Recap: Finite Fields

$\mathbb{Z}_p$ (p prime) with + and * mod p, is a **finite** field.

1. $(\mathbb{Z}_p, +)$ is an **abelian group** (0 is identity)
2. $(\mathbb{Z}_p \setminus 0, *)$ is an **abelian group** (1 is identity)
3. **Distribution**: a*(b+c) = a*b + a*c
4. **Cancellation**: a*0 = 0

We denote this by $\mathbb{F}_p$ or GF(p)

What about ones that fit nicely into bits, bytes and words (i.e with 2^k elements)?

GF(2^n)

Another notation: $\mathbb{F}_{2^n}$

Has 2^n elements

Natural correspondence with bits in $\{0,1\}^n$

E.g., Elements of $\mathbb{F}_{2^8}$ can be represented as **a byte**, one bit for each term.

Linear Codes

If $\mathbb{F}$ is a finite field, then $\mathbb{F}^n$ is a vector space

Definition: C is a linear code if it is a linear subspace of $\mathbb{F}^n$ of dimension k.

This means that there is a set of k independent vectors $v_i \in \mathbb{F}^n$ $(1 \le i \le k)$ that span the subspace.

i.e. every codeword can be written as:

$$c = a_1 v_1 + a_2 v_2 + \ldots + a_k v_k \quad \text{where } a_i \in \mathbb{F}$$

"Basis (or spanning) Vectors"

Some Properties of Linear Codes

1. Linear combination of two codewords is a codeword.

2. Minimum distance (d) = weight of least weight (non-zero) codewords

(Weight of a vector refers to the Hamming weight of a vector, which is equal to the number of non-zero symbols in the vector)

$$d = \min_{\substack{c_i, c_j \in C \\ i \neq j}} |c_i - c_j|$$

$$= \min_{\substack{c \in C \\ c \neq 0}} |c|$$

Generator and Parity Check Matrices

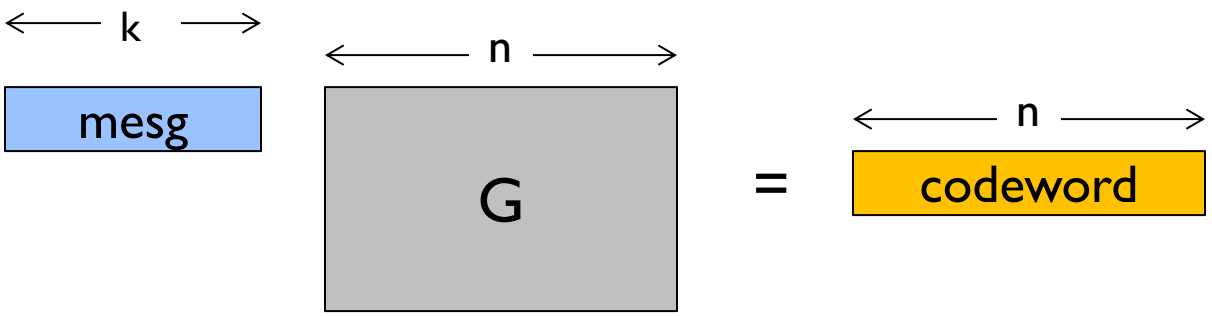
3. Every linear code has two matrices associated with it.

1. Generator Matrix:

A k x n matrix **G** such that: C = { mG | m $\in \mathbb{F}^k$ }

(Note: Here vectors are "row vectors".)

Made from stacking the spanning vectors

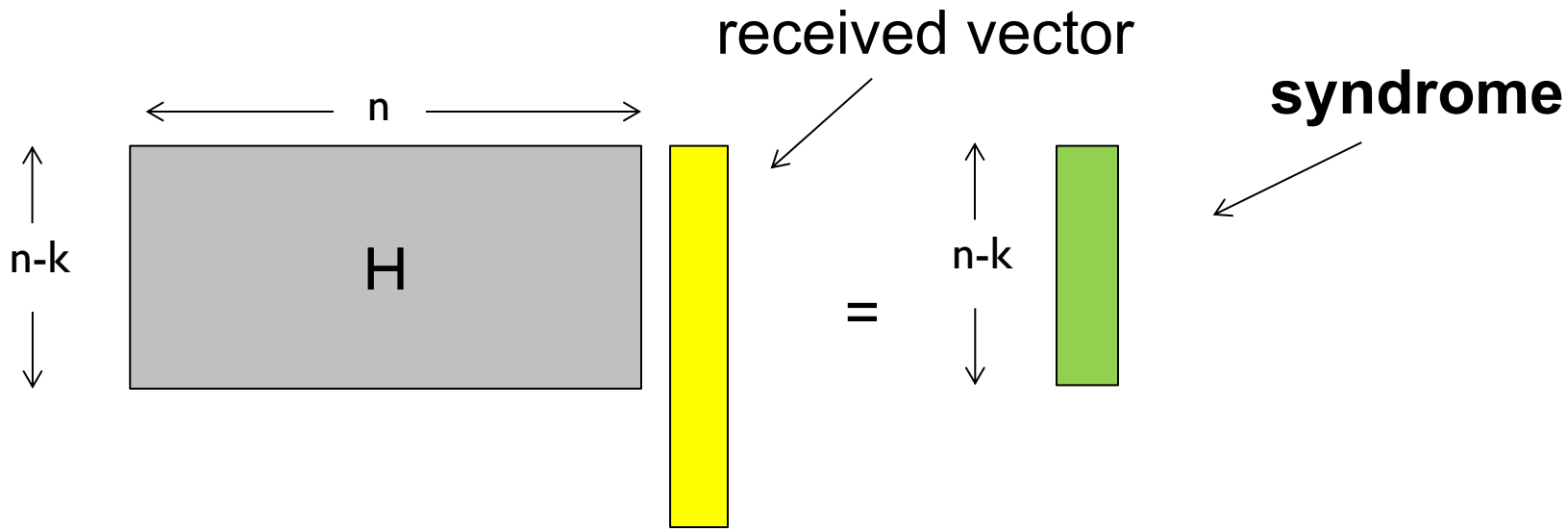


Generator and Parity Check Matrices

2. Parity Check Matrix:

An (n – k) x n matrix **H** such that: $C = \{y \in \mathbb{F}^n \mid Hy^T = 0\}$

(Codewords are the null space of H.)



if syndrome = 0, received vector = codeword

else have to use syndrome to get back codeword (“decode”)

Advantages of Linear Codes

- Encoding is efficient (vector-matrix multiply)
- Error detection is efficient (vector-matrix multiply)
- **Syndrome** (Hy^T) has error information
- How to decode? In general, have q^{n-k} sized table for decoding (one for each syndrome).
  Useful if n-k is small, else want (and there exist) other more efficient decoding algorithms.

The distance of linear codes

Theorem: Linear codes have distance d if
every set of (d-1) columns of **H** are linearly independent,
but there is a set of d columns that are linearly dependent.

Proof sketch: Ideas?

For linear codes, distance equals least weight
of non-zero codeword.
Each codeword gives some collection of columns that
must sum to zero.

Example and "Standard Form"

"Standard form" of G for systematic codes: $[I_k \; A]$.

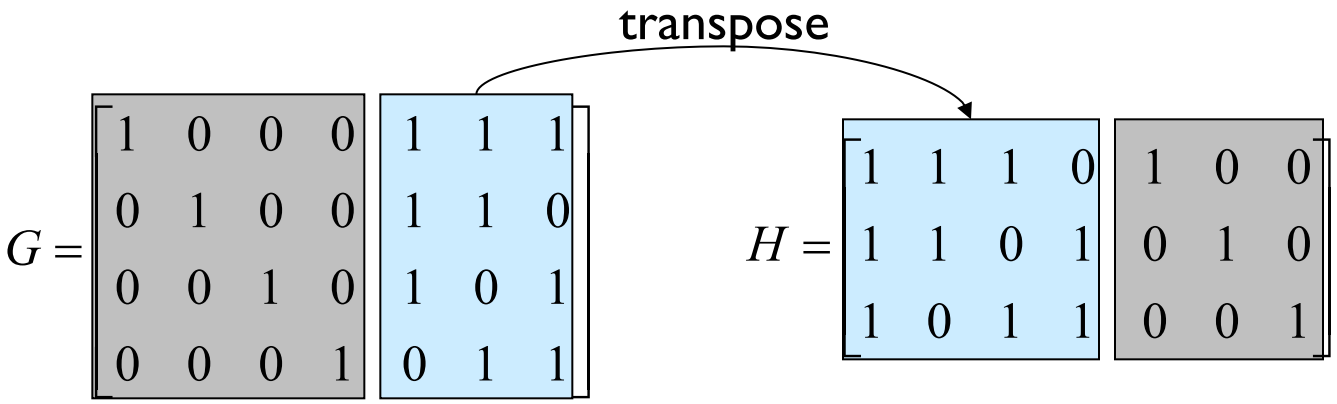
$$G = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right]$$

(7,4,3) Hamming code

Relationship of G and H

Theorem: For binary codes, if G is in standard form $[I_k \; A]$ then $H = [-A^T \; I_{n-k}]$

Example of (7,4,3) Hamming code:



Relationship of G and H

Proof:

Two parts to prove: (exercise)

1. Suppose that m is a message. Then $H(mG)^T = 0$.

2. Conversely, suppose that $Hy^T = 0$. Then y is a codeword.

Singleton bound

Theorem: For every $(n, k, d)_q$ code, $n \geq (k + d - 1)$

Another way to look at this: $d \leq (n - k + 1)$

(We will not go into the proof of this theorem in this course due to limited time on this topic.)

Codes that meet Singleton bound with equality are called

Maximum Distance Separable (MDS)

Maximum Distance Separable (MDS)

Only two binary MDS codes!

Q: What are they?

1. Repetition codes (k = 1)
2. Single-parity check codes (n-k = 1)

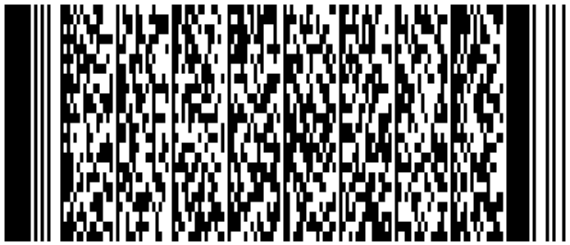
Need to go beyond the binary alphabet.

Finite fields!

Reed-Solmon (RS) codes

One of the most widely codes

- Storage systems, communication systems
- Bar codes (2-dimensional Reed-Solomon bar codes)



PDF-417

QR code

Aztec code

DataMatrix code

images: wikipedia

RS code: Polynomials viewpoint

Message: $[a_0, a_1, \ldots, a_{k-1}]$ where $a_i \in GF(q)$

Consider the polynomial of degree k-1

$$P(x) = a_{k-1} x^{k-1} + \cdots + a_1 x + a_0$$

RS code: Codeword: $[P(\alpha_1), P(\alpha_2), \ldots, P(\alpha_n)]$

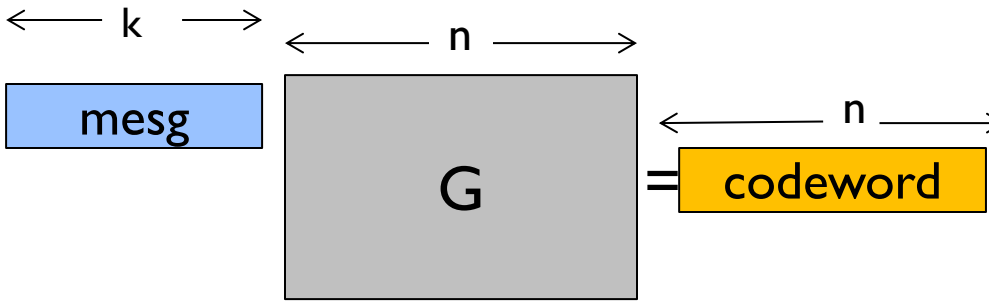
(distinct α_i's)

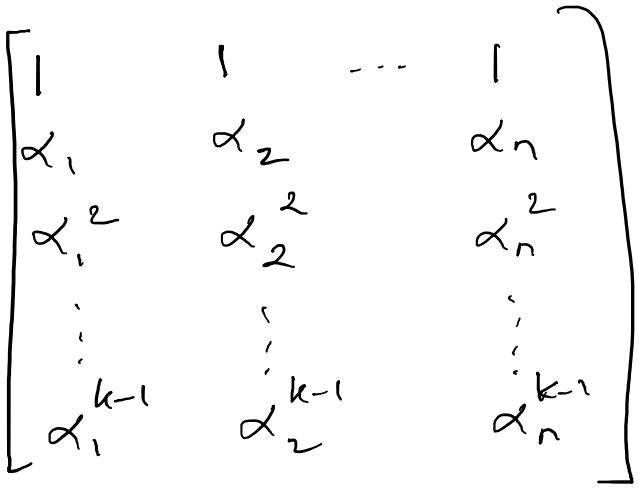
To make the α_i's in $P(\alpha_i)$ distinct, need field size $q \geq n$

That is, need sufficiently large field size for desired codeword length.

Generator matrix of RS code

What is the generator matrix?





"Vandermonde matrix"

Special property of Vandermonde matrices: Full rank (columns linearly independent)

Very useful in constructing codes.