>
<td>6498</td>
<td>294</td>
<td>18871</td>
<td>313</td>
<td>20037</td>
<td>1024</td>
</tr>
<tr>
<td>Falcon</td>
<td>57613</td>
<td>1172</td>
<td>75020</td>
<td>15</td>
<td>1004</td>
<td>4744</td>
</tr>
<tr>
<td>Dilithium</td>
<td>11664</td>
<td>465</td>
<td>29788</td>
<td>53</td>
<td>3407</td>
<td>3658</td>
</tr>
<tr>
<td>Dilithium</td>
<td>26672</td>
<td>135</td>
<td>8655</td>
<td>23</td>
<td>1510</td>
<td>19504</td>
</tr>
<tr>
<td>LMS</td>
<td>12864</td>
<td>9224</td>
<td>590354</td>
<td>123</td>
<td>7908</td>
<td>1580</td>
</tr>
</tbody>
</table>

Table 5.5: Relative costs of adding quantum-resistance to RIOT firmware updates with SUIT on the Cortex M4 platform in Table 5.4. The SUIT manifest is the metadata including the signature. “No crypto” and “With crypto” refer to updates not including and including the signature library code in the update, respectively.

<table>
<thead>
<tr>
<th>Signature in SUIT</th>
<th>Memory</th>
<th>Data Transfer</th>
<th>SUIT manifest</th>
<th>Stack</th>
<th>Stack</th>
<th>With crypto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ed25519 / SHA256</td>
<td>52.4kB</td>
<td>47kB</td>
<td>16.3kB</td>
<td>483B</td>
<td>+120%</td>
<td>+120%</td>
</tr>
<tr>
<td>Falcon / SHA3-256</td>
<td>+30%</td>
<td>+224%</td>
<td>+18%</td>
<td>+1.1%</td>
<td>+1.1%</td>
<td>+34%</td>
</tr>
<tr>
<td>Dilithium / SHA3-256</td>
<td>+34%</td>
<td>+587%</td>
<td>+210%</td>
<td>+4.3%</td>
<td>+4.3%</td>
<td>+34%</td>
</tr>
<tr>
<td>LMS / SHA3-256</td>
<td>+3%</td>
<td>+984%</td>
<td>+1.2%</td>
<td>+9%</td>
<td>+9%</td>
<td>+43%</td>
</tr>
</tbody>
</table>

If the software update is just a small module update, or a small firmware update without crypto libraries, then speed and signature size are more important than flash. In these cases, Falcon has an advantage over LMS and Dilithium.

The case of small firmware updates including crypto libraries is more complicated, with flash playing more important role. Since we must transfer the update over a low-power network, the package size has a higher impact on energy costs. It takes 30-60s to transfer 50kB on a low-power IEEE802.15.4 radio link (depending on link quality and network load), but signature verification only varies by ±2s across all the candidates. In this case, LMS presents the best tradeoff between flash size, network transfer costs, verification time, and stack size.

For larger updates, the large network transfer costs overwhelm the other costs, minimising the advantage of one post-quantum signature over another.

5.3 Wavelet: adventures in code-based cryptography

Wave [DST19] is a post-quantum signature scheme built from a novel trapdoor based on a hard problem in coding theory. It has the attractive feature of having relatively short signatures and fast verification, but at the cost of an enormous public key. This, together with the unusual feature of being based on codes over
\( \mathbb{F}_3 \), creates some interesting algorithmic challenges.

In [Ban+21b] we define Wavelet, the first full implementation of Wave, including the first constant-time Wave key-generation and signing algorithms.² Most of the features of Wavelet have been absorbed into the Wave submission [Ban+23a] to the NIST post-quantum standardization signature on-ramp [NIST-sigs].

The contributions of [Ban+21b] include truncated signatures and fast signature compression, close to the information-theoretic optimum. Wavelet offers sub-kilobyte signatures for 128-bit classical and NIST Level-1 postquantum security: these are the smallest signatures in code-based cryptography, and are competitive with lattice-based schemes. We also define a much faster verification algorithm based on a simple transformation that allows us to skip reading almost half of the public key for any given signature—and given the size of Wave public keys, reading them from memory is a major performance bottleneck.

Table 5.6 shows timings for our open C implementation based on high-speed bitsliced \( \mathbb{F}_3 \)-vector arithmetic (following [Coo13]). We benchmark Wave key generation and signing, and both classic Wave and faster Wavelet verification, on an Intel Core 64-bit platform. We implemented both verifications with AVX instructions for speed, and without AVX for portability. Improving the painfully slow constant-time signing algorithm is the subject of current research.

We also tested Wavelet verification on an ARM Cortex M4 platform. The very large public keys (over 3MB) had to be key pre-loaded into extended flash memory and accessed via QSPI. We see that Wavelet’s short signatures and fast verification make it an interesting option for post-quantum signatures on microcontrollers, at least where long-term keys can be stored in a large amount of external memory.

Table 5.6: Wavelet performance (averaged over 100 runs) on an Intel Core i5-1135G7 processor with Turbo Boost and SpeedStep disabled, compiled with g++ 12.2.0 –O3 (cycles counted with the SUPERCOP toolkit [BL]); and on a nRF52840 Development Kit with an ARM Cortex M4 MCU running at 64 MHz with 256 kB RAM, 1 MB flash, and 64 MB of external memory, compiled with GNU ARM Embedded Toolchain v11.2.0 –O2s. CT denotes constant-time algorithms. Key generation and signing were not implemented with AVX or for Cortex M4, and classic Wave verification was not implemented for Cortex M4.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Intel Core Time (ms)</th>
<th>Cycles</th>
<th>Intel Core (AVX) Time (ms)</th>
<th>Cycles</th>
<th>ARM Cortex M4 Time (ms)</th>
<th>Ticks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Generation</td>
<td>1844.1</td>
<td>447 \times 10^7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key Generation (CT)</td>
<td>50199.5</td>
<td>566 \times 10^8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sign</td>
<td>303.95</td>
<td>105 \times 10^7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sign (CT)</td>
<td>4911.19</td>
<td>92 \times 10^8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave Verification</td>
<td>1.67</td>
<td>365 \times 10^4</td>
<td>1.09 \times 10^3</td>
<td>76 \times 10^4</td>
<td>402</td>
<td>13 172</td>
</tr>
</tbody>
</table>
6 Looking forward

Elliptic curve cryptography is dead; long live elliptic curve cryptography.

Classical ECC has reached a high state of maturity both in theory and in implementation. While the rise of quantum computers and Shor’s algorithm might suggest that there is little point working on new topics in ECC, elliptic-curve research continues to be pushed forward in IBC—but also in pairing-based cryptography and (by extension) the Zero-Knowledge community, where there are few (or no) post-quantum alternatives. The Zero-Knowledge sphere is creating fascinating new problems for curve-based cryptographers, from cycles of curves \[AHG22\] to implementations with arithmetization in systems like PLONK [GWC19] as a target “architecture” rather than traditional binary hardware.

While the results in §4.6 suggest that higher-dimensional supersingular PPAVs might be an inefficient setting for isogeny-based cryptosystems, the theory of isogenies in higher dimension has turned out to be the crucial ingredient in the recent attacks on SIDH and SIKE [CD23; Mai+23; Rob23]. Dartois, Leroux, Robert, and Wesolowski [Dar+23] suggest that high-dimensional isogenies are a powerful path to optimizations for elliptic isogeny-based schemes like SQISign [De +20; De +23]. Further optimizing isogeny computations, and solving cryptographic and algorithmic problems in isogeny graphs, will be an important topic for years to come—not only in IBC, but also in computational number theory.

IBC only forms a small part of the spectrum of post-quantum cryptography, and will likely only play a small part in the transition to quantum-safe cryptography. One focus of my current research is rendering a variety of post-quantum cryptosystems—not only isogeny-based, but also code-based and lattice-based—suitable for applications in constrained environments. Even in code-based cryptosystems like Wave [Ban+23a], which makes no use of elliptic curves, my experience in computational number theory has been a valuable source of optimizations and mathematical transformations.

To continue and extend this work, I proposed an Inria Action Exploratoire in the 2022 call on verified implementation of new post-quantum cryptosystems for low-end microcontrollers. Inria has a strong tradition of software verification, which has already been applied to cryptographic software, but the focus is usually on verified implementations of standardized algorithms (or standardization candidates). The AEx CACHACA (Compact Asymmetric Cryptography with High Assurance for Constrained Applications), now based at Campus Cyber in La Défense, aims to include verification techniques from the very start of the development of new and highly experimental cryptographic algorithms. CACHACA will also exploit and showcase RIOT OS as a platform for developing and deploying state-of-the-art post-quantum asymmetric cryptosystems.
Publication list

Sixteen journal articles, nineteen articles in international refereed conference proceedings, one book chapter, and five preprints. Joint publications list authors in alphabetical order¹ (except [Ren+16] and [Ban+22]).

International journal articles


¹See https://www.ams.org/profession/leaders/culture/JointResearchandItsPublicationfinal.pdf.


Articles in reviewed international conference proceedings


**Book chapters**


**Preprints**


Ph.D. thesis

Bibliography


BIBLIOGRAPHY


