Alcom-FT Technical Report Series
ALCOMFT-TR-01-2

The complete analysis of a polynomial factorization algorithm over finite fields

Philippe Flajolet

Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay, France
E-mail: Philippe.Flajolet@inria.fr

Xavier Gourdon

Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay, France
E-mail: Xavier.Gourdon@inria.fr

and

Daniel Panario

School of Mathematics and Statistics, Carleton University, Ottawa, K1S 5B6, Canada.
E-mail: daniel@math.carleton.ca

Communicated by H. Prodinger

Received 2000; revised 2000; accepted 2001

A unified treatment of parameters relevant to factoring polynomials over finite fields is given. The framework is based on generating functions for describing parameters of interest and on singularity analysis for extracting asymptotic values. An outcome is a complete analysis of the standard polynomial factorization chain that is based on elimination of repeated factors, distinct degree factorization, and equal degree separation. Several basic statistics on polynomials over finite fields are obtained in the course of the analysis.

Key Words: Analysis of algorithms, analytic combinatorics, computer algebra, finite fields, factorization algorithms, random polynomials

INTRODUCTION

Factoring polynomials over finite fields intervenes in many areas of computer science and computational mathematics like symbolic computation at large [24], polynomial factorization over the integers [12, 40], cryptography [10, 44, 48], number theory [5], or coding theory [4]. The implications include finding complete

partial fraction decompositions, designing cyclic redundancy codes, computing the number of points on elliptic curves, and building arithmetic public key cryptosystems. In particular, the factorization of *random* polynomials over finite fields is directly needed in the randomized index calculus method for computing discrete logarithms over finite fields [48].

This paper derives basic probabilistic properties of random polynomials over finite fields that are of interest in the study of polynomial factorization algorithms. We show that the main characteristics of random polynomial can be treated systematically by methods of "analytic combinatorics" based on the combined use of generating functions and of singularity analysis. Our object of study is the classical factorization chain which is described in Fig. 1 and which, despite its simplicity, does not appear to have been totally analysed so far. In this paper, we provide a complete average-case analysis.

In asymptotic terms, the algorithm that we study need not be the fastest available at the moment; compare with [25, 26, 34, 55, 56]. For instance, Shoup [55] provides the average-case analysis of an algorithm he proposes and his highly interesting methods are based on estimates for the number of solutions of equations over finite fields and Weil's bounds—the scope is however quite different from the framework of this paper. One of the good reasons for studying the classical chain is that it is representative of what is often implemented in general purpose computer algebra systems and of what is likely to be practically relevant for many problems of "moderate" size. In effect, the book by Geddes *et al.* [28] describes a chain very much similar to ours on which the design of polynomial factorization in the MAPLE system is based—a fact easily confirmed by tracing facilities available under the system. Knuth's authoritative treatise [40] and the remarkable recent book of von zur Gathen and Gerhard [24] expose the fundamental aspects of the classical chain in detail. We refer globally to [40, p. 449] and [24, pp. 393–397] for the historical context of this and closely related algorithms, with some of the originators being Legendre, Gauß, Galois, Berlekamp, Cantor, and Zassenhaus.

An important reason for our interest in these questions is methodological. Indeed, the discipline of analysing completely an algorithm, which is in the line of Knuth's works [39, 40, 41, 42], reveals parameters that are of general interest for polynomial factorization algorithms and for problems that deal with polynomials over finite fields at large. A specific illustration is the paper of Panario and Richmond [52] that deals with testing polynomial irreducibility along lines similar to those of the present paper.

Algorithmic framework. Let $\mathbb{F}_q$ be the finite field with q elements. The algorithmic problem that we address is as follows: given a monic univariate polynomial $f \in \mathbb{F}_q[x]$, find the complete factorization $f = f_1^{e_1} \cdots f_r^{e_r}$, where $f_1, \ldots, f_k$ are pairwise distinct monic irreducible polynomials and $e_1, \ldots, e_r$ are (strictly) positive integers. The basic factorization chain (see Fig. 1) operates in three stages.

```
procedure factor(f : polynomial);
1.      a:=ERF(f);
        {"a" is squarefree}
2.      b:=DDF(a);
        {"b" is a partial fact. into distinct degrees}
        F:=1;
3.      for k from 1 to n do
            F:=F.EDF(b[k],k);
            {refine distinct degree fact. for deg. "k"}
        od;
4.      return(F.factor(f/a));
end;
```

FIG. 1. The complete factorization chain.

ERF: elimination of repeated factors replaces a polynomial

$$f = f_1^{e_1} \cdots f_r^{e_r}$$

by a squarefree one that contains all the irreducible factors of the original polynomial but with exponents reduced to 1:

$$f \mapsto a = \prod_j f_j.$$

DDF: distinct-degree factorization splits a squarefree polynomial into polynomials whose irreducible factors all have the same degree:

$$a \mapsto b_1 \cdot b_2 \cdots b_n \quad \text{where} \quad b_k = \prod_{\deg(f_j)=k} f_j.$$

EDF: equal-degree factorization completely splits a polynomial whose irreducible factors are already constrained to have the same degree:

$$b_k \mapsto b_{k,1} \cdots b_{k,\ell_k} \quad \text{where} \quad \deg(b_{k,j} = k).$$

An often used variant of the first stage is:

SFF: squarefree factorization determines directly the decomposition $f = \prod_i g_i^i$ where the g_i are squarefree and pairwise coprime so that $g_i = \prod_{e_j=i} f_j$.

As implied by the results of Section 2 the difference in costs induced by the two versions, ERF and SFF, is marginal, while consideration of ERF greatly simplifies the whole analysis. We refer globally to [24, Ch. 14] for a complete discussion of this and other algorithmic aspects.

Given a polynomial over the integers, like

$$g = x^9 - 3x^8 + 4x^7 - 3x^6 + 2x^5 - 2x^4 + x^3 + x^2 - 2x + 1,$$

most general-purpose computer algebra will approach the factorization via modular reductions. For instance, the reduction modulo 5 gives

$$f \equiv (g \bmod 5) = x^9 + 2x^8 + 4x^7 + 2x^6 + 2x^5 + 3x^4 + x^3 + x^2 + 3x + 1.$$

Elimination of repeated factors (ERF) is easily achieved by gcd computations and it produces, modulo 5, the partial factorization

$$f = (x+4)a = (x+4)(x^8 + 3x^7 + 2x^6 + 4x^5 + x^4 + 4x^3 + x + 4). \quad (1)$$

The next phase of distinct-degree factorization (DDF) then uncovers a partial factorization of the eighth degree factor as

$$a = b_1 \cdot b_2 = \left[x^2 + 2x + 2\right] \cdot \left[x^6 + x^5 + 3x^4 + x^3 + 3x^2 + x + 2\right]. \quad (2)$$

There, in two successive steps, the groups of factors of degree 1 (there must be two such factors) and of degree 2 (there must be three such factors) have been isolated. Finally, the second and sixth degree polynomials are split by two separate calls to the equal-degree factorization (EDF) algorithm into their linear and quadratic factors:

$$b_1 = (x+3)(x+4), \qquad b_2 = (x^2 + x + 1)(x^2 + x + 2)(x^2 + 4x + 1). \quad (3)$$

In this way, the complete factorization of f modulo 5 is obtained from (1), (2), (3):

$$f = (x+3)(x+4)^2(x^2 + x + 1)(x^2 + x + 2)(x^2 + 4x + 1).$$

Various methods, not discussed in this paper, then permit one to lift the factorization to the integers; see [24, Ch. 15]. Here, this eventually produces the complete factorization over $\mathbb{Z}[x]$,

$$f = (x-1)^2(x^2 - x + 1)(x^2 + x + 1)(x^3 - x^2 + 1).$$

In this case, the squarefree factorization (SFF) approach would only lead to extracting the repeated factor $(x-1)^2$, or $(x+4)^2 \bmod 5$, at an early stage.

Computational model. We *fix* a finite field $\mathbb{F}_q$ with $q = p^m$ (p prime) and consider the ring of polynomials $\mathbb{F}_q[x]$; see [4, 24, 28, 40, 45] for background. The *probabilistic model* assumes all q^n monic polynomials of degree n to be equally likely and all average-case analyses are expressed as asymptotic forms in n, the degree of the polynomial to be factored. The *complexity model* assumes that a basic field operation has cost $\mathcal{O}(1)$, the cost of a sum is $\mathcal{O}(n)$, and the cost of a product, a division or a gcd is $\mathcal{O}(n^2)$, when applied to polynomials of degree $\leq n$. For *dominant asymptotics*, we can freely restrict our attention to polynomial products and gcds. We take as $M(n) = \tau_1 n^2$ the cost of multiplying two polynomials of degree less than n modulo a polynomial of degree n, and as $G(n) = \tau_2 n^2$ the cost of a gcd between a polynomial of degree n and a polynomial of degree at most n. (There, τ_1 and τ_2 are system and implementation dependent constants.) What we have in mind is a general purpose factorization algorithm typically applied to polynomials of moderate size and degree, where operations are often implemented by quadratic algorithms. Similar studies could be conducted using FFT (fast Fourier transform) based algorithms.

Summary of results. Figure 2 summarizes the interplay between probabilistic properties of random polynomials and polynomial factorization. We offer here a few comments.

A random polynomial of degree n is irreducible with a small probability of about $1/n$ and has close to $\log n$ factors on average and with a high probability (Section 1). Thus, the factorization of a random polynomial over a finite field is almost surely nontrivial. Each of the various phases of polynomial factorization has its own "physics" with implications on the corresponding costs. Here is a brief summary.

ERF: The first phase of the factorization is the elimination of repeated factors, ERF. It is deterministic and described in Section 2. This ERF stage returns the *squarefree part* of the original polynomial in which each irreducible factor of the original polynomial appears exactly once (the remaining factors form the *non-squarefree part*). In fact, the squarefree polynomials have a positive density (asymptotically) and the non-squarefree part of a random polynomial is with high probability of degree $\mathcal{O}(1)$ as expressed by Theorem 2.1. In a precise technical sense, the cost of the ERF phase is asymptotically that of a single gcd (Theorem 2.2), so that most of the factorization cost results from the subsequent phases, namely, DDF and EDF.

DDF: The second phase DDF, also deterministic and described in Section 3, splits the squarefree part a of the polynomial to be factored into a product $a = b_1 \cdot b_2 \cdots b_n$, where b_k is formed by the product of all the irreducible factors of a

Phase	*Properties*	*Refs.*
(Sec. 1)	Fraction of irreducibles is $\sim \frac{1}{n}$.	Eq. (8); [4, 40]
	The number of irreducible factors has mean $\sim \log n$ and a limiting Gaussian law.	Eq. (13); [4, 7, 21, 40]
ERF (Sec. 2)	Non squarefree part has size $\mathcal{O}(1)$ on average and with high probability.	Thm. 2.1; [4, 9, 18]
	Algorithmic cost of ERF is $\sim G(n)$ on average and with high probability	Thm 2.2; [18]
DDF (Sec. 3)	Largest degree is $\Theta(n)$ with Dickman distribution and mean $\sim \mathrm{g}\, n$ where $\mathrm{g} \doteq 0.62432$.	Thm 3.1; [30, 50]
	Modified Dickman laws hold for two largest degrees.	Thm 3.2; [30, 50]
	Algorithmic cost of DDF is $\mathcal{O}(nM(n)\log q)$; three stopping rules are compared.	Thm 3.3; [18]
(Sec. 4)	DDF yields a complete factorization with probability close to $e^{-\gamma}$.	Thm 4.1; [18, 32, 38]
	Total degree of unfactored part after DDF has mean $\mathcal{O}(\log n)$, but "soft" distribution tails and standard deviation $\mathcal{O}(\sqrt{n})$.	Thm 4.2; [18, 35]
ERF (Sec. 5)	The probability of repeated irreducible factors of degree k is approximately Poisson$(1/k)$	Thm 5.1; [37]
	Algorithmic cost of ERF is $\mathcal{O}(n^2 \log q)$ on average, as opposed to a worst-case $\mathcal{O}(n^3 \log q)$.	Thm 5.2; [18]

FIG. 2. Main properties of random polynomials of degree n and of corresponding factorizations, together with some relevant references. Here $M(n) := \tau_1 n^2$ represents the cost of a polynomial multiplication and $G(n) := \tau_2 n^2$ the cost of a gcd.

that have degree k. A random polynomial has with high probability several large factors, that is, of degrees $\Theta(n)$. This is quantified by Theorems 3.1 and 3.2 where the Dickman function and some of its relatives serve to express the corresponding probability distributions. Such estimates form the basis of a precise comparison of three stopping rules: the "naïve" rule, the "half-degree" rule and the "early abort" rule whose costs are found to be in the approximate proportion $1 :: \frac{3}{4} :: \frac{2}{3}$; see Theorem 3.3. At the end of the DDF phase, the factorization is complete with a probability ranging asymptotically between 0.56 and 0.67 (Theorem 4.1). In addition, the number of degree values such that more than one irreducible factor

of that degree occurs is typically $\mathcal{O}(1)$, and the total degree passed to EDF is $\mathcal{O}(\log n)$ on average, though it has a large variability (Theorem 4.2).

EDF: The third phase, EDF, is a randomized procedure described in Section 5. It involves a recursive refinement process based on randomized splittings that turns out to be closely related to digital trees, also known as "tries", see [41]. The analysis combines properties of polynomials (Theorem 5.1) with properties of the splitting process (Lemma 5.1). As a consequence, the expected cost of EDF is proved to be comparatively small, being $\mathcal{O}(n^2 \log q)$ (Theorem 5.2).

Precise statements are given in the next sections with an explicit dependence on the field cardinality q. (Some of them involve number-theoretic functions that can be both evaluated and estimated easily.) A simplified picture is as follows. The ERF phase involves with high probability little more than a single polynomial gcd, so that its expected cost is $\mathcal{O}(n^2)$. The DDF phase of cost $\mathcal{O}(n^3 \log q)$ (both on average and in the worst-case) is the one that is most intensive computationally, where control by the "early-abort" strategy is expected to bring gains close to 36%. The last phase of EDF is typically executed less than 50% of the time and its cost, $\mathcal{O}(n^2 \log q)$ on average, is again small compared to that of DDF. A comparison between worst-case costs and average-case costs for each phase is drawn in Section 6.

Note on methodology. An earlier but almost identical version of this paper submitted to another journal dedicated to computing was met with sharp criticism from two referees. One criticism had to do with the asymptotically suboptimal character of the algorithms that we analyse. However, record-breaking computations are only one facet of the story and, from the authors' experience with computer algebra systems, much usage involves polynomials of moderate degree having moderately large coefficients. For this, good implementations are definitely wanted; thus, we claim some justification for interest in algorithms that are suboptimal in the strict asymptotic sense but may very well turn out in many practical contexts to perform better than asymptotically optimal algorithms.

Another criticism had to do with the model of random polynomials that we adopt, namely the uniform distribution over the q^n polynomials of degree n. One referee said: "*most polynomials are not sampled at random over the set of all polynomials*". This is certainly true in a narrow-minded perspective. However, we make the following observations: (i) a large body of algorithms building on top of factorization over finite fields *do* already involve a randomization that is expected to "propagate", see for example Ben-Or's construction of irreducible polynomials [3] or the index calculus method described in [48]; (ii) no one is aware of any explicit construction law that would bias polynomials towards being more likely to be irreducible than factorizable [47, Problem 27]—indeed such a law would be a major

discovery!—so that the randomness assumption, even if somewhat heuristic[1], is one decent way of coping with the current lack of our knowledge; (iii) simulations amply confirm that many varieties of polynomials produced by all sorts of processes behave "as expected" with respect to factorization (see below for a striking example); (iv) accordingly, there exist several theoretical results demonstrating rigorously the fact that various systematic "laws" produce polynomials whose behaviour with respect to factorization is just as predicted by the uniform randomness model. As a typical illustration of the latter point, Hensley [33] has established a "Dirichlet's theorem" for the ring of polynomials over a finite field to the effect that the distribution of irreducibles in any arithmetic progression (*i.e.*, a sequence $a(x)n(x) + b(x)$ with a, b fixed) conforms to "normality".

To demonstrate our point, here is an easily reproducible experiment. We build quite specific polynomials by a deterministic process and examine the mean number of irreducible factors that they contain. Let p_j be the jth prime and let $M(n, r)$ be the Hankel matrix filled with primes starting with p_r: the (i, j) entry of M is $p_{i+j+r-1}$; the test polynomial $\Xi_{n,r}(x) \in \mathbb{Z}[x]$ is the characteristic polynomial of $M(n, r)$. In the experiment, we fix the dimension $n = 20$ and reduce the Ξ polynomials modulo various primes p (actually the primes of rank $4, 16, 64 \ldots$) upon averaging over values $r = 0, \ldots, 99$. Here is the observed mean over each sample of size 100 as compared to the values predicted by the uniform model and given in [40, Ex. 4.6.2.5].

p	7	53	311	1619	8161	38873
Exact mean	3.690	3.607	3.599	3.598	3.597	3.597
Sample mean	3.84	3.57	3.60	3.53	3.40	3.50

We purposely took here a *small* sample of *special* polynomials (characteristic polynomials) associated with *structured* matrices (of Hankel type) built on *particular* coefficients (here the prime numbers). We indeed verify that the empirical data conform quite well to what the uniform randomness model predicts.

This paper constitutes the journal version of an extended abstract presented at the ICALP'96 Conference [18]. It is also organized as a survey of major probabilistic properties of polynomials that are relevant to basic factorization algorithms.

1. ANALYTIC–COMBINATORIAL METHODS

[1] For instance the design of Pollard's "rho method" for integer factoring is entirely based on similar heuristics regarding integers; these were later largely vindicated by Bach [2].

This section gathers basic tools needed to analyse properties of random polynomials. It centres around the use of generating functions, either univariate or multivariate, whose functional relations reflect the algebraic decompositions of various classes of polynomials. The asymptotic analysis of coefficients of generating functions is then attained by means of singularities. The results in this section are classical, but they are needed to set the stage for subsequent analyses. General references for this section are Chapter 3 of Berlekamp's book [4], the exercise section 4.6.2 of Knuth's book [40], the paper [19] or Odlyzko's survey [49] for asymptotic techniques, and the book [20] for the interplay between combinatorial and analytic methods.

1.1. Generating functions

We present first a few general principles that enable one to set up "symbolically" equations for generating functions starting from combinatorial specifications. Given a combinatorial object ω, the formal identity

$$\frac{1}{1-\omega} = 1 + \omega + \omega^2 + \cdots$$

generates symbolically arbitrary sequences composed of ω. Let $\mathcal{I}$ be a family of "primitive" combinatorial objects. Then, the products

$$\mathcal{Q} = \prod_{\omega\in\mathcal{I}} (1+\omega), \qquad \text{and} \qquad \mathcal{P} = \prod_{\omega\in\mathcal{I}} (1-\omega)^{-1}. \tag{4}$$

generate respectively the class of all finite sets and multisets (sets with possible repetitions) of elements taken from $\mathcal{I}$. (This results simply from expansion by distributivity and the remark above concerning the meaning of $(1-\omega)^{-1}$.)

We now specialize the discussion to *monic* polynomials[2] and take $\mathcal{I}$ to be the collection of all monic irreducible polynomials. Then, by the unique factorization property of polynomials, $\mathcal{P}$ gives rise to the collection of all monic polynomials and $\mathcal{Q}$ becomes the collection of all monic squarefree polynomials over $\mathbb{F}_q$. Let z be a formal variable, and $|\omega|$ be the degree of the polynomial ω. The substitution $\omega \mapsto z^{|\omega|}$ in formal sums and products of objects gives rise to counting generating function. For $\mathcal{I}$ itself identified with the formal sum $\sum_{\omega\in\mathcal{I}} \omega$, the generating function is

$$I(z) = \sum_{\omega\in\mathcal{I}} z^{|\omega|} = \sum_n I_n z^n,$$

[2] All subsequent enumeration results are for polynomials normailized so as to be monic.

where I_n is the number of polynomials in $\mathcal{I}$ having degree n. Similarly, the generating functions of $\mathcal{Q}$ and $\mathcal{P}$ are obtained as reflexes of the decompositions (4):

$$\begin{aligned} Q(z) &= \prod_{\omega\in\mathcal{I}}(1+z^{|\omega|}) &&= \prod_{n=1}^{\infty}(1+z^n)^{I_n} \\ P(z) &= \prod_{\omega\in\mathcal{I}}(1-z^{|\omega|})^{-1} &&= \prod_{n=1}^{\infty}(1-z^n)^{-I_n}. \end{aligned} \tag{5}$$

The coefficients $Q_n = [z^n]Q(z)$ and $P_n = [z^n]P(z)$ evaluate to the number of monic squarefree polynomials of degree n, and to the number of monic polynomials of degree n, respectively. Obviously, $P_n = q^n$, and therefore we have *a priori*

$$P(z) = \frac{1}{1-qz}. \tag{6}$$

In other words, we know $P(z)$ elementarily while $I(z)$ and $Q(z)$ are bound by their relations to $P(z)$.

First and foremost, the number of irreducible polynomials, I_n, is determined implicitly from the second equation of (5) that relates it to $P(z)$. A well-known process based on the Moebius inversion formula gives it in explicit form; see [9, p. 41], and Theorem 3.43 in [4, p. 84]. Indeed, taking logarithms and rearranging the sums leads to

$$\log\frac{1}{1-qz} = \sum_{k=1}^{\infty}\frac{I(z^k)}{k}, \quad \text{and} \quad \frac{q^n}{n} = \sum_{k|n}\frac{I_{n/k}}{k}. \tag{7}$$

The relation is then solved for I_n by means of Moebius inversion, which yields

$$I_n = \frac{1}{n}\sum_{k|n}\mu(k)q^{n/k}, \qquad \text{and} \qquad I(z) = \sum_{k=1}^{\infty}\frac{\mu(k)}{k}\log\frac{1}{1-qz^k}.$$

In summary:

FACT. The number I_n of irreducible polynomials over $\mathbb{F}_q[x]$ satisfies

$$I_n = \frac{1}{n}\sum_{k|n}\mu(k)q^{n/k} = \frac{q^n}{n} + \mathcal{O}\left(\frac{q^{n/2}}{n}\right). \tag{8}$$

Thus, a fraction very close to $1/n$ of the polynomials of degree n is irreducible.

This theorem is an analogue of the prime number theorem for integers. As an aside, it was first proven by Gauss in the case of prime fields and it appeared in his posthumous opus [27]; see also [15, 53] for early references.

Next the number Q_n of squarefree polynomials is entirely determined by the knowledge of I_n, given Equation (5). A direct way to make Q_n explicit goes as follows: each polynomial f factors uniquely as $f = s \cdot t^2$, where s is a squarefree polynomial and t is an arbitrary polynomial (separate the irreducible factors of f according to the parity of their exponents). We thus have $P(z) = Q(z) \cdot P(z^2)$, so that

$$Q(z) = \frac{P(z)}{P(z^2)} = \frac{1 - qz^2}{1 - qz}.$$

Thus, the number of squarefree polynomials of degree n in $\mathbb{F}_q[x]$ is

$$Q_n = \begin{cases} q^n & \text{if } n = 0, 1; \\ q^{n-1}(q-1) & \text{if } n \geq 2. \end{cases} \tag{9}$$

This result seems to have appeared for the first time in Carlitz's study [9].

1.2. Parameters

We need extensions of the symbolic method in order to take care of characteristic parameters of polynomial factorization. Let Φ be a class of monic polynomials, and χ some integer–valued parameter on Φ. Then, the bivariate generating function

$$\phi(z, u) = \sum_{\omega \in \Phi} z^{|\omega|} u^{\chi(\omega)}$$

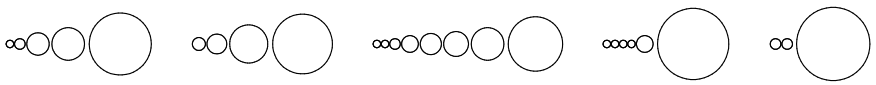
is such that the coefficient $[z^n u^k]\phi(z, u)$ represents the number of polynomials of degree n in Φ with χ–parameter equal to k. Bivariate generating functions contain all information related to the distribution of a parameter. Averages and standard deviations are obtained by successive differentiations of these bivariate generating functions with respect to u (the variable marking the parameter), and then by setting $u = 1$.

If the parameter χ is additive, meaning that $\chi(f \cdot g) = \chi(f) + \chi(g)$ for relatively prime $f, g \in \mathbb{F}_q[x]$, then product decompositions of the form (5) generalize under the translation rule $\omega \mapsto z^{|\omega|} u^{\chi(\omega)}$. The technique of rearranging logarithms of infinite products employed in (7) is then useful in simplifying such expressions. For instance, the bivariate generating function of all polynomials with χ the number of their irreducible factors (multiplicity being counted) is

$$P(z, u) = \prod_{n \geq 1} (1 - uz^n)^{-I_n} = \exp\left(\sum_{k=1}^{\infty} u^k \frac{I(z^k)}{k} \right). \tag{10}$$

Much use is made of this technique in the next four sections.

A graphical rendering of the factorization types of 5 polynomials of degree 100 over $\mathbb{F}_7$:



Each circle represents the degree of an irreducible factor with the circle areas being proportional to degrees. On this experiment, the 5 factorization types found are

$$[1,2,9,19,69],\ [3,7,26,64],\ [1,1,2,5,8,11,19,53],\ [1,1,1,1,5,91],\ [2,2,96].$$

This illustrates the property that factorizations over finite fields are usually nontrivial and that there tends to be one or a few irreducible factors of comparatively large degree (Theorems 3.1 and 3.2).

FIG. 3. Simulations of random polynomial factorizations.

1.3. Asymptotic analysis

Generating functions encode complete information on their coefficients. Furthermore, the behaviour of a generating function near its dominant positive singularity ("dominant" means in this context "of smallest modulus") is an important source of coefficient asymptotics.

The generating functions $f(z)$ to be studied in this paper are singular at $z = 1/q$ and most have there an isolated singularity. Consequently, their coefficients $f_n = [z^n]f(z)$ satisfy an estimate of the form $f_n \sim q^n\theta(n)$, where $\limsup |\theta(n)|^{1/n} = 1$ is a subexponential factor that reflects the nature of the singularity at $z = 1/q$. In particular, an expansion near $z = 1/q$ of the form

$$f(z) = \frac{1}{(1-qz)^\alpha}\left(\log\frac{1}{1-qz}\right)^k (1+o(1)) \tag{11}$$

translates into coefficients by the method known as singularity analysis [19, 49]:

$$f_n = [z^n]f(z) = q^n\frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^k\,(1+o(1)). \tag{12}$$

The transition from (11) to (12) is ensured by transfer theorems that require analytic continuation of $f(z)$ outside its circle of convergence, a condition that is verified by inspection in most cases considered here.

A typical instance is the analysis of the number of irreducible factors in a random polynomial of degree n. The bivariate generating function appears already in (10). Differentiating with respect to u, setting $u = 1$ and then analysing the singularity at $u = 1$ provides moment estimates [40, Ex. 4.6.2.5]; methods that build further on singularity analysis even give access to a limiting distribution [7, 21].

FACT. Let X_n be the random variable that represents the number of irreducible factors in a random polynomial of degree n. Then the mean $\mathrm{E}(X_n)$ and variance $\mathrm{Var}(X_n)$ satisfy

$$\mathrm{E}(X_n) = \log n + \mathcal{O}(1), \qquad \mathrm{Var}(X_n) = \log n + \mathcal{O}(1). \tag{13}$$

In addition the distribution of $(X_n - \mathrm{E}(X_n))/\sqrt{\mathrm{Var}(X_n)}$ is asymptotically normal.

Thus, the factorization of a polynomial is with high probability nontrivial; see Figure 3 for a graphic illustration.

Another useful asymptotic coefficient extraction method is Darboux's method [13, 32] whose principle is as follows: if an analytic function $f(z)$ defined in the closed disk $|z| \leq 1$ is k times continuously differentiable ($\mathcal{C}^k$) on $|z| = 1$, then its coefficients satisfy

$$[z^n] f(z) = o(1/n^k). \tag{14}$$

A recourse to Darboux's method is needed in Section 4, given the existence of natural boundaries for generating functions that occur specifically there.

1.4. The permutation model and large fields

The following fact is well-known: As the cardinality q of the field $\mathbb{F}_q$ goes to infinity (n staying fixed!), the joint distribution of the degrees of the irreducible factors in a random polynomial of degree n converges to the joint distribution of the lengths of cycles in a random permutation of size n. This property is visible at the generating functions level, whenever a generating function of polynomials taken at z/q converges (as $q \to \infty$) to the corresponding exponential generating function of permutations. For instance, the generating function of all monic polynomials, when normalized with the change of variable $z \mapsto z/q$, is

$$P\left(\frac{z}{q}\right) = \frac{1}{1-z} = \sum_{n=0}^{\infty} n! \frac{z^n}{n!},$$

which is also the exponential generating function of all permutations. Similarly, we have

$$I\left(\frac{z}{q}\right) \underset{(q\to\infty)}{\to} \log \frac{1}{1-z} = \sum_{n=1}^{\infty} (n-1)! \frac{z^n}{n!},$$

the exponential generating function of cyclic permutations. The relation (10) between $P(z)$ and $I(z)$ that expresses the unique factorization property for poly-

```
procedure ERF(f : polynomial);
       g := gcd(f,f'); h := f/g; k := gcd(g,h);
       while k<>1 do  g := g/k; k := gcd(g,h)  od;
       if  g <> 1 then h := h*ERF(g^(1/p)) fi;
       return(h);    {"h" is squarefree}
end;
```

FIG. 4. The elimination of repeated factors algorithm (ERF).

nomials reduces as $q \to \infty$ to

$$\frac{1}{1-z} = \exp\left(\log \frac{1}{1-z}\right),$$

which is well-kknown to express the unique decomposition of permutations into cycles; see [20, 29] for background.

This gives rise to a useful heuristic: probabilistic properties of polynomial factorization are expected to have a shape resembling that of corresponding properties of the cycle decomposition of permutations[3]. A precise instance of this fact is mentioned by Greene and Knuth [32] in connection with the probability that a random polynomial admits factors of distinct degrees, which, for large q and large n is found to approach $e^{-\gamma}$. Such phenomena are clearly useful in understanding the behaviour of polynomial factorizations in fields of "large" cardinalities—the case of "small" fields then appearing as a "q–deformation" with a typically mild effect. Section 4 provides several examples related to quantifying the output of the DDF factorization.

2. ELIMINATION OF REPEATED FACTORS (ERF)

The first step in the factorization chain summarized in Figure 4 is the elimination of its repeated factors. In characteristic zero, this is achieved thanks to the following property: *A gcd between a polynomial f and its derivative f' extracts all the repeated factors of f.*

In $\mathbb{F}_q$ with $q = p^m$ and p a prime number, additional control is needed in order to deal with pth powers whose derivatives are identically 0. The first line of the algorithm in Figure 4 collects in h one copy of each of the irreducible factors of f, except the ones whose multiplicity is a multiple of p. The while

[3] There is a similar "resemblance" between properties of natural numbers whose representation in base q has length n and polynomials of degree n in $\mathbb{F}_q[x]$. For additive properties, this corresponds to a rough equivalence between additions with or without carries; for mixed additive-multiplicative properties, the analogy lies however far deeper and is accordingly much less understood.

loop stores in g the factors whose multiplicity is a power of p, without eliminating their repetitions. The last part of the algorithm adds to h the factors in g with repetitions eliminated. The auxiliary computation of pth roots, $g^{1/p}$, is performed in the classical way [28, p. 344], using the identity $(a^p + b^p)^{1/p} = (a + b)$. There exist alternative algorithms giving the full squarefree factorization (see [40], Ex. 4.6.2.36); however, as Theorem 2.1 below shows, the reduction of degree induced by the additional computational effort is only $\mathcal{O}(1)$. We have opted for the elimination of repeated factors by ERF (rather than the more common squarefree factorization SFF) as the analysis develops more transparently while the overall costs of the complete factorization chain are only *very* marginally affected by the algorithm chosen for the first stage.

THEOREM 2.1. (i) *A random polynomial of degree $n \geq 2$ in $\mathbb{F}_q[x]$ has a probability $1 - 1/q$ to be squarefree.*

(ii) *The total degree ρ of the non-squarefree part of a random polynomial of degree n has an expected value that tends to the limit*

$$C_q = \sum_{k \geq 1} \frac{k I_k}{q^{2k} - q^k};$$

in addition, the quantity C_q satisfies $C_q \sim 1/q$ as $q \to \infty$.

(iii) *The tail probabilities of ρ decay exponentially fast:*

$$\Pr(\rho(f) = k,\ |f| = n) = \mathcal{O}\left(\left(\frac{2}{3}\right)^k\right).$$

Proof. Part (i) is a classical result that we have already derived in (9). As for part (ii), the bivariate generating function of the degree of the non-squarefree part of monic polynomials in $\mathbb{F}_q[x]$ is, by the symbolic method of Section 1,

$$P(z,u) = \prod_{k \geq 1} \left(1 + \frac{z^k}{1 - u^k z^k}\right)^{I_k}.$$

The mean degree of the non-squarefree part is obtained by setting $u = 1$ in the derivative of $P(z,u)$ with respect to u and the asymptotic estimate follows by singularity analysis,

$$\begin{aligned} \left.\frac{\partial P(z,u)}{\partial u}\right|_{u=1} &= P(z) \cdot \sum_{k \geq 1} I_k \frac{z^{2k}}{1 - z^k} \\ &\underset{z \to 1/q}{\sim} \frac{1}{1 - qz} \sum_{k \geq 1} I_k \frac{q^{-2k}}{1 - q^{-k}}, \end{aligned}$$

hence the stated value of C_q. The asymptotic limit of C_q as $q \to \infty$ is obtained from there by the expansion $k\, I_k = q^k + \mathcal{O}(q^{k/2})$.

As regards part (iii), consider the function $P(z, 3/2)$ and compare it with $Q(z)$, the generating function of squarefree polynomials:

$$\frac{P(z,3/2)}{Q(z)} = \prod_{k\geq 1} \left(1 + \frac{z^{2k}(\frac{3}{2})^k}{(1+z^k)(1-(\frac{3}{2})^k z^k)}\right)^{I_k}.$$

The infinite product is convergent and analytic in a disk that properly contains $|z| \leq 1/q$. Thus, $P(z, 3/2)$ has a simple pole at $z = 1/q$, and by singularity analysis, one has

$$\frac{1}{q^n}[z^n]P(z,3/2) \equiv \frac{1}{q^n} \sum_{|f|=n} \left(\frac{3}{2}\right)^{\rho(f)} = \mathcal{O}(1).$$

Thus the expectation $\mathrm{E}((3/2)^\rho)$ remains bounded, which implies the exponential tail bound. ▮

Theorem 2.1 has meaningful consequences for the recursive structure of the `factor` procedure. The exponential tail of Part (iii) implies that any polynomially bounded function of the non-squarefree part stays of order $\mathcal{O}(1)$ on average. In particular, the overall cost of the recursive calls (Step 4 in the top-level factor procedure of Figure 1) is $\mathcal{O}(1)$ on average. Accordingly, the ERF phase has a cost entirely dominated by that of its first gcd, namely $\gcd(f, f')$.

THEOREM 2.2. *The expected cost of the ERF phase applied to a random polynomial of degree n is asymptotically that of a single gcd,*

$$\overline{\tau ERF}_n \sim \tau_2 n^2.$$

3. DISTINCT-DEGREE FACTORIZATION (DDF)

The second stage of the factorization chain is the distinct-degree factorization and it involves splitting a squarefree polynomial into polynomials whose irreducible factors all have the same degree. This means expressing the squarefree polynomial a in the form $b_1 \cdot b_2 \cdots b_n$ where b_k is the product of all the irreducible factors of degree k that figure in a (assumed to be squarefree). The basic algorithm is described in Fig. 5 and its design relies on the following property (see, e.g., [45],

```
procedure DDF(a : polynomial);
["a" is a monic squarefree polynomial]
       n := deg(a); g := a; h := x;
       for k := 1 to n do
1.         h := h^q mod g;
2.         b[k] := gcd(h-x,g);
3.         g := g/b[k];
           {"a" without factors of deg <=k}
4.         if b[k] <> 1 then h := h mod g fi;
       od;
       return(b[1].b[2]...b[n]);
       {b[k] is prod. of irred. of deg.=k}
end;
```

FIG. 5. The "basic strategy" of the distinct-degree factorization algorithm (DDF).

p. 91, Theorem 3.20): *For $k \geq 1$, the polynomial $\Pi_k := x^{q^k} - x \in \mathbb{F}_q[x]$ is the product of all monic irreducible polynomials in $\mathbb{F}_q[x]$ whose degree divides k.* Thus sweeping over successive values of k and computing gcd's with Π_k leads to a partial factorization. Three different stopping rules can be considered:

— The *basic strategy* explores all the values $k = 1 \,..\, n$ and corresponds to the version given in the algorithm in Fig. 5.

— The *half-degree strategy* consists in stopping the DDF loop when $k = n/2$, since at that moment the remaining factor is either 1 or irreducible.

— The *early-abort strategy* stops the main loop of DDF as soon as $2k$ exceeds the degree of the remaining factor g. In this case, the remaining factor is by necessity irreducible.

The first rule is too naïve to be of algorithmic interest but it serves the purpose of introducing the basic methods needed. The analyses are carried out in Subsection 3.2. They benefit from informations regarding the two largest irreducible factors of a random polynomial, a topic that we address first in Subsection 3.1. See Figures 6 and 7 for numerical data.

3.1. Distribution of largest degrees of factors

A random polynomial has with high probability several irreducible factors whose degree is of the order of n. The distribution of the largest degree among the irreducible factors of a random polynomial over $\mathbb{F}_q$ underlies many problems dealing with polynomials over finite fields. For instance, information on

this distribution is useful when computing discrete logarithms in order to discard polynomials that cannot be written in terms of smooth ones [48].

Specifically, we consider the two largest degrees $D_n^{[1]}$, $D_n^{[2]}$ of the distinct factors of a random polynomial of degree n in $\mathbb{F}_q$. Statistics on the largest degree of the irreducible factors of a random polynomial in $\mathbb{F}_q[x]$ have already been considered in the literature. Car [8] first obtained an asymptotic expression for the cumulative distribution function of $D_n^{[1]}$ in terms of the *Dickman function*. This function is a classical number-theoretic function that was originally introduced to describe the distribution of the largest prime divisor of a random integer. We refer to [14, 16, 41, 43, 57] as general references to the Dickman function in relation to arithmetic problems, especially Tenebaum's analytic treatment of the Dickman function [57], the paper of Knuth and Trabb Pardo on integer factorization [43], and Knuth's account in [41, Sec. 4.5.4].

FACT. The distribution of the largest degree $D_n^{[1]}$ satisfies for all $x \in (0,1)$:

$$\lim_{n\to\infty} \Pr\{D_n^{[1]} \le x\} = F(x),$$

where $F(x) = \rho(1/x)$ and ρ denotes the Dickman function. The Dickman function is defined as the unique continuous solution of the difference-differential equation

$$\rho(u) = 1 \qquad (0 \le u \le 1), \qquad\qquad u\rho'(u) = -\rho(u-1) \qquad (u > 1). \quad (15)$$

In particular, one has

$$\mathrm{E}(D_n^{[1]}) \sim \mathrm{g}\, n,$$

where $\mathrm{g} \doteq 0.62432$ is known as Golomb's constant; see [17].

In the context of this paper, fine properties of $D_n^{[1]}$, $D_n^{[2]}$ provide insight on the stopping condition for the DDF stage. The statements of Theorems 3.1 and 3.2 that follow are borrowed[4] from [50]. The proofs given in [50] rely on an original approach of Gourdon [30, 31] that leads to local limit theorems. Indeed, Theorems 3.1, 3.2 below, give the asymptotic distribution of the two largest degrees $D_n^{[1]}, D_n^{[2]}$ where the limit density functions are accessed via their Laplace transforms. A key rôle is played by the exponential integral function E defined by

$$E(t) = \int_t^{+\infty} \frac{e^{-v}}{v}\, dv.$$

[4] The authors take this opportunity to correct mistakes in the statements of Theorems 1 and 2 of [50]: the error terms inadvertently written there as multiplicative terms should be read as additive correction terms (in accordance to the proofs that stand). See Tenenbaum's informative review of [50] in *Mathematical Reviews*, 2001, in press.

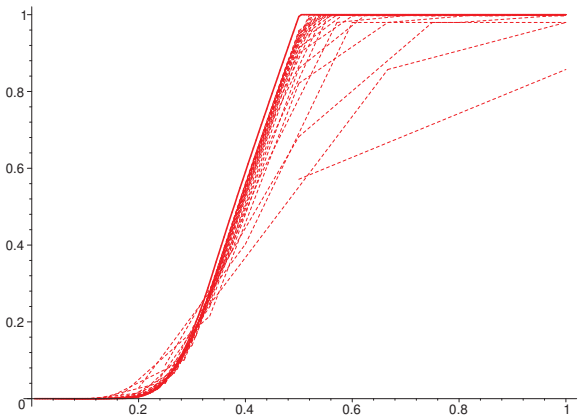


FIG. 6. A plot of the "density" of the largest irreducible factor degree for polynomials of $\mathbb{F}_7[x]$: the representation of $(xn)\Pr(D_1 = xn)$, for degrees $n = 1, \ldots, 25$ and $x \in [0, 1]$ (dashed lines) shows fast convergence against the limit $f(x)$ (continuous curve).

THEOREM 3.1. *The largest degree $D_n^{[1]}$ among the irreducible factors of a random polynomial of degree n over $\mathbb{F}_q$ satisfies the "local limit" estimate*

$$\Pr(D_n^{[1]} = m) = \frac{1}{m} f\left(\frac{m}{n}\right) + \mathcal{O}\left(\frac{\log n}{m^2}\right),$$

where $f(\mu)$ is a continuous function that is defined by the inverse Laplace integral:

$$f(\mu) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} \frac{e^{-E(\mu h)}}{h}\, e^{(1-\mu)h}\, dh. \tag{16}$$

In other words, the largest degree is on average $\mathcal{O}(n)$, and the probability that its value equals m is about $1/m$ with a modulation factor that involves the Dickman function. (It can be verified by direct Laplace transform calculations that $f(\mu) = \rho(1/\mu - 1)$, with ρ the Dickman function as defined by (15).) The next theorem shows that similar estimates involving "Dickman-like" functions hold if a gap is imposed or if one considers the joint distribution of the two largest factors.

THEOREM 3.2. *The two largest degrees $D_n^{[1]}$ and $D_n^{[2]}$ of the distinct factors of a random polynomial of degree n in $\mathbb{F}_q$ satisfy*

(i) for $0 \le m \le n$,

$$\Pr(D_n^{[1]} = m, D_n^{[2]} \le m/2) = \frac{1}{m} g_1\left(\frac{m}{n}\right) + \mathcal{O}\left(\frac{\log n}{m^2}\right),$$

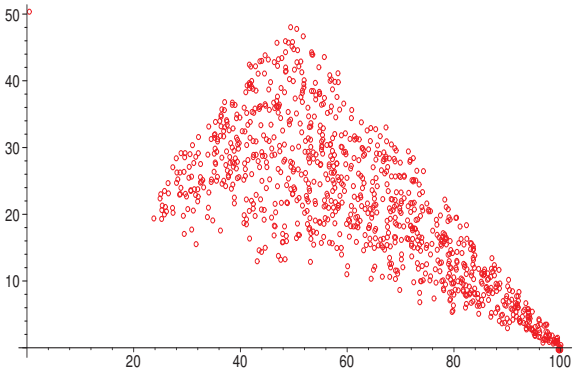


FIG. 7. Two largest degrees for 1,000 random polynomials of degree $n = 100$ in $\mathbb{F}_7[x]$: each circle represents a value of the pair (D_1, D_2).

where $g_1(\mu)$ is defined by the inverse Laplace integral,

$$g_1(\mu) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} \frac{e^{-E(\mu h/2)}}{h}\, e^{(1-\mu)h}\, dh;$$

(ii) for $0 \le m_2 < m_1 \le n$,

$$\Pr(D_n^{[1]} = m_1, D_n^{[2]} = m_2) = \frac{1}{m_1 m_2} g_2\left(\frac{m_1}{n}, \frac{m_2}{n}\right) + \mathcal{O}\left(\frac{\log n}{m_1 m_2^2}\right),$$

where $g_2(\mu_1, \mu_2)$ is defined by

$$g_2(\mu_1, \mu_2) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} \frac{e^{-E(\mu_2 h)}}{h}\, e^{(1-\mu_1-\mu_2)h}\, dh.$$

Figure 7 exemplifies the phenomena at stake by means of simulations. Estimates similar to Theorems 3.1 and 3.2 hold for largest cycles in permutations [54] (this is in agreement with the earlier discussion of the random permutation model of Section 1.4) and for integers as shown by Knuth and Trabb Pardo in [43]. Arratia, Barbour, and Tavaré [1] present an interesting asymptotic model, the Poisson-Dirichlet process, that puts these facts in perspective and covers random mappings, integer partitions, permutations, integers, as well as polynomials.

3.2. Analysis of the distinct-degree factorization

The main result of this section, Theorem 3.3, provides a quantitative comparison of the three stopping rules for the DDF algorithm. It is based on three lemmas, one for each of the strategies to be analysed.

We first fix some notation. The DDF algorithm in its basic version is specified in Fig. 5. The computation in Step 1 is done by means of the classical *binary powering* method [40, p. 441-442] that leads to introducing two number-theoretic functions.

DEFINITION 3.1. The function $\nu(q)$ is the number of ones in the binary representation of q. The function $\lambda(q)$ is defined as

$$\lambda(q) = \lfloor \log_2 q \rfloor + \nu(q) - 1, \tag{17}$$

and it represents the number of products needed to compute $h^q \mod g$ by binary powering.

By the exponential tail result of Theorem 2.1, we need only consider the cost of DDF applied to the squarefree part a of the input polynomial f, and our subsequent analyses are all relative to the statistics induced by a random input f of degree n.

Let d_k denote the degree of polynomial g when the kth iteration of the main loop of DDF starts. The parameter d_k is also the sum of the degrees of the distinct factors of f with degree $\geq k$.

THEOREM 3.3. *The expected cost of the DDF phase satisfies*

$$\overline{\tau DDF}_n^{(S)} \sim \mu^{(S)} \left(\lambda(q)\tau_1 + \tau_2\right) n^3 \qquad \lambda(q) = \lfloor \log_2 q \rfloor + \nu(q) - 1,$$

where $\mu^{(S)}$ is a constant depending on the strategy S,

$$\begin{aligned} \mu^{(B)} &= \frac{5}{12} \doteq 0.46667, \qquad \mu^{(HD)} = \frac{5}{16} = 0.3125, \\ \mu^{(EA)} &= \frac{5}{12} - \frac{1}{3}\int_0^\infty e^{-2x} \exp\left(-\int_x^\infty \frac{e^{-u}}{u}\,du\right) \frac{1-x^2}{x}\,dx \doteq 0.2668903307, \end{aligned}$$

for the basic (B), half-degree (HD) *or* early-abort (EA) *strategy, respectively.*

Proof. The cost of the basic DDF is $C_1 + C_2 + C_3 + C_4$, where C_j denotes the cost of line number j in the algorithm of Fig 5. Let $\overline{C_j}$ be the expectation of C_j. The mean number of irreducible factors of f is $\mathcal{O}(\log n)$, so that $\overline{C_3} + \overline{C_4} = \mathcal{O}(n^2 \log n)$; thus it suffices to consider $C_1 + C_2$.

The quantity d_k is the sum of the degrees of the distinct factors of f with degree $\geq k$. Then, the quantity $C_1 + C_2$ is equal to $(\lambda(q)\tau_1 + \tau_2)\sigma$, where

$$\sigma = \sum_{1 \leq k \leq N} d_k^2, \tag{18}$$

with N the stopping value for the DDF loop. We have $N = n$ for the basic strategy, $N = n/2$ for the half degree rule, $N = \max\{D_1/2, D_2\}$ for the early abort strategy, where D_1, D_2 are the largest and second largest degree of the distinct irreducible factors of the polynomial. The theorem follows from the three lemmas below. ∎

By (18) taken with $N = n$, the analysis of the basic strategy only involves an additive parameter of polynomial factorizations. It is thus dealt with directly by bivariate generating functions and singularity analysis, as summarized in Section 1. The estimate also serves (by difference) in the analysis of the other two strategies.

LEMMA 3.1. *The expected value of σ in the basic strategy satisfies, as $n \to \infty$,*

$$\overline{\sigma}_n^{(B)} \sim \frac{5}{12}\, n^3.$$

Proof. The parameter d_k is by definition the sum of the degrees of the distinct factors of f with individual degree $\geq k$ and the parameter $\sigma^{(B)}$ of (18) is $\sigma^{(B)} = \sum_{k=1}^{n} d_k^2$. The bivariate generating function of d_k is, by the basic decompositions,

$$P_k(z,u) = \prod_{j<k} \left(\frac{1}{1-z^j}\right)^{I_j} \prod_{j\geq k} \left(1 + u^j \frac{z^j}{1-z^j}\right)^{I_j}.$$

The expected value of $\sigma^{(B)}$ is then given by

$$\overline{\sigma}_n^{(B)} = \frac{1}{q^n}[z^n]\xi(z), \qquad \xi(z) = \sum_{k\geq 1} \left(\frac{\partial^2 P_k}{\partial u^2}(z,u) + \frac{\partial P_k}{\partial u}(z,u)\right)\Big|_{u=1}.$$

The basic relation (5) and a computation of $\xi(z)$ by logarithmic derivatives imply that $\xi(z) = P(z)\,(\xi_1(z) + \xi_2(z))$ where

$$\xi_1(z) = \sum_{j\geq 1} j^3 I_j (z^j - z^{2j}) \qquad \text{and} \qquad \xi_2(z) = \sum_{k\geq 1} \left(\sum_{j\geq k} j I_j z^j\right)^2.$$

The function $\xi_1(z)$ is a simple variant of $I(z)$, namely

$$\xi_1(z) = \left(z\frac{d}{dz}\right)^3 \left(I(z) - \frac{1}{8} I(z^2)\right).$$

Thus, $\xi_1(z)$ has its dominant singularity at $z = 1/q$ and near this point,

$$P(z)\xi_1(z) \sim \frac{2}{(1-qz)^4}.$$

Singularity analysis then yields $[z^n]P(z)\xi_1(z) \sim q^n\, n^3/3$.

We next turn to $\xi_2(z)$. The estimate $jI_j = q^j + \mathcal{O}(q^{j/2})$ entails, for $|z| < 1+\eta$,

$$\sum_{j\geq k} jI_j \left(\frac{z}{q}\right)^j = \frac{z^k}{1-z} + S_k(z), \quad S_k(z) = \mathcal{O}\left(\frac{z^k}{q^{k/2}}\right),$$

so that

$$\xi_2\left(\frac{z}{q}\right) = \frac{z^2}{1+z}\,\frac{1}{(1-z)^3} + S(z), \quad S(z) = \sum_{k\geq 1}\left(\frac{2z^k S_k(z)}{1-z} + S_k(z)^2\right).$$

The bounds satisfied by $S_k(z)$ make $S(z)$ regular beyond the unit disk. Thus, singularity analysis can be applied to the function $P(z/q)\xi_2(z/q)$. There are two dominant singularities at 1 and -1, but the one at $z = 1$ has greater weight. The resulting estimate, as $z \to 1$,

$$P\left(\frac{z}{q}\right)\xi_2\left(\frac{z}{q}\right) \sim \frac{1}{2(1-z)^4},$$

implies that the nth coefficient is asymptotic to $n^3/12$. Thus, finally, $\overline{\sigma}_n^{(B)} \sim (1/3 + 1/12)n^3 = 5n^3/12$. ▮

The analysis of the half-degree rule is related to Theorem 3.1 since it involves the distribution of $D^{[1]}$. In fact, it only depends on areas where the Dickman function simplifies so that a direct argument can be applied.

LEMMA 3.2. *The expected value of σ for the half-degree rule satisfies, as $n \to \infty$,*

$$\overline{\sigma}_n^{(HD)} \sim \frac{5}{16}\, n^3.$$

Proof. We now have $\sigma^{(HD)} = \sum_{k\leq n/2} d_k^2$. It suffices to consider the difference $\sigma' = \sigma^{(B)} - \sigma^{(HD)} = \sum_{n/2<k\leq n} d_k^2$. If the largest degree $D_n^{[1]}$ satisfies $D_n^{[1]} \leq n/2$, we have $\sigma' = 0$. Otherwise we have $\sigma' = (D_n^{[1]} - \lfloor n/2 \rfloor)\,(D_n^{[1]})^2$ since there can be only one factor of degree larger than $n/2$, namely $D_n^{[1]}$. Thus, the mean value of σ' is given by

$$\overline{\sigma'}_n = \sum_{n/2<k\leq n} \Pr(D_n^{[1]} = k)\left(k - \left\lfloor \frac{n}{2} \right\rfloor\right) k^2. \tag{19}$$

By the symbolic method of Section 1, the generating function of polynomials for which all factors have degree $\leq m$ is

$$\chi_m(z) = \prod_{k=1}^{m} \left(\frac{1}{1-z^k} \right)^{I_k}. \tag{20}$$

Thus, the generating function of polynomials for which $D_n^{[1]} = m$ is

$$\phi_m(z) = \chi_m(z) - \chi_{m-1}(z) = \chi_m(z) \left(1 - (1-z^m)^{I_m}\right), \tag{21}$$

and the probability $\Pr(D_n^{[1]} = k)$ is related to this generating function by

$$\Pr(D_n^{[1]} = k) = \frac{1}{q^n}[z^n]\phi_k(z).$$

When $k > n/2$, the nth coefficient of $\phi_k(z)$ is obtained from

$$\phi_k(z) = P(z)\left(1 - (1-z^k)^{I_k}\right) \prod_{j>k} (1-z^j)^{I_j} = P(z)\left(I_k z^k + \mathcal{O}(z^{n+1})\right)$$

which entails $\Pr(D_n^{[1]} = k) = I_k/q^k \sim 1/k$ for $n/2 < k \leq n$. Plugging this estimate into (19) gives $\overline{\sigma'}_n \sim \frac{5}{48}\, n^3$. Thus $\overline{\sigma}_n^{(HD)} = \overline{\sigma}_n^{(B)} - \overline{\sigma'}_n \sim \frac{5}{16}\, n^3$. ▮

The early-abort strategy needs to be handled in a more technical way. There is a striking parallel with the analysis of integer factoring by trial divisions, as given by Knuth and Trabb-Pardo [43]. The joint distribution of D_1, D_2 stated in Theorem 3.2 now intervenes.

LEMMA 3.3. *The expected value of σ in the early-abort rule satisfies, as $n \to \infty$,*

$$\overline{\sigma}_n^{(EA)} \sim \delta\, n^3,$$

where

$$\delta = \frac{5}{12} - \frac{1}{3}\int_0^\infty e^{-2x} \exp\left(-\int_x^\infty \frac{e^{-u}}{u}\, du \right) \frac{1-x^2}{x}\, dx \doteq 0.2668903307. \tag{22}$$

Proof. (i) *Algebra.* As above, we denote by D_1 the degree of the largest irreducible factor of f, and by D_2 the degree of the second largest irreducible factor of f (set $D_2 = 0$ if f is irreducible). The iteration is now aborted at step

$k_0 = \max\{\lfloor D_1/2 \rfloor, D_2\}$ and $\sigma^{(EA)} = \sum_{k \le k_0} d_k^2$. Consider the difference

$$\sigma'' = \sigma^{(B)} - \sigma^{(EA)} = \sum_{\max\{\lfloor D_1/2 \rfloor, D_2\} < k \le n} d_k^2.$$

We need to prove the mean-value estimate $\overline{\sigma''}_n \sim (\frac{5}{12} - \delta) n^3$.

We have $\sigma'' = (D_1 - \lfloor D_1/2 \rfloor) D_1^2$ if $D_1/2 \ge D_2$, and $\sigma'' = (D_1 - D_2) D_1^2$ if $D_1/2 < D_2$. Thus, the mean value of σ'' is

$$\begin{aligned} \overline{\sigma''}_n &= \sum_{D_1=1}^{n} D_1^2 \left(D_1 - \left\lfloor \frac{D_1}{2} \right\rfloor \right) \Pr(D_n^{[1]} = D_1, D_n^{[2]} \le D_1/2) \\ &+ \sum_{D_2 < D_1 \le 2D_2} D_1^2 (D_1 - D_2) \Pr(D_n^{[1]} = D_1, D_n^{[2]} = D_2). \end{aligned} \tag{23}$$

Hence, the generating function $\Psi(z)$ of the cumulated values of the parameter σ'' is expressible in terms of the generating function $\widetilde{\phi}_{D_1}(z)$ of polynomials for which $D_n^{[1]} = D_1, D_n^{[2]} \le D_1/2$, and the generating function $\phi_{D_1, D_2}(z)$ of polynomials for which $D_n^{[1]} = D_1, D_n^{[2]} = D_2$. By symbolic methods, the generating function of polynomials for which $D_n^{[1]} = m$ and $D_n^{[2]} \le m/2$ is

$$\widetilde{\phi}_m(z) = \chi_{\lfloor m/2 \rfloor}(z) \frac{I_m z^m}{1 - z^m}, \tag{24}$$

with $\chi_m(z)$ defined in (20). In the same way, the generating function of polynomials with $D_1^{[n]} = m_1$ and $D_2^{[n]} = m_2$, $m_2 < m_1$ is

$$\phi_{m_1, m_2}(z) = \phi_{m_2}(z) \frac{I_{m_1} z^{m_1}}{1 - z^{m_1}}, \tag{25}$$

with $\phi_m(z)$ defined in (21). Thus,

$$\Psi(z) = \sum_{D_1} \left(D_1 - \left\lfloor \frac{D_1}{2} \right\rfloor \right) D_1^2 \, \widetilde{\phi}_{D_1}(z) + \sum_{D_2 < D_1 \le 2D_2} (D_1 - D_2) D_1^2 \, \phi_{D_1, D_2}(z).$$

(ii) *Analysis.* The analysis of the generating function $\Psi(z)$ near the positive singularity $z = 1/q$ is done by approximating sums with integrals (Euler-Maclaurin summation). The following property summarizes what is needed.

LOGARITHMIC REMAINDER ESTIMATE. The remainders of the logarithmic series, $r_m(z) := \sum_{k > m} z^k/k$, are approximable in terms of the exponential integral,

$$r_m(e^{-h}) = E(mh) + \mathcal{O}\left(\frac{1}{m} \right), \tag{26}$$

where the $\mathcal{O}$–error term is uniform with respect to $h > 0$.

Justifying this for any $h > 0$ only requires the Euler Maclaurin formula and "subtraction of singularities",

$$\begin{aligned} r_m(e^{-h}) &= \int_h^{+\infty} \left(\sum_{k>m} e^{-ku} \right) du = \int_{mh}^{+\infty} e^{-v} \frac{v/m}{e^{v/m}-1} \frac{dv}{v} \\ &= E(mh) + R_m(mh), \qquad R_m(u) = \frac{1}{m} \int_u^{+\infty} e^{-v} \eta\left(\frac{v}{m}\right) dv, \end{aligned}$$

with $\eta(z) = 1/(e^z - 1) - 1/z$ that is continuous over the positive half-line.

The estimate of Equation (26) applied to $\chi_m(z)$ and $\phi_m(z)$ entails

$$\chi_m\left(\frac{e^{-h}}{q}\right) = \frac{e^{-E(mh)+\mathcal{O}(1/m)}}{1-e^{-h}}, \qquad \phi_m\left(\frac{e^{-h}}{q}\right) = \frac{e^{-E(mh)+\mathcal{O}(1/m)}}{1-e^{-h}} \frac{e^{-mh}}{m}. \tag{27}$$

When $z = e^{-t}/q$ with $t > 0$, the evaluation (27) together with the expressions (24) and (25) of the intervening generating functions give

$$\widetilde{\phi}_{D_1}(z) \approx \frac{e^{-tD_1}}{D_1} \frac{e^{-E(\lfloor D_1/2 \rfloor t)}}{t}, \qquad \phi_{D_1,D_2}(z) \approx \frac{e^{-(D_1+D_2)t}}{D_1 D_2} \frac{e^{-E(D_2 t)}}{t}.$$

Approximating sums by integrals, when $z = e^{-t}/q$ with $t \to 0^+$, yields

$$\Psi(z) \sim \frac{1}{2t} \int_0^{\infty} x^2 e^{-tx} e^{-E(xt/2)}\, dx + \frac{1}{t} \int_{y<x<2y} (x-y) x^2 \frac{e^{-(x+y)t}}{xy} e^{-E(ty)}\, dx\, dy.$$

The change of variables $tx \mapsto x$ and $ty \mapsto y$, rephrases the double integral as

$$\Psi\left(\frac{e^{-t}}{q}\right) \sim \frac{4}{t^4} \int_0^{\infty} x^2 e^{-2x-E(x)}\, dx + \frac{1}{t^4} \int_0^{\infty} e^{-2y} \frac{2+y-e^{-y}(2+3y+2y^2)}{y}\, dy.$$

These integrals simplify under partial integration, and one finds

$$\Psi\left(\frac{e^{-t}}{q}\right) \sim \frac{c}{t^4} \qquad (t \to 0^+), \qquad c = \frac{5}{2} - 6\delta. \tag{28}$$

(iii) *Coefficients.* The asymptotic form (28) of $\Psi(z)$ as $t \to 0$ is consistent with the assertion that

$$[z^n]\Psi(z) \sim \frac{c}{6} q^n n^3, \tag{29}$$

though it does not imply it. Indeed, an asymptotic estimate like (28) is confined to the vicinity of the real line, since it can be proved that the function $\Psi(z)$ admits its

circle of convergence as a natural boundary. Thus singularity analysis cannot be applied. (A Tauberian theorem could be tried, but Tauberian side conditions appear to be delicate to establish.) We then proceed instead by an Abelian argument that is based on a direct proof of existence of the limit

$$\varpi := \lim_{n\to\infty} \frac{1}{q^n n^3}[z^n]\Psi(z). \tag{30}$$

Indeed, Theorem 3.2 applied to Formula (23) guarantees the existence of the limit in (30) since

$$\begin{aligned}
\overline{\sigma''}_n \;&\sim\; \sum_{D_1=1}^{n} D_1 \left(D_1 - \left\lfloor \frac{D_1}{2} \right\rfloor\right) g_1\left(\frac{D_1}{n}\right) \\
&\qquad + \sum_{D_2 < D_1 \le 2D_2} D_1 \left(\frac{D_1}{D_2} - 1\right) g_2\left(\frac{D_1}{n}, \frac{D_2}{n}\right) \\
&\sim\; \left[\int_0^1 \frac{x^2}{2} g_1(x)\,dx + \int_0^1 x \left(\int_{x/2}^{x} \left(\frac{x}{y} - 1\right) g_2(x,y)dy\right) dx\right] n^3,
\end{aligned}$$

where, in the second line, Riemann sums have been approximated by integrals. This is sufficient to conclude on the existence of ϖ in (30). This value must then be identical to $c/6$ in accordance with (28) and (29). ▌

The constant δ in the above proof is a close relative of the famous Golomb constant that intervenes in the expectation of the longest cycle in a random permutation [54].

The global savings of the early abort strategy is thus of 36% compared to the basic strategy, and of 15% compared to the half-degree rule. The expected cost of DDF is $\mathcal{O}(n^3 \log q)$ and this cost dominates in the whole factorization chain.

4. THE OUTPUT CONFIGURATION OF DDF

The DDF procedure does not completely factor a polynomial that has different irreducible factors of the same degree. However, as shown by the following results, "most" of the factoring has been completed after DDF. First, the DDF procedure produces a complete factorization with asymptotic probability greater than $1/2$ (Theorem 4.1). Next, the number of calls to the subsequent phase of EDF, that is to say the number of degree values for which more than one factor occurs, is only $\mathcal{O}(1)$, and the sum of the degrees where this happens (the total degree of the fragments passed to EDF) is $\mathcal{O}(\log n)$. However, this total degree has a fairly large variability so that the cost of EDF (to be analysed in the next section)

is comparatively small but not entirely negligible. Theorem 4.2 quantifies some of these phenomena. They are established here by means of a hybridization of singularity analysis and Darboux's method, a general technique that we explain in some detail when we first encounter it in the next theorem.

THEOREM 4.1. *The asymptotic probability for the distinct-degree factorization to be the complete factorization is*

$$c_q = \prod_{k\geq 1}\left(1+\frac{I_k}{q^k-1}\right)(1-q^{-k})^{I_k}.$$

In particular: $c_2 \doteq 0.6656$, $c_3 \doteq 0.6123$, $c_5 \doteq 0.5861$, $c_{47} \doteq 0.5635$, $c_{257} \doteq 0.5618$, *and* $c_\infty = e^{-\gamma} \doteq 0.5614$, *where* γ *is Euler's constant.*

Proof. (i) *Permutation model.* We start with the analysis of the permutation model, as this illustrates a "bare-bones" version of the method. By remarks above, this corresponds to the limit case $q=\infty$. The probability that a permutation of length n has all its cycles of different lengths is $[z^n]F(z)$, where the generating function $F(z)$ is susceptible to a variety of expressions obtained by the technique of convergence factors:

$$\begin{aligned} F(z) &:= \prod_{k=1}^{\infty}\left(1+\frac{z^k}{k}\right) \\ &= e^{-z}\frac{1+z}{1-z}\prod_{k=2}^{\infty}\left(1+\frac{z^k}{k}\right)e^{-z^k/k} \\ &= \left(\frac{1+z}{1-z}\exp\left(-\frac{1}{2}\mathrm{Li}_2\left(z^2\right)\right)\right) \\ &\qquad \cdot\left(e^{-z+z^2/2}\prod_{k=2}^{\infty}\left(1+\frac{z^k}{k}\right)e^{-z^k/k+z^{2k}/(2k^2)}\right) \\ &= S(z)\cdot R(z). \end{aligned} \tag{31}$$

Here $\mathrm{Li}_2(z) := \sum_{k\geq 1} z^k/k^2$ is the classical *dilogarithm* function. The first factor, $S(z)$, in the bottom equality of (31) satisfies the conditions of singularity analysis, while the second one, $R(z)$ is continuously differentiable (of class $\mathcal{C}^1$) on the closed unit disc $\overline{D}$, since it is of the form $e^{r(z)}$ where the coefficient $[z^n]r(z)$ is $\mathcal{O}(n^{-3})$.

We thus have a situation where the generating function of interest is the product of a singular part $S(z)$ that satisfies strong analyticity properties outside of $z=\pm 1$, and of a function $R(z)$ of the Darboux type that is smooth on the closed unit disc $\overline{D}$.

We only need to justify the fact that dominant asymptotics of the coefficients $[z^n]F(z)$ can be extracted "as though" $R(z)$ were itself analytic on $\overline{D}$.

The local expansions of $S(z)$ at its singularities ± 1 are readily found to be

$$\begin{aligned} S_{+1}(z) &= 2e^{-\zeta(2)/2}\left(\frac{1}{1-z} + \log\frac{e^{1/2}}{2(1-z)} + \mathcal{O}((1-z)\log^2(1-z))\right) \\ S_{-1}(z) &= \frac{1}{4}e^{-\zeta(2)/2}\left(2(z+1) + \mathcal{O}((z+1)\log(z+1)\right), \end{aligned}$$

with the error terms being $\mathcal{C}^0$ on $\overline{D}$. In summary, we have found that

$$F(z) = \left(2e^{-\zeta(2)/2}\frac{1}{1-z} + \log(1-z) + t(z)\right)\cdot R(z), \tag{32}$$

for some $t(z)$ that is $\mathcal{C}^0$, and $R(z)$ that is $\mathcal{C}^1$ on $\overline{D}$.

The expansion $R(z) = R(1) + U(z)(z-1)$ with a function $U(z)$ that is $\mathcal{C}^0$, can then be inserted in (32). Darboux's method applies to the resulting form for $F(z)$ (see the discussion relative to Equation (14) in Section 1.3), and one has

$$[z^n]F(z) = 2R(1)e^{-\zeta(2)/2} + o(1),\ R(1) := e^{-1/2}\prod_{k=2}^{\infty}\left(1+\frac{1}{k}\right)e^{-1/k+1/(2k^2)}.$$

This finally simplifies using the infinite product formula for $\Gamma(1)$:

$$[z^n]F(z) = \prod_{k=1}^{\infty}\left(1+\frac{1}{k}\right)e^{-1/k} + o(1) = e^{-\gamma} + o(1).$$

This last estimate has been already established by Greene and Knuth [32] by means of a Tauberian argument combined with bootstrapping. The method used here is in contrast a hybrid of singularity analysis and of Darboux's method. It can be employed to derive complete asymptotic expansions, with roots of unity that intervene in successive asymptotic terms corresponding to smaller and smaller singularity weights. (This leads to fluctuating terms involving successive roots of unity.)

(ii) *Finite field model.* We next turn to the case of a finite field of fixed cardinality q to which the same principles apply. The generating function of polynomials with irreducible factors all of distinct degrees (but with single factors possibly repeated) is, by standard decomposition formulæ,

$$F(z) = \prod_{k\geq 1}\left(1 + I_k\left(z^k + z^{2k} + z^{3k} + \cdots\right)\right) = \prod_{k\geq 1}\left(1 + I_k\frac{z^k}{1-z^k}\right). \tag{33}$$

An equivalent form that reveals the pole-like singularity at $z = 1/q$ is obtained by multiplying each term of the product (33) by $(1 - z^k)^{I_k}$,

$$F(z) = \frac{1}{1-qz} \prod_{k\geq 1} \left(1 + I_k \frac{z^k}{1-z^k}\right) (1-z^k)^{I_k}.$$

Thus, as $z \to q^{-1}$, we have

$$F(z) \sim \frac{c_q}{1-qz}, \qquad c_q = \prod_{k=1}^{\infty} \left(1 + I_k \frac{1}{q^k - 1}\right) (1-q^{-k})^{I_k}, \tag{34}$$

and the preceding discussion applies, with the rôle of the dilogarithm function now played by

$$\Lambda_2(z) = \sum_{k=1}^{\infty} \left(\frac{I_k}{q^k - 1}\right)^2 z^k.$$

The hybrid method then yields,

$$[z^n] F\left(\frac{z}{q}\right) = c_q + o(1),$$

which, by (34), is our statement. ▮

Theorem 4.1 was obtained by Knopfmacher and Warlimont [38] and independently by the authors in [18]. The methods used in [18] as well as in the present paper are however rather different from those of [38]. The paper [38] uses elementary techniques and derives constructive bounds. The methods developed here are geared towards full asymptotic expansions and have been successfully used by Gourdon [30] to solve the Golomb-Knuth conjecture [39, Ex. 1.3.3.23] regarding the expectation of maximal cycle lengths in random permutations.

THEOREM 4.2. *The number N_0 of degree values for which there is more than one irreducible factor produced by DDF has an average that is asymptotic to the constant*

$$\mu_0 = \sum_{k\geq 1} (1-q^{-k})^{I_k} \left((1-q^{-k})^{-I_k} - 1 - \frac{I_k q^{-k}}{1-q^{-k}}\right).$$

The total degree N_1 of the corresponding polynomials has expectation $\log n + \mathcal{O}(1)$ and standard deviation of order $\sqrt{n}$.

Proof. Given a family $\mathcal{F}$ of elements, the expression

$$\prod_{\omega\in\mathcal{F}} \frac{1}{1-\omega}$$

formally generates all (finite) multisets of elements taken from $\mathcal{F}$. The expression

$$1 + \sum_{\omega\in\mathcal{F}} \frac{\omega}{1-\omega} + u\left(\prod_{\omega\in\mathcal{F}} \frac{1}{1-\omega} - 1 - \sum_{\omega\in\mathcal{F}} \frac{\omega}{1-\omega}\right)$$

formally generates all multisets each affected by a coefficient of u if there are different elements comprising the multiset, and by a coefficient of 1 otherwise. This applies to the class of irreducible polynomials of each degree n, taking $\mathcal{F} = \mathcal{I}_n$. Thus, the bivariate generating function of the number N_0 of degree values for which there are repeated elements is

$$P_0(z,u) = \prod_{k\geq 1}\left(1 + \frac{I_k z^k}{1-z^k} + u\left((1-z^k)^{-I_k} - 1 - \frac{I_k z^k}{1-z^k}\right)\right). \qquad (35)$$

The logarithmic derivative with respect to u satisfies

$$\frac{P_0'(z,1)}{P_0(z,1)} = \sum_{k\geq 1}(1-z^k)^{I_k}\left((1-z^k)^{-I_k} - 1 - \frac{I_k z^k}{1-z^k}\right). \qquad (36)$$

By our general discussion of the hybrid singularity analysis and Darboux method, the quantity N_0 has an expectation that is asymptotic to the limit μ_0 of $P_0'(z,1)/P_0(z,1)$ as $z \to 1/q$. This quantity is thus nothing but the value of the right hand side of (36) at $z = 1/q$.

For the sum N_1 of the degrees of these polynomials, an adaptation of (35) yields the bivariate generating function,

$$P_1(z,u) = \prod_{k\geq 1}\left(\left(1 + u^k\frac{z^k}{1-z^k}\right)^{I_k} - (u^k-1)I_k\,\frac{z^k}{1-z^k}\right).$$

We only discuss briefly the first moment of N_1. The mean value is $q^{-n}[z^n]R(z)$, where $R(z)$ equals $P_1'(z,u)|_{u=1}$. Thanks to the expansion $kI_k = q^k + \mathcal{O}(q^{k/2})$, near $z = 1/q$, $R(z)$ is asymptotic to $(1-qz)^{-1}\log(1-qz)^{-1}$. Thus, the expectation of N_1 taken over polynomials of degree n is $q^{-n}[z^n]R(z) \sim \log n$. The second factorial moment of N_1 is obtained by a further differentiation of $P_1(z,u)$ at $u = 1$. ▮

The analysis of N_1 in Theorem 4.2 was given in [18] and Knopfmacher [35] has independently obtained an estimate of the first two moments of N_0.

It should be clear that the hybrid asymptotic method has great flexibility. As a final illustration, we discuss a question of von zur Gathen and consider the quantity $\widetilde{N}$ that is the largest degree for which two or more factors occur. The generating function of polynomials such that $\widetilde{N} \leq r$ is in this case

$$F^{(r)}(z) = \prod_{k \leq r} (1 - z^k)^{-I_k} \prod_{k > r} \left(1 + I_k \frac{z^k}{1 - z^k}\right).$$

Thus, the probability of $\widetilde{N} \leq r$ is, for large degree n and fixed r, asymptotic to

$$c_q^{(r)} = \prod_{k > r} (1 - q^{-k})^{-I_k} \left(1 + I_k \frac{q^{-k}}{1 - q^{-k}}\right), \tag{37}$$

and for large field cardinalities, these constants have a limit,

$$c_\infty^{(r)} = e^{-\gamma} \prod_{k \leq r} \left(1 + \frac{1}{k}\right)^{-1} e^{1/k}. \tag{38}$$

We have $1 - c_q^{(r)} = \mathcal{O}(1/r)$ for all fixed q, some representative values with $q = \infty$ being:

$$c_\infty \doteq 0.5614,\ c_\infty^{(1)} \doteq 0.7631,\ c_\infty^{(2)} \doteq 0.8387,\ c_\infty^{(5)} \doteq 0.9179,\ c_\infty^{(10)} \doteq 0.9549.$$

Thus, (37) and (38) give the following simplified picture in the asymptotic limit (n and q large).

FACT. A random polynomial has a small number, $\mathcal{O}(1)$, of "colliding" degrees; the largest colliding degree has a probability distribution tail that decays like $\mathcal{O}(1/r^2)$ (for $r \leq n/2$). Because of this slow tail decay, the largest colliding degree alone has a first moment that is $\mathcal{O}(\sum_r r^{-1}) = \mathcal{O}(\log n)$, but a second moment that is $\mathcal{O}(\sum_r 1) = \mathcal{O}(n)$.

These observations are seen to be consistent with what Theorem 4.2 asserts.

5. EQUAL-DEGREE FACTORIZATION (EDF)

After the first two stages of the general algorithm, the factorization problem has been eventually reduced to factoring a collection of monic squarefree polynomials b_k all of whose irreducible factors have the same (known) degree k. The third step in the factorization process, the equal-degree factorization algorithm (EDF),

```
procedure EDF(c : polynomial, k : integer);
{"c" is a product of irreducibles of degree "k"}
       if degree(c) <= k then return(c) fi;
       h := randpoly(degree(c)-1);
       {draw a random polynomial}
1.     a := h^((q^k-1)/2)-1 mod c;
2.     d := gcd(a,c);
       return(EDF(d,k).EDF(c/d,k));
end;
```

FIG. 8. The equal-degree factorization algorithm (EDF).

focuses on polynomials with this special form. Our reference chain uses the classical Cantor-Zassenhaus algorithm [6] for this purpose. The analysis combines a recursive partitioning problem akin to digital trees —also known as "tries" [41, 46]— together with estimates on the degrees of irreducible factors of random polynomials [37]. The net result is that the global cost of EDF is quadratic, a sharp contrast with the cubic cost of DDF. For convenience, we first assume that q is odd, and relegate to Section 5.4 the case of a characteristic equal to 2.

The EDF algorithm is described in Fig. 8, and we briefly recall the principle here.

PRINCIPLE OF EDF. Let c be a polynomial that is a product of j irreducible factors $c = f_1, \ldots, f_j$, with each f_i of degree k. The Chinese remainder theorem implies the ring homomorphism,

$$\mathbb{F}_q[x]/(c) \cong \mathbb{F}_q[x]/(f_1) \times \cdots \times \mathbb{F}_q[x]/(f_j),$$

and a random element h of $\mathbb{F}_q[x]/(c)$ is associated to a j-tuple $(h_1, \ldots, h_j)$, where each h_i is a random element of $\mathbb{F}_q[x]/(f_i)$.

The following splitting principle makes it possible to isolate the various f_i. Since each f_i is irreducible, the multiplicative group of each component $\mathbb{F}_q[x]/(f_i)$ is a field isomorphic to $\mathbb{F}_{q^k}$. Such a group being cyclic, there are the same number $(q^k - 1)/2$ of squares and nonsquares. The test $h_i^{(q^k-1)/2} = 1$ discriminates the squares in this multiplicative group. Thus, taking a random h and computing $a := h^{(q^k-1)/2} - 1 \bmod c$, we have that $\gcd(a, c)$ "extracts" the product of all the f_i for which h is a square in $\mathbb{F}_q[x]/(f_i)$.

From the algorithmic standpoint, taking random polynomials h, leads to successive refinements of each factor $c = b_k$ known to be composed solely of irreducible polynomials of degree k. The computation develops as a tree (Figure 9). From

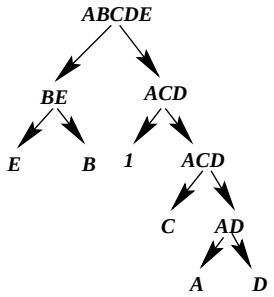


FIG. 9. A tree of successive refinements by EDF of the factorization $b = ABCDE$.

the probabilistic point of view, each component h_i that is random in $\mathbb{F}_q[x]/(f_i)$ has probability $\alpha = \frac{1}{2} - \frac{1}{2q}$ of being discriminated by the gcd test and the dual probability, $\beta = \frac{1}{2} + \frac{1}{2q}$ of being a nonsquare—the (small) difference between α and β is accounted for by the possibility of having noninvertible components.

Then, the analysis of the complete EDF phase (Section 5.3) requires a purely combinatorial analysis of what takes place at each degree k (Section 5.2) combined with an estimate of the probability that there are j irreducible factors of degree k in a random polynomial of degree n. These probabilities give interesting information on random polynomials and have been obtained by Knopfmacher and Knopfmacher [37] whose results we recall in Section 5.1.

5.1. Irreducible factors of each degree

Let $\kappa_n(k)$ be the random variable counting the number of distinct irreducible factors of degree k in a random polynomial of degree n. We consider here k as fixed. The corresponding probability distribution was given in [37]. It can be easily computed by the decomposition techniques of Section 1, as we now show.

THEOREM 5.1 (Knopfmacher and Knopfmacher). *The probability that there are j distinct irreducible factors of degree k in a random polynomial of degree n is, for n large enough ($n \geq kI_k$), given by the binomial distribution $\mathcal{B}(I_k, q^{-k})$, namely,*

$$\Pr\{\kappa_n(k) = j\} = \binom{I_k}{j} (q^{-k})^j (1 - q^{-k})^{I_k - j}.$$

Proof. The bivariate generating function for the number of irreducibles of degree k is, by the basic decomposition,

$$\begin{aligned} Q_k(z,u) &= \left(1+u\frac{z^k}{1-z^k}\right)^{I_k}\prod_{\ell\neq k}(1-z^\ell)^{-I_\ell} \\ &= \frac{1}{1-qz}(1+(u-1)z^k)^{I_k}. \end{aligned}$$

The asymptotics as $n\to\infty$ are derived from the polar singularity at $z=1/q$: the probability generating function of the distribution is in the limit $n\to\infty$,

$$(1+(u-1)q^{-k})^{I_k},$$

that is to say, the probability generating function of a binomial distribution $\mathcal{B}(I_k, q^{-k})$. As is easily observed, this asymptotic formula is even exact as soon as $n\geq kI_k$. ▮

Since a binomial distribution corresponding to rare events converges to a Poisson distribution, one has:

FACT. The probability distribution of the number of factors of degree k in a random polynomial of degree n is approximately a Poisson law of parameter $1/k$,

$$\Pr\{\kappa_n(k)=j\}\approx e^{-1/k}\frac{k^{-j}}{j!}.$$

This fact is in accordance with the known distribution of cycle lengths in the random permutation model [54]. The table below provides numerical values of $\Pr\{\kappa_n(k)=j\}$ for $j=1,2,3$, when $k=1,2,3$. The data corresponding to degree $n=20$ (for values of $q=2,17$) show an excellent fit with the Poisson law ($q=\infty$), even in the non-asymptotic regime.

q	$k=1$			$k=2$			$k=3$		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
2	0.250	0.250	0.187	0.750	0.187	0.046	0.765	0.191	0.035
17	0.356	0.356	0.188	0.624	0.293	0.069	0.717	0.238	0.0396
∞	0.367	0.367	0.183	0.606	0.303	0.075	0.716	0.238	0.039

5.2. EDF and splitting trees

For each value of k, we can regard the EDF phase as an abstract splitting process as follows. Start with a group G formed of m individuals. (In EDF, if the degree of b_k is kj, then, $m = j$.) By flipping coins, separate G randomly into two subgroups, G_0 and G_1, with the probabilities for each element to be sent to G_0 and G_1 being α and β. The process is repeated recursively until all elements have been isolated. Clearly, any such recursive execution is described by a binary tree (Figure 9). The corresponding randomness model

$$\begin{aligned} &\Pr\left(|G_0| = m_0 \;\middle|\; |G| = m\right) = \binom{m}{m_0} \alpha^{m_0} \beta^{m-m_0}, \\ &\alpha = \frac{1}{2} - \frac{1}{2q}, \qquad \beta = \frac{1}{2} + \frac{1}{2q}, \end{aligned} \tag{39}$$

that is induced by independent splittings then coincides with the one underlying *digital trees*. Given the importance of the digital tree in the design and analysis of algorithms many properties are known. We cite here:

FACT. The expectation of the number of binary nodes in a splitting tree, relative to m individuals and with probabilities (α, β), is

$$\frac{m}{H}(1 + \epsilon(m)) + o(m), \qquad H = \alpha \log_2 \frac{1}{\alpha} + \beta \log_2 \frac{1}{\beta},$$

with $\epsilon(m)$ a fluctuating function of amplitude typically $< 10^{-5}$. The expectation of the height of the tree is

$$\frac{2}{K} \log_2 m + \mathcal{O}(1), \qquad K = \log_2(\alpha^2 + \beta^2)^{-1}.$$

The size estimate due to Knuth and De Bruijn around 1965 (published in the 1973 edition of [41]) involves the entropy function H; the height estimate first appeared in [22] (a paper already motivated by polynomial factorization) and it involves the "coincidence probability" $\alpha^2 + \beta^2$. Nowadays, these results are best understood in the context of Vallée's general theory of dynamical sources; see [11, 58]. As a consequence of these estimates, splitting trees tend to be fairly well balanced so that the cost of an EDF phase is expected to be close to that of a perfect splitting. The lemma below provides an explicit expression for the costs induced by the computational model at hand.

LEMMA 5.1. *The expected cost $C_{j,k}$ of the EDF algorithm applied to any product of j irreducible factors of degree k is*

$$\left(\frac{j(j-1)}{2\alpha\beta} + j \sum_{m=0}^{\infty} \sum_{\ell=0}^{m} \binom{m}{\ell} \alpha^{m-\ell} \beta^{\ell} \left(1 - (1 - \alpha^{m-\ell}\beta^{\ell})^{j-1}\right) \right) (\mu_k \tau_1 + \tau_2) k^2,$$

where $\mu_k = \lambda((q^k-1)/2) = \left\lfloor \log_2 \frac{q^k-1}{2} \right\rfloor + \nu\left(\frac{q^k-1}{2}\right) - 1.$

Proof. It is convenient to regard an execution of the splitting process as a tree t and to consider, with t_0, t_1 the root subtrees, a general cost function of the additive type,

$$C[t] = e_{|t|} + C[t_0] + C[t_1]. \tag{40}$$

Here $e_{|t|}$ is a (problem specific) "toll" function that depends on the size $|t|$, that is to say, the number of irreducible factors (of degree k) to be separated.

The subtree sizes obey the Bernoulli probability of (39). Also the subproblems described by t_0, t_1 have, by design, the same characteristics as the whole tree. Thus, the expectation c_j of $C[t]$ over trees of size j satisfies the recurrence

$$c_j = e_j + \sum_{\ell=0}^{j} \binom{j}{\ell} \alpha^\ell \beta^{j-\ell} (c_\ell + c_{j-\ell}) = e_j + \sum_{\ell=0}^{j} \binom{j}{\ell} (\alpha^\ell \beta^{j-\ell} + \alpha^{j-\ell}\beta^\ell) c_\ell.$$

This translates in terms of the exponential generating functions

$$C(z) = \sum_j c_j z^j/j!, \qquad E(z) = \sum_j e_j z^j/j!,$$

into the functional equation

$$C(z) = E(z) + e^{\beta z} C(\alpha z) + e^{\alpha z} C(\beta z).$$

This difference equation iterates, leading to the explicit generating function solution

$$C(z) = \sum_{j=0}^{\infty} \sum_{\ell=0}^{j} \binom{j}{\ell} E(\alpha^{j-\ell}\beta^\ell z) e^{z(1-\alpha^{j-\ell}\beta^\ell)}. \tag{41}$$

The analysis is completed by specializing this discussion to the EDF costs. The toll function that arises from the top-level execution of the EDF procedure is then

$$\widehat{e}_j = (\mu_k \tau_1 + \tau_2)(kj)^2,$$

for $j \geq 2$. There, μ_k is the number of multiplications of the binary powering method, and the quadratic costs are induced by the naïve multiplication and gcd algorithms considered here. The toll function $e_j = j^2(1-\delta_{1,j})$ corresponds to the generating function $E(z) = z(e^z(1+z)-1)$ in (41). Extracting coefficients of the resulting generating function $C(z)$ in (41) and rescaling by $\widehat{e}_j/e_j$ then yields the statement. ∎

5.3. Complete analysis

Completing the analysis of EDF only requires weighting the costs given by Lemma 5.1 by the probability $\Pr(\kappa_n(k) = j)$ of finding j irreducible factors of degree k given by Theorem 5.1. By Lemma 5.1, the cost is of the form $\mathcal{O}(j^2k^3)$, and by Theorem 5.1, the probabilities are approximately $e^{-1/k}k^{-j}/j!$; thus, we expect the total cost of the DDF phase to be of the order of

$$\sum_{k,j} \left(j^2k^3\right) \cdot \left(\frac{k^{-j}}{j!}\right) = \mathcal{O}(n^2).$$

The main result of this section gives a firm basis to this heuristic computation and determines the implied constant. In order to prove the final estimate of Theorem 5.2 below, we need two technical lemmas.

LEMMA 5.2. *When $k \to \infty$, one has for all j such that $kj \le n$*

$$\Pr(\kappa_n(k) = j) = \frac{\binom{I_k}{j}}{q^{kj}} \left(1 + \mathcal{O}(1/k)\right),$$

where the $\mathcal{O}(1/k)$ is uniform in j. Moreover, the following uniform estimate holds

$$\Pr(\kappa_n(k) = j) = \mathcal{O}\left(\frac{1}{j!k^j}\right).$$

Proof. When $kj \le n$, Theorem 5.1 yields

$$\Pr(\kappa_n(k) = j) = \frac{\binom{I_k}{j}}{q^{kj}}(1+\upsilon),\ \upsilon = \sum_{\ell=1}^{N} (-1)^\ell \frac{\binom{I_k-j}{\ell}}{q^{k\ell}},\ N = \min(\lfloor n/k \rfloor - j, I_k - j).$$

When k is large, one has $\upsilon = \mathcal{O}(1/k)$ since

$$|\upsilon| \le \sum_{\ell=1}^{I_k-j} \frac{\binom{I_k-j}{\ell}}{q^{k\ell}} = (1+q^{-k})^{I_k-j} - 1 \le (1+q^{-k})^{I_k} - 1 = \mathcal{O}(1/k).$$

This proves the first estimate. As for the second one, it suffices to note that

$$\binom{I_k}{j} \le \frac{I_k^j}{j!} = \mathcal{O}\left(\frac{q^{kj}}{j!k^j}\right),$$

and to use the first estimate of the lemma. ∎

LEMMA 5.3. *The average costs $C_{j,k}$ of Lemma 5.1 satisfy for all k*

$$C_{0,k} = C_{1,k} = 0, \qquad C_{2,k} = \frac{2}{\alpha\beta}\,(\mu_k\,\tau_1 + \tau_2)\,k^2,$$

and, uniformly,

$$C_{j,k} = \mathcal{O}(j^2 k^3).$$

Proof. The first relations are direct applications of Lemma 5.1. For the estimate of $C_{j,k}$, the inequality $1-(1-u)^{j-1} \leq (j-1)u$ implies

$$\begin{aligned} C_{j,k} &\leq \left(\frac{j(j-1)}{2\alpha\beta} + j \sum_{m\geq 0} \sum_{\ell=0}^{m} \binom{m}{\ell} (j-1)\alpha^{2(m-\ell)}\beta^{2\ell} \right) (\mu_k\,\tau_1 + \tau_2)\,k^2 \\ &= \frac{j(j-1)}{\alpha\beta}\,(\mu_k\,\tau_1 + \tau_2)\,k^2. \end{aligned}$$

The last equality holds since

$$\sum_{m\geq 0} \sum_{\ell=0}^{m} \binom{m}{\ell} \alpha^{2(m-\ell)}\beta^{2\ell} = \sum_{m\geq 0} (\alpha^2+\beta^2)^m = \frac{1}{2\alpha\beta}.$$

Since $\mu_k = \mathcal{O}(k)$, the statement follows. ∎

We are now ready to prove the main result of this section.

THEOREM 5.2. *The expected cost of the EDF phase satisfies*

$$\overline{\tau EDF}_n \sim \frac{\tau_1}{\alpha\beta} \sum_{k=1}^{\lceil n/2 \rceil} \mu_k, \qquad \mu_k = \left\lfloor \log_2 \frac{q^k-1}{2} \right\rfloor + \nu\left(\frac{q^k-1}{2}\right) - 1.$$

In addition, this cost is $\mathcal{O}(n^2 \log q)$ and

$$\overline{\tau EDF}_n \sim \left(\frac{3}{4}\tau_1 \frac{q^2}{q^2-1} \log_2 q\right)\,(1+\xi_n)\cdot n^2, \quad -\frac{1}{3} + o(1) \leq \xi_n \leq \frac{1}{3} + o(1). \tag{42}$$

Proof. The intuition behind the proof is that the major contribution comes from situations where just 2 factors are present, the other cases having globally a very small probability of occurrence. Let $\overline{E_k}$ be the expected value of the

cost of the EDF algorithm corresponding to degree k. By definition, we have $\overline{E_k} = \sum_{j\geq 2} \Pr(\kappa_n(k) = j)\, C_{j,k}$, where $C_{j,k}$ is given by Lemma 5.1.

When $2k \leq n$, Lemma 5.2 and Lemma 5.3 entail, as $k \to \infty$,

$$\overline{EDF_k} = C_{2,k}\frac{\binom{I_k}{2}}{q^{2k}}(1+\mathcal{O}(1/k)) + \sum_{j\geq 3}\mathcal{O}\left(\frac{k^{-j}}{j!}\, j^2k^3\right) = \frac{\tau_1}{\alpha\beta}\,\mu_k + \mathcal{O}(1).$$

When $2k > n$, we have $\overline{EDF_k} = 0$. Thus, the overall cost of the EDF component is $\sum_k \overline{E_k} = \frac{\tau_1}{\alpha\beta}\sum_{k=1}^{\lceil n/2\rceil}\mu_k + \mathcal{O}(n)$. The second form of the cost is obtained from the general inequality $1 \leq \nu(m) \leq 1 + \log_2 m$, upon subtracting from $\nu(m)$ its "mean value" $\frac{1}{2}\log_2 m$. ▮

The quantity ξ_n in the statement measures the default of uniformity in binary representations of numbers related to the powers of q. Under the unproven assumption that such representations behave like random integers, the arithmetic function ξ_n should be close to 0. This assumption is well supported by empirical evidence: for instance, with $q = 17$, we have

$$\xi_5 = -0.425,\ \ \xi_{10} = -0.060,\ \ \xi_{20} = -0.024,\ \ \xi_{50} = -0.016.$$

For all practical purposes, we may safely regard ξ_n as being asymptotic to 0.

5.4. Equal-degree factorization in characteristic 2

In the previous sections, we have analysed in detail the equal-degree factorization over finite fields with odd characteristic. For these cases, we have followed the algorithm by Cantor and Zassenhaus [6] who also provide a solution for the even case that relies on factoring the polynomial in a quadratic extension. Ben-Or [3] showed that this detour is not needed while proposing a method based on trace computations.

Trace computations introduce only a small change in the EDF algorithm of Fig. 8. Let m be such that $q = 2^m$. In order to compute the traces, we replace line 1 by

```
1'.  a := h+h^2+h^(2^2)+...+h^(2^(km-1)) mod b;
```

We observe that the analysis for the odd case is valid for the even case. First, the splitting process is the same (with probabilities $\alpha = \beta = 1/2$). Then, the cost of computing line `1'` is the same as the cost of computing line 1 in Fig. 8. Indeed, the trace computations can be determined using basically km products of a polynomial containing j factors of degree k. This costs essentially $km(jk)^2 = k^3j^2\log q$, the same cost as in the odd case.

Phase	Worst-case	Average-case
ERF	$\mathcal{O}(n^2)$	$\tau_2 n^2$
DDF	$\mathcal{O}(n^3 \log q)$	$0.26689\,(\lambda(q)\tau_1 + \tau_2)\,n^3$
EDF	$\mathcal{O}(n^3 \log q)$	$\left(\frac{3}{4}\tau_1 \frac{q^2}{q^2-1} \log q \cdot n^2\right)\,(1 \pm \frac{1}{3} + o(1))$

FIG. 10. A comparison of the worst cases and the average cases of the three phases of polynomial factorization. (The cost of multiplying two polynomials of degree less than n modulo a polynomial f of degree n is $\tau_1 n^2$, and the cost of a gcd between f and a polynomial of degree less than n is $\tau_2 n^2$. The number of products needed to compute h^q mod f is $\lambda(q) = \lfloor \log_2 q \rfloor + \nu(q) - 1$.)

6. CONCLUSIONS

In this paper we have shown how analytic combinatorics adapts well to the case of polynomials over finite fields. A systematic usage of this methodology leads not only to the derivation of basic probabilistic properties of random polynomials over finite fields but also to the average-case analysis of a complete polynomial factoring algorithm. Figure 10 summarizes the main results of the paper in terms of the average-case analysis of the factoring algorithm and it provides a comparison with worst-case behaviour.

It should be clear that a large number of variants of the factorization chain can be analysed by our methods. For instance, specifics of the elimination of repeated factors stage are largely immaterial from the expected complexity standpoint, since they lead to identical results in asymptotic terms. A radical possibility is then to bypass completely the first stage. In this variant, DDF not only produces the polynomials for the EDF part but also returns a polynomial containing the non-squarefree part of the original polynomial. Once more, there is no difference in asymptotic terms.

Several authors [26, 34] have stated that from a worst-case perspective, and even when considering fast arithmetic instead of the classical one, DDF is the bottleneck for the factorization process. Our results confirm that such is also the case from an average-case perspective. Chapter 15 of the recent book [24] further illustrates this point with striking graphics.

Finally, we should mention here a few other algorithms that have been subjected to a "Knuthian" analysis similar to what was conducted here. Panario and Richmond analyse in [52] the irreducibility test of Ben-Or that is based on trial DDF: the analysis involves statistics on smallest degrees of irreducible factors and the Buchstab function, a dual of the Dickman function. Rabin's probabilistic construc-

tion of irreducible polynomials is also dealt with in [51]. Last but not least, the Euclidean algorithm turns out to be somewhat easier to analyse for polynomials than for integers: see the analyses by Knopfmacher and Knopfmacher [36] as well as by Friesen and Hensley [23].

ACKNOWLEDGMENT

The work of Philippe Flajolet was supported in part by the ALCOM-FT Project (number IST-1999-14186) of the European Union. The work of Daniel Panario was done for the most part while with the Department of Computer Science of the University of Toronto. We are grateful to Joachim von zur Gathen for having put us in contact and for having incited us to analyse polynomial factorization in detail. Thanks to Brigitte Vallée for many constructive suggestions regarding the general organization of the paper. Thanks also to Helmut Prodinger and Allan Borodin for much appreciated support.

REFERENCES

1. ARRATIA, R., BARBOUR, A. D., AND TAVARÉ, S. Random combinatorial structures and prime factorizations. *Notices of the American Mathematical Society 44* (1997), 903–910.
2. BACH, E. Toward a theory of Pollard's rho method. *Information and Computation 90* (1991), 139–155.
3. BEN-OR, M. Probabilistic algorithms in finite fields. In *Proc. 22nd IEEE Symp. Foundations Computer Science* (1981), pp. 394–398.
4. BERLEKAMP, E. *Algebraic Coding Theory*. McGraw Hill, New York NY, 1968.
5. BUCHMANN, J. Complexity of algorithms in algebraic number theory. In *Number Theory. Proc. First Conf. Canadian Number Theory Assoc.* Walter de Gruyter, 1990, pp. 37–53.
6. CANTOR, D., AND ZASSENHAUS, H. A new algorithm for factoring polynomials over finite fields. *Math. Comp. 36* (1981), 587–592.
7. CAR, M. Factorisation dans $\mathbb{F}_q[x]$. *C. R. Acad. Sci. Paris Ser. I 294* (1982), 147–150.
8. CAR, M. Théorèmes de densité dans $\mathbf{F}_q[X]$. *Acta Arith. 48* (1987), 145–165.
9. CARLITZ, L. The arithmetic of polynomials in a Galois field. *Amer. J. Math. 54* (1932), 39–50.
10. CHOR, B., AND RIVEST, R. A knapsack-type public key cryptosystem based on arithmetic in finite field. *IEEE Trans. Inform. Theory. 34* (1988), 901–909.
11. CLÉMENT, J., FLAJOLET, P., AND VALLÉE, B. Dynamical sources in information theory: a general analysis of trie structures", *Algorithmica 29* (2001), 307–369.
12. COLLINS, G. Factoring univariate integral polynomials in polynomial average time. In *Proc. EUROSAM 79* (1979), vol. 72 of *Lecture Notes in Computer Science*, pp. 317–329.
13. COMTET, L. *Advanced Combinatorics*. Reidel, Dordrecht, 1974.
14. DE BRUIJN, N. On the number of positive integers $\leq x$ and free of prime factors $> y$. *Indag. Math. 13* (1951), 2–12.
15. DEDEKIND, R. Abriss einer Theorie der höhern Congruenzen in Bezug auf einen reellen Primzahlmodulus. *J. reine u. angew. Math. 54* (1857), 1–26.
16. DICKMAN, K. On the frequency of numbers containing prime factors of a certain relative magnitude. *Ark. Mat. Astr.Fys. 22* (1930), 1–14.
17. FINCH, S. Golomb-Dickman constant. In *Favorite Mathematical Constants* (2000). Published electronically at `http://algo.inria.fr/bsolve/constant/constant.html`.

18. FLAJOLET, P., GOURDON, X., AND PANARIO, D. Random polynomials and polynomial factorization. In *Automata, Languages, and Programming* (1996), F. Meyer auf der Heide and B. Monien, Eds., vol. 1099 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 232–243. Proceedings of the 23rd ICALP Conference, Paderborn, July 1996.

19. FLAJOLET, P., AND ODLYZKO, A. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics 3:2* (1990), 216–240.

20. FLAJOLET, P., AND SEDGEWICK, R. *Analytic Combinatorics.* Book in preparation (2000).

21. FLAJOLET, P., AND SORIA, M. Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory,* Series A *53* (1990), 165–182.

22. FLAJOLET, P., AND STEYAERT, J. A branching process arising in dynamic hashing, trie searching and polynomial factorization. In *Proc. 9th ICALP Symp. 1982* (1982), vol. 140 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 239–251.

23. FRIESEN, C. AND HENSLEY, D., The statistics of continued fractions for polynomials over a finite field. *Proc. Amer. Math. Soc. 124* (1996), 2661–2673.

24. VON ZUR GATHEN, J., AND GERHARD J. *Modern Computer Algebra*. Cambridge University Press, 1999.

25. VON ZUR GATHEN, J., AND PANARIO, D. Factoring polynomials over finite fields: a survey. To appear in *J. of Symb. Comp.*, 2000.

26. VON ZUR GATHEN, J., AND SHOUP, V. Computing Frobenius maps and factoring polynomials. *Comput. Complexity 2* (1992), 187–224.

27. GAUSS, C. *Untersuchungen über höhere Mathematik*. Chelsea, New York, 1889.

28. GEDDES, K., CZAPOR, S., AND LABAHN, G. *Algorithms for Computer Algebra*. Kluwer Academic Publishers, Boston, 1992.

29. GOULDEN, I., AND JACKSON, D. *Combinatorial Enumeration*. John Wiley, New York, 1983.

30. GOURDON, X. *Combinatoire, algorithmique et géométrie des polynômes*. Thèse, École Polytechnique, 1996.

31. GOURDON, X. Largest components in random combinatorial structures. *Discrete Mathematics 180* (1998), 185–209.

32. GREENE, D., AND KNUTH, D. *Mathematics for the Analysis of Algorithms*, 3 ed. Birkhäuser, Boston, 1990.

33. HENSLEY, D. Dirichlet's theorem for the ring of polynomials over $\mathrm{GF}(2)$. *Pacific Journal of Mathematics 123* (1986), 93–101.

34. KALTOFEN, E., AND SHOUP, V. Subquadratic-time factoring of polynomials over finite fields. In *Proc. 27th ACM Symp. Theory of Computing* (1995), pp. 398–406.

35. KNOPFMACHER, A. On the degrees of irreducible factors of polynomials over a finite field. *Disc. Math. 196* (1999), 197–206.

36. KNOPFMACHER, J., AND KNOPFMACHER, A. The exact length of the Euclidean algorithm in $F_q[X]$. *Mathematika 35* (1988), 297-304.

37. KNOPFMACHER, J., AND KNOPFMACHER, A. Counting irreducible factors of polynomials over a finite field. *SIAM Journal on Discrete Mathematics 112* (1993), 103–118.

38. KNOPFMACHER, A., AND WARLIMONT, R. Distinct degree factorizations for polynomials over a finite field. *Trans. Amer. Math. Soc. 347* (1995), 2235–2243.

39. KNUTH, D. *The Art of Computer Programming, Vol.1: Fundamental Algorithms*, 3 ed. Addison-Wesley, Reading MA, 1997.

40. KNUTH, D. *The Art of Computer Programming, Vol.2: Seminumerical Algorithms*, 3 ed. Addison-Wesley, Reading MA, 1997.

41. KNUTH, D. *The Art of Computer Programming, Vol.3: Sorting and Searching*, 2 ed. Addison-Wesley, Reading MA, 1998.

42. KNUTH, D. *Selected Papers on Analysis of Algorithms.* CSLI Publications, Stanford, CA, 2000.

43. KNUTH, D., AND TRABB-PARDO, L. Analysis of a simple factorization algorithm. *Theoretical Computer Science 3* (1976), 321–348.

44. LENSTRA, H. On the Chor-Rivest knapsack cryptosystem. *J. of Cryptology 3* (1991), 149–155.

45. LIDL, R., AND NIEDERREITER, H. *Finite Fields*, vol. 20 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1983.

46. MAHMOUD, H. *Evolution of Random Search Trees*. John Wiley, New York, 1992.

47. MULLEN, G. L., AND SHPARLINSKI, I. Open problems in finite fields. In *Proc. 3rd Conference of Finite Fields and their Applications.* London Math. Soc., *Lect. Note Series 233* (1996), 243–268.

48. ODLYZKO, A. Discrete logarithms and their cryptographic significance. In *Advances in Cryptology, Proceedings of Eurocrypt 1984* (1985), vol. 209 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 224–314.

49. ODLYZKO, A. Asymptotic enumeration methods. In *Handbook of Combinatorics*, R. Graham, M. Grötschel, and L. Lovász, Eds., vol. 2. Elsevier, 1995, pp. 1063–1229.

50. PANARIO, D., GOURDON, X., AND FLAJOLET, P. An analytic approach to smooth polynomials. In *Algorithmic Number Theory Symposium* (1998), vol. 1423 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 226-236.

51. PANARIO, D., PITTEL, B., RICHMOND, B., AND VIOLA, A. Analysis of Rabin's irreducibility test for polynomials over finite fields. Preprint, 2000.

52. PANARIO, D., AND RICHMOND, B. Analysis of Ben-Or's polynomial irreducibility test. *Random Struct. Alg. 13* (1998), 439–456.

53. SCHÖNEMANN, T. Grundzüge einer allgemeinen Theorie der höheren Congruenzen, deren Modul eine reelle Primzahl ist. *J. f. d. reine u. angew. Math. 31* (1846), 269–325.

54. SHEPP, L., AND LLOYD, S. Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc. 121* (1966), 340–357.

55. SHOUP, V. On the deterministic complexity of factoring polynomials over finite fields. *Inform. Process. Lett. 33* (1990), 261–267.

56. SHOUP, V. A new polynomial factorization algorithm and its implementation. *J. Symb. Comp. 20* (1996), 363–397.

57. TENENBAUM, G. *Introduction to Analytic and Probabilistic Number Theory*. Cambridge University Press, 1995.

58. VALLÉE, B. Dynamical sources in information theory: fundamental intervals and word prefixes, *Algorithmica 29* (2001), 262–306.