J. H. van Lint

Introduction to Coding Theory

J. H. van Lint

Introduction to Coding Theory

With 8 Illustrations

Springer-Verlag
New York Heidelberg Berlin

J. H. van Lint
Eindhoven University of Technology
Department of Mathematics
Den Dolech 2, P.O. Box 513
5600 MB Eindhoven
The Netherlands

AMS Classification (1980): 05-01, 68E99, 94A24

Library pf Congress Cataloging in Publication Data
Lint, Jacobus Hendricus van, 1932–
Introduction to coding theory.

(Graduate texts for mathematics; 86)
Bibliography: p.
Includes index.
1. Coding theory. I. Title. II. Series.
QA268.L57 519.4 82-3242
AACR2

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-11284-7 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-11284-7 Springer-Verlag Berlin Heidelberg New York

Preface

Coding theory is still a young subject. One can safely say that it was born in 1948. It is not surprising that it has not yet become a fixed topic in the curriculum of most universities. On the other hand, it is obvious that discrete mathematics is rapidly growing in importance. The growing need for mathematicians and computer scientists in industry will lead to an increase in courses offered in the area of discrete mathematics. One of the most suitable and fascinating is, indeed, coding theory. So, it is not surprising that one more book on this subject now appears. However, a little more justification and a little more history of the book are necessary. A few years ago it was remarked at a meeting on coding theory that there was no book available which could be used for an introductory course on coding theory (mainly for mathematicians but also for students in engineering or computer science). The best known textbooks were either too old, too big, too technical, too much for specialists, etc. The final remark was that my Springer Lecture Notes (#201) were slightly obsolete and out of print. Without realizing what I was getting into I announced that the statement was not true and proved this by showing several participants the book *Inleiding in de Coderingstheorie*, a little book based on the syllabus of a course given at the Mathematical Centre in Amsterdam in 1975 (M.C. Syllabus 31). The course, which was a great success, was given by M. R. Best, A. E. Brouwer, P. van Emde Boas, T. M. V. Janssen, H. W. Lenstra Jr., A. Schrijver, H. C. A. van Tilborg and myself. Since then the book has been used for a number of years at the Technological Universities of Delft and Eindhoven.

The comments above explain why it seemed reasonable (to me) to translate the Dutch book into English. In the name of Springer-Verlag I thank the Mathematical Centre in Amsterdam for permission to do so. Of course it turned out to be more than a translation. Much was rewritten or expanded, problems were changed and solutions were added, and a new chapter and several new proofs were included. Nevertheless the M.C. Syllabus (and the Springer Lecture Notes 201) are the basis of this book.

The book consists of three parts. Chapter 1 contains the prerequisite mathematical knowledge. It is written in the style of a memory-refresher. The reader who discovers topics which he does not know will get some idea about them but it is recommended that he also looks at standard textbooks on those topics. Chapters 2 to 6 provide an introductory course in coding theory. Finally, Chapters 7 to 11 are introductions to special topics and can be used as supplementary reading or as a preparation for studying the literature.

Despite the youth of the subject, which is demonstrated by the fact that the papers mentioned in the references have 1974 as the average publication year, I have not considered it necessary to give credit to every author of the theorems, lemmas, etc. Some have simply become standard knowledge.

It seems appropriate to mention a number of textbooks which I use regularly and which I would like to recommend to the student who would like to learn more than this introduction can offer. First of all F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes* (reference [46]), which contains a much more extensive treatment of most of what is in this book and has 1500 references! For the more technically oriented student with an interest in decoding, complexity questions, etc. E. R. Berlekamp's *Algebraic Coding Theory* (reference [2]) is a must. For a very well-written mixture of information theory and coding theory I recommend: R. J. McEliece, *The Theory of Information and Coding* (reference [51]). In the present book very little attention is paid to the relation between coding theory and combinatorial mathematics. For this the reader should consult P. J. Cameron and J. H. van Lint, *Graphs, Codes and Designs* (reference [11]).

I sincerely hope that the time spent writing this book (instead of doing research) will be considered well invested.

Eindhoven
July 1981

J. H. VAN LINT

Contents

CHAPTER 4

Some Good Codes

CHAPTER 5

Bounds on Codes

CHAPTER 6

Cyclic Codes

CHAPTER 7

Perfect Codes and Uniformly Packed Codes

Chapter 1

Mathematical Background

In order to be able to read this book a fairly thorough mathematical background is necessary. In different chapters many different areas of mathematics play a role. The most important one is certainly algebra but the reader must also know some facts from elementary number theory, probability theory and a number of concepts from combinatorial theory such as designs and geometries. In the following sections we shall give a brief survey of the prerequisite knowledge. Usually proofs will be omitted. For these we refer to standard textbooks. In some of the chapters we need a large number of facts concerning a not too well-known class of orthogonal polynomials, called Krawtchouk polynomials. These properties are treated in Section 1.2. The notations which we use are fairly standard. We mention a few which may not be generally known. If C is a finite set we denote the number of elements of C by $|C|$. If the expression B is the definition of concept A then we write $A := B$. We use "iff" for "if and only if". An identity matrix is denoted by I and the matrix with all entries equal to one is J. Similarly we abbreviate the vector with all coordinates 0 (resp. 1) by $\mathbf{0}$ (resp. $\mathbf{1}$). Instead of using $[x]$ we write $\lfloor x \rfloor := \max\{n \in \mathbb{Z} \mid n \leq x\}$ and use the symbol $\lceil x \rceil$ for rounding upwards.

§1.1 Algebra

We need only very little from elementary number theory. We assume known that in $\mathbb{N}$ every number can be written in exactly one way as a product of prime numbers (if we ignore the order of the factors). If a divides b then we write $a \mid b$. If p is a prime number and $p^r \mid a$ but $p^{r+1} \nmid a$ then we write $p^r \| a$. If

$k \in \mathbb{N}$, $k > 1$ then a representation of n in the base k is a representation

$$n = \sum_{i=0}^{l} n_i k^i,$$

$0 \le n_i < k$ for $0 \le i \le l$. The largest integer n such that $n|a$ and $n|b$ is called the greatest common divisor of a and b and denoted by g.c.d.(a, b) or simply (a, b). If $m|(a - b)$ we write $a \equiv b \pmod{m}$.

(1.1.1) Theorem. *If*

$$\varphi(n) := |\{m \in \mathbb{N} \mid 1 \le m \le n, (m, n) = 1\}|,$$

then

(i) $\varphi(n) = n \prod_{p|n} (1 - 1/p)$,
(ii) $\sum_{d|n} \varphi(d) = n$.

The function φ is called the *Euler indicator.*

(1.1.2) Theorem. *If* $(a, m) = 1$ *then* $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Theorem 1.1.2 is called the Euler–Fermat theorem.

(1.1.3) Definition. The *Moebius function* μ is defined by

$$\mu(n) := \begin{cases} 1, & \text{if } n = 1, \\ (-1)^k, & \text{if } n \text{ is the product of } k \text{ distinct prime factors,} \\ 0, & \text{otherwise.} \end{cases}$$

(1.1.4) Theorem. *If and g are functions defined on* $\mathbb{N}$ *such that*

$$g(n) = \sum_{d|n} f(d),$$

then

$$f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right).$$

Theorem 1.1.4 is known as the *Moebius inversion formula.*

Algebraic Structures

We assume that the reader is familiar with the basic ideas and theorems of linear algebra although we do refresh his memory below. We shall first give a sequence of definitions of algebraic structures with which the reader must be familiar in order to appreciate algebraic coding theory.

(1.1.5) Definition. A *group* $(G,\ )$ is a set G on which a product operation has been defined satisfying

(i) $\forall_{a\in G}\forall_{b\in G}[ab \in G]$,
(ii) $\forall_{a\in G}\forall_{b\in G}\forall_{c\in G}[(ab)c = a(bc)]$,
(iii) $\exists_{e\in G}\forall_{a\in G}[ae = ea = a]$,
(the element e is unique),
(iv) $\forall_{a\in G}\exists_{b\in G}[ab = ba = e]$,
(b is called the inverse of a and also denoted by a^{-1}).

If furthermore

(v) $\forall_{a\in G}\forall_{b\in G}[ab = ba]$,

then the group is called *abelian* or *commutative.*

If $(G,\ )$ is a group and $H \subset G$ such that $(H,\ )$ is also a group then $(H,\ )$ is called a subgroup of $(G,\ )$. Usually we write G instead of $(G,\ )$. The number of elements of a finite group is called the *order* of the group. If $(G,\ )$ is a group and $a \in G$ then the smallest positive integer n such that $a^n = e$ (if such an n exists) is called the *order* of a. In this case the elements $e, a, a^2, \ldots, a^{n-1}$ form a so-called *cyclic* subgroup with a as *generator*. If $(G,\ )$ is abelian and $(H,\ )$ is a subgroup then the sets $aH := \{ah \mid h \in H\}$ are called *cosets* of H. Since two cosets are obviously disjoint or identical the cosets form a partition of G. An element chosen from a coset is called a *representative* of the coset. It is not difficult to show that the cosets again form a group if we define multiplication of cosets by $(aH)(bH) := abH$. This group is called the *factor group* and indicated by G/H. As a consequence note that if $a \in G$ then the order of a divides the order of G (also if G is not abelian).

(1.1.6) Definition. A set R with two operations, usually called addition and multiplication, denoted by $(R, +,\ )$, is called a *ring* if

(i) $(R, +)$ is an abelian group,
(ii) $\forall_{a\in R}\forall_{b\in R}\forall_{c\in R}[(ab)c = a(bc)]$,
(iii) $\forall_{a\in R}\forall_{b\in R}\forall_{c\in R}[a(b + c) = ab + ac \wedge (a + b)c = ac + bc]$.

The identity element of $(R, +)$ is usually denoted by 0.

If the additional property

(iv) $\forall_{a\in R}\forall_{b\in R}[ab = ba]$

holds, then the ring is called *commutative.*

The integers $\mathbb{Z}$ are the best known example of a ring.

(1.1.7) Definition. If $(R, +,\ )$ is a ring and $0 \neq S \subseteq R$, then S is called an *ideal* if

(i) $\forall_{a\in S}\forall_{b\in S}[a - b \in S]$,
(ii) $\forall_{a\in S}\forall_{b\in R}[ab \in S \wedge ba \in S]$.

It is clear that if S is an ideal in R then $(S, +, \ )$ is a subring, but requirement (ii) says more than that.

(1.1.8) Definition. A *field* is a ring $(R, +, \ )$ for which $(R\backslash\{0\}, \ )$ is an abelian group.

(1.1.9) Theorem. *Every finite ring R with at least two elements such that*

$$\forall_{a\in R}\forall_{b\in R}[ab = 0 \Rightarrow (a = 0 \vee b = 0)]$$

is a field.

(1.1.10) Definition. Let $(V, +)$ be an abelian group, $\mathbb{F}$ a field and let a multiplication $\mathbb{F} \times V \to V$ be defined satisfying

(i) $\forall_{\mathbf{a}\in V}[1\mathbf{a} = \mathbf{a}]$,
$\forall_{\alpha\in \mathbb{F}}\forall_{\beta\in \mathbb{F}}\forall_{\mathbf{a}\in V}[\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}]$,
(ii) $\forall_{\alpha\in \mathbb{F}}\forall_{\mathbf{a}\in V}\forall_{\mathbf{b}\in V}[\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}]$,
$\forall_{\alpha\in \mathbb{F}}\forall_{\beta\in \mathbb{F}}\forall_{\mathbf{a}\in V}[(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}]$.

Then the triple $(V, +, \mathbb{F})$ is called a *vector space* over the field $\mathbb{F}$. The identity element of $(V, +)$ is denoted by $\mathbf{0}$.

We assume the reader to be familiar with the vector space $\mathbb{R}^n$ consisting of all n-tuples $(a_1, a_2, \ldots, a_n)$ with the obvious rules for addition and multiplication. We remind him of the fact that a *k-dimensional subspace C* of this vector space is a vector space with a *basis* consisting of vectors $\mathbf{a}_1 := (a_{11}, a_{12}, \ldots, a_{1n})$, $\mathbf{a}_2 := (a_{21}, a_{22}, \ldots, a_{2n}), \ldots, \mathbf{a}_k := (a_{k1}, a_{k2}, \ldots, a_{kn})$, where the word basis means that every $\mathbf{a} \in C$ can be written in a unique way as $\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \cdots + \alpha_k\mathbf{a}_k$. The reader should also be familiar with the process of going from one basis of C to another by taking combinations of basis vectors, etc. We shall usually write vectors as *row vectors* as we did above. The *inner product* $\langle\mathbf{a}, \mathbf{b}\rangle$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ is defined by

$$\langle\mathbf{a}, \mathbf{b}\rangle := a_1b_1 + a_2b_2 + \cdots + a_nb_n.$$

The elements of a basis are called *linearly independent*. In other words this means that a linear combination of these vectors is $\mathbf{0}$ iff all the coefficients are 0. If $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are k linearly independent vectors, i.e. a basis of a k-dimensional subspace C, then the system of equations $\langle\mathbf{a}_i, \mathbf{y}\rangle = 0$ $(i = 1, 2, \ldots, k)$ has as its solution all the vectors in a subspace of dimension $n - k$ which we denote by $C^\perp$. So,

$$C^\perp := \{\mathbf{y} \in \mathbb{R}^n \mid \forall_{\mathbf{x}\in C}[\langle\mathbf{x}, \mathbf{y}\rangle = 0]\}.$$

These ideas play a fundamental role later on, where $\mathbb{R}$ is replaced by a finite field $\mathbb{F}$. The theory reviewed above goes through in that case.

(1.1.11) Definition. Let $(V, +)$ be a vector space over $\mathbb{F}$ and let a multiplication $V \times V \to V$ be defined which satisfies

(i) $(V, +, \ )$ is a ring,
(ii) $\forall_{\alpha \in \mathbb{F}} \forall_{\mathbf{a} \in V} \forall_{\mathbf{b} \in V}[(\alpha\mathbf{a})\mathbf{b} = \mathbf{a}(\alpha\mathbf{b})]$.

Then we say that the system is an *algebra* over $\mathbb{F}$.

Suppose we have a finite group $(G, \cdot)$ and we consider the elements of G as basis vectors for a vector space $(V, +)$ over a field $\mathbb{F}$. Then the elements of V are represented by linear combinations $\alpha_1 g_1 + \alpha_2 g_2 + \cdots + \alpha_n g_n$, where

$$\alpha_i \in \mathbb{F}, \qquad g_i \in G, \qquad (1 \le i \le n = |G|).$$

We can define a multiplication $*$ for these vectors in the obvious way, namely

$$\left(\sum_i \alpha_i g_i\right) * \left(\sum_j \beta_j g_j\right) := \sum_i \sum_j (\alpha_i \beta_j)(g_i \cdot g_j),$$

which can be written as $\sum_k \gamma_k g_k$, where γ_k is the sum of the elements $\alpha_i \beta_j$ over all pairs (i, j) such that $g_i \cdot g_j = g_k$. This yields an algebra which is called the *group algebra* of G over $\mathbb{F}$ and denoted by $\mathbb{F}G$.

EXAMPLES. Let us consider a number of examples of the concepts defined above.

If $A := \{a_1, a_2, \ldots, a_n\}$ is a finite set, we can consider all one-to-one mappings of S onto S. These are called *permutations*. If σ_1 and σ_2 are permutations we define $\sigma_1\sigma_2$ by $(\sigma_1\sigma_2)(a) := \sigma_1(\sigma_2(a))$ for all $a \in A$. It is easy to see that the set S_n of all permutations of A with this multiplication is a group, known as the *symmetric group of degree n*. In this book we shall often be interested in special permutation groups. These are subgroups of S_n. We give one example. Let C be a k-dimensional subspace of $\mathbb{R}^n$. Consider all permutations σ of the integers $1, 2, \ldots, n$ such that for every vector $\mathbf{c} = (c_1, c_2, \ldots, c_n) \in C$ the vector $(c_{\sigma(1)}, c_{\sigma(2)}, \ldots, c_{\sigma(n)})$ is also in C. These clearly form a subgroup of S_n. Of course C will often be such that this subgroup of S consists of the identity only but there are more interesting examples! Another example of a permutation group which will turn up later is the *affine permutation group* defined as follows. Let $\mathbb{F}$ be a (finite) field. The mapping $f_{u,v}$, where $u \in \mathbb{F}$, $v \in \mathbb{F}$, $u \ne 0$, is defined on $\mathbb{F}$ by $f_{u,v}(x) := ux + v$ for all $x \in \mathbb{F}$. These mappings are permutations of $\mathbb{F}$ and clearly they form a group under composition of functions.

A *permutation matrix* P is a (0, 1)-matrix which has exactly one 1 in each row and column. We say that P corresponds to the permutation σ of $\{1, 2, \ldots, n\}$ if $p_{ij} = 1$ iff $i = \sigma(j)$ $(i = 1, 2, \ldots, n)$. With this convention the product of permutations corresponds to the product of their matrices. In this way one obtains the so-called matrix representation of a group of permutations.

A group G of permutations acting on a set Ω is called *k-transitive* on Ω if for every ordered k-tuple $(a_1, \ldots, a_k)$ of distinct elements of Ω and for every k-tuple $(b_1, \ldots, b_k)$ of distinct elements of Ω there is an element $\sigma \in G$ such that $b_i = \sigma(a_i)$ for $1 \le i \le k$. If $k = 1$ we call the group transitive.

Let S be an ideal in the ring $(R, +, \ )$. Since $(S, +)$ is a subgroup of the abelian group $(R, +)$ we can form the factor group. The cosets are now called *residue classes mod S*. For these classes we introduce a multiplication in the obvious way: $(a + S)(b + S) := ab + S$. The reader who is not familiar with this concept should check that this definition makes sense (i.e. it does not depend on the choice of representatives a resp. b). In this way we have constructed a ring, called the *residue class ring* R mod S and denoted by R/S. The following example will surely be familiar. Let $R := \mathbb{Z}$ and let p be a prime. Let S be $p\mathbb{Z}$, the set of all multiples of p, which is sometimes also denoted by (p). Then R/S is the ring of integers mod p. The elements of R/S can be represented by $0, 1, \ldots, p - 1$ and then addition and multiplication are the usual operations in $\mathbb{Z}$ followed by a reduction mod p. For example, if we take $p = 7$ then $4 + 5 = 2$ because in $\mathbb{Z}$ we have $4 + 5 \equiv 2 \pmod 7$. In the same way $4.5 = 6$ in $\mathbb{Z}/7\mathbb{Z} = \mathbb{Z}/(7)$. If S is an ideal in $\mathbb{Z}$ and $S \neq \{0\}$ then there is a smallest positive integer k in S. Let $s \in S$. We can write s as $ak + b$, where $0 \le b < k$. By the definition of ideal we have $ak \in S$ and hence $b = s - ak \in S$ and then the definition of k implies that $b = 0$. Therefore $S = (k)$. An ideal consisting of all multiples of a fixed element is called a *principal ideal* and if a ring R has no other ideals than principal ideals it is called a *principal ideal ring*. Therefore $\mathbb{Z}$ is such a ring.

(1.1.12) Theorem. *If p is a prime then $\mathbb{Z}/p\mathbb{Z}$ is a field.*

This is an immediate consequence of Theorem 1.1.9 but also obvious directly. A finite field with n elements is denoted by $\mathbb{F}_n$ or GF(n) (Galois field).

Rings and Finite Fields

More about finite fields will follow below. First some more about rings and ideals. Let $\mathbb{F}$ be a finite field. Consider the set $\mathbb{F}[x]$ consisting of all polynomials $a_0 + a_1x + \cdots + a_nx^n$, where n can be any integer in $\mathbb{N}$ and $a_i \in \mathbb{F}$ for $0 \le i \le n$. With the usual definition of addition and multiplication of polynomials this yields a ring $(\mathbb{F}[x], +, \ )$, which is usually just denoted by $\mathbb{F}[x]$. The set of all polynomials which are multiples of a fixed polynomial $g(x)$, i.e. all polynomials of the form $a(x)g(x)$ where $a(x) \in \mathbb{F}[x]$, is an ideal in $\mathbb{F}[x]$. As before, we denote this ideal by $(g(x))$. The following theorem states that there are no other types.

(1.1.13) Theorem. *$\mathbb{F}[x]$ is a principal ideal ring.*

The residue class ring $\mathbb{F}[x]/(g(x))$ can be represented by the polynomials whose degree is less than the degree of $g(x)$. In the same way as our example $\mathbb{Z}/7\mathbb{Z}$ given above, we now multiply and add these representatives in the usual way and then reduce mod $g(x)$. For example, we take $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$ and $g(x) = x^3 + x + 1$. Then $(x + 1)(x^2 + 1) = x^3 + x^2 + x + 1 = x^2$. This example is a useful one to study carefully if one is not familiar with finite fields. First observe that $g(x)$ is *irreducible*, i.e., there do not exist polynomials $a(x)$ and $b(x) \in \mathbb{F}[x]$, both of degree less than 3, such that $g(x) = a(x)b(x)$. Next, realize that this means that in $\mathbb{F}_2[x]/(g(x))$ the product of two elements $a(x)$ and $b(x)$ is 0 iff $a(x) = 0$ or $b(x) = 0$. By Theorem 1.1.9 this means that $\mathbb{F}_2[x]/(g(x))$ is a field. Since the representatives of this residue class ring all have degrees less than 3, there are exactly eight of them. So we have found a field with eight elements, i.e. $\mathbb{F}_{2^3}$. This is an example of the way in which finite fields are constructed.

(1.1.14) Theorem. *Let p be a prime and let $g(x)$ be an irreducible polynomial of degree r in the ring $\mathbb{F}_p[x]$. Then the residue class ring $\mathbb{F}_p[x]/(g(x))$ is a field with p^r elements.*

PROOF. The proof is the same as the one given for the example $p = 2$, $r = 3$, $g(x) = x^3 + x + 1$. □

(1.1.15) Theorem. *Let $\mathbb{F}$ be a field with n elements. Then n is a power of a prime.*

PROOF. By definition there is an identity element for multiplication in $\mathbb{F}$. We denote this by 1. Of course $1 + 1 \in \mathbb{F}$ and we denote this element by 2. We continue in this way, i.e. $2 + 1 = 3$, etc. After a finite number of steps we encounter a field element which already has a name. Suppose, e.g. that the sum of k terms 1 is equal to the sum of l terms 1 ($k > l$). Then the sum of $(k - l)$ terms 1 is 0, i.e. the first time we encounter an element which already has a name, this element is 0. Say 0 is the sum of k terms 1. If k is composite, $k = ab$, then the product of the elements which we have called a resp. b is 0, a contradiction. So k is a prime and we have shown that $\mathbb{F}_p$ is a subfield of $\mathbb{F}$. We define linear independence of a set of elements of $\mathbb{F}$ with respect to (coefficients from) $\mathbb{F}_p$ in the obvious way. Among all linearly independent subsets of $\mathbb{F}$ let $\{x_1, x_2, \ldots, x_r\}$ be one with the maximal number of elements. If x is any element of $\mathbb{F}$ then the elements $x, x_1, x_2, \ldots, x_r$ are not linearly independent, i.e. there are coefficients $0 \neq \alpha, \alpha_1, \ldots, \alpha_r$ such that $\alpha x + \alpha_1 x_1 + \cdots + \alpha_r x_r = 0$ and hence x is a linear combination of x_1 to x_r. Since there are obviously p^r distinct linear combinations of x_1 to x_r the proof is complete. □

From the previous theorems we now know that a field with n elements exists iff n is a prime power, providing we can show that for every $r \geq 1$ there is an irreducible polynomial of degree r in $\mathbb{F}_p[x]$. We shall prove this by calculating the number of such polynomials. Fix p and let I_r denote the number of irreducible polynomials of degree r which are *monic*, i.e. the coefficient of x^r is 1. We claim that

$$(1.1.16) \qquad (1 - pz)^{-1} = \prod_{r=1}^{\infty} (1 - z^r)^{-I_r}.$$

In order to see this, first observe that the coefficient of z^n on the left-hand side is p^n which is the number of monic polynomials of degree n with coefficients in $\mathbb{F}_p$. We know that each such polynomial can be factored uniquely into irreducible factors and we must therefore convince ourselves that these products are counted on the right-hand side of (1.1.16). To show this we just consider two irreducible polynomials $a_1(x)$ of degree r and $a_2(x)$ of degree s. There is a 1-1 correspondence between products $(a_1(x))^k(a_2(x))^l$ and terms $z_1^{kr}z_2^{ls}$ in the product of $(1 + z_1^r + z_1^{2r} + \cdots)$ and $(1 + z_2^s + z_2^{2s} + \cdots)$. If we identify z_1 and z_2 with z, then the exponent of z is the degree of $(a_1(x))^k(a_2(x))^l$. Instead of two polynomials $a_1(x)$ and $a_2(x)$ we now consider all irreducible polynomials and (1.1.16) follows.

In (1.1.16) we take logarithms on both sides, then differentiate, and finally multiply by z to obtain

$$(1.1.17) \qquad \frac{pz}{1 - pz} = \sum_{r=1}^{\infty} I_r \frac{rz^r}{1 - z^r}.$$

Comparing coefficients of z^n on both sides of (1.1.17) we find

$$(1.1.18) \qquad p^n = \sum_{r|n} rI_r.$$

Now apply Theorem 1.1.4 to (1.1.18). We find

$$(1.1.19) \qquad I_r = \frac{1}{r}\sum_{d|r} \mu(d)p^{r/d} > \frac{1}{r}\{p^r - p^{r/2} - p^{r/3} - \cdots\}$$

$$> \frac{1}{r}\left(p^r - \sum_{i=0}^{r/2} p^i\right) > \frac{1}{r}p^r(1 - p^{-r/2+1}) > 0.$$

Now that we know for which values of n a field with n elements exists we wish to know more about these fields. The structure of $\mathbb{F}_{p^r}$ will play a very important role in many chapters of this book. As a preparation consider a finite field $\mathbb{F}$ and a polynomial $f(x) \in \mathbb{F}[x]$ such that $f(a) = 0$, where $a \in \mathbb{F}$. Then by dividing we find that there is a $g(x) \in \mathbb{F}[x]$ such that $f(x) = (x - a)g(x)$. Continuing in this way we establish the trivial fact that a polynomial $f(x)$ of degree r in $\mathbb{F}[x]$ has at most r zeros in $\mathbb{F}$.

If α is an element of order e in the multiplicative group $(\mathbb{F}_{p^r}\backslash\{0\},\ )$ then α is a zero of the polynomial $x^e - 1$. In fact, we have

$$x^e - 1 = (x - 1)(x - \alpha)(x - \alpha^2)\cdots(x - \alpha^{e-1}).$$

It follows that the only elements of order e in the group are the powers α^i where $1 \leq i < e$ and $(i, e) = 1$. There are $\varphi(e)$ such elements. Hence, for every e which divides $p^r - 1$ there are either 0 or $\varphi(e)$ elements of order e in the field. By (1.1.1) the possibility 0 never occurs. As a consequence there are elements of order $p^r - 1$, in fact exactly $\varphi(p^r - 1)$ such elements. We have proved the following theorem.

(1.1.20) Theorem. *In $\mathbb{F}_q$ the multiplicative group $(\mathbb{F}_q\backslash\{0\},\ )$ is a cyclic group.*

This group is often denoted by $\mathbb{F}_q^*$.

(1.1.21) Definition. A generator of the multiplicative group of $\mathbb{F}_q$ is called a *primitive element* of the field.

Note that Theorem 1.1.20 states that the elements of $\mathbb{F}_q$ are exactly the q distinct zeros of the polynomial $x^q - x$. An element β such that $\beta^k = 1$ but $\beta^l \neq 1$ for $0 < l < k$ is called a *primitive* kth root of unity. Clearly a primitive element α of $\mathbb{F}_q$ is a primitive $(q - 1)$th root of unity. If e divides $q - 1$ then α^e is a primitive $((q - 1)/e)$th root of unity. Furthermore a consequence of Theorem 1.1.20 is that $\mathbb{F}_{p^r}$ is a subfield of $\mathbb{F}_{p^s}$ iff r divides s. Actually this statement could be slightly confusing to the reader. We have been suggesting by our notation that for a given q the field $\mathbb{F}_q$ is unique. This is indeed true. In fact this follows from (1.1.18). We have shown that for $q = p^n$ every element of $\mathbb{F}_q$ is a zero of some irreducible factor of $x^q - x$ and from the remark above and Theorem 1.1.14 we see that this factor must have a degree r such that $r|n$. By (1.1.18) this means we have used all irreducible polynomials of degree r where $r|n$. In other words, the product of these polynomials is $x^q - x$. This establishes the fact that two fields $\mathbb{F}$ and $\mathbb{F}'$ of order q are isomorphic, i.e. there is a mapping $\varphi: \mathbb{F} \to \mathbb{F}'$ which is one-to-one and such that φ preserves addition and multiplication.

The following theorem is used very often in this book.

(1.1.22) Theorem. *Let $q = p^r$ and $0 \neq f(x) \in \mathbb{F}_q[x]$.*

(i) *If $\alpha \in \mathbb{F}_{q^k}$ and $f(\alpha) = 0$, then $f(\alpha^q) = 0$.*
(ii) *Conversely, if $f(\alpha^q) = 0$ for every α for which $f(\alpha) = 0$ then $f(x) \in \mathbb{F}_q[x]$.*

PROOF.

(i) By the binomial theorem we have $(a + b)^p = a^p + b^p$ because p divides $\binom{p}{k}$ for $1 \leq k \leq p - 1$. It follows that $(a + b)^q = a^q + b^q$. If $f(x) = \sum a_i x^i$ then $(f(x))^q = \sum a_i^q (x^q)^i$.