Jahresber Dtsch Math-Ver (2012) 114:57–58
DOI 10.1365/s13291-012-0043-1

PREFACE

Preface Issue 2-2012

Hans-Christoph Grunau

Distinguishing between pure and applied mathematics has, in my opinion, never been particularly helpful. The current issue of the "Jahresbericht der DMV", focussing on topics from Algebra, shows once more that such a distinction, if possible at all, would be most difficult.

"Elliptic curves are beautiful mathematical objects that again and again appear in the most surprising places" is the first sentence of the survey article on "The Magic of Elliptic Curves and Public-Key Cryptography" by Florian Heß, Manfred Lochter, Andreas Stein, and Sandra Stein. Public-key cryptography is based on "one-way functions", of which the inverse is hard to compute. The discrete logarithm problem for the abelian group of points of an elliptic curve over a finite field appears to be in general computationally hard to solve and provides an efficient example of such a one-way function. For this reason elliptic curves have become ubiquitous in modern public-key cryptography. Since one of the authors, Manfred Lochter, is working for the German Federal Office for Information Security and so very much involved in the development of cryptographic standards, explicit technical solutions and their security, the article also gives in some detail real world applications.

"Mathematicians love to count things." This is the first phrase of Michael Vaughan-Lee's survey article on "Graham Higman's PORC conjecture". It is the number of groups of order p^n which has to be counted here, and it is known that this number can be estimated and is approximately $p^{\frac{2}{27}n^3}$. Higman's PORC conjecture states that for fixed n, these numbers do not only enjoy polynomial bounds but that it should be possible to calculate them by means of a finite set of polynomials in p. "PORC" stands for **P**olynomial **O**n **R**esidue **C**lasses. It is known since 2005 that the

H.-Ch. Grunau (✉)
Institut für Analysis und Numerik, Fakultät für Mathematik, Otto-von-Guericke-Universität,
Postfach 4120, 39016 Magdeburg, Deutschland
e-mail: hans-christoph.grunau@ovgu.de

conjecture holds true for $n \leq 7$. In a very comprehensible way, Michael Vaughan-Lee gives an introduction to this topic, provides some historical background, and sketches Higman's proof of a special case of the PORC conjecture. In the second part of this survey article, however, the author summarises a recent work by Marcus du Sautoy and himself on properties of a specific family of groups, which does not yet disprove the PORC conjecture, but which does obstruct the envisaged strategy of proof, and so gives some evidence that the conjecture might indeed be false.

As usual we try to present book reviews focussing on subjects different from those of the survey articles. For the current issue this means that one finds a review of a "synopsis" of numerical methods for nonlinear elliptic differential equations as well as of a book on ergodic theory and its applications in number theory.

Jahresber Dtsch Math-Ver (2012) 114:59–88
DOI 10.1365/s13291-012-0038-y

SURVEY ARTICLE

The Magic of Elliptic Curves and Public-Key Cryptography

Florian Heß · Andreas Stein · Sandra Stein · Manfred Lochter

Received: 18 January 2012 / Published online: 6 April 2012

Abstract Elliptic curves are beautiful mathematical objects that again and again appear in the most surprising places. Their history certainly originates at least in ancient Greece, whereas the study of arithmetic properties of elliptic curves as objects in algebra, geometry, and number theory traces back to the nineteenth century. Curiously, the earliest use of the term "elliptic curve" in the literature seems to have been by James Thomson in 1727 in "A Poem sacred to the Memory of Sir Isaac Newton":

> "He, first of Men, with awful Wing pursu'd the Comet tro' the long Elliptic Curve."

In 1985, Koblitz and Miller independently proposed to use elliptic curves in cryptography which can only be described as a magnificent and practical application of elliptic curves. This paper intends to mostly present a low-brow introduction of elliptic curves and their use in real-world applications of public-key cryptography.

Keywords Cryptography · Elliptic curves · Discrete logarithm problem · Public-key cryptography · Pairing-based cryptosystem · Weil pairing · Tate-Lichtenbaum pairing · Side channel analysis

F. Heß · A. Stein (✉) · S. Stein
Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany
e-mail: andreas.stein1@uni-oldenburg.de

F. Heß
e-mail: florian.hess@uni-oldenburg.de

S. Stein
e-mail: sandra.stein@uni-oldenburg.de

M. Lochter
Bundesamt für Sicherheit in der Informationstechnik (BSI), Postfach 200363, 53133 Bonn, Germany
e-mail: manfred.lochter@bsi.bund.de

Mathematics Subject Classification 94A60 · 14H52 · 11T71 · 14G50 · 68P25 · 11Y40 · 11Y16

1 Introduction

This paper is a short survey on the use of elliptic curves in public-key cryptography. It contains merely topics from well selected areas. Mostly, our aim is to give a low-brow introduction for non-experts and a motivation for further research. We will also provide new results in pairing-based cryptography. The main relevant tools from arithmetic geometry are nicely summarized in a recent article by Frey [30]. In order to obtain a thorough understanding of the constructive and the destructive aspects of elliptic curve cryptography (ECC), tools from the following research areas are required: Number theory, algebra, geometry, cryptology, theoretical computer science, efficient implementations, measurement technology, stochastics, lattices and many others.

In their seminal article in 1976, Diffie and Hellman [24] introduced public-key cryptography through a key agreement protocol which uses arithmetic in the multiplicative group $\mathbb{F}_p^*$ of a finite field $\mathbb{F}_p$ of large prime order p. With this protocol two communication partners, usually called A and B or Alice and Bob, agree on a secret key for a symmetric encryption algorithm over an insecure channel. The security of the Diffie-Hellman scheme is related to the presumed computational difficulty of computing discrete logarithms in $\mathbb{F}_p^*$. In 1985, ElGamal [27] then invented a public-key encryption protocol and a signature scheme whose security also relies on the presumed computational intractability of the discrete logarithm problem in $\mathbb{F}_p^*$.

These concepts and their subsequent refinements can be extended to arbitrary finite groups G of order n as long as there exist efficient ways of representing group elements and computing the group law. Furthermore, one should select the group G so that the following discrete[1] logarithm problem (DLP) in G is computationally infeasible: Given two group elements $g, h \in G$, where h lies in the subgroup of G generated by g, determine an integer ℓ such that $h = g^\ell$ and $0 \leq \ell < n$.

In addition to the DLP, and depending on the cryptographic application, the following weaker computational problems are also required to be computationally infeasible: The computational Diffie-Hellman problem (CDH) is to compute $g^{\ell_1\ell_2}$ from g, h_1, h_2, where ℓ_1 and ℓ_2 are the discrete logarithms of h_1 and h_2 respectively to the base g, and the decisional Diffie-Hellman (DDH) problem is to decide whether $h = g^{\ell_1\ell_2}$ given g, h_1, h_2, h, with ℓ_1 and ℓ_2 as above.

A very popular and efficient choice of G is the group of points on an elliptic curve over a finite field as independently suggested by Koblitz [50] and Miller [52]. A thorough and comprehensive discussion of elliptic curve cryptography can be found in [12, 14, 18, 44, 73]. For the arithmetic of elliptic curves, we refer to [63, 65].

In this article we will only consider elliptic curve variants of the relevant protocols. It is very advantageous that with an elliptic curve cryptosystem the corresponding elliptic curve discrete logarithm problem (ECDLP) appears to be significantly harder

[1] Sometimes the term DLP is exclusively reserved for the discrete logarithm problem in $\mathbb{F}_p^*$.

than the DLP in conventional discrete logarithm systems if the underlying elliptic curve is properly chosen. We will discuss the parameter choices for ECC in this survey according to their realization in standards which is based on current attacks to the ECDLP and computer technology. Note that there is no mathematical proof that the ECDLP is intractable.

1.1 Domain Parameters

For the choice of the curve, we mainly follow the recommendations of the ECC Brainpool [26, 51], a group consisting of universities, industry and government with the goal of promoting elliptic curve cryptography.[2] There are slight deviations of the recommendations depending on the requirements for the applications and the intended level of security. Let E be an elliptic curve defined over the finite field $\mathbb{F}_q$ with q elements, where q is a power of a prime p. We then consider the group $G = E(\mathbb{F}_q)$ of $\mathbb{F}_q$-rational points on E plus a point P on E of order n with coefficients in $\mathbb{F}_q$ and impose the following *domain parameters* in order to prevent mathematical attacks.

- q satisfies
  - either $q = 2^{p'}$, where p' is prime.
  - or $q = p$, where p is prime.
- $q \sim 2^{224}$, 2^{256}, 2^{320}, 2^{384}, or 2^{512}, dependent on the application.
- $\#E(\mathbb{F}_q) = \lambda n$, where n is a prime and $\lambda = 1, 2, 3$, or 4.
- E is not anomalous. (See Sect. 5.3; for $q = p > 3$, this means $\#E(\mathbb{F}_q) \neq p$.)
- E fulfills the Menezes-Okamoto-Vanstone condition, i.e. $n-1$ divided by the order of p modulo n is less than 10000. That is, the ECDLP in G must not be reducible to the DLP in a multiplicative group $\mathbb{F}^*_{p^k}$ of a finite field $\mathbb{F}_{p^k}$ of p^k elements for a 'small' integer k. This includes the case that the elliptic curve must not be supersingular. For $q = p > 3$, this means $n \neq p + 1$.
- The class number of the fundamental order of the endomorphism ring of E should be at least 200. This condition counters (hypothetical) lifting attacks. It was first introduced by Spallek in her Diploma thesis, supervised by Frey. When the condition was first introduced it de-facto prohibited the CM method explained below for the construction of cryptographically strong curves. This has changed during the last years, due to improved construction methods. A method for checking the class number condition is described in [51].

The choice of the domain parameters seems arbitrary at first sight. We will attempt to justify these values in the course of this paper and refer to Sects. 5 and 6 for further explanations.

1.2 What is Cryptography?

Historically, cryptography was identified with the design and implementation of secrecy systems. In the last years cryptography has become more than that and covers

[2] There also exist various technical guidelines and standards published by the German Federal Office for Information Security (BSI) and the National Institute of Standards and Technology (NIST), an agency of the U.S. Department of Commerce, e.g. [10, 56, 57].

a broader range of topics related to information security such as confidentiality, integrity, authenticity, and non-repudiation. A cryptosystem is formally defined (see e.g. [15, 67]) as a quintuple $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, where $\mathcal{P}$ is a finite set of admissible plaintext messages, $\mathcal{C}$ is a finite set of admissible ciphertext messages, $\mathcal{K}$ is a finite set of possible keys $K \in \mathcal{K}$, i.e. the key space, $\mathcal{E} = \{E_K \mid K \in \mathcal{K}\}$ is a set of encryption functions $E_K : \mathcal{P} \to \mathcal{C}$, and $\mathcal{D} = \{D_K \mid K \in \mathcal{K}\}$ is a set of decryption functions $D_K : \mathcal{C} \to \mathcal{P}$. For each $K \in \mathcal{K}$, there exist an $E_K \in \mathcal{E}$ and a $D_K \in \mathcal{D}$ such that the following condition is satisfied:

$$D_K\big(E_K(M)\big) = M \quad \text{for all } M \in \mathcal{P}.$$

A symmetric cryptosystem is constructed in a way so that either D_K and E_K are the same or can be easily derived from each other. Exposure of either D_K or E_K will immediately reveal both and the cryptosystem with the key K will be completely insecure. Necessarily K must be secret and a prior communication of the key between the communicants needs to take place. An asymmetric cryptosystem is constructed so that for each $K \in \mathcal{K}$, it is infeasible to determine D_K given E_K.

1.3 Elliptic Curves over Finite Fields

We only mention some basic properties of elliptic curves over finite fields that are needed to understand the use of elliptic curves in cryptography. For details we refer to [12, 18, 44, 65, 73].

Let $\mathbb{F}_q$ be a finite field of q elements and let p be its prime characteristic. An *elliptic curve* E over $\mathbb{F}_q$ is a smooth projective curve over $\mathbb{F}_q$ given by a homogeneous equation $F(x, y, z) \in \mathbb{F}_q[x, y, z]$ of degree 3. Let r be any positive integer. The projective solutions $P = [x : y : z] \in \mathbb{P}^2(\mathbb{F}_{q^r})$ to this homogeneous polynomial F of degree 3 form the set of $\mathbb{F}_{q^r}$-rational points of E. By remembering that there is one point $\mathcal{O} = [0 : 1 : 0]$ on E at infinity and identifying the finite points $P = [x : y : 1]$ with affine points (x, y), one represents an elliptic curve E as an affine Weierstrass equation

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \quad (a_i \in \mathbb{F}_q). \tag{1.1}$$

The set of $\mathbb{F}_{q^r}$-rational points of E is then

$$E(\mathbb{F}_{q^r}) = \big\{(x, y) \in \mathbb{F}_{q^r} \times \mathbb{F}_{q^r} : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\big\} \cup \{\mathcal{O}\}.$$

Note that for $p > 3$ the affine equation (1.1) of an elliptic curve E can be transformed into an isomorphic short Weierstrass form

$$E : y^2 = x^3 + ax + b, \tag{1.2}$$

where $a, b \in \mathbb{F}_q$ and $\Delta = -16(4a^3 + 27b^2) \neq 0$.

According to a famous theorem of Hasse the cardinality of $E(\mathbb{F}_{q^r})$ is

$$\#E(\mathbb{F}_{q^r}) = \big(q^r + 1\big) - t \quad \text{where } |t| \leq 2q^{\frac{r}{2}}. \tag{1.3}$$

t is called the trace of E over $\mathbb{F}_{q^r}$. For large values of q^r this means in particular that $\#E(\mathbb{F}_{q^r}) \approx q^r$. If the characteristic p of $\mathbb{F}_q$ divides t then the elliptic curve E is called supersingular, otherwise ordinary.

For cryptographic applications it is fundamental that $E(\mathbb{F}_{q^r})$ forms an abelian group with neutral element $\mathcal{O}$. The group law can be interpreted geometrically by a tangent and chord method and is usually written additively. For any positive integer m we denote by $E(\mathbb{F}_{q^r})[m]$ the subgroup of $E(\mathbb{F}_{q^r})$ consisting of points whose order divides m.

The main reasons why the group $E(\mathbb{F}_{q^r})$ of $\mathbb{F}_{q^r}$-rational points of E is a wonderful choice of a finite group G for applications in public-key cryptography are the following. Firstly, points $P \in E(\mathbb{F}_{q^r}) \setminus \{\mathcal{O}\}$ are easily representable in affine coordinates and the group law of $E(\mathbb{F}_{q^r})$ is efficiently computable in the coordinates of points. Secondly, we know that $\#E(\mathbb{F}_{q^r}) \approx q^r$ for large values of q^r. This means that the group size is roughly the size of the finite field $\mathbb{F}_{q^r}$ and parameters can be selected accordingly. In addition, there exist algorithms for efficiently computing $\#E(\mathbb{F}_{q^r})$ for the sizes given in Sect. 1.1. Thirdly, the discrete logarithm problem in $E(\mathbb{F}_{q^r})$ for the domain parameters in Sect. 1.1 appears to be computationally infeasible.

1.4 Selected Applications of Elliptic Curves

In the internet age elliptic curves are almost ubiquitous. Everyone who opens a secure session using his web browser has a good chance to use an elliptic curve based protocol. The protocols SSL/TLS which are used for https sessions support elliptic curves. Also packet-based communication based on the IPv4 or IPv6 protocol supports elliptic curve based security mechanisms. Details are for example given with the IPSec standard, especially with IKEv2. In most web browsers one can verify the certificate of websites by clicking left from the web address. Note that in these applications one often only needs one-sided authentication. We want to be sure to communicate with the right web server, but want to stay anonymous. For efficiency reasons web servers prefer to use the ephemeral-static version of the Diffie-Hellman key agreement (see Sect. 2.2).

The security of the new German identity documents relies on elliptic curve cryptography as well. One of the core elements is the elliptic curve based protocol for password authenticated connection establishment (PACE). We refer to [3, 11] for details.

The PACE protocol is an advanced security mechanism for MRTDs (machine readable travel documents) and the respective reader terminals. It is also a framework for authenticated key exchange between the MRTD chip (of user A) and the terminal (user B). The purpose of PACE is to establish a secure channel based on shared passwords with low entropy, where the domain parameters of the MRTD chip are authenticated by a governmental authority. The PACE protocol is secure in a certain real-or-random sense. It is currently under standardization of ISO/IEC. One interesting fact is that the PACE protocol computes an intermediate point on an elliptic curve that has to be kept secret. Its knowledge breaks the scheme. We only mention two additional ECC-based functions of the new German identity card.

- It offers an eID function (see [28]) that allows users to identify themselves on the internet. For each authorization process the user determines what information he is willing to transmit to the service provider. One example is age verification. Age verification confirms that a user has reached a certain age. Another application is a pseudonym function that allows the user to communicate anonymously in social networks.
- It allows to use qualified electronic signatures according to the German signature law. For this functionality, an appropriate certificate has to be acquired from a certification service provider.

Furthermore, the high-security encryption solutions used by the German government make heavy use of elliptic curve cryptography.

2 Elliptic Curve Cryptography

The use of elliptic curves in cryptography was independently proposed by Koblitz [50] and Miller [52]. Let us start with some interesting protocols based on the arithmetic of elliptic curves over finite fields. We present the textbook versions. Some problems arising for real-world applications are briefly mentioned. Assume we have two parties called Alice and Bob and they want to send messages to each other over an insecure channel. An eavesdropper Eve is able to intercept the messages sent over the channel.

2.1 Elliptic Curve Discrete Logarithm Problem (ECDLP)

Let E be an elliptic curve defined over a finite field $\mathbb{F}_q$ with q elements, where q is a power of a prime. We define the ECDLP which is the discrete logarithm problem in the group $E(\mathbb{F}_q)$ of $\mathbb{F}_q$-rational points on E, now written additively: Given a point $P \in E(\mathbb{F}_q)$ of prime order n and let $Q \in E(\mathbb{F}_q)$ be another point such that $Q = \ell P$ for some integer ℓ; find $\ell \in [0, n-1]$. In general, the quantities q, E, and P should be selected so that the ECDLP is presumably a computationally difficult problem (see Sects. 1.1 and 5).

2.2 Elliptic Curve Key Agreement Protocol

Diffie and Hellman introduced in 1976 a key agreement protocol based on the arithmetic in the multiplicative group of a finite field of large prime characteristic. With this protocol Alice and Bob are able to agree on a secret key over an insecure network, whereas Eve is not able to find out the key. In the elliptic curve analogue of this protocol, Alice and Bob first agree on q, an elliptic curve E defined over $\mathbb{F}_q$, and a point $P \in E(\mathbb{F}_q)$. Then Alice and Bob choose a secret integer d_A respectively d_B at random and compute the points $Q_A = d_A P$ respectively $Q_B = d_B P$. Now the communication over the unsecured channel can take place. Alice sends Q_A to Bob and Bob sends Q_B to Alice. In the end, they can both compute the common secret point, namely $K_{AB} = d_A d_B P \in E(\mathbb{F}_q)$ by

$$K_{AB} = d_A Q_B = d_A(d_B P) = d_B(d_A P) = d_B Q_A.$$

Eve knows $Q_A = d_A P$ and $Q_B = d_B P$ but neither d_A nor d_B and therefore she cannot compute $K_{AB} = d_A d_B P$ from the known parameters unless she is able to solve the elliptic curve Diffie-Hellman problem (ECDHP) which is: Given P, $d_A P$ and $d_B P$, determine $d_A d_B P$. Note that if one can solve the ECDLP one can solve the ECDHP.

Nevertheless the Diffie-Hellman scheme is insecure if the messages exchanged between Alice and Bob are not authenticated. In this case Eve can impersonate Bob against Alice and Alice against Bob. She can then agree on a secret point with both of them and act as a man-in-the-middle. Authentication can be achieved by the use of digital signatures (see Sect. 2.3) and certificates—which have to be distributed in a trusted way.

In practice often only one of the communication partners changes his ephemeral key (the client), the other one uses a static key (the server). This is called static-ephemeral DH. There are also static-static versions of DH.

2.3 Elliptic Curve Digital Signature Algorithm (ECDSA)

Similar to handwritten signatures, one uses digital signatures in todays communication in order to achieve three services: Authentication (assurance of identity), data-integrity (assurance that data has not been modified), and non-repudiation (providing evidence to a third-party that a specific party participated in a transaction). The ECDSA achieves these properties with the help of elliptic curves. Assume that Alice and Bob have agreed on q, an elliptic curve E defined over $\mathbb{F}_q$, and a point $P \in E(\mathbb{F}_q)$ of prime order n. For the ECDSA they also have to agree on a collision-free cryptographic hash function H. We simply interpret H as a one-way function that takes an arbitrary binary string as input and returns an integer less than n. We assume that Alice has a public key $Q_A = d_A P$ and a randomly chosen secret key $d_A \in [1, n-1]$.

In order to sign a message m Alice performs the following operations: First she selects a secret integer[3] $k_e \in [1, n-1]$ at random, computes $k_e P$ and transforms the x-coordinate of $k_e P$ into an integer x_1. Then she calculates $r \equiv x_1 \pmod{n}$. If r happens to be 0, then a new ephemeral key k_e has to be selected. Next, Alice computes the value of the hash function $h = H(m)$ and also $s \equiv k_e^{-1}(h + d_A r) \pmod{n}$. In the unlikely case that $s = 0$ she has to start over with a new value of k_e. If not, the signature generation is completed and the signature is (r, s) which Alice sends along with the message m to Bob.

Now Bob wants to verify the validity of the signature. He looks up Alice's public key Q_A and checks if r and s are in the interval $[1, n-1]$. Then Bob computes the hash value $h = H(m)$ as well as the values $v_1 \equiv h s^{-1} \pmod{n}$ and $v_2 \equiv r s^{-1} \pmod{n}$ where s^{-1} denotes the multiplicative inverse element of s modulo n. Finally he computes $v_1 P + v_2 Q_A$ which is a point in $E(\mathbb{F}_q)$. If $v_1 P + v_2 Q_A = \mathcal{O}$ then Bob simply rejects the signature. In fact, this situation should obviously be avoided. Otherwise $v_1 P + v_2 Q_A$ has coordinates in $\mathbb{F}_q$ and the x-coordinate of this point can be transformed into an integer x_2. Bob has to verify that x_2 is congruent to $r \pmod{n}$. If this is the case the signature is accepted, otherwise it is rejected.

[3] This randomly selected, secret integer is usually called the ephemeral key.

In practice the way of choosing ephemeral keys is crucial. For example there is an unpublished attack on the ECDSA and similar schemes that uses the fact that ephemeral keys will be biased depending on the output of a pseudorandom number generator. This attack led to a change of NIST's original Digital Signature Standard (DSS). In Germany the use of legally binding digital signatures is regulated by law.

2.4 Elliptic Curve ElGamal Public-Key Cryptosystem

Another protocol of elliptic curve cryptography is the elliptic curve version of the ElGamal public-key cryptosystem which works as follows. In a precomputational step, Alice and Bob agree on q, an elliptic curve E defined over $\mathbb{F}_q$, and a point $P \in E(\mathbb{F}_q)$.

Alice randomly chooses a secret multiplier $d_A \in [1, n-1]$, computes the point $Q_A = d_A P$, publishes the point Q_A as her public key, and keeps d_A secret.

Bob wishes to send a message m to Alice. In order to do so, he first looks up Alice's public key Q_A and then converts the plaintext message m into a point $M \in E(\mathbb{F}_q)$. Then Bob selects an integer $k_e \in [1, n-1]$ at random and computes

$$C_1 = k_e P \quad \text{and} \quad C_2 = M + k_e Q_A.$$

Finally, the two points (C_1, C_2) are sent to Alice so that she can recover the plaintext by computing

$$C_2 - d_A C_1 = (M + k_e Q_A) - d_A(k_e P) = M + k_e(d_A P) - d_A(k_e P) = M$$

and extracting m from M.

Formally, this cryptosystem is the quintuple $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, where $\mathcal{P} = E(\mathbb{F}_q)$, $\mathcal{C} = E(\mathbb{F}_q) \times E(\mathbb{F}_q)$, and $\mathcal{K} = \{(q, E, P, n, Q, d) \mid Q = dP\}$. For each $K = (q, E, P, n, Q, d) \in \mathcal{K}$ and $k_e \in [1, n-1]$, $E_K(M) := (k_e P, M + k_e Q)$. For $(C_1, C_2) \in E(\mathbb{F}_q) \times E(\mathbb{F}_q)$, $D_K(C_1, C_2) := C_2 - dC_1$.

We just mention that there are more refined encryption systems such as the Elliptic Curve Integrated Encryption Scheme (ECIES) which are popular choices for state-of-the-art implementations. For details on practical considerations and vulnerabilities of the plain version of the ElGamal cryptosystem described here, we refer for instance to [5].

3 RSA

We now present the ultimate example for an asymmetric cryptosystem, namely the RSA cryptosystem, named after its inventors Rivest, Shamir, and Adleman. For details, we refer to any textbook in cryptography (see e.g. [67] or simply the original research paper [61]).

The communication problem is the following: The sender Alice wishes to transmit a message m securely over a public channel to the receiver Bob. This can be accomplished by the following steps.

1. Bob
   - generates two large primes p and q.
   - computes $N = pq$ and $\varphi(N) = (p-1)(q-1)$, where φ denotes Euler's phi function.
   - chooses a random integer e, $1 < e < \varphi(N)$, so that $\gcd(e, \varphi(N)) = 1$.
   - computes $d \equiv e^{-1} \pmod{\varphi(N)}$ using the Euclidean algorithm.
   - publishes (N, e) and keeps d, p, q secret.
2. Alice wants to send a message to Bob. For the purpose of simplicity, we assume the message m to be already encoded as an integer M such that $0 < M < N$. She
   - looks up Bob's public key (N, e).
   - computes $C \equiv M^e \pmod{N}$.
   - sends C to Bob.
3. Bob recovers the message M via

$$C^d \equiv M^{ed} \equiv M^{1+j\varphi(N)} \equiv M \pmod{N},$$

   where j is an integer such that $ed = 1 + j\varphi(N)$.

We point out that the following original version of RSA corresponds to the textbook version and is insufficient for secure implementations. As presented, this protocol is the basic frame of the RSA cryptosystem and not the version that is used in implementations. For details on practical considerations, we refer for instance to [5]. For example the pure scheme is homomorphic and can be attacked by lattice-methods if d is to small.

In order to analyze RSA, we mention the idea of trapdoors. For given N, e, we define the trapdoor one-way RSA function[4] as

$$f_{N,e}(M) := M^e \pmod{N} \quad \text{for } M \in \mathbb{Z}_N.$$

This function is clearly related to the above protocol. It is

(a) easy to evaluate $M \mapsto M^e \pmod{N}$.
(b) difficult to invert $C \mapsto C^{\frac{1}{e}} \pmod{N}$ for integers C with $1 < C < N$ and $\gcd(C, N) = 1$.
(c) possible to invert $f_{N,e}(M)$ with the "trapdoor" d.

The protocol immediately produces a digital signature scheme for free, since it is known that trapdoor one-way functions yield digital signatures. In our case, this works as follows:

(a) Bob signs the message M by computing $S \equiv M^d \pmod{N}$ and sends S to Alice.
(b) Alice verifies that $S^e \equiv M \pmod{N}$.

[4]For any positive integer N we let $\mathbb{Z}_N = \{a \in \mathbb{Z} \mid 0 \le a < N\}$ be the set of representatives of $\mathbb{Z}/N\mathbb{Z}$. Then $\mathbb{Z}_N$ is a group under addition modulo N with identity 0. Its group of units $\mathbb{Z}_N^* = \{a \in \mathbb{Z} \mid 1 \le a < N, \gcd(a, N) = 1\}$ forms a group under multiplication modulo N with identity 1.

In the formal description, given a positive integer N and primes p, q such that $N = pq$, the corresponding RSA asymmetric cryptosystem is the quintuple $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, where $\mathcal{P} = \mathbb{Z}_N = \mathcal{C}$, and

$$\mathcal{K} = \left\{(N, p, q, e, d) \mid e,\ d \in \mathbb{Z}^*_{\varphi(N)},\ ed \equiv 1 \pmod{\varphi(N)}\right\}.$$

For each $K = (N, p, q, e, d) \in \mathcal{K}$ and $M, C \in \mathbb{Z}_N$, we define

$$E_K(M) := M^e \pmod{N}, \qquad D_K(C) := C^d \pmod{N}.$$

We obtain the following complexity-theoretic problems.

1. *Breaking RSA*: Inverting $f_{N,e}(M)$; that is, given $N = pq$ and e, C with $\gcd(e, \varphi(N)) = 1$ and $C = f_{N,e}(M)$, compute $C^{\frac{1}{e}} \pmod{N}$. This is precisely the problem that has to be solved by an adversary in order to break the RSA system.
2. *Special Integer Factorization Problem (SIFP)*: Given a positive integer N with $N = pq$, where p and q are primes; determine p and q.

It is easy to see that SIFP $\Rightarrow$ Breaking RSA. Suppose we are given N, e and we are able to efficiently factor N. Then we can determine p and q and thus we are able to compute $\varphi(N) = (p-1)(q-1)$. Obviously, we are now in the position to determine the secret key d by applying the extended Euclidean algorithm. With the knowledge of d and N, we have complete control over the RSA system.

The other direction Breaking RSA $\overset{?}{\Rightarrow}$ SIFP is quite unclear yet. However, one can show that SIFP $\Leftrightarrow$ Computing d from (N, e). Even though breaking RSA means to invert $f_{N,e}(M)$, one often identifies the mathematical problem of RSA to be the SIFP. This is a hard and well-studied problem in computational number theory, and also in all of number theory. The fastest known algorithmic solution to this problem is given by the general number field sieve. Heuristically, under the generalized Riemann hypothesis and various other assumptions, the general number field sieve has a subexponential running time of

$$L(N) = O\left(\mathrm{e}^{(1.923+o(1))(\ln N)^{1/3}(\ln\ln N)^{2/3}}\right) \tag{3.1}$$

operations, where $o(1) = \theta(N) \to 0$ for $N \to \infty$. For details, we refer to [21, 22]. Many experts believe that the SIFP is a computationally hard problem. Even Gauss mentioned:

> The problem of distinguishing prime numbers from composites, and of resolving composite numbers into their prime factors, is one of the most important and useful in all of arithmetic... The dignity of science seems to demand that every aid to the solution of such an elegant and celebrated problem be zealously cultivated.

It is the current belief of the community that despite the enormous efforts of many excellent researchers over the last decades, and despite the most sophisticated computer equipment, the problem of efficiently factoring $N = pq$ is not even close to be solved, and currently not even for 2048-bit numbers N.

4 ECC Versus RSA

Elliptic curve cryptography is the most attractive alternative to RSA. Why? Most importantly, ECC ought to be faster than RSA in certain applications because the elliptic curve Diffie-Hellman key agreement protocol as well as the elliptic curve ElGamal public-key cryptosystem should actually need smaller keys and are ad hoc therefore more efficient than RSA. According to the currently best known attacks on RSA and the ECDLP, the cryptographic security of ECC grows exponentially in the length of the input parameters (see Table 1). The only theoretical disadvantages are that the underlying mathematical problem is still considered "new" and that the patent situation is confusing.

What reasons do exist to pursue further research on the still important problems in ECC? Clearly, the current absence of a subexponential-time algorithm for solving the ECDLP implies significantly smaller parameters in ECC than with competing, established technologies such as RSA, but with equivalent levels of security. Thus the key-per-bit-strength of ECC is somewhat better than that of RSA. An immediate consequence from having smaller parameters is the potential gain in speed and the use of smaller certificates. These advantages are especially important for long-term security and in environments where at least one of the following resources is limited: Storage space, bandwidth, or power. Popular examples in Germany are the national ID cards and passports. Summarizing, ECC is especially well-suited for constrained environments such as smart cards, cellular phones, pagers, digital postal marks, and personal digital assistants (PDAs).

5 ECDLP Attacks

Here, we sketch the ideas of some attacks on the ECDLP which ultimately led to the recommended domain parameters introduced in Sect. 1. We investigate the ECDLP as defined in Sect. 2.1. For further details and explicit formulations of the attacks we refer to [18, 36, 44].

5.1 Generic Attacks

The usual generic attacks for finite abelian groups apply to solving the ECDLP. The most notable ones in this context are the Pohlig-Hellman attack and Pollard's rho method. Let E be an elliptic curve defined over $\mathbb{F}_q$.

The *Pohlig-Hellman attack* is successful if the group order $\#E(\mathbb{F}_q)$ is known, if $\#E(\mathbb{F}_q)$ can be factored into primes by using an integer factorization method, and if $\#E(\mathbb{F}_q)$ is smooth, i.e. splits into small primes only. In that case, the ECDLP can be solved by first solving the ECDLP in small groups of prime order in parallel and by then using a Chinese remainder technique to put the pieces together. In order to prevent this attack, it is therefore recommended to choose the parameters so that $\#E(\mathbb{F}_q) = \lambda n$, where n is a prime and $\lambda = 1, 2, 3,$ or 4.

Pollard's rho method can always be applied. It uses a generic pseudo-random walk à la Pollard or Teske in $E(\mathbb{F}_q)$ together with a distinguished point method with

very little space. It has an overall expected running time of $O(\sqrt{q})$. In particular, this algorithm can be parallelized effectively with a linear speed-up in the number of processors. If additional knowledge is given, then the expected running time of another attack is measured with respect to Pollard's rho method. Also, the size of the underlying finite field in the domain parameters is determined by current and predicted running times of implementations of the parallelized rho method with up-to-date computer equipment.

5.2 Weil Pairing and Tate-Lichtenbaum Pairing Attack

The basic idea of these attacks stems from the so-called Menezes-Okamoto-Vanstone reduction [55]. Let E be an elliptic curve defined over $\mathbb{F}_q$. One makes use of the fact that there exists a non-degenerate pairing, namely the Weil pairing, and similarly the Tate-Lichtenbaum pairing, from $E(\mathbb{F}_q)[n] \times E(\mathbb{F}_q)[n]$ to the multiplicative group $\mathbb{F}_{q^k}^*$ of $\mathbb{F}_{q^k}$ for some positive integer k, where $n|(q^k - 1)$. More details about these pairings and their constructive use in cryptography are given in Sect. 7. Here, the ECDLP can thus be reduced to the discrete logarithm problem in $\mathbb{F}_{q^k}^*$. Now, in $\mathbb{F}_{q^k}^*$ a variation of the general number field sieve method for factoring exists which solves the corresponding DLP and thus the ECDLP in subexponential-time $L(q^k)$ according to (3.1). In general k is expected to be very large. However, for special curves such as supersingular curves the embedding degree k is comparably small. This attack is therefore successful if $n|(q^k - 1)$ for some small value of k so that finding discrete logarithms in $\mathbb{F}_{q^k}^*$ is computationally feasible. This yields the Menezes-Okamoto-Vanstone condition in Sect. 1.1, i.e. that $n - 1$ divided by the order of p modulo n is less than 10000. Note that on the other hand, a small value of k leads to another very productive branch of cryptography (see Sect. 7).

5.3 Anomalous Curves Attack

Let E be an elliptic curve defined over $\mathbb{F}_q$. This attack is successful for $\mathbb{F}_q$-anomalous elliptic curves, which are curves with a large p-subgroup. For instance, let $q = p$. Then elliptic curves with $\#E(\mathbb{F}_p) = p$ are $\mathbb{F}_p$-anomalous. By Hasse's theorem we have $\#E(\mathbb{F}_p) = p + 1 - t$, where t denotes the trace of E over $\mathbb{F}_p$. Thus $\#E(\mathbb{F}_p) = p$ is equivalent to $t = 1$. The idea of the attack in this case is in principle to lift the points Q and P to points $\tilde{Q}$ and $\tilde{P}$ on the lifted elliptic curve $\tilde{E}/\mathbb{Q}_p$, where $\mathbb{Q}_p$ denotes the field of p-adic numbers. One then uses the concept of formal p-adic elliptic logarithm on $\tilde{E}/\mathbb{Q}_p$ to solve the ECDLP altogether in polynomial time complexity in the input size. Consequently, one requires the domain parameters in 1.1 selected so that E over $\mathbb{F}_q$ is not $\mathbb{F}_q$-anomalous.

5.4 General Index Calculus Attacks

We first sketch the generic index calculus technique in a finite abelian group G adapted to the additive setting. The following steps should be performed if terms such as "norm", "prime element", "smoothness" are meaningful and hopefully the group G can be generated by elements of small norm.

Let G be a finite abelian group and let $D_1 \in G$ of order n. We assume the group order and n are known. Given D_2 such that $D_2 = \ell D_1$, we wish to find $\ell \in [0, n-1]$. In particular, an index calculus attack to the ECDLP as in Sect. 2.1 can be formulated for $G = E(\mathbb{F}_{q^r})$, $D_1 = P$, and $D_2 = Q = \ell P$.

1. One selects a smoothness bound B and chooses a *factor base* $\mathcal{F}_B$ for the relation generation

$$\mathcal{F}_B = \{\mathcal{P}_1, \ldots, \mathcal{P}_s\} = \{\text{prime elements in } G \text{ of norm } \leq B\}.$$

2. Construct enough different relations and create the relation matrix $A = (a_{ij}) \in \mathbb{Z}_n^{s\times(s+5)}$. This could for instance be done as follows: Perform a pseudo-random walk à la Pollard or Teske in the group to find smooth group elements of the form

$$\alpha_i D_1 + \beta_i D_2 = \sum_{j=1}^{s} a_{ji}\mathcal{P}_j \quad (1 \leq i \leq s+5),$$

   where the α_i's and β_i's are integers. Store A and the coefficients α_i, β_i.
3. Use linear algebra to determine an element $\gamma \in \ker(A)$, i.e. a solution $\gamma = (\gamma_1, \ldots, \gamma_{s+5})$ to $Ax = 0$.
4. If $\sum \beta_i\gamma_i \not\equiv 0 \pmod{n}$, then $\ell \equiv -(\sum \alpha_i\gamma_i)/(\sum \beta_i\gamma_i) \pmod{n}$.

5.5 Xedni Attack

We first discuss the basic idea of the relevant version of an index calculus method (see [52, 64]). Let E be an elliptic curve defined over $\mathbb{F}_p$, where p is a prime. The direct approach as in the previous section applied to $G = E(\mathbb{F}_p)$, $D_1 = P$, and $D_2 = Q$ is unsuccessful. An explanation is as follows: First one lifts the curve $E/\mathbb{F}_p$ to an elliptic curve $\mathcal{E}/\mathbb{Q}$. Then, one attempts to lift various points from $E/\mathbb{F}_p$ to $\mathcal{E}/\mathbb{Q}$. Finally, one uses relationships among those lifted points to recover the ECDLP. However, this method fails since lifting of the points is difficult and one needs many rational points of small height.

Silverman [64] suggested to proceed conversely. For this reason he called his attack xedni attack. This specific attack has been subsequently analyzed in detail in [46]. Interestingly, Koblitz pointed out that if Silverman's xedni algorithm were successful, then RSA could be attacked by an extension of the method as well. Silverman's idea was the following: First, one chooses points $P_1, \ldots, P_s$ in $E(\mathbb{F}_p)$ and lifts them to points $Q_1, \ldots, Q_s$ having integer coefficients. This can easily be accomplished. Then one chooses by linear algebra an elliptic curve $\mathcal{E}/\mathbb{Q}$ that goes through the lifted points $Q_1, \ldots, Q_s$. The whole point is that one hopes that the lifted curve $\mathcal{E}/\mathbb{Q}$ has smaller Mordell-Weil rank than expected. Silverman imposed an additional idea of Mestre to make this probability even higher, namely lift the curve $E/\mathbb{F}_p$ to $\mathcal{E}/\mathbb{Q}$ so that the reduced curve $E/\mathbb{F}_u$ for various small primes u satisfies $\#E(\mathbb{F}_u) \approx u + 1 - 2\sqrt{u}$. If there are nontrivial relations among the points $Q_1, \ldots, Q_s$, the ECDLP is solved. However, the analysis in [46] shows that this attack fails, since mainly the absolute bound on the size of the coefficients of a relation satisfied by the lifted points is too small.

5.6 Semaev's Index Calculus Attack

Even though this attack is not immediately successful, ideas of this attack led to important attacks on elliptic curves defined over finite field extensions (see Sects. 5.7 and 5.8). Again let E be an elliptic curve defined over $\mathbb{F}_p$ and let $\overline{\mathbb{F}}_p$ denote an algebraic closure of $\mathbb{F}_p$. Semaev suggested to use the so-called summation polynomials for the generation of the relation in the general index calculus algorithm. For integers $j \geq 2$ these are recursively defined symmetric polynomials $f_j \in \mathbb{F}_p[X_1, \ldots, X_j]$ with $f_2(X_1, X_2) := X_1 - X_2$ and $\deg_{X_i} f_j = 2^{j-2}$. The following important property holds true: For $(x_1, \ldots, x_j) \in \overline{\mathbb{F}}_p^j$ we have $f(x_1, \ldots, x_j) = 0$ if and only if there exists $(y_1, \ldots, y_j) \in \overline{\mathbb{F}}_p^j$ such that

$$P_1 + P_2 + \cdots + P_{j-1} + P_j = \mathcal{O}$$

and $P_i = (x_i, y_i) \in E(\overline{\mathbb{F}}_p)$.

As usual we now identify $\mathbb{F}_p$ with its set of representatives $\mathbb{Z}_p = \{a \in \mathbb{Z} \mid 0 \leq a < p\}$. Semaev's index calculus attack uses for a cyclic group $E(\mathbb{F}_p)$ and an integer $j \geq 2$ the factor base

$$\mathcal{F}_j = \left\{(x, y) \in E(\mathbb{F}_p) : 0 \leq x \leq p^{\frac{1}{j}}\right\}.$$

A relation is constructed by generating a random point $R = k_1 P + k_2 Q \in E(\mathbb{F}_p)$ and by then expressing $R = (x_R, y_R)$ as a sum of points in $\mathcal{F}_j$ by solving the multivariate polynomial congruence

$$f_{j+1}(x_1, \ldots, x_j, x_R) \equiv 0 \pmod{p} \quad \text{and} \quad x_1, \ldots, x_j \leq p^{\frac{1}{j}}.$$

If this is solvable, determine the corresponding y-coordinates $\pm y_i \in \mathbb{F}_p$ of points. If all $y_i \in \mathbb{F}_p$, then each $P_i = (x_i, y_i) \in \mathcal{F}_j$ and we have a relation

$$s_1 P_1 + s_2 P_2 + \cdots + s_j P_j = R = k_1 P + k_2 Q \quad (s_i = \pm 1)\ .$$

A detailed analysis shows that the approximate heuristic running time of Semaev's index calculus is

$$O\left(t_{j,p}\, j!\, p^{\frac{1}{j}} + p^{\frac{2}{j}}\right),$$

where $t_{j,p}$ is the expected heuristic running time for solving the multivariate polynomial congruence for small values of x_i. However, there is no indication why solving the multivariate polynomial congruence for small values of x_i should be easy and why this attack would work faster than exhaustive search.

5.7 Gaudry's Index Calculus Attack

Now, let E be an elliptic curve defined over $\mathbb{F}_{q^r}$, where $r > 1$, and for simplicity let $E(\mathbb{F}_{q^r})$ be cyclic. Gaudry [33] suggested to use the ideas of Semaev's index calculus method and especially the summation polynomials (see Sect. 5.6) in order to derive an

attack that works for elliptic curves over $\mathbb{F}_{q^r}$ for certain values of r. Gaudry suggested to use the factor base

$$\mathcal{F} = \{(x, y) \in E(\mathbb{F}_{q^r}) : x \in \mathbb{F}_q\}.$$

For the relation generation step in the general index calculus algorithm one tries to express a random point $R = (x_R, y_R) \in E(\mathbb{F}_{q^r})$ as a sum of points in $\mathcal{F}$. As in Semaev's approach one uses the summation polynomials and finds solutions $(x_1, \ldots, x_r)$ to

$$f_{r+1}(x_1, \ldots, x_r, x_R) = 0 \quad \text{and} \quad x_i \in \mathbb{F}_q.$$

Now, let $\{\beta_1, \ldots, \beta_r\}$ be a basis for $\mathbb{F}_{q^r}$ over $\mathbb{F}_q$. One represents x_R and the coefficients of the equations of the elliptic curve in terms of this basis. Inserting this into the original equation for $f_{r+1}(x_1, \ldots, x_r, x_R)$ and equating coefficients at β_i yields r equations in r variables $x_1, \ldots, x_r$. Eventually, this leads to the following system of polynomial equations

$$g_i(x_1, \ldots, x_r, x_R) = 0 \quad (1 \leq i \leq r).$$

By using optimized algorithms and a so-called double large prime variant one obtains an expected running time of $O(q^{2-\frac{2}{r}})$ for r fixed and small. Comparing this with the expected complexity of Pollard's rho algorithm of $O(q^{\frac{r}{2}})$ we can conclude: Gaudry's algorithm for solving the ECDLP for an elliptic curve E defined over $\mathbb{F}_{q^3}$ or $\mathbb{F}_{q^4}$ is asymptotically faster than Pollard's rho algorithm. These techniques are independent of the arithmetic properties of the elliptic curves.

5.8 Diem's Index Calculus Attack

Diem [25] generalized and confirmed Gaudry's results in various ways. We mention only certain aspects of his work. The principal idea goes back to Semaev's index calculus idea by making use of the summation polynomials (see Sect. 5.6). Let E be an elliptic curve defined over a finite field extension $\mathbb{F}_{q^r}$. In the first step, a factor base has to be selected. Let e be an integer with $e \geq 3$ and put $v := \lceil r/e \rceil$. Randomly select v linear independent elements $\alpha_1, \ldots, \alpha_v \in \mathbb{F}_{q^r}$ and define the subspace $F_v = \langle \alpha_1, \ldots, \alpha_v \rangle$ of dimension v. The factor base is defined as

$$\mathcal{F} = \{(x, y) \in E(\mathbb{F}_{q^r}) : x \in F_v\}.$$

The relation generation is performed as follows: For a random $R \in E(\mathbb{F}_{q^r})$, solve

$$f_{e+1}(x_1, \ldots, x_e, x_R) = 0 \quad \text{and} \quad x_i \in F_v.$$

Reformulation with respect to a basis $\{\beta_1, \ldots, \beta_r\}$ for $\mathbb{F}_{q^r}$ over $\mathbb{F}_q$, inserting everything in the polynomial equation and equating coefficients yields r polynomial equations in ev variables.

We mention an important consequence of Diem's analysis. Let a, b be real numbers such that $0 < a < b$. Then Diem's algorithm has an expected subexponential running time $L_{q^r}[\frac{3}{4}, c]$ for $a \log q \leq r \leq b \log q$ and $e \sim \sqrt{\log q}$, where $c = c(a, b)$

is a constant. Notice that Diem's results are remarkable since he proved that a subexponential-time algorithm exists for solving the ECDLP of a certain small class of elliptic curves over finite field extensions and the attack is independent of the arithmetic properties of the curves.

5.9 Weil Descent Attack

Let E be an elliptic curve defined over $\mathbb{F}_{q^r}$, where $r > 1$. This attack is successful for some elliptic curves with special arithmetic properties. We refer to [18, 35, 40, 42] for details on the theoretical results. An explicit realization of the attack as well as recent computational examples with running times over $\mathbb{F}_{2^{124}}$ and $\mathbb{F}_{2^{155}}$ can be found in [47, 72].

The idea of the attack is to first embed E into the Weil restriction of scalars $\mathcal{W}_E(\mathbb{F}_q)$ over the smaller field $\mathbb{F}_q$. One then tries to find a curve $\mathcal{X} \subseteq \mathcal{W}_E$ with good properties and one constructs an efficiently computable group homomorphism $\Phi : E(\mathbb{F}_{q^r}) \longrightarrow J_{\mathcal{X}}(\mathbb{F}_q)$ such that $\#\ker(\Phi) \leq \lambda$. Then one computes $D_1 = \Phi(P)$ and $D_2 = \Phi(Q)$. At this point, one has reduced the ECDLP of $E/\mathbb{F}_{q^r}$ for points P and $Q = \ell P$ to the discrete logarithm problem in the Jacobian of a higher genus curve $\mathcal{X}$ over $\mathbb{F}_q$; that is, one solves $D_2 = \ell D_1$ on $J_{\mathcal{X}}(\mathbb{F}_q)$ by known index calculus methods as in Sect. 5.4 if the key size is appropriate.

This method works for a small but significant part of all elliptic curves over $\mathbb{F}_{q^r}$. As a consequence of this result, the results in Sect. 5.7 as well as Sect. 5.8, and other results for elliptic curves E defined over finite field $\mathbb{F}_{q^r}$, one recommends to use domain parameters (see (1.1)) such that E is an elliptic curve defined over a finite field $\mathbb{F}_q$ and either q is a prime or $q = 2^{p'}$, where p' is a prime.

6 Parameter Selections

6.1 Elliptic Curves for Security Applications

In this section we focus on the selection of elliptic curves for security applications, such as digital signatures, MRTDs, or government use. The requirements in the low-cost area (e.g. RFIDs for copyright protection) or for applications of pairing-based cryptography may differ. There are several aspects that have to be taken into account when selecting elliptic curves for security applications. The explicit selection of the parameters has been listed in Sect. 1. We refer to [18, 44, 51] for details.

Implementation Issues Several choices of curve parameters may ease the implementation of elliptic curves and prevent implementation errors.

Resistance Against Side Channel Attacks There are side channel attacks that make use of the existence of curve points with distinguished properties. For instance, it is easier to attack curves with a $\mathbb{F}_p$-rational point whose x- or y-coordinate equals zero. In the short Weierstrass form (1.2) these conditions translate to: b should be a quadratic nonresidue modulo p and there should be no point of order 2 in the subgroup of E under consideration. For a comprehensive overview of implementational issues we refer to [49] and Sect. 8.

Appropriate Key Length When used in a hybrid scheme the asymmetric key length should be roughly twice the underlying symmetric key length. However, it is considered appropriate to use 384-bit curves in conjunction with 256-bit symmetric algorithms. For signature applications that are compliant with the German digital signature law, the German Federal Network Agency publishes algorithms and parameter lengths that are considered secure for at least six years. Their catalogue is updated annually and the current version considers a group order of magnitude $\sim 2^{224}$ as secure for qualified electronic signatures until 2015, and $\sim 2^{250}$ until 2018. There is no restriction on the size of the base field.

German identity cards and MRTDs use elliptic curves as well. For identification purposes two different curves are used. The country signing certificate authority uses the elliptic curve brainpoolP384r1, whereas for document signing the curve brainpoolP256r1 is used.

Representation of the Curve and Choice of the Arithmetic There are many popular representations of elliptic curves that allow fast or secure implementations. To implement the multiplication with scalars also many methods exist, e.g. double-and-add or use of the non-adjacent form (NAF). On the other hand methods against SCA involve randomization and may slow down the arithmetic. One may benefit from curve representations with uniform addition and doubling formulas, such as curves in Edwards form. Note that for curves in short Weierstrass form the standard formulas for doubling are different form the formulas for adding. It can also be helpful to implement scalar multiplication by using the Montgomery ladder which makes doubling and adding of points indistinguishable. It is still of interest to implement algorithms in a way that makes timing behaviour and power consumption independent of the data processed.

Quantum Computers On (hypothetical) quantum computers Shor's algorithm allows to easily solve the ECDLP. One countermeasure is to keep the curve equation secret. For curves over prime fields three curve points suffice to find the characteristic of the base field and the equation of the curve. As a consequence one would also have to keep curve points secret. In the case of an unknown curve equation, this can be achieved by point-compression methods that also may help to save on bandwidth.

6.2 Generation of Suitable Elliptic Curves

There are two different approaches to generate suitable curves for security applications.

- The CM method uses class field theory to construct curves with a prescribed number of points whose endomorphism ring has a relatively small class number. There are recent variations which allow to generate curves with relatively large class number. Interestingly the same ideas are used in Atkin's deterministic primality proving algorithm.
- Point counting. By using the Schoof-Elkies-Atkin algorithm (SEA) the number of points of elliptic curves over $\mathbb{F}_q$ in the cryptographic range can easily be computed.

Computer Algebra systems such as Magma [1] offer quick implementations. Interestingly, point counting on elliptic curves over the Ring $\mathbb{Z}/N\mathbb{Z}$ and factoring N are equivalent

When constructing an elliptic curve one starts with the choice of a base field. A popular choice are prime fields with pseudo Mersenne characteristic. Such fields offer fast arithmetic which can in turn be used to develop fast implementations. This approach has been taken by NIST. When following this approach one has to carefully consider the patent situation regarding fast arithmetic. One should also consider that the use of special primes may make implementations more vulnerable to side-channel attacks. Another approach is to choose prime fields as well as curve equations pseudo randomly and deterministically.

7 Pairing-Based Cryptography

The Weil pairing on supersingular elliptic curves was the first pairing to occur in cryptography, and its use has been of a destructive nature as discussed in Sect. 5.2. This attack lowered the efficiency and security of supersingular elliptic curves to a level comparable to RSA. As a consequence, the general advantages of elliptic curves over RSA were lost in this special case and supersingular elliptic curves were banned from further cryptographic research for about a decade. Only around 2000 it was then observed that revolutionary new cryptographic primitives could be realized using supersingular elliptic curves and the Weil pairing. This marked the start of pairing-based cryptography.

7.1 Pairing-Based Cryptographic Protocols

Pairing-based cryptographic protocols have taken a central position in cryptographic research in the past twelve years. These protocols are based on classical discrete logarithm based protocols and use a bilinear, non-degenerate map

$$e : G_1 \times G_2 \to G_3$$

of cyclic groups G_1, G_2, and G_3 with prime order n as a key additional feature. It is customary to write the groups laws of G_1, G_2, and G_3 multiplicatively, while in practice G_1 and G_2 will be subgroups of some $E(\mathbb{F}_{q^k})$ with additive group law and G_3 a subgroup of $\mathbb{F}_{q^k}^*$ with multiplicative group law. Bilinearity thus means $e(xy, u) = e(x, u)e(y, u)$ and $e(x, uv) = e(x, u)e(x, v)$ for all $x, y \in G_1$ and $u, v \in G_2$. Non-degeneracy means that there are $x \in G_1$ and $u \in G_2$ with $e(x, u) \neq 1$.

The papers [2, 48, 66] on identity based cryptography and tripartite key exchange have been the starting point for pairing-based cryptography. Much attention was in particular payed to [2], which solved an open research problem posed in [62] back in 1984. Since the publication of these three papers a very large number of applications of pairings in cryptography have been exhibited which exceed by far simple encryption or signature protocols. An early survey can be found in [59].

In the rest of this subsection the protocols from [2] and [7] are described in a simplified way to give the reader a rough impression how pairings are used in cryptography and what security notions are put in place.

Identity-Based Cryptography Identity-based cryptography and in particular identity-based encryption offer an alternative approach to a traditional public key infrastructure. The idea is that public keys can be arbitrarily prescribed instead of being derived from a secret. For example, public keys could be email addresses, and the corresponding private key would be derived from the public key by a trustworthy third party, called trust center. There is then no need for Bob to obtain the public key of his communication partner Alice from Alice herself, but Bob can immediately encrypt messages for Alice using the email address of Alice. On the other hand, Alice can only decrypt when she has queried the trust center for her private key. The trust center is assumed to ensure that private keys are only handed out to the corresponding eligible person. A man-in-the-middle or impersonation attack on Alice is then not possible per assumption on the trust center. Note also that the trust center has access to all data encrypted with Alice's identity. In summary, identity-based cryptography offers some interesting advantages, but also disadvantages over a traditional public key infrastructure. A discussion of relevant aspects in view of practical employment can for example be found in [60].

We now describe the basic protocol from [2]. Assume $G = G_1 = G_2$ and consider a pairing $e : G \times G \to G_3$ and let $g \in G$ be a generator of G. Each person X has an associated identity string ID_X. Furthermore, let $H_{ID} : \{0,1\}^* \to G$ and $H : G_3 \to \{0,1\}^s$ be two cryptographic hash functions, where $s \approx \log_2(n)$. If $m_1, m_2 \in \{0,1\}^s$ then $m_1 \oplus m_2$ denotes the sum of m_1 and m_2 as elements of $\mathbb{F}_2^s$ (equivalently, $m_1 \oplus m_2$ is the bitwise xor of m_1 and m_2). The trust center is denoted by T. The key generation for identity-based cryptography proceeds as follows: The trust center T chooses $x \in \mathbb{Z}/n\mathbb{Z}$ uniformly at random and computes $g_{pub} = g^x$. The public key of T is g, g_{pub}. The private (secret) key of T is x. In addition, T computes $y_A = H_{ID}(ID_A)$ and $s_A = y_A^x$. The public key of Alice is y_A. The private key of Alice is s_A. In the identity-based encryption protocol from [2], Bob computes the ciphertext $(u, v) = \mathcal{E}(g, g_{pub}, y_A, m)$ as follows: The plaintext m is encoded as $w \in \{0,1\}^s$. Then $u = g^r$ and $v = w \oplus H(e(y_A, g_{pub}^r))$ are computed for $r \in \mathbb{Z}/n\mathbb{Z}$ chosen uniformly at random. The ciphertext is (u, v). To decrypt (u, v), Alice computes the plaintext $m = \mathcal{D}(s_A, (u, v))$ via $w = v \oplus H(e(s_A, u))$ and decodes w to the plaintext m.

The standard security model for identity-based encryption is as follows: The attacker has access to the secret keys of identities different from the target identity and the attacker can also obtain decryptions of arbitrary ciphertexts under the target identity by querying an oracle. The attacker is first required to output two different target plaintexts and a target identity. The attacker is then given the encryption of one of the target plaintexts, chosen uniformly at random, under the target identity. Finally, the attacker is deemed successful, if he can guess the corresponding target plaintext with probability significantly different from $1/2$ without querying for the decryption of the given encryption (we leave a precise definition of "significantly" open).

An identity-based encryption protocol is called IND-ID-CCA secure if there is no successful attacker in the above sense. The application of a variant of the Fujisaki-Okamoto transformation to the above basic identity-based encryption protocol yields an improved identity-based encryption protocol that is IND-ID-CCA secure under the following assumptions: The attackers are supposed to work independently of randomly chosen cryptographic hash functions H_{ID} and H (the so called random oracle

model) and the bilinear (computational) Diffie-Hellman problem (BDH) is hard. This basic computational problem is: Given uniformly at random chosen g, g^a, g^b, g^c in G compute $e(g,g)^{abc}$. The BDH can easily be reduced to the CDH in G_3 or the CDH in G.

Short Deterministic Signatures The signature scheme from [7] uses a cryptographic hash function $H : \{0,1\}^* \to G$. The key generation is the same as in classical discrete logarithm based systems. The signer chooses $x \in \mathbb{Z}/n\mathbb{Z}$ uniformly at random and computes $y = g^x$. The public key is y, the private key x. For a signature computation $\sigma = \mathcal{S}(x,m)$ with the message m the signer computes $\sigma = H(m)^x$. The signature is (m,σ). For a signature verification $b = \mathcal{V}(y,m,\sigma)$ the verifier first computes $v = e(g,\sigma)$ and then $v' = e(y,H(m))$. If $v' = v$ then $b = 1$ and the signature is accepted by the verifier, and if $v' \neq v$ then $b = 0$ and the signature is rejected by the verifier.

The security model for signature schemes considers attackers which can obtain signatures for arbitrary messages of their choice by querying an oracle. The attacker is deemed successful if he can compute a message and a valid signature for the message without having queried for the signature of this message. If there is no such attacker then the signature scheme is called secure with respect to existential forgery.

The signature scheme from [7] is secure with respect to existential forgery in the random oracle model if the CDH in G is hard. Incidentally, the DDH is easy in this case: The verification step of the signature scheme actually uses the pairing to check whether $(g, g^x, H(m), \sigma)$ is a Diffie-Hellman tupel. In comparison to previous discrete logarithm based signature schemes such as ECDSA this signature scheme is deterministic and requires only about half of the bandwith of the previous signature schemes.

Choice of Pairings Suitable pairings are the Weil and Tate-Lichtenbaum pairings and modifications of these pairings for very carefully chosen elliptic (or hyperelliptic) curves. There are no pairings in other mathematical contexts known which would appear to be both efficiently computable and secure, and it is an interesting open problem to find such pairings. In practice G, G_1, G_2 are thus point groups of elliptic curves (or Picard groups of hyperelliptic curves), and G_3 is a subgroup of the multiplicative group of a finite field.

In order to balance and thus optimize security and efficiency, the groups G, G_1, G_2, and G_3 respectively need to be chosen such that the DLP has roughly the same complexity in each of these groups.

Protocols often require further properties of pairings. Cryptographers usually classify pairings in three types, see [19, 38]. Pairings of type 1 come with an isomorphism $G_1 \to G_2$ that can be efficiently computed in either direction. In this situation $G_1 = G_2$ can be assumed. Pairings of type 2 come with a one-way isomorphism $G_2 \to G_1$, and for pairings of type 3 there is no efficiently computable isomorphism from $G_1 \to G_2$ or $G_2 \to G_1$. Some further remarks about the mathematical realization of such types of pairings are be made below.

7.2 Weil Pairing and Tate-Lichtenbaum Pairing

We now focus on the mathematics behind the pairings. In the following let E denote an elliptic curve over $\mathbb{F}_q$. We assume that its cardinality $\#E(\mathbb{F}_q)$ has a sufficiently

large prime divisor $n \neq q$. Let $k \in \mathbb{Z}^{\geq 1}$ be minimal with $n|(q^k - 1)$. We assume in addition, that $k \geq 2$ and that $q^k - 1$ is divisible by n but not divisible by n^2. The number k is called the embedding degree of E.

Under these assumptions we have that $E(\mathbb{F}_{q^k})[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ and that the group of n-th roots of unity μ_n is contained in $\mathbb{F}_{q^k}$.

The Tate-Lichtenbaum pairing

$$\langle \cdot, \cdot \rangle_n : E(\mathbb{F}_{q^k})[n] \times E(\mathbb{F}_{q^k})/nE(\mathbb{F}_{q^k}) \to \mathbb{F}_{q^k}^*/\left(\mathbb{F}_{q^k}^*\right)^n$$

is defined as follows. For every $P \in E(\mathbb{F}_{q^k})$ and $j \in \mathbb{Z}$ let $f_{j,P} \in \mathbb{F}_{q^k}(E)$ denote a rational function on E with divisor $(f_{j,P}) = j((P) - (\mathcal{O})) - ((jP) - (\mathcal{O}))$, where (P) is the prime divisor on E defined by P. The function $f_{j,P}$ is defined only up to non zero multiples from $\mathbb{F}_{q^k}$ (see [32, 65, 68] for more details on rational functions, divisors and the theorem of Riemann-Roch on curves, and Example 1 how such rational functions can look like). Now let $P \in E(\mathbb{F}_{q^k})[n]$ and $Q \in E(\mathbb{F}_{q^k})$. We choose $R \in E(\mathbb{F}_{q^k})$ with $\{Q + R, R\} \cap \{P, \mathcal{O}\} = \emptyset$ and define

$$\left\langle P, Q + nE(\mathbb{F}_{q^k}) \right\rangle_n := f_{n,P}(Q + R) \cdot f_{n,P}(R)^{-1} \cdot \left(\mathbb{F}_{q^k}^*\right)^n.$$

With the help of Weil reciprocity one can easily show that $\langle \cdot, \cdot \rangle_n$ is well-defined [32, 41].

In applications one usually uses the reduced Tate-Lichtenbaum pairing for simplification reasons. The reduced Tate-Lichtenbaum pairing

$$t_n : E(\mathbb{F}_{q^k})[n] \times E(\mathbb{F}_{q^k})[n] \to \mu_n$$

is defined by

$$t_n(P, Q) = \left\langle P, Q + nE(\mathbb{F}_{q^k}) \right\rangle_n^{(q^k-1)/n} = \left(f_{n,P}(Q + R) \cdot f_{n,P}(R)^{-1}\right)^{(q^k-1)/n}.$$

The Tate-Lichtenbaum pairing and the reduced Tate-Lichtenbaum pairing are bilinear and non-degenerate [29, 41]. For an explicit example of a variant of the reduced Tate pairing see Example 1.

The Weil pairing

$$e_n : E(\mathbb{F}_{q^k})[n] \times E(\mathbb{F}_{q^k})[n] \to \mu_n$$

can be defined in a fashion similar to the Tate-Lichtenbaum pairing, but using four instead of two function evaluations [32], [65, Exerc. 3.16]. A shorter and computationally more efficient description is the following [53]: For $P, Q \in E(\mathbb{F}_{q^k})[n]$ choose $f_{n,P}$ and $f_{n,Q}$ with $(f_{n,P} \cdot f_{n,Q}^{-1})(\mathcal{O}) = 1$. Then

$$e_n(P, Q) = \begin{cases} 1 & \text{for } P = Q,\ P = \mathcal{O},\ Q = \mathcal{O}, \\ (-1)^n f_{n,P}(Q) \cdot f_{n,Q}(P)^{-1} & \text{else.} \end{cases}$$

The Weil pairing e_n as above is bilinear and non-degenerate since this holds true over the algebraic closure $\overline{\mathbb{F}}_q$ and here $E(\overline{\mathbb{F}}_q)[n] = E(\mathbb{F}_{q^k})[n]$. It is furthermore alternating, i. e. $e_n(P, P) = 1$, which is for example useful for tests whether a given point

Table 1 Comparison for bit lengths at roughly equal security

Symm	ECC	RSA	k
80	160	1024	6
128	256	3072	12
256	512	15360	30

lies in a given cyclic subgroup. The relation between the Weil and Tate-Lichtenbaum pairings is

$$e_n(P, Q)^{(q^k-1)/n} = t_n(P, Q) \cdot t_n(Q, P)^{-1}.$$

Types 1–3 The application of these pairings in cryptography requires the choice of suitable subgroups of $E(\mathbb{F}_{q^k})$ which is discussed now.

The Frobenius endomorphism π_q of E is given by $(x, y) \mapsto (x^q, y^q)$. It induces an automorphism of the two-dimensional $\mathbb{F}_n$-vector space $E(\mathbb{F}_{q^k})[n]$. Let $P \in E(\mathbb{F}_q)[n]$ be a point of order n. Since $\pi_q(P) = P$ we have that P is an eigenvector of π_q for the eigenvalue 1. The characteristic polynomial of π_q is of the form $x^2 - tx + q$ and has the roots 1 and q modulo n. Hence there is another point $Q \in E(\mathbb{F}_{q^k})$ of order n, such that Q is an eigenvector of π_q for the eigenvalue q, or equivalently such that $\pi_q(Q) = qQ$ holds. In summary we have $E(\mathbb{F}_{q^k})[n] = \langle P \rangle \times \langle Q \rangle$.

For these subgroups $t_n(P, P) = t_n(Q, Q) = 1$ and $t_n(P, Q) \neq 1$, similarly for the Weil pairing. The endomorphism $\mathrm{Tr} = c \sum_{i=0}^{k-1} \pi_q^i$ with $kc \equiv 1 \pmod{n}$ defines a surjective projection $\langle P \rangle \times \langle Q \rangle \to \langle P \rangle$ with kernel $\langle Q \rangle$, the trace zero subgroup.

A distortion map for $T = \lambda P + \mu Q \neq \mathcal{O}$ is an endomorphism ψ of E with $\psi(T) \notin \langle T \rangle$. If λ and μ are not zero, then Tr is a distortion map for T. It can be shown that there is a distortion map for $T = P$ and $T = Q$ if and only if E is supersingular [39, 70].

Using these mathematical objects the three types of pairings described at the end of Sect. 7.1 can be realized. Type 1 uses supersingular elliptic curves, a distortion map ψ and $Q = \psi(P)$. Thus $G_1 = G_2 = \langle P \rangle$ holds true. Type 2 uses ordinary elliptic curves, $G_1 = \langle P \rangle$ and $G_2 = \langle \lambda P + \mu Q \rangle$ with $\lambda, \mu \neq 0$. Here $G_1 \neq G_2$ and Tr yields a one-way isomorphism $G_2 \to G_1$. Type 3 uses ordinary elliptic curves with $G_1 = \langle P \rangle$ and $G_2 = \langle Q \rangle$. Thus $G_1 \neq G_2$ and there is (as far as one knows) no efficient computable isomorphism between G_1 and G_2. We refer to [19, 38] for more details.

Choice of Parameters The embedding degree is the parameter that is most important for security and efficiency. The complexities of the DLP in $E(\mathbb{F}_q)$ and $(\mathbb{F}_{q^k})^*$ are roughly $\exp(1/2 \log q)$ and $\exp((k \log q)^{1/3})$ respectively, the latter similar to (3.1). To balance these complexities for growing q we need to choose $k \approx (\log q)^{2/3}$. The embedding degree thus grows with q. Table 1 gives an overview over the bit-lengths of the keys for symmetric cryptosystems, the value $\log_2(q)$ for elliptic curves, the value $\log_2(N)$ for the RSA cryptosystem or the value $\log_2(q^k)$ for $\mathbb{F}_{q^k}^*$ respectively, and the corresponding embedding degree for rows of comparable security.

Construction Supersingular curves yield embedding degrees $k \in \{2, 3, 4, 6\}$ only [55]. In view of Table 1 larger embedding degrees k are highly desirable. This re-

quires the construction of suitable ordinary elliptic curves. The following conditions on $q, n, t = q + 1 - \#E(\mathbb{F}_q)$, and k, called MNT conditions, have to be observed (ϕ_k is the k-th cyclotomic polynomial):

1. $q + 1 - t = cn$.
2. $\phi_k(q) \equiv 0 \pmod{n}$.
3. q is a prime power, n is a prime, $|t| \leq 2\sqrt{q}$.
4. $4q - t^2 = Df^2$ with D small.
5. $\rho = \log(q)/\log(n) \approx 1$.

Condition 2 implies $n|(q^k - 1)$. Condition 3 is required to enable the efficient computation of the elliptic curve by the theory of complex multiplication. Condition 5 means that c from Condition 1 is about as small as possible.

Solutions to the conditions 1–5 for arbitrary k can be found rather easily by a search strategy of Cox and Pinch, if one allows $\rho \approx 2$, see [32]. But the resulting cryptosystems will be rather inefficient. On the other hand it can be shown that solutions with $\rho \approx 1$ are rare [69] and accordingly difficult to find.

Constructions of ordinary elliptic curves with $\rho = 1$ have been found for the embedding degrees $k \in \{3, 4, 6\}$, $k = 10$ and $k = 12$, see [8, 31, 54]. For constructions with $1 < \rho < 2$ see for example [9, 17, 20, 23, 37]. For given k solutions to the conditions 1–5 can often be given in parametrized form $q = q(z)$ and $n = n(z)$ for $z \in \mathbb{Z}$. The methodology for these constructions makes partial use of algebraic number theory and diophantine geometry.

As an example consider the particularly nice elliptic curves from [8] with embedding degree 12 and $\rho = 1$, which are very well suited for 128 bit security. Let

$$p(z) = 36z^4 + 36z^3 + 24z^2 + 6z + 1,$$
$$t(z) = 6z^2 + 1,$$
$$n(z) = p(z) + 1 - t(z).$$

Then $\phi_{12}(p(z)) \equiv 0 \pmod{n(z)}$ and $4p(z) - t(z)^2 = 3(6z^2 + 4z + 1)^2$. The construction of the corresponding elliptic curves is as follows:

Algorithm

1. Find $x \in \mathbb{Z}$ such that $p(x)$ and $n(x)$ are primes.
2. Choose an elliptic curve $E : y^2 = x^3 + b$ with $b \in \mathbb{F}_p$ uniformly at random.
3. If $\#E(\mathbb{F}_p) = n(x)$ then output E. Otherwise repeat from Step 2.

At least at first sight this is an amazingly simple construction: The computed curves E automatically satisfy the Conditions 1–5 for $k = 12$. Furthermore, the construction of E via the complex multiplication method is not necessary and the check in Step 3 will be successful after expected 6 choices of E, so that the construction is also very efficient. It is also possible to prove these statements formally. The explicit construction of such an elliptic curve is most conveniently done using a computer algebra system such as Magma [1].

Efficient Computation It is not possible to discuss all the details of the efficient computation of the Weil- and Tate-Lichtenbaum pairings here. Instead we wish to focus on some more conceptual aspects regarding the Tate-Lichtenbaum pairing.

The efficient computation of the Tate-Lichtenbaum pairing has been investigated in many publications, among the first [4, 6, 34]. The basic and essential building block, the Ate pairing, for the most efficient pairings today has been introduced in [45]: A particularly efficient recent family of such pairings for embedding degree 24, $\rho = 1.25$ and high security levels has been investigated in [20].

We consider the reduced Tate-Lichtenbaum pairing t_n, the generators P, and Q of the eigenspaces G_1 and G_2 as above. Then it can be shown that t_n restricted to $G_1 \times G_2$ or $G_2 \times G_1$ respectively can already be described by $t_n(P, Q) = f_{n,P}(Q)^{(q^k-1)/n}$ and $t_n(Q, P) = f_{n,Q}(P)^{(q^k-1)/n}$ respectively if $f_{n,P}$ and $f_{n,Q}$ are chosen among all possible scalar multiples in a fixed suitably normalized form. This means that only one function evaluation is necessary. Recall that $f_{n,P}$ and $f_{n,Q}$ have been defined above as very specific rational functions on E. An example is given below.

But it is in fact possible to give a further significant simplification if the pairing is restricted to $G_2 \times G_1$. This yields a new, bilinear and non-degenerate pairing, called Ate pairing in [45], with a significantly simplified defining rational function. Let $T = t - 1$, where $\#E(\mathbb{F}_q) = q + 1 - t$, and suppose $T^k \not\equiv 1 \pmod{n^2}$. The Ate pairing

$$\hat{t}_n : G_2 \times G_1 \to \mu_n$$

is defined by

$$\hat{t}_n(Q, P) = f_{T,Q}(P)^{(q^k-1)/n}.$$

The main point here is that the function $f_{T,Q}$ has degree about $|T|$, which roughly lies between $n^{1/\varphi(k)}$ and $q^{1/2}$. On the other hand, the degree of the functions $f_{n,P}$ and $f_{n,Q}$ of the Tate-Lichtenbaum pairing is about $n \approx q$. This implies a drastic improvement in terms of efficiency for the Ate pairing, in particular for low absolute values of T.

Example 1 Let $E : y^2 = x^3 + 4$ over $\mathbb{F}_q$ with $q = p = 41761713112311845269$, $n = 715827883$, $k = 31$ and $T = -2$. Then

$$\hat{t}_n : G_2 \times G_1 \to \mu_n,$$
$$(Q, P) \mapsto \left(y_P - \left(3x_Q^2/(2y_Q)\right)x_P - \left(-x_Q^3 + 8\right)/(2y_Q)\right)^{(q^k-1)/n}$$

defines a non-degenerate bilinear pairing, where $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$.

Combinations of t_n and $\hat{t}_n$ together with lattice techniques yield pairings with defining rational functions of degree always about equal to $n^{1/\varphi(k)}$, even if T has large absolute value. Pairing functions of such degrees appear to be best possible or at least close to best possible [43, 71].

Using the theory of twists the memory requirements for elements from G_2 can be decreased and efficiency be increased. An elliptic curve E' over $\mathbb{F}_q$ is called twist

of degree d of E, if there is an isomorphism $\psi : E' \to E$, defined over $\mathbb{F}_{q^d}$ with d minimal. If E is ordinary, $k = ed$, and if E has a twist over $\mathbb{F}_{q^e}$ of degree $d > 1$, then E' and ψ can be chosen such that $E'(\mathbb{F}_{q^e})[n] = \langle \psi^{-1}(Q) \rangle$ holds. We let $Q' = \psi^{-1}(Q)$, $G_2' = \langle Q' \rangle$ and obtain the modified Ate pairing

$$\hat{t}_n' : G_2' \times G_1 \to \mu_n$$

via

$$\hat{t}_n'\big(Q', P\big) = \hat{t}_n\big(\psi\big(Q'\big), P\big).$$

In the case of the elliptic curves discussed above with $k = 12$ from [8] we get the following: We have $E : y^2 = x^3 + b$ with $b \in \mathbb{F}_p$ and $p \equiv 1 \pmod 6$. Let $\lambda \in \mathbb{F}_{p^2} \backslash (\mathbb{F}_{p^2})^3$ and $\mu \in \mathbb{F}_{p^2} \backslash (\mathbb{F}_{p^2})^2$. The curve $E' : \mu y^2 = \lambda x^3 + b$ is a twist of E of degree 6 and $\psi : E' \to E$, $\psi(x, y) = (\lambda^{1/3} x, \mu^{1/2} y)$ is the corresponding isomorphism. In the example the degree of $f_{T,Q}$ is about $n^{1/2}$, and using Q' and E' over $\mathbb{F}_{p^2}$ instead of Q and E over $\mathbb{F}_{p^{12}}$ yields an improvement by a factor of 6 in terms of bandwidth.

Finally, it is instructive to interpret the Ate pairing in a broader mathematical context from number theory. It is well known that the Tate-Lichtenbaum pairing gives an algebraic description of the Artin symbol on unramified abelian extensions of exponent n of a global function field containing the n-th roots of unity. As it turns out, the Ate pairing gives an algebraic description of the Artin symbol in the general case where the global function field is not required to contain the n-th roots of unity.

8 Side Channel Attacks

In recent years side channel analysis (SCA) of physical devices implementing asymmetric and symmetric cryptographic algorithms and protocols and operating on secret data have become a very active area of research, both mathematically and technologically. SCA uses passive and active attacks by exploiting the intended interface of a cryptographic device. This section contains a very important example for actual realization problems which is contrast to the mere theoretical descriptions of the protocols in Sects. 2, 3, and 5.

Side channel cryptanalysis uses physical observables resulting from internal states and processes of a cryptographic computation as additional source for cryptanalysis. The outcome of the measurement of physical observables are real-valued vectors.

Internal state changes of the cryptographic device including the state change caused by operations with secret or ephemeral keys cause instantaneous leakage, that can be exploited. Examples of information sources are: Varying execution times of operations, varying power consumption during operation, varying electro-magnetic emanation during operation, enforced unexpected behaviour as a consequence of induced transient or permanent device faults, enforced error messages of a cryptographic product, or photon emissions.

The exploitation of the above mentioned information sources leads to interesting new mathematical questions and results. Many of these can be formulated and solved

as lattice problems in number theory. The use of stochastic modeling in the interpretation of measurements can not be underestimated. Countermeasures include avoiding key dependent power profiles and timing behaviour by uniformizing and randomizing computations. A full overview is given in [49].

Example 2 (The Nguyen-Shparlinsky Attack) We refer to [16, 58] for further details. Suppose that during the generation of ECDSA signatures some bits of the ephemeral key are leaked for many signatures. Then the secret user key can easily be recovered. The Nguyen-Shparlinski attack works when three bits are leaked for each of about a hundred signatures and consists of a reduction of the ECDSA-problem to the so called Hidden Number Problem (HNP) introduced by Boneh and Venkatesan. In this case the recovery of the secret key can be accomplished via lattice methods.

Remark Note that in real world implementations modified versions of the ECDSA scheme are being used. We mention three main differences:

- Depending on the bitlength of the chosen curve hash values are truncated.
- Smart cards may be fed with hash values by an external source or only perform some rounds of the hash calculation internally.
- Blinding measures may lead to the use of modified ephemeral keys which are longer than the original ephemeral keys. Typically k_e is replaced by $k_e + \lambda \cdot n$, where n is the order of the cyclic subgroup under consideration and λ is a small random number. This increases the attack complexity but does not prevent the lattice attack.

This implies that security proofs for the pure scheme may no longer hold for real world implementations.

Acknowledgements The authors wish to thank several colleagues for carefully proofreading several versions of this paper. We also wish to thank Gabriele Nebe for her patience with this project and for making valuable suggestions.

References

1. Bosma, W., Cannon, J., Playoust, C.: The Magma algebra system I: The user language. J. Symbolic Comp. **24**(3/4), 235–265 (1997)
2. Boneh, D., Franklin, M.: Identity-based encryption from the Weil pairing. In: Kilian, J. (ed.) Advances in Cryptology (CRYPTO 2001). Lecture Notes in Computer Science, vol. 2139, pp. 213–229. Springer Berlin (2001)
3. Bender, J., Fischlin, M., Kügler, D.: Security analysis of the PACE key-agreement protocol. In: Proceedings of the 12th International Conference on Information Security, Pisa, Italy, 7–9 September, 2009. Lecture Notes in Computer Science, vol. 5735, pp. 33–48. Springer Berlin (2009)
4. Barreto, P., Galbraith, S., O'hEigeartaigh, C., Scott, M.: Efficient pairing computation on supersingular abelian varieties. Designs, Codes and Cryptography **42**(3), 239–271 (2007)
5. Boneh, D., Joux, A., Nguyen, P.Q.: Why textbook ElGamal and RSA encryption are insecure. In: Proceedings of the 6th International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT'00), pp. 30–43. Springer Berlin (2000)
6. Barreto, P., Kim, H., Lynn, B., Scott, M.: Efficient algorithms for pairing-based cryptosystems. In: Yung, M. (ed.) Advances in Cryptology (CRYPTO 2002), Santa Barbara. Lecture Notes in Computer Science, vol. 2442, pp. 354–369. Springer, Berlin (2002)

7. Boneh, D., Lynn, B., Shacham, H.: Short signatures from the Weil pairing. In: Boyd, C. (ed.) Advances in Cryptology (ASIACRYPT 2001), Gold Coast, Australia. Lecture Notes in Computer Science, vol. 2248, pp. 514–532. Springer, Berlin (2001)
8. Barreto, P., Naehrig, M.: Pairing-friendly elliptic curves of prime order. In: Preneel, B., Tavares, S. (eds.) Selected Areas in Cryptography (SAC 2005), Kingston, ON, Canada. Lecture Notes in Computer Science, vol. 3897, pp. 319–331. Springer, Berlin (2006)
9. Bisson, G., Satoh, T.: More discriminants with the Brezing-Weng method. In: Chowdhury, D.R., Rijmen, V., Das, A. (eds.) Progress in Cryptology (INDOCRYPT 2009), Kharagpur, India. Lecture Notes in Computer Science, vol. 5365, pp. 389–399. Springer, Berlin (2008)
10. BSI: Elliptic Curve Cryptography. Technical guideline TR-03111, Version 1.11 (2009)
11. BSI: Advanced security mechanisms for machine readable travel documents—extended access control (EAC), password authenticated connection establishment (PACE), and restricted identification (RI). Technical guideline TR-03110 (2010)
12. Blake, I., Seroussi, G., Smart, N.: Elliptic Curves in Cryptography. London Mathematical Society, vol. 265. Cambridge University Press, Cambridge (2000)
13. Blake, I., Seroussi, G., Smart, N. (eds.): Advances in Elliptic Curve Cryptography. Cambridge University Press, Cambridge (2005)
14. Blake, I., Seroussi, G., Smart, N., Cassels, J. W. S. (eds.): Advances in Elliptic Curve Cryptography. London Mathematical Society Lecture Note Series, vol. 317. Cambridge University Press, Cambridge (2005)
15. Buchmann, J.: Introduction to Cryptography, 2 edn. Springer, Berlin (2004)
16. Boneh, D., Venkatesan, R.: Hardness of computing the most significant bits of secret keys in Diffie-Hellman and related schemes. In: CRYPTO. Lecture Notes in Computer Science, vol. 1109, pp. 129–142. Springer, Berlin (1996)
17. Brezing, F., Weng, A.: Elliptic curves suitable for pairing based cryptography. Designs, Codes and Cryptography **37**, 133–141 (2005)
18. Cohen, H., Frey, G., Avanzi, R., Doche, C., Lange, T., Nguyen, K., Vercauteren, F. (eds.): Handbook of Elliptic and Hyperelliptic Curve Cryptography. Discrete Mathematics and Its Applications, vol. 34. Chapman & Hall/CRC, London (2005)
19. Chatterjee, S., Hankerson, D., Menezes, A.: On the efficiency and security of pairing-based protocols in the type 1 and type 4 settings. In: Hasan, M., Helleseth, T. (eds.) Arithmetic of Finite Fields, Istanbul, Turkey. Lecture Notes in Computer Science, vol. 6087, pp. 114–134. Springer, Berlin (2010)
20. Costello, C., Lauter, K., Naehrig, M.: Attractive subfamilies of BLS curves for implementing high-security pairings. In: Bernstein, D., Chatterjee, S. (eds.) Progress in Cryptology (INDOCRYPT 2011), Chennai, India. Lecture Notes in Computer Science, vol. 7107, pp. 320–342. Springer, Berlin (2011)
21. Cohen, H.: A Course in Computational Algebraic Number Theory. Springer, Berlin (1993)
22. Crandall, R., Pomerance, C.: Prime Numbers: A Computational Perspective, 2nd edn. Springer, Berlin (2005)
23. Duan, P., Cui, S., Chan, C.: Special polynomial families for generating more suitable elliptic curves for pairing-based cryptosystems. In: Proceedings of the 5th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications EHAC'06 (2006)
24. Diffie, W., Hellman, M.E.: New directions in cryptography. IEEE Trans. Inf. Theory **22**, 644–654 (1976)
25. Diem, C.: On the discrete logarithm problem in elliptic curves. Compositio Mathematica **147**(01), 75–104 (2011)
26. ECC Brainpool: ECC brainpool standard curves and curve generation. Internet Draft, http://www.ecc-brainpool.org/download/Domain-parameters.pdf (October 2005)
27. ElGamal, T.: A public-key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans. Inf. Theory **IT-31**, 469 (1985)
28. Fumy, W., Paeschke, M. (eds.): Handbook of eID Security. Publicis Publishing, Erlangen (2011)
29. Frey, G., Rück, H.-G.: A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves. Math. Comp. **62**, 865–874 (1994)
30. Frey, G.: The arithmetic behind cryptography. Notices Am. Math. Soc. **57**(3), 366–374 (2010)
31. Freeman, D., Scott, M., Teske, E.: A taxonomy of pairing-friendly elliptic curves. J. Cryptology **23**(2), 224–280 (2010)
32. Galbraith, S.: Pairings (book chapter). In Blake et al. [13]
33. Gaudry, P.: Index calculus for abelian varieties of small dimension and the elliptic curve discrete logarithm problem. J. Symb. Comput. **44**, 1690–1702 (2009)

34. Galbraith, S., Harrison, K., Soldera, S.: Implementing the Tate pairing. In: Fieker, C., Kohel, D.R. (eds.) Proceedings of the Fifth Symposium on Algorithmic Number Theory (ANTS-V), Sydney, Australia. Lecture Notes in Computer Science, vol. 2369, pp. 324–337. Springer, Berlin (2002)
35. Gaudry, P., Hess, F., Smart, N.P.: Constructive and destructive facets of Weil descent on elliptic curves. J. Cryptology **15**(1), 19–46 (2002)
36. Galbraith, S., Menezes, A.: Algebraic curves and cryptography. Finite Fields and Applications **11**(3), 544–577 (2005)
37. Galbraith, S.D., McKee, J.F., Valenca, P.C.: Ordinary abelian varieties having small embedding degree. Finite Fields Appl. **13**(4), 800–814 (2007)
38. Galbraith, S.D., Paterson, K.G., Smart, N.P.: Pairings for cryptographers. Discrete Appl. Math. **156**(16), 3113–3121 (2008)
39. Galbraith, S., Rotger, V.: Easy decision Diffie-Hellman groups. LMS J. Comput. Math. **7**, 201–218 (2004)
40. Galbraith, S., Smart, N.P.: A cryptographic application of Weil descent. In: Walker, M. (ed.) Cryptography and Coding, Cirencester. Lecture Notes in Computer Science, vol. 1746, pp. 191–200. Springer, Berlin (1999)
41. Hess, F.: A note on the Tate pairing of curves over finite fields. Arch. Math. **82**, 28–32 (2004)
42. Hess, F.: Weil descent attacks. In: Blake et al. [14]
43. Hess, F.: Pairing lattices. In: Galbraith, S., Paterson, K. (eds.) Progress in Cryptology (INDOCRYPT 2009), Egham, UK. Lecture Notes in Computer Science, vol. 5208, pp. 18–38. Springer, Berlin (2008)
44. Hankerson, D., Menezes, A., Vanstone, S.: Guide to Elliptic Curve Cryptography. Springer Professional Computing (2004)
45. Hess, F., Smart, N., Vercauteren, F.: The Eta pairing revisited. IEEE Transactions on Information Theory **52**(10), 4595–4602 (2006)
46. Jacobson, M.J. Jr., Koblitz, N., Silverman, J.H., Stein, A., Teske, E.: Analysis of the xedni calculus attack. Designs, Codes and Cryptography **20**(1), 41–64 (2000)
47. Jacobson, M.J. Jr., Menezes, A.J., Stein, A.: Hyperelliptic curves and cryptography. In: High Primes and Misdemeanours: Lectures in Honour of the 60th Birthday of Hugh Cowie Williams. Fields Institute Communications Series, vol. 41, pp. 255–282. Am. Math. Soc., Providence (2004)
48. Joux, A.: A one round protocol for tripartite Diffie–Hellman. In: Bosma, W. (ed.) Proceedings of the Fourth Symposium on Algorithmic Number Theory (ANTS-IV), Leiden, Netherlands. Lecture Notes in Computer Science, vol. 1838, pp. 385–394. Springer, Berlin (2000)
49. Killmann, W., Lange, T., Lochter, M., Thumser, W., Wicke, G.: Minimum requirements for evaluating side-channel attack resistance of elliptic curve implementations. Downloadable via http://www.bsi.bund.de (2011)
50. Koblitz, N.: Elliptic curve cryptosystems. Math. Comput. **48**, 203–209 (1987)
51. Lochter, M., Merkle, J.: Elliptic curve cryptography (ecc) brainpool standard curves and curve generation. IETF internet draft, RFC 5639 (March 2010)
52. Miller, V.: Use of elliptic curves in cryptography. In: Advances in Cryptology (CRYPTO'85). Lecture Notes in Computer Science, vol. 218, pp. 417–426. Springer, Berlin (1986)
53. Miller, V.: The Weil pairing, and its efficient calculation. J. Cryptology **17**, 235–261 (2004)
54. Miyaji, A., Nakabayashi, M., Takano, S.: New explicit conditions of elliptic curve traces for FR-reduction. IEICE Trans. Fundamentals E **84**, 1234–1243 (2001)
55. Menezes, A.J., Okamoto, T., Vanstone, S.A.: Reducing elliptic curve logarithms to a finite field. IEEE Trans. on Inform. Theory **39**, 1639–1646 (1993)
56. NIST: Digital signature standard. FIPS publication 186-3 (2009)
57. NIST: Recommendation for key derivation through extraction-then-expansion. NIST special publication 800-56C (November 2011)
58. Nguyen, P.Q., Shparlinski, I.E.: The insecurity of the elliptic curve digital signature algorithm with partially known nonces. Des. Codes Cryptography **30**(2), 201–217 (2003)
59. Paterson, K.: Cryptography from pairings (book chapter). In: Blake et al. [13]
60. Paterson, K.: Identity-based cryptography—panacea or pandemonium? Invited talk at 9th Workshop on Elliptic Curve Cryptography (ECC 2005). Available under http://www.cacr.math.uwaterloo.ca/conferences/2005/ecc2005/paterson.pdf, 2005
61. Rivest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**, 120–126 (1978)
62. Shamir, A.: Identity based cryptosystems and signature schemes. In: Blakley, G.R., Chaum, D. (eds.) Advances in Cryptology (CRYPTO 1984), Santa Barbara. Lecture Notes in Computer Science, vol. 196, pp. 47–53. Springer, Berlin (1985)

63. Silverman, J.H.: Advanced Topics in the Arithmetic of Elliptic Curves. Graduate Texts in Mathematics, vol. 151. Springer, Berlin (1994)
64. Silverman, J.H.: The xedni calculus and the elliptic curve discrete logarithm problem. Des. Codes Cryptography **20**, 5–40 (2000)
65. Silverman, J.H.: The Arithmetic of Elliptic Curves, 2nd edn. Graduate Texts in Mathematics, vol. 106. Springer Berlin (2009)
66. Sakai, R., Ohgishi, K., Kasahara, M.: Cryptosystems based on pairing. In: Symposium on Cryptography and Information Security (SCIS2000), Okinawa (2000)
67. Stinson, D.R.: Cryptography: Theory and Practice, 3rd edn. Chapman & Hall/CRC, London (2005)
68. Stichtenoth, H.: Algebraic Function Fields and Codes, 2 edn. Springer, Berlin (2008)
69. Urroz, J., Luca, F., Shparlinski, I.: On the number of isogeny classes of pairing-friendly elliptic curves and statistics of mnt curves. Math. Comput. **81**, 1093–1110 (2012)
70. Verheul, E.: Evidence that XTR is more secure than supersingular elliptic curve cryptosystems. J. Cryptology **17**, 277–296 (2004)
71. Vercauteren, F.: Optimal pairings. IEEE Trans. Inf. Theory **56**(1), 455–461 (2010)
72. Velichka, M.D., Jacobson, M.J. Jr., Stein, A.: Computing discrete logarithms in the jacobian of high-genus hyperelliptic curves over even characteristic finite fields. IACR Cryptol. ePrint Arch. **2011**, 98 (2011)
73. Washington, L.C.: Elliptic Curves. Number Theory and Cryptography, 2nd edn. Chapman & Hall/CRC, Boca Raton (2008). xviii, 513 p.

Florian Heß is a full professor of mathematics at the University of Oldenburg. He studied mathematics and computer science at the Technical University of Berlin where he received his Ph.D. in 1999. During 1999–2001 and 2001–2003 he worked as a visiting scholar in the Magma group of John Cannon at the University of Sydney and as a research associate in the cryptology group of Nigel Smart at the University of Bristol respectively. In 2003 Heß became an assistant professor at the Technical University of Berlin and in 2009 an associate professor at the University of Magdeburg. He joined the faculty at Oldenburg in 2010. Heß serves in the steering committees of the Fachgruppe Kryptologie of the Gesellschaft für Informatik (GI) and the Fachgruppe Computeralgebra of the DMV, GI and GAMM. His research interests are in computational mathematics with a particular emphasis on cryptography, number theory and algebraic geometry.

Andreas Stein is a full professor of mathematics at the University of Oldenburg where he joined the faculty in 2008. He studied mathematics and computer science at the University of Saarland where he received his Ph.D. in 1997 as a Siemens scholar. As a postdoctoral fellow he worked from 1997 to 2000 in Canada at the Universities in Winnipeg and Waterloo. Before he came to Oldenburg in 2008 he was from 2000 to 2004 a tenure track assistant professor at the University of Illinois at Urbana-Champaign, USA, and an associate professor at the University of Wyoming, USA. In addition to cryptography his research interests are in number theory, computational arithmetic geometry, and computer algebra.

Sandra Stein is a research assistent as well as a Ph.D. student at the Carl von Ossietzky University Oldenburg. In 2009 she received her diploma in mathematics and physics at the University of Bayreuth. Her research interests are in number theory, cryptography, and coding theory.

Manfred Lochter studied mathematics in Cologne and Saarbrücken. In 1992 he received his Ph.D. from the University of Cologne. In his thesis he investigated connections between prime-decomposition, group theory and representation theory. Since 1994 he works for the German Federal Office for Information Security (BSI), section *Fundamentals of Cryptography*. His interests include number theory, elliptic curve cryptography, factoring, side-channel analysis and the implementation of cryptographic mechanisms in hardware and software.

Jahresber Dtsch Math-Ver (2012) 114:89–106
DOI 10.1365/s13291-012-0039-x

SURVEY ARTICLE

Graham Higman's PORC Conjecture

Michael Vaughan-Lee

Received: 14 October 2011 / Published online: 13 April 2012

Abstract We survey the history of Graham Higman's PORC conjecture concerning the form of the function $f(p^n)$ enumerating the number of groups of order p^n. The conjecture is that for a fixed n there is a finite set of polynomials in p, $g_1(p)$, $g_2(p)$, $\ldots$, $g_k(p)$, and a positive integer N, such that for each prime p, $f(p^n) = g_i(p)$ for some i $(1 \leq i \leq k)$ with the choice of i depending on the residue class of p modulo N. We describe some properties of a group recently discovered by Marcus du Sautoy which has major implications for the PORC conjecture.

Keywords p-Groups · PORC

Mathematics Subject Classification 20D15

1 Introduction

Mathematicians love to count things. How many possibilities are there for a Sudoku solution grid? How many Latin squares of order 4 are there? How many groups are there of order 8? Answers: 6,670,903,752,021,072,936,960 and 576 and 5. I found the answers to the first two questions in Wikipedia. The answer to the third question has been known to group theorists for well over 100 years. For a short modern analysis of the groups of order 8 see Sect. 4.4 of Hall [8]. Often the answers to this sort of question involve a classification of all the possibilities, and perhaps a complete list of the possibilities. As mentioned above, the five groups of order 8 have been well understood for well over 100 years, but it might be a bit tricky to produce a list of all the Sudoku solution grids. It would be perfectly possible (though tedious!) to draw up a complete list of the 576 Latin squares of order 4 by hand. However, a little thought

M. Vaughan-Lee (✉)
Christ Church, Oxford, OX1 1DP, UK
e-mail: michael.vaughan-lee@chch.ox.ac.uk

enables you to see that there are 576 Latin squares of order 4 without actually listing them all. Let L be a Latin square of order n, and assume that the entries in the cells of the square are integers in the range $\{1, 2, \ldots, n\}$. We say that L is reduced if the entries in the first row and first column are in their natural order $1, 2, \ldots, n$. It is easy to see that the total number of Latin squares of order n is $n!(n-1)!$ times the number of reduced Latin squares of order n. And it is easy to see that there are exactly 4 reduced Latin squares of order 4:

$$\begin{bmatrix} 1&2&3&4\\ 2&1&4&3\\ 3&4&1&2\\ 4&3&2&1 \end{bmatrix}, \begin{bmatrix} 1&2&3&4\\ 2&3&4&1\\ 3&4&1&2\\ 4&1&2&3 \end{bmatrix}, \begin{bmatrix} 1&2&3&4\\ 2&4&1&3\\ 3&1&4&2\\ 4&3&2&1 \end{bmatrix}, \begin{bmatrix} 1&2&3&4\\ 2&1&4&3\\ 3&4&2&1\\ 4&3&1&2 \end{bmatrix}.$$

It immediately follows that there are 576 Latin squares in all. If we also allow ourselves to permute the names of the symbols then we are left with just two Latin squares of order 4:

$$\begin{bmatrix} 1&2&3&4\\ 2&1&4&3\\ 3&4&1&2\\ 4&3&2&1 \end{bmatrix}, \begin{bmatrix} 1&2&3&4\\ 2&3&4&1\\ 3&4&1&2\\ 4&1&2&3 \end{bmatrix}.$$

These two Latin squares give the group multiplication tables for the two groups of order 4, the Klein four-group and the cyclic group of order 4.

$\times$	1	a	b	ab
1	1	a	b	ab
a	a	1	ab	b
b	b	ab	1	a
ab	ab	b	a	1

$\times$	1	a	a^2	a^3
1	1	a	a^2	a^3
a	a	a^2	a^3	1
a^2	a^2	a^3	1	a
a^3	a^3	1	a	a^2

The total number of Latin squares of order n is bounded by n^{n^2}. (There are n^2 cells, and n choices for the entry in each cell.) You can do a little better than this if you note that there are n rows, and $n!$ possibilities for the entries in any given row, so that the total number of Latin squares is bounded by $(n!)^n$. The best bounds do not seem to do a lot better than this. J.H. van Lint and R.M. Wilson [20] show that if $L(n)$ is the total number of Latin squares of order n then

$$\frac{(n!)^{2n}}{n^{n^2}} \leq L(n) \leq \prod_{k=1}^{n} (k!)^{\frac{n}{k}}.$$

There are much tighter bounds for the number $f(n)$ of groups of order n. Pyber [17] has shown that

$$f(n) \leq n^{\frac{2}{27}\mu(n)^2 + O(\mu(n)^{3/2})},$$

where $\mu(n)$ denotes the highest power to which any prime divides n. Note that there is no lower bound on $f(n)$ in this theorem, because the value of $f(n)$ is heavily

dependent on the factorization of n into a product of primes. In particular, if n is prime then $f(n) = 1$. One of the main components in the proof of this result is a bound on the number of groups of prime-power order p^n given by Higman [9], improved by Sims [19], and further improved in unpublished work by Mike Newman and Craig Seeley.

$$p^{\frac{2}{27}n^3-6n^2} \leq f\left(p^n\right) \leq p^{\frac{2}{27}n^3+O(n^{5/2})}.$$

The book *Enumeration of finite groups* by Blackburn, Neumann and Venkatamaran [2] gives a good account of the history of this problem.

Graham Higman's PORC conjecture is a conjecture about the precise form of the function $f(p^n)$, and we will return to the problem of enumerating groups of order p^n later. Meanwhile, as what might seem like a diversion, we will consider the problem of enumerating the algebras of dimension n over a field F.

2 Enumerating Algebras of Dimension *n*

By an algebra over a field F we mean a vector space A over F together with a product: for each pair of elements $a, b \in A$ there is a uniquely defined product $ab \in A$. The product is required to be bilinear, so that if $a, b, c \in A$ and $\lambda, \mu \in F$ then

$$(\lambda a + \mu b)c = \lambda(ac) + \mu(bc),$$
$$c(\lambda a + \mu b) = \lambda(ca) + \mu(cb).$$

We do not require the product to satisfy any other conditions such as commutativity or associativity. If A is an algebra over F, and if we pick a basis $\{a_i \mid i \in I\}$ for A as a vector space over F then for each pair of basis elements a_i, a_j we can express the product $a_i a_j$ as a linear combination

$$a_i a_j = \sum_{k \in I} \lambda_{ijk} a_k$$

for some scalars $\lambda_{ijk} \in F$. These scalars are called *structure constants* for the algebra A. These structure constants completely determine the product on A since if $a = \sum_{i \in I} \alpha_i a_i$ and $b = \sum_{j \in I} \beta_j a_j$ are any two elements of A then using bilinearity we see that

$$ab = \sum_{i,j,k \in I} \alpha_i \beta_j \lambda_{ijk} a_k.$$

Note however that if we pick a different vector space basis for A then we may get a different set of structure constants, so that different sets of structure constants can give the same algebra A. We will return to this point shortly.

If F is an infinite field then there are infinitely many choices for sets of structure constants. But there is a unique finite field $\mathbb{F}_q$ of order q for every prime-power q, and if A is an algebra of dimension n over $\mathbb{F}_q$ then there are exactly q^{n^3} possible sets of structure constants $\{\lambda_{ijk} \mid 1 \leq i, j, k \leq n\}$ for A. So there is an upper bound of

q^{n^3} for the number $g(n,q)$ of n-dimensional algebras over $\mathbb{F}_q$. (But remember that different sets of structure constants can give the same algebra, so $g(n,q)$ is less than this upper bound.) This means that, for fixed n, $g(n,q)$ is bounded by a polynomial in q. Graham Higman [10] proved a much stronger result than this. He showed that, for fixed n, $g(n,q)$ is Polynomial On Residue Classes—PORC. This means that there is a finite set of polynomials in q, $g_1(q)$, $g_2(q)$, …, $g_k(q)$, and a positive integer N, such that for any prime-power q

$$g(n,q) = g_i(q)$$

for some i $(1 \le i \le k)$, with the choice of i depending on the residue class of q modulo N. For example, if $n = 2$ then we have three polynomials. If q is a power of 2 then $g(2,q) = q^4 + q^3 + 4q^2 + 3q + 6$, if q is a power of 3 then $g(2,q) = q^4 + q^3 + 4q^2 + 4q + 6$, and if q is a power of p with $p > 3$ then $g(2,q) = q^4 + q^3 + 4q^2 + 4q + 7$. So we can take $N = 6$, and the choice of polynomial depends on the residue class of q modulo 6. When $n = 3$ there are 22 polynomials of degree 18, with the choice of polynomial depending on the residue class of q modulo $4 \times 3 \times 5 \times 7$.

Higman's proof of this result is far too long and difficult to give here, but we can illustrate many of the key ideas in Higman's proof by looking at how the polynomials above for $n = 2$ can be obtained.

First we investigate how a change of basis affects the structure constants. We might as well do this for general n. So let A be an algebra of dimension n over a field F and let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ be two bases for A as a vector space over F. Let the sets of structure constants for these two bases be $\{\lambda_{ijk} \mid 1 \le i, j, k \le n\}$ and $\{\mu_{ijk} \mid 1 \le i, j, k \le n\}$. We can express the elements of the second basis as linear combinations of elements of the first basis, and vice versa:

$$b_i = \sum_{j=1}^{n} \alpha_{ji} a_j \quad (1 \le i \le n),$$

$$a_j = \sum_{k=1}^{n} \beta_{kj} b_k \quad (1 \le j \le n),$$

where $[\alpha_{ji}]$ and $[\beta_{kj}]$ are $n \times n$ matrices over F which are inverse to each other. So

$$\begin{aligned} b_i b_j &= \sum_{r,s=1}^{n} \alpha_{ri} \alpha_{sj} a_r a_s \\ &= \sum_{r,s,t=1}^{n} \alpha_{ri} \alpha_{sj} \lambda_{rst} a_t \\ &= \sum_{r,s,t,k=1}^{n} \alpha_{ri} \alpha_{sj} \lambda_{rst} \beta_{kt} b_k. \end{aligned}$$

It follows that

$$\mu_{ijk} = \sum_{r,s,t=1}^{n} \alpha_{ri}\alpha_{sj}\lambda_{rst}\beta_{kt}.$$

It is time to simplify the notation a bit! Each set of structure constants consists of n^3 elements of F, and we can think of these sets of structure constants as elements in an n^3-dimensional vector space V over F. The set of all non-singular $n \times n$ matrices over F form a group $\mathrm{GL}(n, F)$, the general linear group of degree n over F. The formula above defines an *action* of $\mathrm{GL}(n, F)$ on V. If $v = \{\lambda_{ijk} \mid 1 \le i, j, k \le n\} \in V$ and $g = [\alpha_{ji}] \in \mathrm{GL}(n, F)$ then we set $vg = \{\mu_{ijk} \mid 1 \le i, j, k \le n\}$, where μ_{ijk} is given by the formula above. (The formula also involves the matrix $[\beta_{kj}]$, but this matrix is the inverse of $[\alpha_{ji}]$ and so depends only on g.) This action of $\mathrm{GL}(n, F)$ on V satisfies three key properties.

1. If $u, v \in V$, $\alpha, \beta \in F$, and $g \in \mathrm{GL}(n, F)$ then $(\alpha u + \beta v)g = \alpha(ug) + \beta(vg)$.
2. If $v \in V$ and if I is the identity matrix in $\mathrm{GL}(n, F)$ then $vI = v$.
3. If $v \in V$ and $g, h \in \mathrm{GL}(n, F)$ then $v(gh) = (vg)h$.

There is also a fourth property which is critical for Higman's argument. This property is that the action of $[\alpha_{ji}]$ on V is given by a matrix in $\mathrm{GL}(n^3, F)$ whose entries are rational functions in the entries α_{ji}.

The three properties given above are easy to check, but I think we have had enough matrix algebra for now! Two elements $u, v \in V$ (i.e. two sets of structure constants) define the same algebra if and only if $u = vg$ for some $g \in \mathrm{GL}(n, F)$. In the jargon of *groups acting on sets* we say that two elements $u, v \in V$ define the same algebra if and only if they lie in the same orbit under the action of $\mathrm{GL}(n, F)$ on V. It is easy to see that "being in the same orbit" is an equivalence relation on V, so that the orbits partition V. The number of algebras of dimension n over F is the number of orbits in V under the action of $\mathrm{GL}(n, F)$.

If we take F to be the field $\mathbb{F}_q$ of q elements, then the number of orbits is $g(n, q)$ and Higman proves that this number is PORC. Actually, Higman proves a much more general result than this, but we can illustrate many of the key ideas that appear in Higman's proof by showing how to compute $g(2, q)$.

So let $n = 2$, and consider the action of $\mathrm{GL}(2, \mathbb{F}_q)$ on V described above. Note that when $n = 2$ then V has dimension 8. The number of orbits of $\mathrm{GL}(2, \mathbb{F}_q)$ on V can be computed using a result which is often called Burnside's Lemma [3], though some people think this is a misnomer. If $g \in \mathrm{GL}(2, \mathbb{F}_q)$ then we define $\mathrm{fix}(g) = \{v \in V \mid vg = v\}$. The number of orbits is then given by the formula

$$\frac{1}{|\mathrm{GL}(2, \mathbb{F}_q)|}\left(\sum_{g \in \mathrm{GL}(2, \mathbb{F}_q)} |\mathrm{fix}(g)|\right).$$

It follows from properties (2) and (3) above that if g and h are conjugate elements of $\mathrm{GL}(2, \mathbb{F}_q)$ (i.e. if $h = x^{-1}gx$ for some $x \in \mathrm{GL}(2, \mathbb{F}_q)$) then $|\mathrm{fix}(g)| = |\mathrm{fix}(h)|$. Two elements in $\mathrm{GL}(n, F)$ are conjugate if and only if they have the same rational canonical form. However we do not have to concern ourselves with the rational canonical form in the case $n = 2$, since two elements in $\mathrm{GL}(2, \mathbb{F}_q)$ are conjugate if and only

if they have the same minimal polynomial. The minimal polynomial of a matrix in $\mathrm{GL}(2, \mathbb{F}_q)$ will have degree one or two. The roots of the minimal polynomial of an element $g \in \mathrm{GL}(2, \mathbb{F}_q)$ are the eigenvalues of g. These may not lie in $\mathbb{F}_q$, but they will lie in some extension field of $\mathbb{F}_q$, so allowing roots in an extension field we have three types of minimal polynomial that can arise:

$$(x-\lambda),\ (x-\lambda)^2,\ (x-\lambda)(x-\mu)$$

with $\lambda, \mu \neq 0$, $\lambda \neq \mu$. The first case corresponds to $g = \left[\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda \end{smallmatrix}\right]$. In the second case g has a repeated eigenvalue, but is not diagonalizable—in this case g is conjugate to $\left[\begin{smallmatrix} \lambda & 1 \\ 0 & \lambda \end{smallmatrix}\right]$. In the third case g has two distinct eigenvalues which may or may not lie in $\mathbb{F}_q$.

For each $g \in \mathrm{GL}(2, \mathbb{F}_q)$ we can compute the matrix $A(g)$ giving the action of g on V. Property (1) above implies that $\mathrm{fix}(g)$ is a subspace of V, and so $|\mathrm{fix}(g)| = q^k$ where k is the dimension of $\mathrm{fix}(g)$. This dimension k is the dimension of the eigenspace of $A(g)$ corresponding to eigenvalue 1.

If $g = \left[\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda \end{smallmatrix}\right]$ then $A(g)$ equals

$$\begin{bmatrix} \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix},$$

so $\mathrm{fix}(g)$ has dimension 0 unless $\lambda = 1$, in which case it has dimension 8.

If g in conjugate to $\left[\begin{smallmatrix} \lambda & 1 \\ 0 & \lambda \end{smallmatrix}\right]$ then $A(g)$ is conjugate to

$$\begin{bmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix},$$

$$\text{or} \quad \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix},$$

depending on whether q is a power of 2, a power of 3, or a power of p for a prime $p > 3$.

Finally if g has distinct eigenvalues λ, μ then $A(g)$ has eigenvalues

$$\lambda, \lambda, \lambda, \mu, \mu, \mu, \lambda^2\mu^{-1}, \lambda^{-1}\mu^2.$$

In this case, if $\lambda, \mu \in \mathbb{F}_q$ then $A(g)$ is conjugate to a diagonal matrix with these eigenvalues along the diagonal. If $\lambda, \mu \notin \mathbb{F}_q$ then $A(g)$ is not conjugate to a diagonal matrix in $\mathrm{GL}(8, \mathbb{F}_q)$, but it is conjugate to a diagonal matrix in $\mathrm{GL}(8, K)$ for any extension field K of $\mathbb{F}_q$ containing λ and μ. In either case, the dimension of $\mathrm{fix}(g)$ is the number of 1's in the sequence $\lambda, \lambda, \lambda, \mu, \mu, \mu, \lambda^2\mu^{-1}, \lambda^{-1}\mu^2$.

It is now easy to check the following.

- If g has minimal polynomial $x - 1$ then $\mathrm{fix}(g)$ has dimension 8.
- If g has minimal polynomial $(x - 1)^2$ then $\mathrm{fix}(g)$ has dimension 4 if q is a power of 2 and dimension 3 if q is not a power of 2.
- If g has minimal polynomial $(x - 1)(x + 1)$ then $\mathrm{fix}(g)$ has dimension 4. (Note that if q is a power of 2 then $(x - 1)(x + 1) = (x - 1)^2$.)
- If g has minimal polynomial $(x - 1)(x - \mu)$ with $\mu \neq 0, 1, -1$ then $\mathrm{fix}(g)$ has dimension 3.
- If g has minimal polynomial $(x - \lambda)(x - \lambda^2)$ with $\lambda \neq 0, \pm 1$, $\lambda^3 \neq 1$ then $\mathrm{fix}(g)$ has dimension 1.
- If g has minimal polynomial $(x - \lambda)(x - \lambda^2)$ with $\lambda \neq 1$, $\lambda^3 = 1$, then $\mathrm{fix}(g)$ has dimension 2. Note that this cannot arise if q is a power of 3.
- In all other cases $\mathrm{fix}(g)$ has dimension 0.

In each case $|\mathrm{fix}(g)|$ is PORC. Note that these 7 cases arise from subdividing the three types of minimal polynomial according to whether or not the eigenvalues satisfy various *monomial* equations, such as $\lambda = 1$, $\lambda^2 = 1$, $\lambda^3 = 1$, $\lambda^2\mu^{-1} = 1$. It is not really necessary here, but you can also distinguish between eigenvalues in $\mathbb{F}_q$ and eigenvalues not in $\mathbb{F}_q$ with monomial equations: λ is a root of an irreducible quadratic over $\mathbb{F}_q$ if $\lambda^{q-1} \neq 1$, $\lambda^{q^2-1} = 1$; and μ is the other root of the same irreducible quadratic if $\lambda^q\mu^{-1} = 1$.

To compute the number of orbits we also need to know how many g lie in each of the seven categories just listed. The numbers are as follows.

- There is 1 element g with minimal polynomial $x - 1$ (the identity element).
- There are $q^2 - 1$ elements with minimal polynomial $(x - 1)^2$.
- If q is odd there are $q^2 + q$ elements with minimal polynomial $(x - 1)(x + 1)$. If q is a power of 2 this case does not arise.
- If q is a power of 2 there are $(q - 2)q(q + 1)$ elements with minimal polynomial $(x - 1)(x - \mu)$ with $\mu \neq 0, 1, -1$; and if q is odd there are $(q - 3)q(q + 1)$ elements.
- If q is a power of 2 and $q \equiv 1 \bmod 3$ then there are $(q - 4)q(q + 1)$ elements with minimal polynomial $(x - \lambda)(x - \lambda^2)$ with $\lambda \neq 0, \pm 1$, $\lambda^3 \neq 1$; if q is a power of 2 and $q \equiv 2 \bmod 3$ then there are $(q - 2)q(q + 1)$ elements; if q is a power of 3 then there are $(q - 3)q(q + 1)$ elements; if q is a power of p for $p > 3$ and if $q \equiv 1 \bmod 3$ there are $(q - 5)q(q + 1)$ elements; and if q is a power of p for $p > 3$ and $q \equiv 2 \bmod 3$ then there are $(q - 3)q(q + 1)$ elements.

- If $q = 3^k$ there are no elements with minimal polynomial $(x-\lambda)(x-\lambda^2)$ with $\lambda \neq 1$, $\lambda^3 = 1$; if $q \equiv 1 \bmod 3$ there are $q(q+1)$ elements; and if $q \equiv 2 \bmod 3$ there are $q(q-1)$ elements.
- The number of elements in this last category is $|\mathrm{GL}(2, \mathbb{F}_q)|$ minus the sum of all the numbers of elements in the other 6 categories.

All these numbers are PORC, and it follows that the number of orbits, $g(2,q)$, is PORC. The three polynomials giving the value of $g(2,q)$ depending on whether q is a power of 2, a power of 3, or a power of p for some $p > 3$ can be obtained by feeding these numbers into the formula for the number of orbits of $\mathrm{GL}(2, \mathbb{F}_q)$ on V.

3 Enumerating the Groups of Order p^n

As we saw in the Introduction, Higman [9] proved that for fixed n the number of groups of order p^n, $f(p^n)$, is bounded by a polynomial in p. Higman conjectured that (for fixed n) $f(p^n)$ is PORC—this is his famous PORC conjecture. The conjecture has been proved correct for $n \leq 7$. The table below gives the number of groups of order p^n for $n \leq 5$.

	$p=2$	$p=3$	$p \geq 5$
p	1	1	1
p^2	2	2	2
p^3	5	5	5
p^4	14	15	15
p^5	51	67	$2p + 61 + 2\gcd(p-1,3) + \gcd(p-1,4)$

There are 267 groups of order 2^6 and 504 groups of order 3^6. For $p \geq 5$ the number of groups of order p^6 is

$$3p^2 + 39p + 344 + 24\gcd(p-1,3) + 11\gcd(p-1,4) + 2\gcd(p-1,5).$$

The numbers of groups of order 2^7, 3^7, 5^7 are respectively 2328, 9310, 34297. For $p > 5$ the number of groups of order p^7 is

$$\begin{aligned} &3p^5 + 12p^4 + 44p^3 + 170p^2 + 707p + 2455 \\ &\quad + \left(4p^2 + 44p + 291\right)\gcd(p-1,3) + \left(p^2 + 19p + 135\right)\gcd(p-1,4) \\ &\quad + (3p + 31)\gcd(p-1,5) + 4\gcd(p-1,7) + 5\gcd(p-1,8) \\ &\quad + \gcd(p-1,9). \end{aligned}$$

So, for example, for $p \geq 5$ the number of groups of order p^6 is one of 8 polynomials in p, where the choice of polynomial depends on the residue class of p modulo 60. The PORC conjecture is still open for $n = 8$.

The groups of order p^2 were classified by Netto [13] in 1882. The groups of order p^3 were independently determined by Cole and Glover [4], Hölder [11] and Young

[21] in 1893. The groups of order p^4 were determined by Hölder [11] and Young [21]. The groups of order p^5 were classified by Bagnera [1] in 1898. However it was not until 2004 that Newman, O'Brien and Vaughan-Lee [14] classified the groups of order p^6. The groups of order p^7 were classified by O'Brien and Vaughan-Lee [16] in 2005.

Higman [10] proves that the number of groups of order p^n with p-class 2 is PORC (for any fixed n). The Frattini subgroup of a p-group G is the subgroup generated by the p-th powers $\{x^p \mid x \in G\}$ and the commutators $\{[x, y] \mid x, y \in G\}$ (where $[x, y]$ denotes $x^{-1}y^{-1}xy$). We say that G has p-class 2 if the Frattini subgroup is elementary Abelian and central, that is to say if

$$\big[x^p, y\big] = 1, \quad x^{p^2} = 1, \qquad \big[[x, y], z\big] = 1, \quad [x, y]^p = 1$$

for all $x, y, z \in G$. (Higman uses the term Φ-class 2.) Evseev [7] has extended Higman's result to the more general class of p-groups in which the derived group is elementary Abelian and central, that is groups satisfying

$$\big[[x, y], z\big] = 1, \qquad [x, y]^p = 1$$

for all $x, y, z \in G$.

4 Immediate Descendants

Nowadays the classification of p-groups of small order makes use of the lower exponent-p-central series of a group. If G is any group then the lower exponent-p-central series of G,

$$G = G_1 \geq G_2 \geq \cdots \geq G_i \geq \cdots,$$

is defined by setting $G_1 = G$, $G_2 = G'G^p$, and in general setting $G_{i+1} = [G_i, G]G_i^p$. If G is a finite p-group then $G_{c+1} = \{1\}$ for some c, and we say that G has p-class c if $G_c \neq \{1\}$, $G_{c+1} = \{1\}$. If G is a finite p-group of p -class $c > 1$ then we say that G is an *immediate descendant* of G/G_c. Apart from the elementary Abelian group of order p^n, every group of order p^n is an immediate descendant of a group of order p^k for some $k < n$. To list the groups of order p^n, first list the groups of order p^k for all $k < n$. Then for each group G of order p^k for $k < n$, find all the immediate descendants of G which have order p^n.

So (for example) the formula given above for the number of p-groups of order p^6 ($p \geq 5$) can be obtained as follows. It turns out that for $p > 3$ there are 42 groups of order at most p^5 which have immediate descendants of order p^6. Each of these 42 groups is given by a presentation involving the prime p symbolically—for example one of the 42 groups has presentation

$$\big\langle a, b \mid a^p = [b, a, a], b^p = 1, \text{class}3\big\rangle. \tag{1}$$

For each of these 42 groups we compute the number of immediate descendants of order p^6, and the formula given above is obtained by adding together each of these

individual contributions. For example, group (1) above has $p + \gcd(p-1, 3) + 1$ descendants of order p^6. Finally, we have to add one to this total to account for the elementary Abelian group of order p^6. Each of the individual contributions is PORC, and as a consequence the formula above is PORC.

Higman does not use the term *immediate descendant*, and does not explicitly mention the lower exponent-p-central series. But nevertheless his theorem can be expressed in these terms. Every group of order p^n and p-class 2 is an immediate descendant of the elementary Abelian group of order p^r for some $r < n$. If G has order p^{r+s}, and if G is an immediate descendant of the elementary Abelian group of order p^r then in Higman's terminology we say that G has Φ-complexion (r, s). Higman defines $g(r, s; p)$ to be the number of groups with Φ-complexion (r, s). So the number of p-class 2 groups of order p^n is

$$\sum_{r+s=n} g(r, s; p).$$

If we let V be a vector space of dimension r over $\mathbb{F}_p$, and if we let $V \wedge V$ be the exterior square of V, then $\mathrm{GL}(r, p)$ induces an action on the direct sum $V \oplus (V \wedge V)$, in much the same way as $\mathrm{GL}(n, p)$ induces an action on sets of structure constants for algebras of dimension n. Higman shows that if $p > 2$ then $g(r, s; p)$ is equal to the number of orbits under this action on subspaces of codimension s in $V \oplus (V \wedge V)$. Higman uses his theorem on the number of orbits in a vector space under the action of general linear groups to show that this number is PORC. In fact his theorem shows that the number of orbits of subspaces of dimension *at most* s is PORC, and he obtains the number of orbits of subspaces of dimension s as the difference between the number of orbits of susbspaces of dimension at most s and the number of orbits of subspaces of dimension at most $s - 1$.

Marcus du Sautoy has found a group G_p of order p^9 with the property that the number of immediate descendants of G_p of order p^{10} is *not* PORC. We will describe this group and some of its properties in Sect. 5 below. However Marcus's example does not disprove the PORC conjecture. As we have seen, the total number of groups of order p^{10} is obtained by adding together the number of immediate descendants of order p^{10} of each group of order less than p^{10}, and then adding 1 to the total to account for the elementary Abelian group of order p^{10}. The grand total might still be PORC, even though we know that one of the individual summands is not PORC. My own view is that this is extremely unlikely. But in any case I believe that Marcus's group provides a counterexample to what I hazard to call the *philosophy* behind Higman's conjecture. Higman obtains the number of groups of order p^n of p-class 2 by adding up the number of immediate descendants of order p^n of all the elementary Abelian groups of order less than p^n. He shows that the grand total is PORC by proving that all the individual summands are PORC. Each of the individual summands is the difference of two PORC functions obtained from his theorem on the action of general linear groups. And, as we saw in Sect. 2, this theorem is obtained by splitting the elements of the general linear group into a number of distinct classes with the property that the number of elements in each class is PORC, and with the property that for each class C there is a single PORC function giving the value of $|\mathrm{fix}(g)|$ for $g \in C$.

5 Marcus du Sautoy's Group

Marcus du Sautoy's group has the following presentation for all $p > 3$:

$$G_p = \left\langle \begin{matrix} x_1, x_2, x_3, x_4, x_5, x_6, y_1, y_2, y_3 : [x_1, x_4] = y_3, [x_1, x_5] = y_1, [x_1, x_6] = y_2 \\ [x_2, x_4] = y_1, [x_2, x_5] = y_3, [x_3, x_4] = y_2, [x_3, x_6] = y_1 \end{matrix} \right\rangle$$

where all other commutators are defined to be 1, and where $g^p = 1$ for all $g \in G_p$.

The group is a class two nilpotent group of order p^9. The quotient group G_p/G'_p is elementary Abelian of order p^6, and G'_p is elementary Abelian of order p^3. It turns out that both the order of the automorphism group of G_p and the number of conjugacy classes of G_p are not PORC.

In [6], du Sautoy and Vaughan-Lee prove the following result:

Let D_p be the number of descendants of G_p of order p^{10} and exponent p. Let V_p be the number of points (x, y) in $\mathbb{F}_p^2$ that satisfy $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$. Then

1. If $p \equiv 5 \bmod 12$ then $D_p = (p+1)^2/4 + 3$.
2. If $p \equiv 7 \bmod 12$ then $D_p = (p+1)^2/2 + 2$.
3. If $p \equiv 11 \bmod 12$ then $D_p = (p+1)^2/6 + (p+1)/3 + 2$.
4. If $p \equiv 1 \bmod 12$ and $V_p = 0$ then $D_p = (p+1)^2/4 + 3$.
5. If $p \equiv 1 \bmod 12$ and $V_p \neq 0$ then $D_p = (p-1)^2/36 + (p-1)/3 + 4$.

They also show that there are infinitely many primes $p \equiv 1 \bmod 12$ for which $V_p > 0$, but that there is no sub-congruence of $p \equiv 1 \bmod 12$ for which $V_p > 0$ for all p in that sub-congruence class.

So the number of descendants of G_p of order p^{10} and exponent p is not PORC. It follows easily from this that the number of descendants of G_p of order p^{10} is not PORC.

5.1 The Conjugacy Classes of G_p

The center of G_p is the derived group G'_p, and most elements outside G'_p have breadth 3 (i.e. they lie in conjugacy classes of size p^3). However some elements outside G'_p have breadth 2 (i.e. they lie in conjugacy classes of size p^2). First we determine the elements of breadth 2 in the subgroup $\langle x_1, x_2, x_3 \rangle$. This subgroup is elementary Abelian of order p^3. If $0 < \alpha < p$ then the elements x_2^α, x_3^α have breadth 2, but if $0 < \alpha, \beta < p$ then $x_2^\alpha x_3^\beta$ has breadth 3. We need to determine the elements of breadth 2 in $\langle x_1, x_2, x_3 \rangle$ which lie outside the subgroup $\langle x_2, x_3 \rangle$, and so we consider an element $x_1 x_2^d x_3^e$. The subgroup $[x_1 x_2^d x_3^e, G_p]$ is generated by

$$y_1^d y_2^e y_3, \qquad y_1 y_3^d, \qquad y_1^e y_2$$

and so $x_1 x_2^d x_3^e$ has breadth 2 if p divides

$$\det \begin{bmatrix} d & e & 1 \\ 1 & 0 & d \\ e & 1 & 0 \end{bmatrix} = de^2 - d^2 + 1.$$

Now if $p|(de^2-d^2+1)$ then $p|((de)^2-d^3+d)$ and so elements of breadth 2 of the form $x_1x_2^dx_3^e$ correspond to points on the elliptic curve $y^2=x^3-x$ over $\mathbb{F}_p$. Let E be the number of points on this curve, including the point at infinity. Then there are $E-2$ elements of breadth 2 of the form $x_1x_2^dx_3^e$. It follows that there are $(p-1)\times E$ elements of breadth 2 in the subgroup $\langle x_1,x_2,x_3\rangle$. There is an automorphism θ of G_p given by

$$x_1\theta=x_4,\qquad x_2\theta=x_5,\qquad x_3\theta=x_6,$$
$$x_4\theta=x_1,\qquad x_5\theta=x_2,\qquad x_6\theta=x_3,$$

and it follows that the elements of breadth 2 in the subgroup $\langle x_4,x_5,x_6\rangle$ are of the form $x_5^\alpha, x_6^\alpha, (x_4x_5^dx_6^e)^\alpha$ with $0<\alpha<p$ and with $p|(de^2-d^2+1)$. A general element $g\in G_p$ can be written in the form $g=abc$ with $a\in\langle x_1,x_2,x_3\rangle$, $b\in\langle x_4,x_5,x_6\rangle$ and $c\in G_p'$. If abc has breadth 2 then a and b must either be trivial, or have breadth 2, and it is straightforward to show that the elements in G_p of breadth 2 are the following

$$x_2^\alpha x_5^\beta c, x_3^\alpha x_6^\beta c, \left(x_1x_2^dx_3^e\right)^\alpha\left(x_4x_5^dx_6^e\right)^\beta c,$$

where $0\le\alpha,\beta<p$ and α,β are not both zero, where $p|(de^2-d^2+1)$, and where $c\in G_p'$. So the total number of elements of breadth 2 in G_p is $(p^2-1)p^3\times E$. It follows that the number of conjugacy classes of G_p is

$$p^6+p^3-1+\left(p^3-p^2-p+1\right)\times E.$$

This number is *not* PORC. It is shown in Sect. 18.4 of [12] that if $p\equiv 3 \bmod 4$ then $E=p+1$, but if $p\equiv 1 \bmod 4$ then $E=p+1-2a$ where $p=a^2+b^2$ with $a+ib\equiv 1 \bmod(2+2i)$. Note that a is uniquely determined by p. The Gaussian integers are a unique factorization domain, and we can write $p=(a+ib)(a-ib)$ where this factorization is unique up to unit factors $\pm1,\pm i$. The choice of a and b so that $a+ib\equiv 1 \bmod(2+2i)$ means that we take a to odd and b even, and we choose the sign of a so that $a-b\equiv 1 \bmod 4$. So the value of a is a function of p. But a (and hence E) cannot be a PORC function of p. To see this note that if a was a polynomial in p then the fact that $|a|<\sqrt{p}$ would imply that a was constant. So a could only be a PORC function of p if a took only finitely many values as p varies. However Dirichlet's theorem on primes in arithmetic progression implies that approximately half the primes are equal to $1 \bmod 4$. Putting this more precisely, if $\pi(x)$ is the number of primes less than x, then asymptotically $\pi(x)\sim\frac{x}{\log x}$, and if we set $\pi'(x)$ equal to the number of primes less than x which are equal to $1 \bmod 4$, then $\pi'(x)\sim\frac{x}{2\log x}$. However for a fixed a there can only be at most $\sqrt{x}$ primes less than x which have the form a^2+b^2. So if a only took K distinct values as p varies, then there could only be at most $K\sqrt{x}$ primes less than x which are equal to $1 \bmod 4$. Asymptotically, $K\sqrt{x}$ is much less than $\pi'(x)$. So E is not PORC, and hence the number of conjugacy classes of G_p is not PORC.

5.2 The Automorphism Group of G_p

Let H be the automorphism group of G_p. Then H has a normal subgroup N of order p^{18} consisting of automorphisms mapping x_i to x_ig_i for $i=1,2,\ldots,6$, with

$g_1, g_2, \ldots, g_6$ arbitrary elements of G'_p. The quotient group H/N acts as a group of automorphisms of G_p/G'_p. The quotient group G_p/G'_p is isomorphic as a group to the additive group of a 6-dimensional vector space over $\mathbb{F}_p$, and we can identify H/N with a subgroup of the general linear group $\mathrm{GL}(6, \mathbb{F}_p)$. If we reorder the generators of G_p in the order $x_1, x_4, x_2, x_5, x_3, x_6$ then it is easy to see that for every $A \in \mathrm{GL}(2, \mathbb{F}_p)$, and for every $u \in \mathbb{F}_p$ satisfying $u^4 = 1$, there is an element of H/N with action on G_p/G'_p given by the matrix

$$\begin{bmatrix} uA & 0 & 0 \\ 0 & u^{-1}A & 0 \\ 0 & 0 & A \end{bmatrix}.$$

There are $\gcd(p-1, 4)$ choices for u here, and $|\mathrm{GL}(2, \mathbb{F}_p)|$ choices for A. So these automorphisms give a subgroup $K/N \leq H/N$ of order $|\mathrm{GL}(2, \mathbb{F}_p)| \cdot \gcd(p-1, 4)$. For most primes ($\frac{11}{16}$ of them!) $H = K$, so that H has order $|\mathrm{GL}(2, \mathbb{F}_p)| \cdot \gcd(p-1, 4) \cdot p^{18}$. But if we can find $x, y \in \mathbb{F}_p$ satisfying $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$, then there are some additional automorphisms. Specifically, if we order the generators of G_p in their original order $x_1, x_2, x_3, x_4, x_5, x_6$, and if $x, y \in \mathbb{F}_p$ satisfy $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$, then if we let $d = x$ and $e = y/x$ and take

$$A = \begin{bmatrix} \frac{u(d^2+1)e}{4} & \frac{u(d^3+9d)e}{4} & \frac{u(d^3+5d)}{2} \\ \frac{u^{-1}(d^3+5d)e}{4} & -\frac{u^{-1}(d^2+5)e}{4} & \frac{u^{-1}(d^2+1)}{2} \\ 1 & d & e \end{bmatrix}$$

for any u with $u^4 = 1$, then there are elements in H/N with action on G_p/G'_p given by the matrix

$$\begin{bmatrix} \alpha A & 0 \\ 0 & \beta A \end{bmatrix}$$

for all $\alpha, \beta \neq 0$ in $\mathbb{F}_p$. In the cases when there do exist $x, y \in \mathbb{F}_p$ satisfying $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$, then these additional automorphisms together with automorphisms in K generate the full automorphism group H. (All this is described in detail in [6].)

The roots of the polynomial $x^2 + 6x - 3$ are $-3 \pm 2\sqrt{3}$. Now if $p > 3$ then 3 is a quadratic residue modulo p if and only if $p \equiv \pm 1 \bmod 12$, so we need $p \equiv \pm 1 \bmod 12$ to have any hope of solutions to our two equations.

The case $p \equiv -1 \bmod 12$ is straightforward. We need to solve $x^2 = -3 \pm 2\sqrt{3}$. Now $(-3 + 2\sqrt{3})(-3 - 2\sqrt{3}) = -3$, which is *not* a quadratic residue modulo p, so one of the two equations $x^2 = -3 \pm 2\sqrt{3}$ has two solutions, and the other has none. So we obtain two solutions $\pm d \in \mathbb{F}_p$ to the equation $x^4 + 6x^2 - 3 = 0$. We then want to solve the equations $y^2 = \pm(d^3 - d)$, and since -1 is *not* a quadratic residue modulo p one of these two equations will have two solutions, and the other will have none. So if $p \equiv -1 \bmod 12$ there are exactly two solutions to the two equations.

The case $p \equiv 1 \bmod 12$ is much trickier. In this case -3 *is* a quadratic residue modulo p, so the equation $x^4 + 6x^2 - 3 = 0$ either has no solutions in $\mathbb{F}_p$, or it has four solutions. It turns out that it has four solutions for approximately half the primes

$p \equiv 1 \bmod 12$. This is because the splitting field of $x^4 + 6x^2 - 3$ has degree 8 over $\mathbb{Q}$, so that, by Chebotarev's density theorem, the primes p for which the polynomial splits over $\mathbb{F}_p$ have density $\frac{1}{8}$. These primes are necessarily equal to $1 \bmod 12$, and the primes equal to $1 \bmod 12$ have density $\frac{1}{4}$ by Dirichlet's theorem on primes in arithmetic progression. In the case when $x^4 + 6x^2 - 3 = 0$ has four solutions $\pm d_1, \pm d_2$ in $\mathbb{F}_p$ we still need to solve the equations $y^2 = \pm(d_1^3 - d_1)$ and $y^2 = \pm(d_2^3 - d_2)$. Since $p \equiv 1 \bmod 4$, -1 is a quadratic residue mod p, and so the equations $y^2 = \pm(d_1^3 - d_1)$ either have 0 solutions or 4 solutions. Similarly the equations $y^2 = \pm(d_2^3 - d_2)$ either have 0 solutions or 4 solutions. In fact the four equations either have 0 solutions or 8 solutions. To see this observe that

$$\begin{aligned}\big(d_1^3 - d_1\big)\big(d_2^3 - d_2\big) &= \big(d_1^2 - 1\big)\big(d_2^2 - 1\big)d_1 d_2 \\ &= (-4 + 2\sqrt{3})(-4 - 2\sqrt{3})\sqrt{-3} = 4\sqrt{-3}.\end{aligned}$$

If we pick $u \in \mathbb{F}_p$ such that $u^2 = -1$ then

$$\left(\frac{1}{4}(1+u)\big(d_1^3 + 5d_1\big)\right)^4 = -3,$$

so $4\sqrt{-3}$ is a square in $\mathbb{F}_p$, and either both the equations $y^2 = d_1^3 - d_1$, $y^2 = d_2^3 - d_2$ have solutions in $\mathbb{F}_p$, or neither does.

It turns out that there are 8 solutions (x, y) to the two equations $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$ over $\mathbb{F}_p$ for approximately half the primes for which $x^4 + 6x^2 - 3$ splits. It is straightforward to see that if d is a root of $x^4 + 6x^2 - 3$ then $d^3 - d$ is a root of $x^4 + 360x^2 - 48$. Furthermore, if $x^4 + 360x^2 - 48$ has a root then so does $x^4 + 6x^2 - 3$. It follows that the two equations $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$ have solutions over $\mathbb{F}_p$ if and only if $x^8 + 360x^4 - 48$ has a root in $\mathbb{F}_p$. The splitting field of $x^8 + 360x^4 - 48$ has degree 16 over $\mathbb{Q}$, and so by Chebotarev's density theorem the primes p for which the polynomial splits over $\mathbb{F}_p$ have density $\frac{1}{16}$. These primes are necessarily equal to $1 \bmod 12$. In particular there are infinitely many primes $p \equiv 1 \bmod 12$ for which the two equations have 8 solutions. However the primes $p \equiv 1 \bmod 12$ for which $x^4 + 6x^2 - 3$ (or equivalently $x^4 + 360x^2 - 48$) have a root are extremely irregular. It is proved in [6] that if $p \equiv 1 \bmod 12$, and if we write $p = a^2 - 12b^2$ with $a, b > 0$ then $x^4 + 6x^2 - 3$ has a root in $\mathbb{F}_p$ if and only if $a \equiv 1 \bmod 3$. This means that you cannot capture the primes $p \equiv 1 \bmod 12$ for which there are roots in a sub-congruence class of $p \equiv 1 \bmod 12$. If $c \equiv 1 \bmod 12$ and $(c, d) = 1$ then a theorem due to Rademacher [18] implies that there are infinitely many primes $p \equiv c \bmod 12d$ where $p = a^2 - 12b^2$ with $a > 0$ and $a \equiv 1 \bmod 3$, and infinitely many primes $p \equiv c \bmod 12d$ where $p = a^2 - 12b^2$ with $a > 0$ and $a \equiv 2 \bmod 3$.

It is proved in [6] that the order of the automorphism group of G_p is as follows:

- If $p \equiv 1 \bmod 12$ and there are no solutions to the equations $x^4 + 6x^2 - 3 = 0$ and $y^2 = x^3 - x$ over $\mathbb{F}_p$ then there are $|\mathrm{GL}(2, p)| \cdot 4p^{18}$ automorphisms.
- If $p \equiv 1 \bmod 12$ and there are solutions to the equations then there are $|\mathrm{GL}(2, p)| \cdot 36p^{18}$ automorphisms.

- If $p \equiv 11 \bmod 12$ there are $|\mathrm{GL}(2, p)| \cdot 6p^{18}$ automorphisms.
- If $p \equiv 5 \bmod 12$ there are $|\mathrm{GL}(2, p)| \cdot 4p^{18}$ automorphisms.
- If $p \equiv 7 \bmod 12$ there are $|\mathrm{GL}(2, p)| \cdot 2p^{18}$ automorphisms.

Since there are infinitely many primes $p \equiv 1 \bmod 12$ for which the equations have solutions, and since we cannot capture these primes in a sub-congruence class of $p \equiv 1 \bmod 12$, it follows that the order of the automorphism group of G_p is not PORC.

6 The p-Group Generation Algorithm

We described in Sect. 5 how the order of the automorphism group of G_p is not PORC, and this is the reason that the number of descendants of G_p of order p^{10} is not PORC. To see why this is so we need to describe O'Brien's p-group generation algorithm [15] for computing the immediate descendants of a p-group G.

So let G be a finite p-group of p-class c. The p-covering group P of G is defined to be the largest finite p-group with a normal subgroup M satisfying the following three properties:

- $P/M \cong G$,
- $M \leq P^p P'$,
- M is central in P and of exponent p.

The p-covering group P is unique, and is actually quite easy to compute. The normal subgroup M is called the p-multiplier of G, and is also unique. Since G has p-class c it follows that $P_{c+1} \leq M$. (Recall from Sect. 3 that P_{c+1} is the $(c+1)^{th}$ term of the lower exponent-p-central series of P.) Since M is central and of exponent p, $P_{c+2} = \{1\}$. (It can happen that $P_{c+1} = \{1\}$.) A proper subgroup $S < M$ is said to be allowable if S is a supplement in M for P_{c+1}, that is if $SP_{c+1} = M$. The immediate descendants of G are the quotient groups P/S where S is an allowable subgroup. Note that if $P_{c+1} = \{1\}$ then there are no allowable subgroups, and hence no immediate descendants. In this case we say that G is terminal.

The automorphism group of G acts on M, and two immediate descendants P/S and P/T are isomorphic if and only if the allowable subgroups S and T are in the same orbit under the action of the automorphism group of G. So the bigger the automorphism group the smaller the number of immediate descendants.

We can describe Higman's analysis of the p-groups of p-class 2 in these terms. The p-groups of p-class 2 are immediate descendants of elementary Abelian groups. Let G be an elementary Abelian group of order p^r, and identify G with a vector space V of dimension r over $\mathbb{F}_p$. The p-multiplier of G is $V \oplus (V \wedge V)$. The p-class of G is 1, and in this case if P is the p-covering group then $P_2 = M = V \oplus (V \wedge V)$. So in this case every proper subgroup of M is allowable, and these subgroups correspond to proper subspaces of $V \oplus (V \wedge V)$. The automorphism group of G is $\mathrm{GL}(r, \mathbb{F}_p)$, and the number of immediate descendants of G is the number of orbits of proper subspaces of $V \oplus (V \wedge V)$ under the action of $\mathrm{GL}(r, \mathbb{F}_p)$.

7 Further Problems

As we have seen, Graham Higman's PORC conjecture has been confirmed for $n \leq 7$. Higman has shown that the number of p-class two groups of order p^n is PORC for all n. Evseev has extended Higman's proof to show that the number of groups of order p^n with derived groups which are elementary Abelian and central is PORC for all n.

What are the possibilities for extending these positive results, and what are the possibilities for actually settling the question completely? It should be possible to classify the groups of order p^8, though I believe that this will be extraordinarily difficult. Classifying the groups of order p^9 or p^{10} seems to be way out of reach for the time being. One possible way of making progress would be to aim for a combination of Higman's methods and classification methods. The hardest part of classifying the groups of order p^6 and p^7 was classifying the p-class two groups of those orders. In contrast, classifying the groups of maximal class of order p^6 and p^7 was relatively easy. You could settle the PORC conjecture for $n = 8$ if you could classify just those groups of order p^8 with p-class greater than two. However classifying the p-class three groups of order p^6 and p^7 was nearly as hard as classifying the p-class two groups. Nevertheless I believe that we could make substantial progress with groups of order p^8 by first classifying the groups of maximal class, and then looking at groups of coclass two, and so on.

It might be possible to extend Higman's result about groups of p-class two to groups of class two (without any restriction on the orders of the elements). But I believe that Marcus's group shows that it would be impossible to directly extend Higman's methods to p-class three groups, or even to class three groups of exponent p. As stated above, I believe that there is no immediate prospect of classifying all the groups of order p^{10}, or even of classifying all class 3 groups of exponent p and order p^{10}. We know that Marcus's group has a non-PORC number of immediate descendants of order p^{10}, but it seems likely that there are other class two groups of order p^9 which also have a non-PORC number of immediate descendants of order p^{10}. It is possible that the grand total of all class three groups of order p^{10} is PORC even though some of the individual contributions to the total are non-PORC. But I do not see how to settle this without classifying this class of groups. Nevertheless it would be useful to find some more examples of class two groups of order p^9 with a non-PORC number of immediate descendants of order p^{10}. Even better would be to find some class two groups of order p^8 with a non-PORC number of immediate descendants of order p^9.

It is perhaps fitting to end this note with a mention of another remarkable result of Marcus du Sautoy [5]. Marcus proves that for each n there are finitely many subvarieties E_i $(i \in T)$ of a variety Y defined over $\mathbb{Q}$ and for each subset $I \subset T$ a polynomial $H_I(x)$ such that for almost all primes p

$$f\left(p^n\right) = \sum_{I \subset T} c_{p,I} H_I(p),$$

where

$$c_{p,I} = \operatorname{card}\left\{a \in \overline{Y}(\mathbb{F}_p) \mid a \in \overline{E}_i(\mathbb{F}_p) \text{ if and only if } i \in I\right\}.$$

Here $\overline{Y}$ means reduction of the variety modulo p, which is defined for almost all p. Marcus's group G_p embeds the elliptic curve $y^2 = x^3 - x$ in its structure and there seems every reason to suppose that much more complicated algebraic varieties can be embedded in the structure of finite p-groups in such a way as to impact on the number of conjugacy classes, the size of the automorphism group and the number of descendants.

References

1. Bagnera, G.: La composizione dei gruppi finiti il cui grado è la quinta potenza di un numero primo. Ann. Mat. Pura Appl. **1**(3), 137–228 (1898)
2. Blackburn, S.R., Neumann, P.M., Venkataraman, G.: Enumeration of Finite Groups. Cambridge Tracts in Mathematics, vol. 173. Cambridge University Press, Cambridge (2007)
3. Burnside, W.: Theory of Groups of Finite Order, 2nd ed. Dover, Cambridge (1911)
4. Cole, F.N., Glover, J.W.: On groups whose orders are products of three prime factors. Am. J. Math. **15**, 1–4 (1893)
5. du Sautoy, M.P.F.: Zeta functions and counting finite p-groups. Electron. Res. Announc. Am. Math. Soc. **5**, 112–122 (1999)
6. du Sautoy, M.P.F., Vaughan-Lee, M.R.: Non-PORC behaviour of a class of descendant p-groups. J. Algebra (to appear)
7. Evseev, A.: Higman's PORC conjecture for a family of groups. Bull. Lond. Math. Soc. **40**, 415–431 (2008)
8. Hall, M.: The Theory of Groups. Macmillan, New York (1959)
9. Higman, G.: Enumerating p-groups. I. Inequalities. Proc. London Math. Soc. **10**(3) 24–30 (1960)
10. Higman, G.: Enumerating p-groups. II. Problems whose solution is PORC. Proc. London Math. Soc. **10**(3), 566–582 (1960)
11. Hölder, O.: Die Gruppen der Ordnungen p^3, pq^2, pqr, p^4. Math. Ann. **43**, 301–412 (1893)
12. Ireland, K., Rosen, M.: A Classical Introduction to Modern Number Theory. Graduate Texts in Mathematics, vol. 84. Springer, Berlin (1993)
13. Netto, E.: Substitutionentheorie und ihre Anwendungen auf die Algebra. Teubner, Leipzig (1882)
14. Newman, M.F., O'Brien, E.A., Vaughan-Lee, M.R.: Groups and nilpotent Lie rings whose order is the sixth power of a prime. J. Algebra **278**, 383–401 (2004)
15. O'Brien, E.A.: The p-group generation algorithm. J. Symb. Comput. **9**, 677–698 (1990)
16. O'Brien, E.A., Vaughan-Lee, M.R.: The groups with order p^7 for odd prime p. J. Algebra **292**, 243–358 (2005)
17. Pyber, L.: Enumerating finite groups of given order. Ann. of Math. **137**(2), 203–220 (1993)
18. Rademacher, H.: Über die Anzahl der Primzahlen eines reell-quadratischen Zahlkörpers, deren Konjugierte unterhalb gegebener Grenzen liegen. Acta Arith. **1**, 67–77 (1935)
19. Sims, C.C.: Enumerating p-groups. Proc. London Math. Soc. (3) **15**, 151–166 (1965)
20. van Lint, J.H., Wilson, R.M.: A Course in Combinatorics. Cambridge University Press, Cambridge (2001)
21. Young, J.W.A.: On the determination of groups whose order is a power of a prime. Am. J. Math. **15**, 124–178 (1893)

Michael Vaughan-Lee is a retired Professor of Mathematics from Oxford University. He studied at Oxford from 1962 to 1968, receiving his doctorate in 1968. From 1968 to 1970 he was Assistant Professor at Vanderbilt University in Nashville, Tennessee, and from 1970 to 1971 he was a Lecturer at The University of Queensland in Brisbane, Australia. In 1971 he returned to Oxford as a college tutor at Christ Church. He was appointed Professor of Mathematics at Oxford in 1996, and retired in 2010. Nowadays his main research interests are in p-groups and in computational group theory.

Jahresber Dtsch Math-Ver (2012) 114:107–111
DOI 10.1365/s13291-012-0040-4

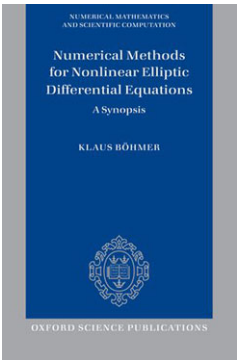
BOOK REVIEW

Klaus Böhmer: "Numerical Methods for Nonlinear Elliptic Differential Equations—A Synopsis"

Oxford University Press, 2010, 776 pp.

Michael Plum

Published online: 5 April 2012



(I) The book is mainly concerned with numerical analysis and convergence theory of various methods for numerically solving boundary value problems with nonlinear elliptic partial differential equations. Particular emphasis is put on a general theoretical discretization framework, which allows the formulation of most of the numerical methods studied here as special cases. This general concept essentially consists of an abstract Petrov-Galerkin method, formulated with suitable projection operators mapping the continuous spaces involved onto their discrete counterparts. For many important discretization methods, convergence is studied in the book on the basis of this general framework, using the classical way via consistency and stability. The exposition includes conforming Finite Element methods (where piece-wise polynomial Galerkin approximations are considered which are globally in the "correct" function space), non-conforming Finite Element methods (where the approximations are allowed to drop out of the "correct" space e.g. by jumps across element edges), Discontinuous Galerkin methods (which are non-conforming Finite Element methods with an additional penalty term punishing non-conformal behavior), Finite Difference methods (using grid functions as approximations and replacing differential by difference operations), and wavelets (which are single generating functions with all their suitably translated and dilated copies forming a basis of the Hilbert space of square integrable functions, finitely many of which are used for approximation purposes).

In each of the corresponding chapters, usually a chain of "increasing nonlinearity" is chosen, i.e. first linear problems are considered (sometimes with Poisson-type

M. Plum (✉)
Karlsruhe, Germany
e-mail: michael.plum@math.uni-karlsruhe.de

equations in the very first step), followed by semilinear, then quasilinear, and finally fully nonlinear problems. An additional subdivision into equations of order 2 and order 2m supplements this structure. Several examples, mainly from physics, are presented for illustration. In particular, the Stokes and the Navier-Stokes equations are studied under various aspects like e.g. proving estimates for the convergence of discrete solutions to the exact one, but also other problems including the von Kármán and the Lamé equations, quasilinear diffusion equations, and the fully nonlinear Monge-Ampère equations, are analyzed. Actual numerical computations are also contained, but play a subordinate role.

The book is meant as the first part of a pair of two books (of roughly equal size) by the author, or in a sense as a preparation for the second one, which is entitled "Numerical Methods for Bifurcation and Center Manifolds in Nonlinear Elliptic and Parabolic Differential Equations", and also published by Oxford University Press. Thus some of the more theoretical results in the present book are declared as prerequisites for the second one, and up to some degree remain as "open ends" here, which however is certainly natural for such a coupled project.

(II) The book starts with two chapters summarizing analytical results which are of importance in particular for the numerical theory following later.

Chapter 1 first introduces some examples from physics and engineering as a motivation for studying nonlinear partial differential equations. In particular, the transition between linear and nonlinear regime, and between a nonlinear model and its linearization, is nicely explained. The rest of the chapter is devoted to some elements of functional analysis, including Hölder and Sobolev spaces, which are needed later in this book, or in the second book mentioned before.

In Chapter 2, analytical results (to be used in both books) on existence, uniqueness, and regularity of solutions to elliptic differential equations are summarized. As a start, linear Poisson-type problems are formulated in classical, strong, and weak form, in order to introduce these concepts, and corresponding results are presented. In order to prepare for more general problems, bilinear forms and related operators are studied next, introducing the concept of Gelfand triples, and aiming at Riesz-Schauder theory and the Fredholm alternative. Existence, uniqueness, regularity, and the Fredholm alternative, in Hölder as well as in Sobolev spaces, are then investigated for linear elliptic operators of order 2m, and, under relaxed smoothness conditions on the coefficients, of order 2. The classes of semilinear, quasilinear, and fully nonlinear elliptic equations are introduced, and results on existence, uniqueness, and regularity are formulated for each of these classes, with monotone operators playing an essential role for fully nonlinear problems and as well for quasilinear problems in divergence form. Also for systems of elliptic equations, results on existence, uniqueness, and regularity (e.g. in Morrey spaces) are presented. In particular, variational approaches for quasilinear elliptic systems are discussed. For semilinear and quasilinear equations and systems, linearization techniques are studied aiming at Fredholm alternatives; the main motivation for these are bifurcation problems investigated in the second book. Finally, the incompressible Navier-Stokes equations are discussed as an application.

Chapter 3 presents the author's general discretization concept which is meant as a unifying approach to various more concrete numerical discretization methods discussed in the subsequent chapters. The concept consists of a Petrov-Galerkin approach relating the original nonlinear operator to its discrete counterpart via suitably

chosen projection operators, mapping the original spaces onto the discrete ones. The goal is to establish, within the frame of this general discretization concept, convergence of the discrete solutions to the original solution, assuming that existence (and uniqueness) of the latter one is at hand. Here, the classical way of proving convergence via consistency and stability is chosen. So first consistency is studied, in variational form as well as in classical form; the latter is needed in particular for fully nonlinear problems, which are lacking a variational structure. Stability is formulated via suitable nonlinear variants of the usual requirement that the inverses of the discrete operators are bounded uniformly in the discretization parameter, with respect to some suitable norm. Methods for proving stability are discussed; the most important approach uses coerciveness of the principal part and compact perturbation techniques. Finally, Newton's and continuation methods are incorporated into the convergence theory.

The remaining Chapters 4 to 9 are devoted to concrete discretization methods, formulated—as far as possible—as special cases of the general discretization concept.

In Chapter 4 conforming Finite Element methods are studied. First the usual element types are introduced, and interpolation in the corresponding Finite Element spaces is investigated, including interpolation error estimates and inverse estimates. Also curved boundaries and isoparametric elements are included in the discussion. For applications of Finite Element methods to elliptic problems, the investigation of convergence starts with linear problems and then extends to quasilinear equations and systems in divergence form, and to monotone nonlinear operators, always using the general concept developed in Chapter 3. Convergence results for Mixed Finite Elements are added, and applied to Stokes and Navier-Stokes problems. Finally, linear eigenvalue problems are included.

Chapter 5 is concerned with nonconforming Finite Elements. These allow violation of the smoothness conditions (contained in the underlying Sobolev space) across element edges, and also of the boundary conditions. Again convergence is studied via consistency, now in the classical (and not the variational) form, and stability. For applying the general discretization concept of Chapter 3, a careful choice of suitable projection operators is required. For fully nonlinear problems, the key is to consider two-component operators, involving the differential operator in one, and the boundary values in the other component. The stability question is attacked by a combination of linearization techniques and regularity of Finite Element solutions. Since the integrals involved in the Finite Element formulation cannot be evaluated exactly in many cases, the convergence analysis is extended to the situation when these integrals are approximated by quadrature formulas.

In Chapter 6 (by W. Dörfler), questions of adaptivity of conforming Finite Element methods are studied for a simple Poisson-type model problem, which is chosen in order to explain the main ideas. Several kinds of singularities, which the exact solution might have, and which give rise to the necessity of mesh refinement near this singularity, are listed. Residual-based a posteriori error estimators are introduced which form the basis of adaptive mesh refinement. Convergence of the resulting adaptive method is proved under suitable assumptions.

Chapter 7 (with V. Dolejsi) contains an investigation of Discontinuous Galerkin methods, which—like the nonconforming elements studied in Chapter 5—allow

violation of smoothness and boundary conditions; here, however, such violations are penalized, with increasing penalty term as the discretization gets finer. Mainly convection-diffusion systems are studied here, in semilinear and quasilinear versions. Fully nonlinear problems are not included. For convergence analysis, broken Sobolev spaces (which do not require smoothness across element edges) and suitable penalty norms are used. The consistency of the various terms involved is proved under suitable conditions, and stability is obtained via discrete coercivity and boundedness of the discrete principal part. Also some hp-variants of Discontinuous Galerkin methods are added to the investigation, and several numerical example computations are presented.

In Chapter 8, a convergence analysis for Finite Difference methods is performed on the basis of the general discretization concept. For this purpose, Finite Difference methods have to be formulated in variational form, which is done here using discrete Sobolev spaces of grid functions. In these spaces, consistency and convergence can be formulated nicely, and stability is obtained via coercivity and compact perturbation arguments. The analysis is essentially restricted to cuboidal domains and to equidistant grids, but on the other hand general quasilinear and fully nonlinear elliptic systems are treated. Also Neumann boundary conditions are included in the analysis.

The final Chapter 9 (with S. Dahlke and partly with T. Raasch) is devoted to wavelet methods, which are again studied on the basis of the general discretization concept of Chapter 3. General elliptic problems, including saddle point problems and in particular Navier-Stokes problems, are included in the analysis, which uses Besov spaces for consistency, stability, and convergence. Again the bounded invertibility of the principal part, combined with compact perturbation techniques, forms the analytical basis of stability and convergence. Also adaptive wavelet methods are investigated by use of a posteriori error estimators.

The exposition ends with a list of 683 references and a comprehensive index.

(III) The book gives an extensive synopsis of convergence analysis for the discretization methods investigated here, in particular for Finite Element methods. It presents research components as well as many general aspects of this field, and thus is also suited for teaching purposes on an advanced level.

It is interesting to see convergence analysis in the light of the author's abstract general discretization concept, which also leads to some new convergence results for nonlinear elliptic problems, and brings new aspects into the convergence analysis of Finite Difference methods. The abstract concept might possibly also have impact on further research on other discretization methods not treated in the book, like e.g. Finite Volume methods or Boundary Element methods.

The great extent of the convergence analysis implies that practical computational aspects and numerical linear algebra, including e.g. conjugate gradient methods, would have been beyond the scope of the book and are not included, with the exception of some practical numerical results at the ends of Chapters 7 and 8. From the reviewer's point of view, it would have been nice if multilevel and multigrid methods, which nowadays are very important tools in practical numerical applications to highly complex problems, could have been added to the investigations. It would be an interesting question if multigrid methods and their convergence analysis can also be covered by the author's general discretization concept.

Summarizing, the book gives a profound insight into the convergence analysis of many important numerical methods for nonlinear elliptic problems, with emphasis on Finite Element Methods in several variants. It should be of significance to all scientists and advanced students who are interested in more theoretical aspects of these numerical methods.

Jahresber Dtsch Math-Ver (2012) 114:113–116
DOI 10.1365/s13291-012-0042-2

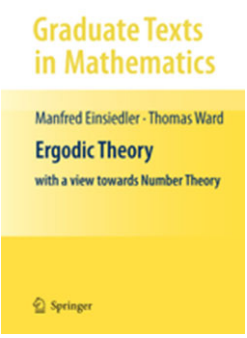
BOOK REVIEW

The remarkable effectiveness… A report on the book Manfred Einsiedler and Thomas Ward: "Ergodic theory, with a view towards number theory"

Springer-Verlag, 2011, 467 pp.

Barak Weiss

Published online: 19 April 2012



The book under review is an introductory textbook on ergodic theory, written with applications to number theory in mind. Among its many merits is its timeliness. It is inspired by dramatic recent developments in which progress on longstanding open questions in number theory is made using ergodic-theoretic tools. Examples of such developments are given in Chapter 1, among these are Margulis' proof of the Oppenheim conjecture, Furstenberg's ergodic proof of Szemerédi's theorem, and the progress of Einsiedler, Lindenstrauss and Katok toward Littlewood's conjecture. I will begin by briefly discussing the fruitful relationship of ergodic theory and number theory, which may seem surprising to some readers, and which motivated the authors in many of their choices. Although this theme is discussed in many interesting survey papers (among them I recommend [1], from whom the title of this review is taken), I hope some readers will find my remarks useful. Then I will focus on the book at hand in more detail.

Ergodic theory has its roots in the study of physical systems, and is classically concerned with the long-term statistical behavior of typical trajectories. To give a naive example, one might consider the motion of the earth around the sun, and ask: *Will the earth ever collide with the sun*? *Will it drift further and further from the sun*? An answer was famously given by Newton in his *Principia*. Modelling the system as two point particles in Euclidean space, and formulating his laws of motion and

B. Weiss (✉)
Be'er Sheva, Israel
e-mail: barakw@math.bgu.ac.il

gravitation, Newton obtained a differential equation which he explicitly solved, recovering Kepler's laws of motion: the earth's orbit describes an ellipse with the sun at a focal point. In particular the long-term behavior is periodic, giving a negative answer to the above questions. Newton's tremendous achievement gave rise to the hope that long-term behavior of deterministic physical systems could always be analyzed successfully by explicitly solving differential equations. However it was gradually realized that many differential equations could not be explicitly solved, and in many cases, solutions depend sensitively on initial conditions. For instance this would be the case if one added a third body (e.g. the moon) to the above model: tiny changes in initial conditions, e.g. adding a milligram to the mass of one of the bodies, can lead to a different answer to questions about long-term behavior. Awareness of such phenomena led many researchers (famously, Poincaré in his memoir of 1890) to the study of *typical* trajectories. Naturally this study developed hand in hand with the emerging theories of probability and measure theory.

One of the resulting theories is ergodic theory, which can be loosely defined as the measure-theoretic analysis of group-actions on measure spaces. The classical setup is an action of the group of integers $\mathbb{Z}$, which is nothing but a measurable invertible transformation $T : X \to X$. In this setup X is a Borel space and one often assumes that T preserves a finite Borel measure on X, and inquires about the long-term statistical behavior of typical trajectories. More generally, one may replace the action of $\mathbb{Z}$ by the action of a topological group, and/or relax the assumption on the preservation of the measure. As the theory acquired a life of its own, two features emerged. It quickly became apparent that difficult questions can be asked concerning systems which are governed by much simpler laws than Newton's laws of motion and gravitation. Moreover it was realized that despite the desirability of a general abstract theory which would apply to all examples, many examples have special features which can be exploited to obtain a much more detailed understanding. An example of such a seemingly innocuous system is when X is the unit circle, and $T : X \to X$ is rotation by an irrational angle. In this example, as shown by Kronecker and Weyl, one could completely understand the behavior of *all* orbits (they are all equidistributed) and classify *all* invariant measures (Lebesgue measure is the only one).

Note that the above example could be rewritten in fancier terms as $X = \mathbb{R}/\mathbb{Z}$, which is equipped with the transitive action of $\mathbb{R}$ (addition mod 1 or rotation by all possible angles) and T is the $\mathbb{Z}$-action obtained by restricting the $\mathbb{R}$-action to a cyclic subgroup of $\mathbb{R}$. This fancier setup leads to a class of algebraic examples (the so-called *homogeneous spaces*) in which the space X is equipped with the transitive action of a topological group G and the action is obtained by restricting to subgroups of G or using suitable endomophisms of G. A complete dynamical analysis of these richly structured actions is a very active field of research, and requires an array of tools from different mathematical fields (Lie theory, representation theory, number theory, Fourier analysis, to name a few). A feature which sets these examples apart from many other dynamical systems is that in certain cases, as in the case of the irrational rotation, one can obtain complete classification results describing the behavior of *all* orbits, invariant measures, etc. It is this feature which is often responsible for the remarkable effectiveness alluded to above.

The book of Einsiedler and Ward aims to equip the reader with the tools with which to study these remarkable applications of ergodic theory. There are several

more specialized books in the same vein, focusing on a particular result (see [2–5]). The book at hand is different as it aims both to provide the reader with a solid comprehensive background in the main results of ergodic theory, and of reaching nontrivial applications to number theory. Loosely speaking, the foundations are laid in Chapters 1–6 (with some number-theoretic asides on continued fractions) and the fruits, in the form of substantial applications, are harvested in Chapters 7–10. This is very ambitious for a book of about 400 pages.

As the reader has presumably surmised, ergodic theory is not a linear theory with a universally accepted axiomatic framework, or to quote the authors, "is a rather diffuse subject with ill-defined boundaries". There are many choices which the authors had to make regarding the first part of the book, namely choosing the scope of their definitions, the generality of their results, and the assumed prerequisites. This is reflected in the flowchart on page (xi) describing the interdependencies of chapters—there is certainly not a unique way to read this book. For most of the first part of the book, Einsiedler and Ward chose to focus on a measure preserving transformation (i.e. a $\mathbb{Z}$-action), postponing the discussion of more general group actions and stationary measures to Chapter 8. Their discussion in the first few chapters is thorough and more comprehensive than in most other textbooks in ergodic theory. For example, in Chapter 5 they characterize the limit functions appearing in ergodic theorems in terms of both Hilbert-space projections onto the space of invariant functions, and conditional expectations with respect to the algebra of invariant sets. Other examples are detailed treatments of martingale theorems and ergodic decomposition. They sometimes give more than one proof of a result, and sometimes sketch alternative routes. By and large, they prefer long proofs giving detailed information over slick proofs giving only the best-known results. The text is interspersed with many footnotes and asides to the literature for possible extensions and refinements.

In writing the second part of the book, the authors were faced with the problem of choosing representative applications of ergodic theory to number theory. The reader should note that all of the celebrated applications listed above require detailed information about specific systems and prerequisites from various fields. It is not an easy task to choose applications which a reader will find sufficiently interesting and at the same time approachable. The authors have chosen two landmark results: Furstenberg's proof of Szemerédi's theorem (Chapter 7), and the results of Dani and Smillie on measure classification and equidistribution of horocycles on finite-volume hyperbolic surfaces (Chapter 10). These are excellent choices and the authors manage to reach important results without sacrificing detail. Along the way the authors present many other important results from the ergodic-theoretic literature which had previously only appeared in the research literature, e.g. Ledrappier's example of a $\mathbb{Z}^2$-action which is mixing but not mixing of all orders, Mozes' result on mixing of all orders for actions of $\mathrm{SL}_2(\mathbb{R})$, or results on translations on nilmanifolds. In all of these choices the authors display their fine taste. One may worry that the authors have tried to include too much material, but the treatment remains consistently solid and thorough. As in the preceding chapters, the authors often include several proofs of the same result (for instance they prove ergodicity of the geodesic flow both by following Hopf's original argument and via the Mautner property).

The authors were also forced to make difficult decisions concerning the mathematical prerequisites for reading the book. They have chosen to rely on a strong

background in measure theory, functional analysis, and harmonic analysis. Much of the relevant background is recalled in the appendices. On the other hand the authors have included efficient crash courses in hyperbolic geometry, elementary Lie theory (via linear groups), and basics of continued fractions and diophantine approximation. The substantial prerequisites should not deter researchers from basing a graduate course on ergodic theory on this book, as the extra effort of the students will be amply rewarded in terms of both depth of coverage and exciting applications. The book should also be very appealing to more advanced readers already conducting research in representation theory or number theory, who are interested in understanding the basis of the recent interaction with ergodic theory.

The authors plan a sequel to the book, focusing on entropy theory and its relations to the recent work on Littlewood's conjecture. Based on the high standard set by this volume, the second volume will be eagerly anticipated by many readers.

References

1. Arbieto, A., Moreira, C., Matheus, C.: The remarkable effectiveness of ergodic theory in number theory. Ens. Mat. **17**, 1–98 (2009)
2. Bekka, B., Mayer, M.: Ergodic Theory and Topological Dynamics for Group Actions on Homogeneous Spaces. Cambridge University Press, Cambridge (2000)
3. Einsiedler, M., Lindenstraus, E.: Diagonal actions on locally homogeneous spaces. In: Homogeneous Flows, Moduli Spaces and Arithmetic. Clay Math. Proc., vol. 10, pp. 155–241. Amer. Math. Soc., Providence (2010)
4. Furstenberg, H.: Recurrence in Ergodic Theory and Combinatorial Number Theory. Princeton Univ. Press, Princeton (1981)
5. Witte Morris, D.: Ratner's Theorems on Unipotent Flows. University of Chicago Press, Chicago (2005)