MATHEMATICS OF COMPUTATION, VOLUME 24, NUMBER 111, JULY, 1970

Factoring Polynomials Over Large Finite Fields

By E. R. Berlekamp

Abstract. This paper reviews some of the known algorithms for factoring polynomials over finite fields and presents a new deterministic procedure for reducing the problem of factoring an arbitrary polynomial over the Galois field GF(p^m) to the problem of finding the roots in GF(p) of certain other polynomials over GF(p). The amount of computation and the storage space required by these algorithms are algebraic in both the degree of the polynomial to be factored and the logarithm of the order of the finite field.

Certain observations on the application of these methods to the factorization of polynomials over the rational integers are also included.

1. Introduction and Summary of Results. This paper presents algorithms for the factorization of a polynomial over a finite field. We are given the polynomial's coefficients, $f_0, f_1, f_2, \cdots, f_n$, which are elements in the finite field GF(q), where q is a power of the prime p, and we wish to find the factors of $f(x) = \sum_{i=0}^{n} f_i x^i$ which are irreducible over GF(q).

The algorithm includes several major steps, which we present in different sections. In Section 3, we obtain a partial factorization,

$$f(x) = \prod_{i=1}^{n} h^{(i)}(x),$$

where $h^{(i)}(x)$ is the product of r_i irreducible-power factors each of which has degree i, and $\sum_{i=1}^{n} ir_i = n$. In Section 4, we reduce the problem of factoring $h^{(i)}(x)$, a polynomial of degree ir_i, to the problem of finding the roots in GF(q) of a new polynomial, $H(x)$, which has degree r_i. If we denote the factorization of $h^{(i)}(x)$ into irreducible-power factors by

$$h^{(i)}(x) = \prod_{j=1}^{r_i} h^{(i,j)}(x)$$

where

$$h^{(i,j)}(x) = \sum_{k=0}^{i} h_k^{(i,j)} x^k,$$

then the roots of $H(x)$ give us the coefficients $h_0^{(i,j)}$. From the roots of $H(x)$, we obtain a partial factorization of $h^{(i)}(x)$. This partial factorization separates the irreducible-power factors according to their $h_0^{(i,j)}$. If this factorization is incomplete, then we may construct another polynomial whose roots are the values of the coefficients $h_1^{(i,j)}$ of the factors of $f(x)$ which have a particular lowest coefficient $h_0^{(i,j)}$. From the roots of this new polynomial, we obtain a further refinement of the factorization of $h^{(i)}(x)$.

Received April 17, 1969, revised October 20, 1969.

AMS Subject Classifications. Primary 1225, 1076; Secondary 1545, 6550, 1009.

Key Words and Phrases. Factorization, computation in finite fields, Kronecker algorithm, Hensel factorizations, root-finding, irreducibility criteria.

The process may be continued until $h^{(i)}(x)$ is factored into the product of irreducible-power factors. In other words, in Sections 3 and 4 we reduce the problem of factoring an arbitrary polynomial, $f(x)$, to the problem of finding the roots of a new polynomial which factors into linear factors over GF(q).

In Sections 5, 6, and 7, we present algorithms for finding the roots of a polynomial over GF(q). In Sections 5 and 6, we present two different methods for converting the root-finding problem in GF(p^m) to a root-finding problem in GF(p), where p is prime. The method of Section 5 is better for small primes, while the method of Section 6 is better for very large primes. Finally, in Section 7 we present an algorithm for finding the roots of a polynomial in a large prime field, thus completing the algorithm for factoring an arbitrary polynomial over a finite field.

The algorithms presented in Sections 3–5 are completely deterministic. Although the amount of time and space these algorithms require to factor a polynomial will depend somewhat on the input polynomial, these costs may be overbounded by an algebraic function of the degree of the input polynomial and the logarithm of q, the order of the finite field.

The algorithm presented in Section 7, on the other hand, is probabilistic rather than deterministic in nature. The algorithm makes a sequence of trials, each of which uses a parameter which is selected at random. Whether or not the trial succeeds in obtaining a factorization depends on the particular choice of this random parameter as well as on the polynomial to be factored. However, each trial succeeds with probability greater than $1/2$, independent of the input polynomial and of all previous trials. Thus, although the number of computations required by the algorithm of Section 7 is a random variable, its mean, variance, and any finite moment may be bounded by an algebraic function of the degree of the input polynomial and $\log p$, where the prime p is the order of the field. For any given $\epsilon > 0$, we may therefore obtain a number N, proportional to $\log 1/\epsilon$, a small power of $\log p$, and a small power of the degree of the input polynomial, such that the probability that the algorithm of Section 7 will require more than N computations is no greater than ϵ. However, for $\epsilon = 0$, the only known general bounds on N are proportional to a root of p rather than a power of $\log p$.

Section 8 reviews a procedure for factoring a polynomial over the rational integers. From the coefficients of the polynomial, we first compute a general bound on the magnitude of any coefficient of any possible factor. We then select a prime, p, larger than twice this bound and factor the polynomial modulo this enormous prime. The factors of the original rational polynomial must then lie among the known factors mod p, so we then try each factor mod p to see whether it is also a factor over the rational integers. The greatest difficulty with this procedure is that a polynomial which has i irreducible factors mod p will have 2^i factors altogether, all but 2 of which will be nontrivial. Consequently, if the original polynomial, of large degree n, factors into $b \times n$ irreducible factors mod p (where perhaps $b = 1/2$ or $1/3$), then the amount of computation required to find which of the 2^{bn} factors mod p are also factors over the rational integers grows exponentially in n, even though the expected amount of computation required to obtain the complete factorization mod p grows only algebraically in n. Fortunately, however, most polynomials of degree n have only about $\ln n$ irreducible factors mod p, and if a particular polynomial which we wish to factor over the rational integers turns out to have unpleasantly many irreducible factors

mod p, we can simply factor it again modulo a still larger prime. While it is easily shown that such a strategy will factor "almost all" polynomials over the rational integers in a very modest amount of effort, the "worst" irreducible polynomials have at least $2^{n/2}$ factors modulo every prime. However, even for one of these polynomials our procedure is substantially better than the classical Kronecker algorithm. (The Kronecker algorithm is presented in Section 25, p. 77 of van der Waerden (1931).)

Some of the material presented in this paper is based on other work. The central notion of this paper, which is the $\mathcal{Q}$ matrix of Section 3, appeared in a previous paper, Berlekamp (1967), a revised version of which was republished as Section 6.1 of Berlekamp (1968). The algorithm presented there succeeded in factoring an arbitrary polynomial of degree n over GF(q) in an amount of computation which grew only algebraically in n, but it was proportional to q rather than algebraic in log q. The fact that the $\mathcal{Q}$ matrix could be used to determine the number of factors had been anticipated by Schwarz (1956), but he gave no procedure for finding the actual factors. The results of Sections 5 and 6 are based on a suggestion of L. Welch (1968), and the results of Section 7 are based on a suggestion of G. Collins (1967) and D. Knuth (1967). The algorithm of Section 7 has apparently been independently discovered by a number of authors, several of whom are listed by Knuth (1969). Indeed, Knuth (1969) gives a more general probabilistic algorithm which finds factors as well as roots over GF(p).

The principal innovation of this paper is the deterministic procedure of Sections 3 and 4, which allows us to reduce an arbitrary factoring problem to a root-finding problem, thereby postponing the probabilistic part of the general factorization algorithm as much as possible. In this manner, we minimize the amount of computation which might be caused by a run of bad luck. We also maximize the opportunities to escape the randomization entirely by the use of special tricks, some of which are discussed in Section 7.

2. **Prerequisites.** In this section we list several results which are required in subsequent sections and which are well known in the theory of finite fields. We also assume that the reader is familiar with the known techniques for performing arithmetic operations on the elements of a finite field. These techniques may be found in Chapter 2 of Berlekamp (1968) and Collins (1969).

LEMMA 2.1. *If $g(x)$ is a polynomial over* GF(q), *then* $(g(x))^q = g(x^q)$.

LEMMA 2.2. *In* GF(q), $x^q - x = \prod_{s \in GF(q)} (x - s)$.

LEMMA 2.3. *In* GF(q), $x^{q^m} - x$ *factors into the product of all monic irreducible polynomials of degrees dividing m.*

LEMMA 2.4. *Let $f(x)$ and $g^{(i)}(x)$ all be monic polynomials over* GF(q), *and suppose that the $g^{(i)}(x)$ are relatively prime. If $f(x) \mid \prod_i g^{(i)}(x)$, then*

$$f(x) = \prod_i \gcd\,(f(x), g^{(i)}(x))$$

where gcd *denotes the monic common divisor of greatest degree.*

LEMMA 2.5. *Every repeated factor of $f(x)$ divides its derivative, $f'(x)$, and* deg $f'(x) <$ deg $f(x)$.

LEMMA 2.6. *A polynomial $f(x)$ over* GF(q) *has zero derivative iff there exists another polynomial, $g(x)$, such that $f(x) = g(x^p)$, where p, the characteristic of the field, is the prime divisor of q.*

Lemmas 2.5 and 2.6 provide a method by which we may reduce the factorization of an arbitrary polynomial, $f(x)$, to the factorization of a polynomial whose irreducible factors are all distinct. We first compute the derivative, $f'(x)$. If it is zero, we then have

$$f(x) = \sum_{i=0}^{n/p} f_i x^{pi} = \left(\sum_{i=0}^{n/p} f_i^{1/p} x^i \right)^p$$

and since the pth root of an element in GF(p^m) is also its p^{m-1} power, which we may calculate with less than $2 \log_2 (p^{m-1})$ multiplications, we may reduce the factorization of $f(x)$ to the factorization of the pth root of $f(x)$. On the other hand, if $f'(x)$ is nonzero, then we may compute the gcd $(f(x), f'(x))$.[1] If this greatest common divisor has positive degree, then it is a proper factor of $f(x)$. After dividing out this factor, we may reapply the same test. Finally, if gcd $(f(x), f'(x)) = 1$, then we know that $f(x)$ has no repeated factors.

In this manner, we may remove the repeated factors of $f(x)$.

Although it may be advisable to eliminate the repeated factors of $f(x)$ immediately, it is not necessary to do so. The algorithm of the following sections will factor an arbitrary polynomial over GF(q) into its irreducible-power factors. Thus if one prefers, he may first use the algorithm of the following sections to factor $f(x)$ into irreducible powers, and then compute the derivative of each irreducible-power factor to factor it into the power of an irreducible polynomial.

3. From Factorization of an Arbitrary Polynomial over GF(q) to the Factorization of a Polynomial Whose Irreducible-Power Factors all have the Same Degree. If $f(x)$ is a polynomial of degree n over GF(q), then for $i = 0, 1, 2, \cdots, n$ we define $h^{(i)}(x)$ as the product of all of the irreducible-power factors of $f(x)$ which have degree i. We then have the factorization

$$f(x) = \prod_{i=1}^{n} h^{(i)}(x). \tag{3.01}$$

In this section, we will present two methods for determining the $h^{(i)}(x)$.

Of course, the degree of each $h^{(i)}$ must be a multiple of i. Most $h^{(i)}$ will be 1, but the factorization of Eq. (3.01) will be nontrivial unless all irreducible-power factors of $f(x)$ have the same degree. Even in that case, the factorization of Eq. (3.01) will reveal the number of factors of $f(x)$ of each degree.

Before presenting our new approach to obtaining the factorization of Eq. (3.01), we review an older, better known approach which works whenever $f(x)$ has no repeated factors. For each successive i, the older algorithm computes $h^{(i)}(x)$, $F^{(i)}(x) = \prod_{j=i+1}^{n} h^{(j)}(x)$, and $R^{(i)}(x)$, the residue of x^{q^i} mod $F^{(i)}(x)$, as follows:

Algorithm 3.02: *Initialization.*

$$R^{(0)}(x) = x, \qquad F^{(0)}(x) = f(x).$$

[1] As described in Chapter 6 of Berlekamp (1968), we may compute the discriminant, $D(f)$, along with gcd $(f(x), f'(x))$. For any given $D \neq 0$, we can invoke Stickelberger's theorem to determine whether the number of irreducible factors of $f(x)$ is odd or even. This tells us only a little bit about the factorization of f, but it requires very little computation.

Recursion.

$$R^{(i)}(x) \equiv (R^{(i-1)}(x))^q \bmod F^{(i-1)}(x),$$
$$\deg R^{(i)} < \deg F^{(i-1)},$$
$$h^{(i)}(x) = \gcd(F^{(i-1)}(x), R^{(i)}(x) - x),$$
$$F^{(i)}(x) = F^{(i-1)}(x)/h^{(i)}(x).$$

Assuming that $f(x)$ has no repeated factors, Lemma 2.3 enables us to verify that each $h^{(i)}(x)$ computed by Algorithm 3.02 is indeed the product of the irreducible factors which have degree i.

In order to obtain $R^{(i+1)}$, Algorithm 3.02 must compute the qth power of $R^{(i)}$ mod $F^{(i)}$. The conventional way of doing this is to compute the residues of $(R^{(i)})^2, (R^{(i)})^4, (R^{(i)})^8, \cdots, (R^{(i)})^{2^{\lfloor \log q \rfloor}}$,[2] and then obtain $R^{(i+1)}$ by multiplying together an appropriate combination of these residues mod $F^{(i)}$. This requires between $\log q$ and $2 \log q$ multiplications mod $F^{(i)}$, and if $f(x)$ is irreducible or the product of two irreducible polynomials each of degree $n/2$, then it will be necessary to calculate $n/2$ successive $R^{(i)}$, each of which is congruent to $(R^{(i-1)})q \bmod f(x)$.

Instead of calculating each of the successive $R^{(i)}$ independently, we might first compute $r^{(0)}(x), r^{(1)}(x), \cdots, r^{(n-1)}(x)$, where each $r^{(i)}(x)$ is the residue of x^{iq} modulo $f(x)$. We could then calculate $R^{(i)}(x)$ from the formula

$$R^{(i)}(x) = \sum_{j=0}^{n-1} R_j^{(i-1)} x^j = \sum_{j=0}^{n-1} R_j^{(i-1)} r^{(j)}(x). \tag{3.03}$$

If we introduce the $n \times n$ matrix, $\mathcal{Q}$, whose n rows are the coefficients of $r^{(0)}, r^{(1)}, \cdots, r^{(n-1)}$, Eq. (3.03) may then be rewritten as

$$[R_0^{(i)}, R_1^{(i)}, \cdots, R_{n-1}^{(i)}] = [R_0^{(i-1)}, R_1^{(i-1)}, \cdots, R_{n-1}^{(i-1)}]\mathcal{Q}. \tag{3.04}$$

Once we have calculated the matrix $\mathcal{Q}$, Eq. (3.04) provides a fast method of calculating $R^{(i)}$ from $R^{(i-1)}$. If we use $\mathcal{Q}$ to calculate several successive $R^{(i)}$, the total savings more than justifies the initial cost of computing $\mathcal{Q}$. For large q, the time required to compute $\mathcal{Q}$ is dominated by the time required to compute the second line, $r^{(1)}(x) = R^{(1)}(x)$. Once $r^{(1)}(x)$ is known, each successive row of $\mathcal{Q}$ may be obtained in only one multiplication and reduction mod $f(x)$.

Although the $\mathcal{Q}$ matrix does serve to expedite the calculation of Algorithm 3.02, it plays a relatively peripheral role. On the other hand, this same matrix lies at the heart of the factorization procedure of Berlekamp (1967). Although conceptually more complicated, this procedure improves on Algorithm 3.02 in several respects. First, and most important, it provides us with tools which prove useful in factoring each $h^{(i)}(x)$. Second, it turns out that even in the case when each irreducible factor of $f(x)$ has a different degree, the new procedure often obtains the factorization in substantially fewer computations than Algorithm 3.02.

The key to the factoring procedure of Berlekamp (1967) lies in finding one or more other polynomials, $g(x)$, such that

$$g(x)^q - g(x) \equiv 0 \bmod f(x), \qquad 0 \leqq \deg g < \deg f. \tag{3.05}$$

[2] Unless another base is explicitly given, "log" means "$\log_2$".

Berlekamp (1967) has shown that if $f(x)$ is the product of powers of r distinct irreducible polynomials, then there are q^r solutions of Eq. (3.05). These solutions and the value of r may be found by solving a system of n linear equations in the n unknown coefficients of $g(x)$ over GF(q), namely

$$g(\mathcal{Q} - \mathcal{I}) = 0 \tag{3.06}$$

where $\mathcal{I}$ is the $n \times n$ identity matrix and $\mathcal{Q}$ is the $n \times n$ matrix whose ith row is the coefficients of $x^{q(i-1)}$ reduced modulo $f(x)$.

If $g(x)$ is any solution of Eq. (3.05), then Lemmas 2.1, 2.2, and 2.4 give us a factorization of $f(x)$, namely

$$f(x) = \prod_{s \in GF(q)} \gcd\,(f(x), g(x) - s). \tag{3.07}$$

This factorization of $f(x)$ is nontrivial unless deg $g(x) = 0$. The factors of $f(x)$ given by Eq. (3.07) may not be irreducible powers, but in that case the factorization may be further refined by computing the gcd's of the composite factors and $g(x) - s$ for other legitimate choices of $g(x)$.

If q is small, this method succeeds in factoring an arbitrary polynomial of degree n. The number of GF(q) computations required is proportional to n^3, most of which are spent finding r linearly independent g's from Eq. (3.06).[3] However, if q is large compared to n, then Eq. (3.07) becomes the bottleneck step of the computation. If q is very large, it becomes impractical to compute the gcd of $f(x)$ and $g(x) - s$ for each $s \in$ GF(q). Of course, most of these computations will prove useless, since at least $q - r$ of these gcd's must be one.

One method of dealing with Eq. (3.07) when q is large was recently proposed by Zassenhaus (1969). If S denotes the subset of GF(q) consisting of those s for which gcd $(f(x), g(x) - s) \neq 1$, then Eq. (3.07) can be simplified to

$$f(x) = \prod_{s \in S} \gcd\,(f(x), g(x) - s)$$

from which

$$f(x) \mid \prod_{s \in S} (g(x) - s).$$

We define

$$G(y) = \prod_{s \in S} (y - s) = \prod_{i=0}^{|S|} G_i y^i.$$

Since $f(x) \mid G(g(x))$, we have the congruence

$$\sum_{i=0}^{|S|} G_i (g(x))^i \equiv 0 \bmod f(x).$$

This congruence enables us to determine the polynomial G by computing the residues mod $f(x)$ of $1, g(x), (g(x))^2, (g(x))^3, \cdots$ until we find a power of $g(x)$ which is linearly dependent on its predecessors. Since deg $G = |S| \leqq r$, the residues of $1, g(x), (g(x))^2, \cdots, (g(x))^r$ cannot be linearly independent. The coefficients of the first linear depend-

[3] If n is large but q and r are small, then the number of GF(q) computations which Algorithm 3.02 requires to find the gcd's is also proportional to n^3, and the constant of proportionality is substantially higher for Algorithm 3.02 than for the procedure of Eqs. (3.06) and (3.07).

ence among these residues mod $f(x)$ are the coefficients of the polynomial G. The values of s which yield nontrivial factors of f via Eq. (3.07) are the roots of G. Thus, the Zassenhaus algorithm transforms the problem of factoring f into the problem of finding the roots of G.

Unfortunately, as we shall see in Section 7, the best algorithms known for finding the roots of a polynomial in a large prime field are probabilistic in nature. Although these algorithms usually run quite quickly, it is difficult to obtain any reasonable upper bound on the amount of time they may require in the most unlucky case. For this reason, we present a new, deterministic algorithm which obtains the partial factorization of Eq. (3.01). Although conceptually complicated, the algorithm runs quite fast. In Section 4 we show how an extension of this algorithm may be used to reduce the problem of factoring the product of r irreducible d-tics to the problem of finding the roots of a polynomial of degree r.

We begin with some definitions.

Definition 3.08. An $r \times r$ *matrix of polynomials*, $\mathfrak{M}$, over GF(q), is a matrix whose entries are polynomials in one indeterminant over GF(q). We say that such a matrix is *invertible*[4] iff its determinant, $|\mathfrak{M}|$, is a nonzero polynomial. We say that a matrix is *unimodular* iff $|\mathfrak{M}|$ is a nonzero scalar.

Definition 3.09. Two matrices of polynomials, $\mathfrak{B}$ and $\mathfrak{C}$ are said to be *equivalent* (written $\mathfrak{B} \cong \mathfrak{C}$) iff there exists a unimodular matrix of polynomials, $\mathfrak{N}$, and an invertible matrix of scalars, $\mathfrak{S}$, such that $\mathfrak{C} = \mathfrak{N}\mathfrak{B}\mathfrak{S}$.

It is trivially verified that the relation "$\cong$" is transitive, reflexive, and symmetric.

THEOREM 3.1. *If* $f(x) = \prod_{i=1}^{r} f^{(i)}(x)$, *where each* $f^{(i)}(x)$ *is the power of a distinct irreducible polynomial, and if* $1, g^{(2)}(x), g^{(3)}(x), \cdots, g^{(r)}(x)$ *are linearly independent monic solutions of the equation*

$$g^{(i)}(x^q) - g^{(i)}(x) \equiv 0 \bmod f(x)$$

and $0 < \deg g^{(i)} < \deg f$, *if*

$$\mathfrak{A} = \begin{bmatrix} f & 0 & 0 & 0 & \cdots & 0 \\ g^{(2)} & -1 & 0 & 0 & \cdots & 0 \\ g^{(3)} & 0 & -1 & 0 & \cdots & 0 \\ g^{(4)} & 0 & 0 & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ g^{(r)} & 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \tag{3.11}$$

and if

$$\mathfrak{F} = \begin{bmatrix} f^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & f^{(2)} & 0 & & 0 \\ 0 & 0 & f^{(3)} & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & & f^{(r)} \end{bmatrix} \tag{3.12}$$

then $\mathfrak{F} \cong \mathfrak{A}$.

[4] If $\mathfrak{M}$ is invertible, it has an inverse, $\mathfrak{M}^{-1}$, whose entries are quotients of polynomials in x. By Cramer's rule, $\mathfrak{M}^{-1}$ will be a matrix of polynomials iff $\mathfrak{M}$ is unimodular.

Proof. Let $g^{(1)} = 1$, and let the $r \times r$ matrix of polynomials $\mathcal{S}$ be defined by

$$\mathcal{S}_{i,j} \equiv g^{(i)}(x) \bmod f^{(j)}(x), \qquad 0 \leqq \deg \mathcal{S}_{i,j} < \deg f^{(j)}.$$

We claim that

LEMMA 3.13. *Every entry in the jth column of the product matrix* $\mathcal{A}\mathcal{S}$ *is a multiple of* $f^{(j)}$.

LEMMA 3.14. $\mathcal{S}$ *is a matrix of scalars.*

LEMMA 3.15. $\mathcal{S}$ *is invertible.*

LEMMA 3.16. $\mathcal{F}(\mathcal{A}\mathcal{S})^{-1}$ *is a unimodular matrix of polynomials.*

Assuming the lemmas, the theorem follows from the formula

$$(\mathcal{F}(\mathcal{A}\mathcal{S})^{-1})\mathcal{A}\mathcal{S} = \mathcal{F}.$$

We now prove the lemmas.

3.13. This lemma is an immediate consequence of the definitions of $\mathcal{A}$ and $\mathcal{S}$.

3.14. In GF(q),

$$g^{(i)}(x^q) - g^{(i)}(x) = \prod_{s \in GF(q)} (g^{(i)}(x) - s).$$

Since $g^{(i)}(x^q) - g^{(i)}(x) \equiv 0 \bmod \prod_{j=1}^{r} f^{(j)}(x)$,

$$f^{(j)}(x) \mid \prod_{s \in GF(q)} (g^{(i)}(x) - s).$$

Since the factors in this product are relatively prime, and $f^{(j)}(x)$ is an irreducible power, there must exist one particular *scalar* $s_{i,j}$ for which $f^{(j)}(x) \mid g^{(i)}(x) - s_{i,j}$ and

$$g^{(i)}(x) \equiv s_{i,j} \mod f^{(j)}(x).$$

3.15. If $\mathcal{S}$ were singular, then $\exists$ scalars $A_1, A_2, \cdots, A_r$, not all zero, such that

$$\sum_i A_i \mathcal{S}_{i,j} = 0 \quad \text{for all } j.$$

This implies that

$$\sum_i A_i g^{(i)}(x) \equiv 0 \bmod f^{(j)}(x) \quad \text{for all } j,$$

so

$$\sum_i A_i g^{(i)}(x) \equiv 0 \bmod f(x),$$

and since $\deg g^{(i)} < \deg f$, we conclude that $\sum_i A_i g^{(i)}(x) = 0$, contradicting the linear independence of $g^{(1)}, g^{(2)}, \cdots, g^{(r)}$.

3.16. In view of Lemma 3.13, the cofactor of the i, j entry in $\mathcal{A}\mathcal{S}$ is a multiple of $\prod_{k \neq j} f^{(k)}(x) = f(x)/f^{(j)}(x)$. Furthermore, $|\mathcal{A}\mathcal{S}| = |\mathcal{A}|\ |\mathcal{S}| = f(x) \cdot$ scalar. Hence, if we evaluate $(\mathcal{A}\mathcal{S})^{-1}$ by Cramer's rule, we find that every element in the jth row of $(\mathcal{A}\mathcal{S})^{-1}$ is of the form polynomial $(x)/f^{(j)}(x)$. It follows that the product $\mathcal{F}(\mathcal{A}\mathcal{S})^{-1}$ is a matrix of polynomials. Its determinant is given by $|\mathcal{F}|\ |\mathcal{S}|^{-1}\ |\mathcal{A}|^{-1} = f(x) \cdot$ scalars$/f(x) =$ scalar. Since $|\mathcal{F}| \neq 0$, $|\mathcal{S}| \neq 0$, and $|\mathcal{A}| \neq 0$, $\mathcal{F}(\mathcal{A}\mathcal{S})^{-1}$ is unimodular. Q.E.D.

Although the proof of Theorem 3.1 considered the matrix $(\mathfrak{A}\mathfrak{S})^{-1}$, whose entries were elements in the field of rational functions, $GF(p^m)(x)$, the statement of Theorem 3.1 involves only matrices over the polynomial ring $GF(p^m)[x]$. Henceforth, all operations we consider are restricted to this ring.

Since the polynomials in the first column of $\mathfrak{A}$ may be found by solving the matrix equation (3.06), Theorem 3.1 shows that the factors of $f(x)$ may be found by diagonalizing the matrix of polynomials, $\mathfrak{A}$, to obtain $\mathfrak{F}$, whose diagonal elements are the factors of $f(x)$. We now consider the problem of diagonalizing a matrix of polynomials.

A square matrix of polynomials may be transformed into another matrix of polynomials by any of the following *elementary operations*:

(1) Permute any pair of rows.

(2) Multiply any row by a nonzero scalar.

(3) Add a scalar times a power of x times a row into any other row, and determine the maximum degree of the elements in the new row.

(4) Permute any pair of columns.

(5) Multiply any column by a nonzero scalar.

(6) Add any scalar multiple of any column into any other column.

The row operations are the elementary unimodular operations.

It is obvious that the new matrix formed by any of these operations is equivalent to the original matrix, because operations (1)–(3) may be performed by premultiplying the orginal matrix by an appropriate unimodular matrix of polynomials, and operations (4)–(6) may be performed by postmultiplying the original matrix by an appropriate invertible matrix of scalars.

Since operation (3) is more powerful than operation (6), we can perform a greater variety of operations on rows than columns. Thus, rows have a special importance.

Definitions 3.17. The *degree of a row* of a matrix of polynomials is the maximal degree of any of the entries in that row. The degree of an all-zero row is conventionally taken as -1.[5] The *total row degree* of a matrix of polynomials is the sum of the degrees of its rows.

An entry whose degree is equal to the maximum degree occurring in its row is said to be a *maxrowdeg* entry. An entry is called a *dominant entry* iff it is a diagonal entry and it is the unique maxrowdeg entry in its row. A row containing a dominant entry is called a *dominated row*. Likewise, a column containing a dominant entry is called a *dominated column*, even though some of the nondiagonal entries in the dominated column may have higher degrees than the diagonal entry.

A matrix of polynomials is said to be a *row-dominated matrix* iff all of its nonzero rows are dominated rows. The *complexity* of a matrix of polynomials is defined as twice its total row degree minus the number of its dominated rows.

THEOREM 3.2. *If an r by r matrix of polynomials is not row-dominated, then it is equivalent to another matrix of smaller complexity, which can be obtained from the original matrix in at most $r - 1$ elementary operations.*

Proof. Any matrix of polynomials which is not row-dominated must contain a diagonal entry which is not the unique maxrowdeg entry in its row. Let such an entry

[5] Notice that this convention violates the usual law, $\deg(fg) = \deg f + \deg g$ if f or g is 0. The law can be preserved only by taking $\deg 0 = \pm\infty$, which would lead to even more difficulties in the present context.

be called the *pivotal entry*, and let the row and column containing the pivotal entry be called the pivotal row and the pivotal column. We then consider two cases:

Case 1. The pivotal row contains a maxrowdeg entry in an undominated column.

In this case we may perform column operations which will convert the pivotal row to a dominated row, thereby increasing the number of dominated rows. Since we will not perform any row operations, we will not increase the degrees of any rows, and since our column operations will never add a dominated column into an undominated column, every row which was dominated in the original matrix will remain dominated in the less complex equivalent matrix.

Specifically, we make the pivotal entry a maxrowdeg entry, by column permutation if necessary. We then decrease the degree of every nonpivotal maxrowdeg entry in the pivotal row by adding to the column containing it an appropriate scalar multiple of the pivotal column.

Case 2. Every maxrowdeg entry in the pivotal row occurs in a dominated column.

In this case we define the *relevant set of rows* as the pivotal row and those rows which contain a dominant entry in the same column as a maxrowdeg entry in the pivotal row. We then select from the relevant set of rows a row having row degree at least as large as any other relevant row, and call this row the *key row*. We then decrease the degree of the key row by adding into the key row appropriate multiples of the other relevant rows. Since all of these operations effect only the key row, they will transform the original matrix into a matrix of smaller complexity, even if the original key row is dominated and the transformed key row is not dominated.

Specifically, if the key row is not the pivotal row, we begin by adding to the key row an appropriate scalar times an appropriate power of x times the pivotal row, chosen so as to reduce the degree of the key row's diagonal entry. If this reduces the degree of the key row, we are finished; if not (or if the key row is the pivotal row), we may assume that each maxrowdeg entry in the key row lies in the same column as a diagonal entry of some row in the relevant set. Adding an appropriate multiple of the corresponding relevant row into the key row will decrease the number of maxrowdeg entries in the key row, etc., until the degree of the key row is decreased. Q.E.D.

THEOREM 3.3. *Let* $\mathfrak{B}$ *be any* $r \times r$ *matrix of polynomials, with total row degree* b. *Then there exists a (possibly nonunique) row-dominated matrix,* $\mathfrak{R}$, *such that* $\mathfrak{B} \cong \mathfrak{R}$. $\mathfrak{R}$ *can be computed from* $\mathfrak{B}$ *in less than* $2(b + r)(r - 1)$ *elementary operations.*

Proof. The complexity of $\mathfrak{B}$ is no greater than $2b$. The complexity of $\mathfrak{R}$ is no less than $-2r$, with equality only if $\mathfrak{R}$ is the all-zero matrix. Theorem 3.3 is therefore a direct consequence of repeated applications of Theorem 3.2. Q.E.D.

If $\mathfrak{R}$ is nonsingular, its complexity is nonnegative and at most $2b(r - 1)$ elementary operations are required.

In general, $\mathfrak{B} \cong \mathfrak{R}$ means only that there exists a unimodular matrix of polynomials, $\mathfrak{C}$, and an invertible matrix of scalars, $\mathfrak{S}$, such that $\mathfrak{B}\mathfrak{S}^{-1} = \mathfrak{C}\mathfrak{R}$. In order to obtain certain information about the relative row degrees of $\mathfrak{B}$ and $\mathfrak{R}$, we begin by considering the special case, $\mathfrak{B} = \mathfrak{C}\mathfrak{R}$.

THEOREM 3.4. *If* $\mathfrak{B}$, $\mathfrak{C}$, *and* $\mathfrak{R}$ *are* $r \times r$ *matrices of polynomials, such that* $\mathfrak{B} = \mathfrak{C}\mathfrak{R}$ *and* $\mathfrak{R}$ *is row-dominated, then any row of* $\mathfrak{B}$ *which has row degree* d *is a linear combination only of rows of* $\mathfrak{R}$ *which have row degrees* $\leqq d$.

Proof. The degree of the ith row of $\mathfrak{B}$ is given by

$$\begin{aligned}\max_j \deg \mathfrak{B}_{i,j} &= \max_j \deg \sum_k \mathfrak{C}_{i,k}\mathfrak{R}_{k,j} \\ &\leqq \max_j \max_k \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,j} \\ &\leqq \max_k \max_j \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,j} \\ &\leqq \max_k \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k}.\end{aligned}$$

Let m be chosen so that

$$\deg \mathfrak{C}_{i,m}\mathfrak{R}_{m,m} = \max_k \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k}.$$

If $k \neq m$, then

$$\deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,m} < \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k} \leqq \deg \mathfrak{C}_{i,m}\mathfrak{R}_{m,m}$$

so

$$\begin{aligned}\deg \sum_k \mathfrak{C}_{i,k}\mathfrak{R}_{k,m} &= \deg \left(\mathfrak{C}_{i,m}\mathfrak{R}_{m,m} + \sum_{k\neq m} \mathfrak{C}_{i,k}\mathfrak{R}_{k,m}\right) \\ &= \deg \mathfrak{C}_{i,m}\mathfrak{R}_{m,m} \\ &= \max_k \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k}\end{aligned}$$

and

$$\max_j \deg \mathfrak{B}_{i,j} \geqq \deg \mathfrak{B}_{i,m} = \max \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k}.$$

Therefore,

$$\boxed{\max_j \deg \mathfrak{B}_{i,j} = \max \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k}.}$$

Hence, if $\max_j \deg \mathfrak{B}_{i,j} = d$, then $\max \deg \mathfrak{C}_{i,k}\mathfrak{R}_{k,k} = d$, whence $\deg \mathfrak{R}_{k,k} > d$ implies that $\mathfrak{C}_{i,k} = 0$. Q.E.D.

Definition 3.41. A *canonically-ordered matrix* is one whose successive rows have nonincreasing degrees. A *normalized* matrix is a canonically ordered row-dominated matrix in which all diagonal entries are monic polynomials. A *uniform* matrix is a normalized matrix in which each polynomial on the diagonal has the same degree.

THEOREM 3.5. *If $\mathfrak{B}$ is a canonically ordered $r \times r$ matrix, and $\mathfrak{R}$ is a normalized matrix, and $\mathfrak{B} \cong \mathfrak{R}$, then the degree of every row of $\mathfrak{B}$ is at least as great as the degree of the corresponding row of $\mathfrak{R}$.*

Proof. $\mathfrak{B} = \mathfrak{C}\mathfrak{R}\mathfrak{S}$, so $\mathfrak{B}\mathfrak{S}^{-1} = \mathfrak{C}\mathfrak{R}$. Since row degrees are unaffected by column operations, there is no loss of generality in assuming that $\mathfrak{B} = \mathfrak{C}\mathfrak{R}$. Since all nonzero rows of $\mathfrak{R}$ are dominated, they are linearly independent, and the rank of $\mathfrak{R}$ is equal to the number of its nonzero rows. If the degree of the ith row of $\mathfrak{B}$ is less than the degree of the ith row of $\mathfrak{R}$, then Theorem 3.4 implies that the ith, $(i+1)$st, $\cdots$ rows of $\mathfrak{B}$ are all linear combinations of the last $r - i$ rows of $\mathfrak{R}$. The rank of $\mathfrak{B}$ is therefore no greater than $i - 1 +$ the dimension of the space spanned by the last $r - i$ rows

of $\mathfrak{R}$. But the rank of $\mathfrak{R}$ is i + the dimension of the space spanned by the last $r - i$ rows of $\mathfrak{R}$, so $\mathfrak{C}$ must be singular, contradicting the definition of equivalence. Q.E.D.

THEOREM 3.6. *If two normalized matrices are equivalent, their corresponding rows have equal degrees.*

Proof. This is a direct consequence of Theorem 3.5.

A normalized matrix may be partitioned into various submatrices, such that $\mathfrak{R}_{i,j}$ and $\mathfrak{R}_{k,l}$ are in the same submatrix iff $\deg \mathfrak{R}_{i,i} = \deg \mathfrak{R}_{k,k}$ and $\deg \mathfrak{R}_{j,j} = \deg \mathfrak{R}_{l,l}$. This is called the *standard partition* corresponding to $\mathfrak{R}$. The partitioned matrix is said to be *triangular* with respect to this partition iff all submatrices below the main diagonal are zero. (See Fig. 1.) Notice that each diagonal submatrix is uniform.

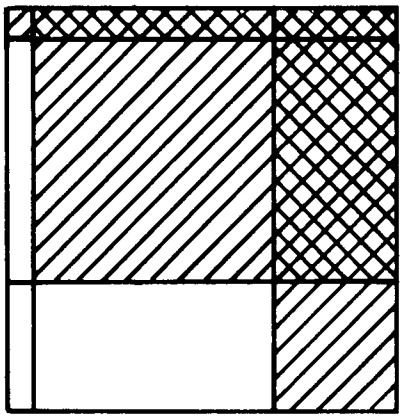

FIGURE 1. *A Partitioned Matrix With Diagonal Submatrices Shaded And Above-Diagonal Submatrices Crosshatched.*

THEOREM 3.7. *If $\mathfrak{R}$ and $\mathfrak{I}$ are equivalent normalized matrices, with $\mathfrak{I} = \mathfrak{C}\mathfrak{R}\mathfrak{S}$, then $\mathfrak{C}$ is triangular with respect to the standard partition of $\mathfrak{R}$ and $\mathfrak{I}$.*

Proof. With $\mathfrak{I}\mathfrak{S}^{-1} = \mathfrak{C}\mathfrak{R}$, the theorem follows directly from Theorem 3.4.

THEOREM 3.8. *If $\mathfrak{R}$ and $\mathfrak{I}$ are equivalent invertible normalized matrices, and $\mathfrak{I}$ is triangular with respect to its standard partition, then so is $\mathfrak{R}$. Furthermore, every uniform diagonal submatrix of $\mathfrak{R}$ (with respect to the standard partition) is equivalent to the corresponding uniform diagonal submatrix of $\mathfrak{I}$.*

Proof. We have $\mathfrak{R} = \mathfrak{C}\mathfrak{I}\mathfrak{S}$. $\mathfrak{I}$ is triangular by hypothesis, and $\mathfrak{C}$ is triangular by Theorem 3.7. Since the product of two triangular matrices is triangular, we deduce that $(\mathfrak{C}\mathfrak{I})$ is triangular with respect to the standard partition.

If $\mathfrak{D}$ is any matrix of polynomials, we may define $\hat{\mathfrak{D}}$ as the matrix of scalars obtained by setting

$$\hat{\mathfrak{D}}_{i,j} = \text{leading coefficient of } \mathfrak{D}_{i,j}, \quad \text{if } \mathfrak{D}_{i,j} \text{ is a maxrowdeg},$$

$$= 0, \quad \text{otherwise.}$$

Then since the equation $\mathfrak{R} = (\mathfrak{C}\mathfrak{I})\mathfrak{S}$ implies, among other things, that the leading coefficients of the entries of maximum degree in each row must be equal, we deduce that $\hat{\mathfrak{R}} = ((\mathfrak{C}\mathfrak{I})\hat{})\mathfrak{S}$. Since $\hat{\mathfrak{I}}$ is truly diagonal (without respect to any partition, even nonstandard ones), it is invertible and we have $\mathfrak{S}^{-1} = \hat{\mathfrak{R}}^{-1}((\mathfrak{C}\mathfrak{I})\hat{})$, which is triangular with respect to the standard partition. Therefore $\mathfrak{S}^{-1}$ and $\mathfrak{S}$ are also triangular. We

conclude that $\mathfrak{R} = (\mathfrak{C}\mathfrak{I})\mathfrak{S}$ is the product of triangular matrices, so it must be triangular too.

When one takes the product of matrices which are triangular with respect to the standard partition, a diagonal submatrix in the multiplier matrix is multiplied only by the corresponding diagonal submatrix of the multiplicand matrix. It follows that corresponding diagonal submatrices of equivalent normalized matrices are equivalent. Q.E.D.

Let us now review what we have shown. Given a polynomial $f(x)$, we may construct $\mathfrak{Q}$, solve Eq. (3.06), and thereby determine r [the number of distinct irreducible factors of $f(x)$] and a certain $r \times r$ matrix of polynomials, called $\mathfrak{A}$. Using Theorems 3.2 and 3.3, we may transform $\mathfrak{A}$ to $\mathfrak{R}$, which is a normalized matrix. From Theorem 3.1 we know that $\mathfrak{R} \cong \mathfrak{F}$, a truly diagonal matrix whose diagonal entries are the r monic irreducible-power factors of $f(x)$. Since $\mathfrak{F}$ is triangular, $\mathfrak{R}$ will automatically be triangular with respect to the standard partition. From Theorems 3.4–3.6, we may determine the degrees of the various irreducible-power factors, and the number of factors of each degree. If $f(x)$ has r_i irreducible-power factors of degree i, then according to Theorems 3.7 and 3.8, $\mathfrak{R}$ will have a corresponding $r_i \times r_i$ submatrix on its diagonal, and this submatrix will be uniform of degree i. Its determinant will be the polynomial $h^{(i)}(x)$ of Eq. (3.01).

Thus, we have obtained a decomposition of the matrix corresponding to the factorization of Eq. (3.01). By calculating the determinants of the corresponding submatrices, we could obtain each factor $h^{(i)}(x)$. However, if we wish to factor $h^{(i)}(x)$ into the product of irreducible-powers of degree i, then it is easier to proceed directly with the further manipulations on the corresponding r_i by r_i matrix of polynomials which are described in Section 4. There is no real need to evaluate the determinant of this matrix explicitly.

4. From Factorization of the Product of r Irreducible-Power d-tics to the Factorization of the Product of r Linear Factors. If we wish to factor a polynomial, $f(x)$, which is the product of r irreducible-power factors, $f^{(1)}(x)$, $f^{(2)}(x)$, $\cdots$, $f^{(r)}(x)$, each of which has the same degree, d, then we might employ the Zassenhaus algorithm, which is described in Section 3 between Eq. (3.07) and Definition (3.08). An alternative procedure continues with $\mathfrak{R}$, the $r \times r$ uniform[6] matrix of polynomials with which the new algorithm of Section 3 terminates. We know that $\mathfrak{R} \cong \mathfrak{F}$, where $\mathfrak{F}$ is the truly diagonal matrix of polynomials whose diagonal elements are $f^{(1)}(x)$, $f^{(2)}(x)$, $\cdots$, $f^{(r)}(x)$. In order to find $\mathfrak{F}$ from the given $\mathfrak{R}$, we require some additional results.

THEOREM 4.1. *If $\mathfrak{R}$ and $\mathfrak{F}$ are uniform matrices of polynomials and $\mathfrak{R} \cong \mathfrak{F}$, then $\mathfrak{R} = \mathfrak{S}^{-1}\mathfrak{F}\mathfrak{S}$ where $\mathfrak{S}$ is a matrix of scalars.*

Proof. Write $\mathfrak{R} = \mathfrak{C}\mathfrak{F}\mathfrak{S}$. Since $\mathfrak{F}$, $\mathfrak{R}$, and $\mathfrak{C}$ are matrices of polynomials, we may write

$$\mathfrak{R} = \mathfrak{R}_0 + \mathfrak{R}_1 x + \mathfrak{R}_2 x^2 + \cdots + \mathfrak{R}_i x^i,$$

$$\mathfrak{C} = \mathfrak{C}_0 + \mathfrak{C}_1 x + \mathfrak{C}_2 x^2 + \cdots + \mathfrak{C}_j x^j,$$

$$\mathfrak{F} = \mathfrak{F}_0 + \mathfrak{F}_1 x + \mathfrak{F}_2 x^2 + \cdots + \mathfrak{F}_k x^k,$$

[6] Recall Definitions 3.08, 3.09, 3.17, and 3.41.

where $\mathcal{R}_n$, $\mathcal{C}_n$, and $\mathcal{F}_n$ are matrices of scalars and $\mathcal{R}_i \neq 0$, $\mathcal{C}_i \neq 0$, and $\mathcal{F}_k \neq 0$. Since $\mathcal{R}$ and $\mathcal{F}$ are uniform, $\mathcal{R}_i = \mathcal{F}_k = \mathcal{I}$, the identity matrix. Equating the scalar matrix coefficients of the leading powers of x on both sides of the equation $\mathcal{R} = \mathcal{C}\mathcal{F}\mathcal{S}$ gives $i = j + k$ and $\mathcal{I} = \mathcal{C}_j\mathcal{I}\mathcal{S}$, where $\mathcal{C}_j = \mathcal{S}^{-1}$. Since $i = k$ by Theorem 3.6, we must have $j = 0$. Q.E.D.

We must now solve the equation $\mathcal{S}^{-1}\mathcal{R}\mathcal{S} = \mathcal{F}$ for $\mathcal{S}$ and $\mathcal{F}$, given $\mathcal{R}$.

Letting $\mathcal{S}^{(i)}$ denote the ith column of $\mathcal{S}$, we write

$$\mathcal{S} = [\mathcal{S}^{(1)}|\mathcal{S}^{(2)}|\mathcal{S}^{(3)}|\cdots|\mathcal{S}^{(r)}]$$

and similarly

$$\mathcal{F} = \begin{bmatrix} f^{(1)} & 0 & 0 & & 0 \\ 0 & f^{(2)} & 0 & & 0 \\ 0 & 0 & f^{(3)} & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f^{(r)} \end{bmatrix}.$$

The equation $\mathcal{R}\mathcal{S} = \mathcal{S}\mathcal{F}$ is equivalent to the equations

$$\mathcal{R}\mathcal{S}^{(i)} = f^{(i)}(x)\mathcal{S}^{(i)} \quad \text{for } i = 1, 2, \cdots, r.$$

Writing out both sides explicitly as polynomials in x gives

$$\sum_{j=0}^{d} \mathcal{R}_j\mathcal{S}^{(i)}x^j = \sum_{j=0}^{d} f_j^{(i)}\mathcal{S}^{(i)}x^j$$

from which we deduce that

$$\mathcal{R}_j\mathcal{S}^{(i)} = f_j^{(i)}\mathcal{S}^{(i)} \quad \text{for} \quad i = 1, 2, \cdots, r; \quad j = 0, 1, 2, \cdots, d$$

and

$$[\mathcal{R}_j - f_j^{(i)}\mathcal{I}]\mathcal{S}^{(i)} = 0.$$

Since $\mathcal{S}^{(i)} \neq 0$,

$$|\mathcal{R}_j - f_j\mathcal{I}| = 0; \qquad f_j = f_j^{(1)}, f_j^{(2)}, \cdots, f_j^{(r)}.$$

If $\mathcal{R}_j$ is diagonal, then the solutions for the scalar f_j are trivially seen to be the diagonal components of $\mathcal{R}_j$. If $\mathcal{R}_j$ is not diagonal, then the determinant $|\mathcal{R}_j - f_j\mathcal{I}|$ is a polynomial in f_j of degree r over GF(q), and it must have r (possibly not distinct) roots in GF(q).

Thus, we may determine the set $f_j^{(1)}, f_j^{(2)}, \cdots, f_j^{(r)}$ from the equation $|\mathcal{R}_j - f_j\mathcal{I}| = 0$.

The details of the problem of finding the roots of this polynomial are discussed in Sections 5–7.

In general, these r eigenvalues of $\mathcal{R}_j$ (and $\mathcal{F}_j$) may not all be distinct. In that case, we may partition them into disjoint sets, and obtain a corresponding partition of $\mathcal{F}_j$. This partition will be trivial (i.e., it will partition all rows and columns together) iff $\mathcal{R}_j$ is a scalar multiple of the identity matrix. It is obvious that this cannot happen if $\mathcal{R}_j$ is not diagonal.

Knowing $f_i^{(1)}, f_i^{(2)}, \cdots, f_i^{(r)}$, we may find a scalar matrix $\mathcal{S}_i$ such that $\mathcal{S}_i^{-1}\mathcal{R}_i\mathcal{S}_i = \mathcal{T}_i$. The ith column of $\mathcal{S}_i$ may be taken as any solution of the equation

$$[\mathcal{R}_i - f_i^{(i)}\mathcal{I}]\mathcal{S}_i^{(i)} = 0.$$

If $f_i^{(i)}$ has multiplicity l and $f_i^{(i)} = f_i^{(i+1)} = \cdots = f_i^{(i+l-1)}$, then the vector solutions of this equation form an l-dimensional subspace, any basis of which may be selected as $\mathcal{S}_i^{(i)}, \mathcal{S}_i^{(i+1)}, \cdots, \mathcal{S}_i^{(i+l-1)}$.

We now assert that the matrix $\mathcal{S}_i$ which truly diagonalizes $\mathcal{R}_i$ also diagonalizes $\mathcal{R}$ with respect to the (known) partition of $\mathcal{T}_i$. This fact follows from the observation that if $f_i^{(i)}$ has multiplicity l and $f_i^{(i)} = f_i^{(i+1)} = \cdots = f_i^{(i+l-1)}$, then the vectors $\mathcal{S}_i^{(i)}, \mathcal{S}_i^{(i+1)}, \cdots, \mathcal{S}_i^{(i+l-1)}$ and $\mathcal{S}^{(i)}, \mathcal{S}^{(i+1)}, \cdots, \mathcal{S}^{(i+l-1)}$ are both bases of the same space, and hence linear combinations of each other. Thus, we have obtained a further decomposition of $\mathcal{R}$. If it is not yet truly diagonal, then we may reapply the same procedure to each diagonal submatrix of $\mathcal{S}_i^{-1}\mathcal{R}\mathcal{S}_i$ until we eventually obtain a matrix whose diagonal entries are the irreducible-power factors of $f(x)$.

5. From Root-Finding in GF(p^m) to Root-Finding in GF(p), p Small. In Sections 3 and 4, we have given a deterministic procedure whereby the problem of factoring an arbitrary polynomial over GF(q) may be reduced to the problem of finding the roots in GF(q) of several other polynomials, each of which has degree no greater than the number of irreducible-power factors of $f(x)$ of a particular degree. We now consider the root-finding problem in GF(p^m). In this section and the next. we give algorithms which reduce the problem of finding the roots of $f(x)$ to the problem of finding the roots of another polynomial which splits in GF(p). Although the algorithms we present here are immediate consequences of the well-known properties of conjugate polynomials and polynomial norms, the algorithms themselves are little known in the subject of error-correcting codes, where computational problems in nonprime finite fields have great practical importance.

In order to represent the elements of GF(p^m), we must begin by specifying an element to be called α, which is a root in GF(p^m) of some polynomial of degree m which is irreducible over GF(p). The minimal polynomial of α is initially selected by some ad hoc procedure. The coefficients of the minimal polynomial of α are often wired into the circuitry for doing computations in GF(p^m). Details are given by Berlekamp (1968).

In GF(p^m),

$$x^{p^m} - x = \prod_{s \in GF(p)} (\mathrm{Tr}(x) - s) \tag{5.01}$$

where

$$\mathrm{Tr}(x) = \sum_{i=0}^{m-1} x^{p^i}.$$

Therefore, if

$$x^{p^m} - x \equiv 0 \bmod f(x),$$

then

$$\prod_{s \in GF(p)} (\mathrm{Tr}(x) - s) \equiv 0 \bmod f(x).$$

Hence, if $f(x)$ is a nonlinear polynomial which splits in GF(p^m), then by Lemma 2.4, we have the factorization

$$f(x) = \prod_{s \in GF(p)} \gcd\,(f(x),\ \mathrm{Tr}(x) - s) \tag{5.02}$$

where gcd denotes the *monic* common divisor of greatest degree. If p is small, Eq. (5.02) enables us to factor $f(x)$ even if m is large.

We proceed as follows:

Compute the residues of $x, x^2, x^{2^2}, \cdots, x^{2^{\lfloor \log p \rfloor}}, x^p, x^{p^2}, \cdots, x^{p^{m-1}}$ modulo $f(x)$. By adding together these residues, compute the residue of $\mathrm{Tr}(x)$. If $\mathrm{Tr}(x)$ is not congruent to a scalar, then $f(x)$ factors according to Eq. (5.02). If $\mathrm{Tr}(x)$ is congruent to a scalar, then the factorization of Eq. (5.02) degenerates into the trivial result: $f(x) = f(x)\cdot\prod 1$.

In this case, additional assaults are required. Let α be the root of an irreducible (not necessarily primitive) polynomial of degree m over GF(p). Then $\alpha^0, \alpha, \alpha^2, \cdots, \alpha^{m-1}$ form a basis for GF(p^m) over GF(p). Substituting $\alpha^i x$ for x in Eq. (5.01) gives

$$(\alpha^i)^{p^m} x^{p^m} - \alpha^i x = \prod_{s \in GF(p)} (\mathrm{Tr}(\alpha^i x) - s).$$

Since $\alpha^i \in$ GF(p^m), $(\alpha^i)^{p^m} = \alpha^i$, and we have

$$x^{p^m} - x = \alpha^{-i} \prod_{s \in GF(p)} (\mathrm{Tr}(\alpha^i x) - s).$$

We thus obtain the following generalization of Eq. (5.02):

$$f(x) = \prod_{s \in GF(p)} \gcd\,(f(x),\ \mathrm{Tr}(\alpha^i x) - s). \tag{5.03}$$

If Eq. (5.03) yields a trivial factorization when $j = 0$, we may reapply Eq. (5.03) with $j = 1, 2, 3, \cdots, m-1$.

We shall now show that the additive property of traces implies that at least one of these m special cases of Eq. (5.03) must yield a nontrivial factorization. Let $\rho_1, \rho_2, \cdots, \rho_n$ be the roots of $f(x)$ in GF(p^m). Then Eq. (5.03) yields a trivial factorization iff there exists an s (which may depend on j) such that $f(x)$ divides $\mathrm{Tr}(\alpha^i x) - s$, which implies that $s = \mathrm{Tr}(\alpha^i \rho_1) = \mathrm{Tr}(\alpha^i \rho_2) = \cdots = \mathrm{Tr}(\alpha^i \rho_n)$. If all m factorizations are trivial then for any j, $0 \leqq j < m$, and any i, k, $1 \leqq i < k \leqq n$, we must have

$$\mathrm{Tr}(\alpha^i \rho_i) = \mathrm{Tr}(\alpha^i \rho_k) \quad \text{or} \quad \mathrm{Tr}(\alpha^i(\rho_i - \rho_k)) = 0$$

and for any $A_0, A_1, \cdots, A_{m-1} \in$ GF(p),

$$\sum_{i=0}^{m-1} A_i \mathrm{Tr}(\alpha^i(\rho_i - \rho_k)) = 0 = \mathrm{Tr}\left(\left(\sum_{i=0}^{m-1} A_i \alpha^i\right)(\rho_i - \rho_k)\right) = 0.$$

Since $\alpha^0, \alpha^1, \alpha^2, \cdots, \alpha^{m-1}$ form a basis of GF(p^m) over GF(p), this means that

$$\mathrm{Tr}(\xi(\rho_i - \rho_k)) = 0 \quad \text{for all } \xi \text{ in } GF(p^m). \tag{5.04}$$

If $\rho_i \neq \rho_k$, then every element in GF(p^m) is of the form $\xi(\rho_i - \rho_k)$, and Eq. (5.04) implies that every element in GF(p^m) has trace 0. But it is evident from Eq. (5.01) that only p^{m-1} elements in GF(p^m) have trace 0, so Eq. (5.04) must be false and Eq. (5.03) must therefore yield a nontrivial factorization for some j, $0 \leqq j < m$.

In practice, there is rarely any need to apply all m versions of Eq. (5.03). For any $\xi \in \mathrm{GF}(p^m)$, we may factor $f(x)$ as

$$f(x) = \prod_{s \in \mathrm{GF}(p)} \gcd\,(f(x), \mathrm{Tr}(\xi x) - s). \tag{5.05}$$

This factorization will be trivial iff $\mathrm{Tr}(\xi\rho_1) = \mathrm{Tr}(\xi\rho_2) = \cdots = \mathrm{Tr}(\xi\rho_n)$. In some cases, this failure may be avoided by wise a priori choice of ξ. Since $f_{n-1} = -\sum_{i=1}^{n} \rho_i$, we know that

$$\mathrm{Tr}(f_{n-1}) = -\sum_{i=1}^{n} \mathrm{Tr}(\rho_i).$$

If n is a multiple of p, we cannot have $\mathrm{Tr}(\rho_1) = \mathrm{Tr}(\rho_2) = \cdots = \mathrm{Tr}(\rho_n)$ unless $\mathrm{Tr}(f_{n-1}) = 0$. Hence, if n is a multiple of p, we may insure a nontrivial factorization in Eq. (5.05) by choosing ξ so that $\mathrm{Tr}(f_{n-1}\xi^{-1}) \neq 0$. If $f_{n-1} \neq 0$, this is easily accomplished.

The methods introduced in this section are the best methods known for factoring $f(x)$ over $\mathrm{GF}(p^m)$ when p is small and $n = \deg f$ is large. When n is small (in particular if $p = 2$ and $\deg f = 2$, 3, or 4), the methods of this section are inferior to those given by Berlekamp, Rumsey, and Solomon (1967), and expanded in Chapter 11 of Berlekamp (1968).

6. From Root-Finding in $\mathrm{GF}(p^m)$ to Root-Finding in $\mathrm{GF}(p)$, p Large. We now consider the problem of finding the roots of the polynomial $f(x)$ which splits in $\mathrm{GF}(p^m)$.

The polynomial whose roots we wish to find is represented as

$$f(x) = f(\alpha, x) = \sum_{i=0}^{m-1} \sum_{j=0}^{n} f_{i,j}\alpha^i x^j,$$

where $f_{i,j} \in \mathrm{GF}(p)$, $f_{0,n} = 1$ and $f_{i,n} = 0$ if $i \neq 0$. Knowing that $x^{p^m} \equiv x \bmod f(x)$, we wish to find $\beta_1, \beta_2, \cdots, \beta_n \in \mathrm{GF}(p^m)$ such that

$$f(\alpha, x) = \prod_{k=1}^{n} (x - \beta_k).$$

To find these roots, we first calculate the new polynomial,

$$F(x) = \prod_{k=0}^{m-1} f(\alpha^{p^k}, x) = \prod_{k=0}^{m-1} \sum_{i=0}^{m-1} \sum_{j=0}^{n} f_{i,j}\alpha^{ip^k} x^j.$$

We shall now show that $F(x)$ is a polynomial of degree mn over $\mathrm{GF}(p)$. Since the jth coefficient of $f(\alpha, x)$ is plus or minus the jth elementary power-sum symmetric function of the β's, we have

$$\sum_i f_{i,j}\alpha^i = (-1)^j \sum_{k_1 < k_2 < \cdots < k_j} \beta_{k_1}\beta_{k_2} \cdots \beta_{k_j}.$$

Taking pth powers gives

$$\sum_i f_{i,j}\alpha^{pi} = (-1)^j \sum_{k_1 < k_2 < \cdots < k_j} \beta_{k_1}^p \beta_{k_2}^p \cdots \beta_{k_j}^p$$

or

$$f(\alpha^p, x) = \prod_{k=1}^{n} (x - \beta_k^p)$$

and therefore

$$F(x) = \prod_{k=0}^{m-1} \prod_{i=1}^{n} (x - \beta_i^{p^k}) = \prod_{i=1}^{n} \left(\prod_{k=0}^{m-1} (x - \beta_i^{p^k}) \right) = \prod_{i} F^{(i)}(x),$$

where each $F^{(i)}$ is a power of a distinct irreducible polynomial over GF(p), and the degree of each $F^{(i)}$ is a multiple of m. We may find these $F^{(i)}$ by factoring $F(x)$ over GF(p) according to the methods indicated in Sections 3 and 4.

The factorization of $f(x)$ is then obtained as

$$f(x) = \prod_{i} \gcd [f(x), F^{(i)}(x)].$$

This factorization is nontrivial unless $F(x)$ is itself an irreducible power. In this case the roots of $f(x)$ are all conjugate. To find them we compute

$$\gcd (f(\alpha, x), f(\alpha^p, x)), \qquad \gcd (f(\alpha, x), f(\alpha^{p^2}, x)),$$

$$\gcd (f(\alpha, x), f(\alpha^{p^3}, x)), \qquad \gcd (f(\alpha, x), f(\alpha^{p^{((m/n)-1)}}, x)).$$

If any of these gcd's is nontrivial, then it gives a nontrivial factor of $f(\alpha, x)$. We claim that all of these gcd's are trivial iff $f(\alpha^{p^{m/n}}, x) = f(\alpha, x)$. To prove this, we observe that $F(x)$ is an irreducible power iff all n roots of $f(x)$ are conjugates. In this case, $\beta_i = \beta_1^{p^{l_i}}$ for each $i = 1, 2, \cdots, n$. Without loss of generality, we may assume that $0 = l_1 < l_2 < l_3 < \cdots < l_n < m$. Define $l_{n+1} = m$, and define $\Delta = \min_{i=1}^{n}(l_{i+1} - l_i)$. Clearly $\Delta \leqq m/n$, with equality iff $f(\alpha^{p^{m/n}}, x) = f(\alpha, x)$. If $\Delta < m/n$, then the sets $\{\beta_1, \beta_2, \cdots, \beta_n\}$ and $\{\beta_1^{p^\Delta}, \beta_2^{p^\Delta}, \cdots, \beta_n^{p^\Delta}\}$ have a nontrivial intersection, and gcd $(f(\alpha, x), f(\alpha^{p^\Delta}, x))$ is nontrivial.

Thus, we need further consider only the case in which $f(\alpha^{p^{m/s}}, x) = f(\alpha, x)$, which happens iff $f(\alpha, x)$ is an irreducible polynomial over GF($p^{m/n}$). In this case (assuming $n > 1$), we transform the polynomial $f(x)$ to $\hat{f}(\alpha, x) = f(\alpha, \alpha x)\alpha^{-n}$. Since the coefficients of $f(x)$ are in GF($p^{m/n}$) but $\alpha \notin$ GF($p^{m/n}$), $\hat{f}(\alpha, x)$ is monic, of degree n, and has at least one coefficient not in GF($p^{m/n}$), so $f(\alpha, x)$ can be factored by the methods of this section. The factorization of $f(x)$ may then be easily recovered from the factorization of $\hat{f}(\alpha, x)$.

7. Finding Roots in GF(p), p a Large Prime. In the previous sections of this paper, we have reduced the factorization of an arbitrary polynomial, $f(x)$ of degree n over GF(p^m), to the special case in which

$$f(x) = \prod_{i=1}^{n} (x - \rho_i)$$

where the ρ_i are distinct elements in GF(p). To solve this problem, we observe that if γ is any element in GF(p), p odd, then

$$f(x - \gamma) = \prod_{i=1}^{n} (x - (\gamma + \rho_i))$$

and $f(x-\gamma) \mid (x^p - x) = x(x^{(p-1)/2}+1)(x^{(p-1)/2}-1)$ and therefore, if $x \nmid f(x-\gamma)$, then

$$f(x-\gamma) = \gcd(f(x-\gamma), x^{(p-1)/2}+1)\gcd(f(x-\gamma), x^{(p-1)/2}-1). \tag{7.01}$$

This equation provides a feasible method of factoring $f(x-\gamma)$, for we may compute the residues (modulo $f(x-\gamma)$) of $x, x^2, x^{2^2}, x^{2^3}, \cdots, x^{2^{[\log p]}}, \cdots, x^{(p-1)/2}$. If $x^{p(-1)/2} \not\equiv \pm 1$, then Eq. (7.01) yields a nontrivial factorization. However, if $x^{(p-1)/2} \equiv \pm 1$, then the factorization of Eq. (7.01) is trivial and we must try again with a new value of γ.

In general, the factorization of Eq. (7.01) will fail iff $\rho_1+\gamma, \rho_2+\gamma, \cdots, \rho_n+\gamma$ are all quadratic residues or all quadratic nonresidues. If $\rho_1, \rho_2, \cdots, \rho_n$ are all residues (or nonresidues), then the theory of cyclotomy[7] leads us to expect that a randomly chosen γ in GF(p) will yield a nontrivial factorization with probability about $(1-2^{-n})$. Thus, any given product of linear factors, $f(x)$, may be factored via Eq. (7.01) and a few randomly chosen values of γ with very high probability. However, there is an unfortunate improbable possibility that each successive choice of γ proves unlucky.

If $n = 2$, then the success or failure of a particular choice of γ depends only on the quadratic character of g_0, the constant term in the polynomial

$$g(x) = \sum_{i=0}^{n} g_i x^i = f(x-\gamma).$$

This is because if n is even, then

$$g_0 = \prod_{i=1}^{n} (\rho_i + \gamma)$$

and hence g_0 is a residue iff an even number of the $\rho_i+\gamma$ are residues. In particular, if $n = 2$, g_0 is a residue iff both $\rho_1+\gamma$ and $\rho_2+\gamma$ have the same quadratic character. Thus, the success of factorization via Eq. (7.01) may be anticipated by evaluating the Legendre symbol, (g_0/p) with aid of Gauss' law of quadratic reciprocity. If $(g_0/p) = -1$, then Eq. (7.01) must yield a nontrivial factorization and the calculation may be continued. However, if $(g_0/p) = 1$, then Eq. (7.01) will yield only a trivial factorization, so the calculation should be aborted and resumed with another candidate value of γ.

In the special case of a quadratic equation over GF(p), $p \equiv -1 \bmod 4$, then the choice of γ which eliminates the linear term in $f(x-\gamma)$ also guarantees a nontrivial factorization via Eq. (7.01). For, in this case, $f(x-\gamma) = x^2 - c$. If $f(x)$ has two roots in GF(p), then c must be a quadratic residue, and since -1 is a quadratic nonresidue, so is the constant term in the polynomial $f(x-\gamma)$.

Certain special classes of quadratic equations modulo primes $\equiv 1 \bmod 4$ may be solved by other methods, such as those given by Schönheim (1956).

In the special cases of cubics and quartics over a field whose order, p, is congruent to $-1 \bmod 12$, we may use the classical formulas of Scipio del Ferro and Ferrari as given on pp. 105–108 of Birkhoff and Mac Lane (1965). Since $p \equiv -1 \bmod 4$, we may extract square roots and fourth roots by the procedure explained above. Since $p \not\equiv 1 \bmod 3$, we may extract cube roots by taking the $(p-1)/3$ power. Thus, the

[7] For details, see pp. 147–166 of Hall (1967).

classical formulas for solving quartic equations in terms of radicals may be applied to obtain the factorization in a small number of steps without any reliance on luck.

In general, however, there is no good deterministic procedure for finding the roots in the large prime field GF(p) of the polynomial $f(x)$, which is known to split into distinct linear factors in GF(p). The best practical procedure is to attempt to factor $f(x - \gamma)$ for several choices of γ, and to hope that you are not too unlucky.

It appears rather difficult to determine the "best" sequence of successive choices of γ, although the sequence $\gamma = 0, 1, 2, 3, \cdots$ seems as plausible as any. However, it is not known how many successive trials are required from this sequence (or any other good sequence) to guarantee a factorization of the "worst" $f(x)$. Burgess (1962) has shown that the maximum number of consecutive quadratic residues or nonresidues modulo a large prime p is no greater than $O(p^{1/4}(\log p)^{3/2})$, but there are probably an infinite number of primes all of whose sequences of consecutive residues or nonresidues have lengths much, much smaller than $p^{1/4}$.

8. From Factorization Over the Integers Mod M to Factorization Over the Rationals. Let $f(x) = \prod_i f^{(i)}(x)$, where $f(x)$ is a given polynomial with integral coefficients and the $f^{(i)}(x)$ are the distinct irreducible factors of $f(x)$. Our problem is to determine the $f^{(i)}(x)$ from the given coefficients of $f(x)$. We may begin by calculating some large integer, C, such that

$$\max_{i,j} |f_j^{(i)}| \leqq C.$$

One method of calculating such an upper bound to the magnitudes of all of the coefficients of the factors of $f(x)$, due to Collins (1967) and Knuth (1969) is based on the inverse of the Vandermonde matrix which arises in the classical Kronecker factorization algorithm. Several expressions for the coefficients of that inverse matrix are given by Gautschi (1962). A more recent method of calculating an upper bound, C, has been presented by Zassenhaus (1969). Either method requires only a modest amount of computation with the coefficients of the original polynomial, $f(x)$.

Once we have found the upper bound, C, we proceed to factor the polynomial $f(x)$ modulo some large integer, $M > 2C$.

If M is a prime p, we may obtain the factorization of $f(x)$ mod M from the factorization algorithms of Sections 3, 4, and 7.

If M is a prime-power, then we first factor $f(x)$ mod p, the prime divisor of M. Following a suggestion of Zassenhaus (1969), we may then extend the complete factorization of $f(x)$ mod p^i to the factorization mod p^{2i} by the classical p-adic lemma of Hensel. Continuing this extension for $i = 1, 2, \cdots, 2^{\log_2 \log_p M}$ we will eventually obtain the complete factorization of $f(x)$ mod M. The number of irreducible factors of $f(x)$ mod M is the same as the number of irreducible factors mod p.

If M is the product of several distinct primes, then we first factor $f(x)$ mod each of these primes and then attempt to reconstruct the factorization of $f(x)$ mod M with the aid of the Chinese remainder theorem. This latter step may require many attempts, because the degrees of the factors of $f(x)$ mod different primes may be compatible. For example, if $M = p_1 p_2$, p_1 and p_1 primes, and $f(x)$ has degree 8, then $f(x)$ may factor into four quadratics mod p_1 and into 2 quartics mod p_2. There are then $\binom{4}{2}$ possible factorizations of $f(x)$ mod M. Consequently, if the degrees of

the factors of $f(x)$ mod several different primes are compatible, it appears unwise to choose an M which is divisible by more than one of these primes. The simpler course is to take M as a power of a single prime.

The polynomial $f^{(i)}(x)$, a factor of $f(x)$ which is irreducible over the rationals, may factor mod M as

$$f^{(i)}(x) \equiv \prod_j f^{(i,j)}(x) \bmod M.$$

Mod M, the original polynomial $f(x)$ then factors as

$$f(x) \equiv \prod_i \prod_j f^{(i,j)}(x) \bmod M.$$

If we know $f^{(i)}(x)$ mod M, then we can easily determine the coefficients of $f^{(i)}(x)$ over the rational integers because we have chosen M so large that we are guaranteed that for all i and j,

$$-M/2 < f_j^{(i)} < M/2.$$

Unfortunately, however, we may have considerable difficulty in determining, mod M, the $f^{(i)}(x)$ from the irreducible $f^{(i,j)}(x)$, for we do not know which irreducible factors mod M to multiply together to obtain a factor over the rationals. For example, the complete factorization of $f(x)$ modulo M might be

$$f(x) \equiv a(x)b(x)c(x)d(x)e(x) \bmod M.$$

Even if we knew that $f(x)$ was the product of exactly two irreducible factors over the rationals, we would still have no easy way of determining $f^{(1)}(x)$ and $f^{(2)}(x)$ modulo M. For example, we might find that $f^{(1)}(x) \equiv b(x)e(x)$ and $f^{(2)}(x) \equiv a(x)c(x)d(x)$. In general, if $f(x)$ is the product of r distinct irreducible factors modulo M, then there are 2^r subsets of these factors whose product might be congruent mod M to the irreducible rational factor $f^{(1)}(x)$. These 2^r subsets occur in 2^{r-1} complementary pairs, but this still leaves 2^{r-1} essentially different candidates for the factor $f^{(1)}(x)$.

Fortunately, for the "typical" polynomial, r will be small. This is because the expected number of irreducible factors of a randomly chosen polynomial of large degree, d, over GF(p) is about $\ln d$, for all large p. For further details, see Problem 3.6, page 86 of Berlekamp (1968).

Unfortunately, however, the "worst" polynomial is much worse than the "typical" polynomial. The Dirichlet density theorem[8] implies that no polynomial which is irreducible over the rationals can have a linear factor modulo every prime, but there do exist polynomials which are irreducible over the rationals and factor into only linear and quadratic factors modulo every prime p. One such class of polynomials, suggested by Swinnerton-Dyer (1969), is constructed as follows: Let the degree, d, be a power of 2, say $d = 2^i$, and let p_i be the ith prime. Consider the monic polynomial, $f(x)$, whose complex roots are given by $\pm\sqrt{-1} \pm \sqrt{2} \pm \sqrt{3} \pm \sqrt{5} \pm \cdots \pm (p_{i-1})^{1/2}$, where the 2^i combinations of signs give the 2^i distinct complex roots. Since $\sqrt{-1}, \sqrt{2}, \sqrt{3}, \cdots, (p_{i-1})^{1/2}$ are rationally independent, it follows that $f(x)$ is

[8] For reference, see pp. 227–230 and Exercise 6 on p. 361 of *Algebraic Number Theory*, edited by J. W. S. Cassels and A. Fröhlich, Academic Press, London; Thompson Book Co., Inc., Washington, D.C., 1967, Math. Rev. 35 #6500.

irreducible over the rational integers. On the other hand, if p is any prime, then $\sqrt{-1}, \sqrt{2}, \sqrt{3}, \cdots, (p_{i-1})^{1/2}$ all lie in GF(p^2), from which it follows that $f(x)$ splits into linear factors in GF(p^2) and into quadratic and linear factors in GF(p). Thus, this polynomial has degree d and is irreducible over the rationals, but it has $r \geqq d/2$ distinct irreducible factors modulo every integer M. Although the amount of work required to obtain the irreducible factors modulo M of a polynomial is only algebraic in its degree, d, the r irreducible factors modulo M lead to 2^{r-1} candidate factors over the rationals. When r is large compared to log d, then testing each of these 2^{r-1} candidates becomes the bottleneck step of the factorization algorithm. We now consider programming tricks by which these tests may be speeded up.

If $f^{(1)}(x) \mid f(x)$, then $f^{(1)}(a) \mid f(a)$ for every integer a. If we can find an integer a for which the integer $f(a)$ has only a small number of prime factors, then we may eliminate the candidate factor $f^{(1)}(x)$ by computing $f^{(1)}(a)$ and obtaining an answer not on the small list of factors of $f(a)$. By performing an appropriate translation of $f(x)$ and each of its irreducible factors modulo M, we may assume that $a = 0$. Testing that $f^{(1)}(a)$ divides $f(a)$ is then equivalent to testing that the constant term of $f^{(1)}(x)$ divides the constant term of $f(x)$.

If we have a sufficiently large memory capacity, then we can use the following programming technique suggested by L. Welch (1969) to determine the subset of the 2^{r-1} candidate factors $f^{(1)}(x)$ which survive the test $f^{(1)}(0) \mid f(0)$ in only about $k2^{r/2} \log r$ operations, where k is the total number of integral divisors of $f(0)$, including the trivial divisors ± 1 and $\pm f(0)$. The procedure is as follows. We first discard one of the r irreducible factors of $f(x)$ modulo M[9] and partition the remaining $(r-1)$ irreducible factors into two sets, each consisting of about $r/2$ irreducible factors. Each candidate factor $f^{(1)}(x)$ is of the form $f^{(1)}(x) = g(x)h(x)$, where $g(x)$ is the product of some subset of irreducible factors in the first set and $h(x)$ is the product of some subset of irreducible factors in the second set.

For each of the $2^{r/2}$ candidate g's, we compute the value of $g(0)$ and its inverse modulo M, $(g(0))^{-1}$. This list of $2^{r/2}$ entries, represented in any convenient fashion, is then sorted. The sorting requires about $r2^{r/2}$ operations. Then, for each candidate $h(x)$, we compute and store the value of $h(0)$, and sort this list of $2^{r/2}$ candidates for $h(0)$. We then make one joint scan through the two sorted lists to see if there is any pair of polynomials $g(x)$ and $h(x)$ such that $(g(0))^{-1} \equiv h(0) \bmod M$. In this manner we find all candidate factors $f^{(1)}(x)$ for which $f^{(1)}(0) = 1$. After we have found and tested all such candidates, we then replace the list of values of $h(0)$ by the list of values of $dh(0)$, where d is a rational integral divisor of $f(0)$. The list of candidates of $dh(0)$ is then resorted, and another joint scan reveals all candidate factors $f^{(1)}(x)$ for which $f^{(1)}(0) = d$. Repeating this procedure for each (positive and negative) integer d which divides $f(0)$, we eventually obtain all candidate factors $f^{(1)}(x)$ which survive the test $f^{(1)}(0) \mid f(0)$.

In some cases, there may be a large number of candidate factors $f^{(1)}(x)$ which survive the test $f^{(1)}(0) \mid f(0)$ and still fail to pass the more general test $f^{(1)}(x) \mid f(x)$. For example, it may happen that many irreducible factors of $f(x)$ mod M have con-

[9] To reduce duplication of subsequent work, it is wise to discard an irreducible factor whose constant term is ± 1 or a small divisor of $f(0)$. If there are no such irreducible factors, any choice of discard is good. If there are several such irreducible factors, it may be wise to consider another choice of a.

stant terms congruent to ± 1. In such cases it may be advantageous to test that $f^{(1)}(a_i) \mid f(a_i)$ for more than one value of a_i. There remains the problem of choosing these a_i in a clever manner. If a sufficiently large number of primes or ± 1's or a sufficiently large number of sufficiently small integers occur among the values of $f(a_i)$, then it may be possible to deduce the irreducibility of $f(x)$ immediately, using the criteria of Brown and Graham (1969) or of Brauer and Ehrlich (1946).

Bell Telephone Laboratories
Murray Hill, New Jersey 07974

E. R. BERLEKAMP (1967), "Factoring polynomials over finite fields," *Bell System Tech. J.*, v. 46, 1967, pp. 1853–1859. MR **36** #2314. Revised and expanded version of this paper was republished in Ch. 6 of Berlekamp (1968).

E. R. BERLEKAMP (1968), *Algebraic Coding Theory*, McGraw-Hill, New York, 1968. MR **38** #6873.

E. R. BERLEKAMP, H. RUMSEY & G. SOLOMON (1967), "On the solution of algebraic equations over finite fields," *Information and Control*, v. 10, 1967, pp. 553–564. MR **37** #6266.

G. BIRKHOFF & S. MAC LANE (1965), *A Survey of Modern Algebra*, 3rd ed., Macmillan, New York, 1965. MR **31** #2250.

A. BRAUER & G. EHRLICH (1946), "On the irreducibility of certain polynomials," *Bull. Amer. Math. Soc.*, v. 52, 1946, pp. 844–856. MR **8**, 195.

W. S. BROWN & R. L. GRAHAM (1969), "An irreducibility criterion for polynomials over the integers," *Amer. Math. Monthly*, v. 76, 1969, pp. 795–797.

D. A. BURGESS (1962), "On character sums and primitive roots," *Proc. London Math. Soc.* (3), v. 12, 1962, pp. 179–192. MR **24** #A2569.

G. E. COLLINS (1967), Unpublished communication.

G. E. COLLINS (1969), "Computing multiplicative inverses in GF(p)," *Math. Comp.*, v. 23, 1969, pp. 197–200.

W. GAUTSCHI (1962), "On inverses of Vandermonde and confluent Vandermonde matrices," *Numer. Math.*, v. 4, 1962, pp. 117–128; ibid., v. 5, 1963, pp. 425–430. MR **25** #3059; MR **29** #1734.

M. HALL (1967), *Combinatorial Theory*, Blaisdell, Waltham, Mass., 1967. MR **37** #80.

D. E. KNUTH (1967), Unpublished communication.

D. E. KNUTH (1969), *The Art of Computer Programming*. Vol. II: *Seminumerical Algorithms*, Addison-Wesley, Reading, Mass., 1969.

D. B. LLOYD (1964), "Factorization of the general polynomial by means of its homomorphic congruential functions," *Amer. Math. Monthly*, v. 71, 1964, pp. 863–870.

Š. SCHWARZ (1956), "On the reducibility of polynomials over a finite field," *Quart. J. Math. Oxford Ser.* (2), v. 7, 1956, pp. 110–124. MR **20** #3162.

I. SCHÖNHEIM (1956), "Formules pour résoudre la congruence $x^2 \equiv a \pmod{P}$ dans des cas encore inconnus et leur application pour déterminer directement des racines primitives de certains nombres premiers," *Acad. R. P. Romîne. Fil. Cluj. Stud. Cerc. Mat. Fiz.*, v. 7, 1956, no. 1–4, pp. 51–58. MR **20** #2299.

H. P. F. SWINNERTON-DYER (1969), Unpublished communication.

B. L. VAN DER WAERDEN (1931), *Moderne Algebra*, Springer, Berlin, 1931; English transl., Ungar, New York, 1949. MR **10**, 587.

L. WELCH (1969), Unpublished communication.

H. ZASSENHAUS (1969), "On Hensel factorization. I," *J. Number Theory*, v. 1, 1969, pp. 291–311.