Foreword

This chapter is based on lecture notes from coding theory courses taught by Venkatesan Guruswami at University at Washington and CMU; by Atri Rudra at University at Buffalo, SUNY and by Madhu Sudan at MIT.

This version is dated **April 28, 2013**. For the latest version, please go to

http://www.cse.buffalo.edu/ atri/courses/coding-theory/book/

The material in this chapter is supported in part by the National Science Foundation under CAREER grant CCF-0844796. Any opinions, findings and conclusions or recomendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

Chapter 9

From Large to Small Alphabets: Code Concatenation

Recall Question 8.3.2 that we had asked before: Is there an explicit asymptotically good binary code (that is, rate $R > 0$ and relative distance $\delta > 0$)? In this chapter, we will consider this question when we think of explicit code in the sense of Definition 6.3.1 as well as the stronger notion of a strongly explicit code (Definition 6.3.2).

Let us recall all the (strongly) explicit codes that we have seen so far. (See Table 9.1 for an overview.)

Code	R	δ
Hamming	$1 - O\left(\frac{\log n}{n}\right)$	$O\left(\frac{1}{n}\right)$
Hadamard	$O\left(\frac{\log n}{n}\right)$	$\frac{1}{2}$
Reed-Solomon	$\frac{1}{2}$	$O\left(\frac{1}{\log n}\right)$

Table 9.1: Strongly explicit binary codes that we have seen so far.

Hamming code (Section 2.4), which has rate $R = 1 - O(\log n/n)$ and relative distance $\delta = O(1/n)$ and the Hadamard code (Section 2.7), which has rate $R = O(\log n/n)$ and relative distance $1/2$. Both of these codes have extremely good R or δ at the expense of the other parameter. Next, we consider the Reed-Solomon code (of say $R = 1/2$) as a binary code, which does much better– $\delta = \frac{1}{\log n}$, as we discuss next.

Consider the Reed-Solomon over $\mathbb{F}_{2^s}$ for some large enough s. It is possible to get an $\left[n, \frac{n}{2}, \frac{n}{2}+1\right]_{2^s}$ Reed-Solomon code (i.e. $R = 1/2$). We now consider a Reed-Solomon codeword, where every symbol in $\mathbb{F}_{2^s}$ is represented by an s-bit vector. Now, the "obvious" binary code created by viewing symbols from $\mathbb{F}_{2^s}$ as bit vectors as above is an $\left[ns, \frac{ns}{2}, \frac{n}{2}+1\right]_2$ code[1]. Note that the distance of this code is only $\Theta\left(\frac{N}{\log N}\right)$, where $N = ns$ is the block length of the final binary code. (Recall that $n = 2^s$ and so $N = n \log n$.)

[1] The proof is left as an exercise.

The reason for the (relatively) poor distance is that the bit vectors corresponding to two different symbols in $\mathbb{F}_{2^s}$ may only differ by one bit. Thus, d positions which have different $\mathbb{F}_{2^s}$ symbols might result in a distance of only d as bit vectors.

To fix this problem, we can consider applying a function to the bit-vectors to increase the distance between those bit-vectors that differ in smaller numbers of bits. Note that such a function is simply a code! We define this recursive construction more formally next. This recursive construction is called concatenated codes and will help us construct (strongly) explicit asymptotically good codes.

9.1 Code Concatenation

For $q \geq 2$, $k \geq 1$ and $Q = q^k$, consider two codes which we call *outer* code and *inner* code:

$$C_{\text{out}} : [Q]^K \to [Q]^N,$$

$$C_{\text{in}} : [q]^k \to [q]^n.$$

Note that the alphabet size of C_{out} exactly matches the number of messages for C_{in}. Then given $\mathbf{m} = (m_1, \ldots, m_K) \in [Q]^K$, we have the code $C_{\text{out}} \circ C_{\text{in}} : [q]^{kK} \to [q]^{nN}$ defined as

$$C_{\text{out}} \circ C_{\text{in}}(\mathbf{m}) = (C_{\text{in}}(C_{\text{out}}(\mathbf{m})_1), \ldots, C_{\text{in}}(C_{\text{out}}(\mathbf{m})_N)),$$

where

$$C_{\text{out}}(\mathbf{m}) = (C_{\text{out}}(\mathbf{m})_1, \ldots, C_{\text{out}}(\mathbf{m})_N).$$

This construction is also illustrated in Figure 9.1.

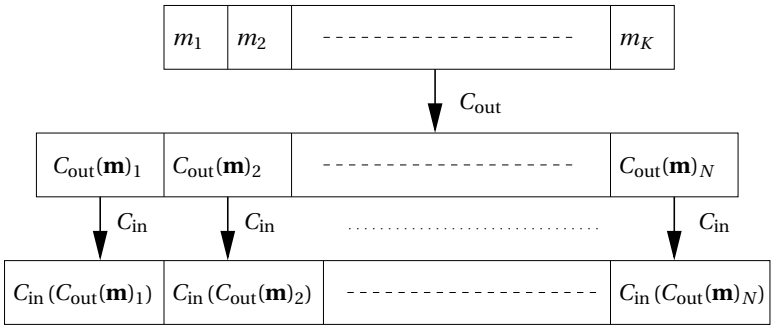

Figure 9.1: Concatenated code $C_{\text{out}} \circ C_{\text{in}}$.

We now look at some properties of a concatenated code.

Theorem 9.1.1. *If C_{out} is an $(N, K, D)_{q^k}$ code and C_{in} is an $(n, k, d)_q$ code, then $C_{\text{out}} \circ C_{\text{in}}$ is an $(nN, kK, dD)_q$ code. In particular, if C_{out} (C_{in} resp.) has rate R (r resp.) and relative distance δ_{out} (δ_{in} resp.) then $C_{\text{out}} \circ C_{\text{in}}$ has rate Rr and relative distance $\delta_{\text{out}} \cdot \delta_{\text{in}}$.*

Proof. The first claim immediately implies the second claim on the rate and relative distance of $C_{\text{out}} \circ C_{\text{in}}$. The claims on the block length, dimension and alphabet of $C_{\text{out}} \circ C_{\text{in}}$ follow from the definition.[2] Next we show that the distance is at least dD. Consider arbitrary $\mathbf{m}_1 \neq \mathbf{m}_2 \in [Q]^K$. Then by the fact that C_{out} has distance D, we have

$$\Delta\left(C_{\text{out}}\left(\mathbf{m}_1\right), C_{\text{out}}\left(\mathbf{m}_2\right)\right) \geq D. \tag{9.1}$$

Thus for each position $1 \leq i \leq N$ that contributes to the distance above, we have

$$\Delta\left(C_{\text{in}}\left(C_{\text{out}}\left(\mathbf{m}_1\right)_i\right), C_{\text{in}}\left(C_{\text{out}}\left(\mathbf{m}_2\right)_i\right)\right) \geq d, \tag{9.2}$$

as C_{in} has distance d. Since there are at least D such positions (from (9.1)), (9.2) implies

$$\Delta\left(C_{\text{out}} \circ C_{\text{in}}\left(\mathbf{m}_1\right), C_{\text{out}} \circ C_{\text{in}}\left(\mathbf{m}_2\right)\right) \geq dD.$$

The proof is complete as the choices of $\mathbf{m}_1$ and $\mathbf{m}_2$ were arbitrary. □

If C_{in} and C_{out} are linear codes, then so is $C_{\text{out}} \circ C_{\text{in}}$, which can be proved for example, by defining a generator matrix for $C_{\text{out}} \circ C_{\text{in}}$ in terms of the generator matrices of C_{in} and C_{out}. The proof is left as an exercise.

9.2 Zyablov Bound

We now instantiate an outer and inner codes in Theorem 9.1.1 to obtain a new lower bound on the rate given a relative distance. We'll initially just state the lower bound (which is called the Zyablov bound) and then we will consider the explicitness of such codes.

We begin with the instantiation of C_{out}. Note that this is a code over a large alphabet and we have seen an optimal code over large enough alphabet: Reed-Solomon codes (Chapter 5). Recall that the Reed-Solomon codes are optimal because they meet the Singleton bound 4.3.1. Hence, let us assume that C_{out} meets the Singleton bound with rate of R, i.e. C_{out} has relative distance $\delta_{\text{out}} > 1 - R$. Note that now we have a chicken and egg problem here. In order for $C_{\text{out}} \circ C_{\text{in}}$ to be an asymptotically good code, C_{in} needs to have rate $r > 0$ and relative distance $\delta_{\text{in}} > 0$ (i.e. C_{in} also needs to be an asymptotically good code). This is precisely the kind of code we are looking for to answer Question 8.3.2! However the saving grace will be that k can be much smaller than the block length of the concatenated code and hence, we can spend "more" time searching for such an inner code.

Suppose C_{in} meets the GV bound (Theorem 4.2.1) with rate of r and thus with relative distance $\delta_{\text{in}} \geq H_q^{-1}(1-r) - \varepsilon$, for some $\varepsilon > 0$. Then by Theorem 9.1.1, $C_{\text{out}} \circ C_{\text{in}}$ has rate of rR and $\delta = (1-R)(H_q^{-1}(1-r) - \varepsilon)$. Expressing R as a function of δ and r, we get the following:

$$R = 1 - \frac{\delta}{H_q^{-1}(1-r) - \varepsilon}.$$

[2] Technically, we need to argue that the q^{kK} messages map to distinct codewords to get the dimension of kK. However, this follows from the fact, which we will prove soon, that $C_{\text{out}} \circ C_{\text{in}}$ has distance $dD \geq 1$, where the inequality follows for $d, D \geq 1$.

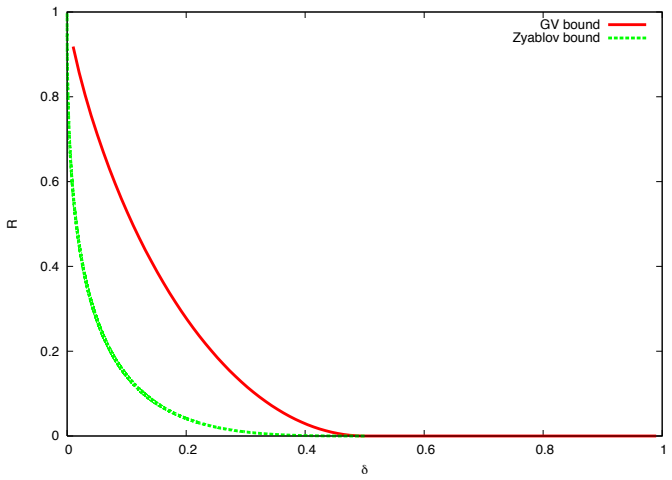

Figure 9.2: The Zyablov bound for binary codes. For comparison, the GV bound is also plotted.

Then optimizing over the choice of r, we get that the rate of the concatenated code satisfies

$$\mathscr{R} \geq \max_{0<r<1-H_q(\delta+\varepsilon)} r\left(1-\frac{\delta}{H_q^{-1}(1-r)-\varepsilon}\right),$$

where the bound of $r < 1-H_q(\delta+\varepsilon)$ is necessary to ensure that $\mathscr{R} > 0$. This lower bound on the rate is called the Zyablov bound. See Figure 9.2 for a plot of this bound for binary codes.

To get a feel for how the bound behaves, consider the case when $\delta = \frac{1}{2} - \varepsilon$. We claim that the Zybalov bound states that $\mathscr{R} \geq \Omega(\varepsilon^3)$. (Recall that the GV bound for the same δ has a rate of $\Omega(\varepsilon^2)$.) The proof of this claim is left as an exercise.

Note that the Zyablov bound implies that for every $\delta > 0$, there exists a (concatenated) code with rate $R > 0$. However, we already knew about the existence of an asymptotically good code by the GV bound (Theorem 4.2.1). Thus, a natural question to ask is the following:

Question 9.2.1. *Can we construct an explicit code on the Zyablov bound?*

We will focus on linear codes in seeking an answer to the question above because linear codes have polynomial size representation. Let C_{out} be an $[N,K]_Q$ Reed-Solomon code where $N = Q-1$ (evaluation points being $\mathbb{F}_Q^*$ with $Q = q^k$). This implies that $k = \Theta(\log N)$. However we still need an efficient construction of an inner code that lies on the GV bound. We do not expect to construct such a C_{in} in time poly(k) as that would answer Open Question 8.3.2! However, since $k = O(\log N)$, note that an exponential time in k algorithm is still a polynomial (in N) time algorithm.

There are two options for this exponential (in k) time construction algorithm for C_{in}:

- Perform an exhaustive search among all generator matrices for one satisfying the required property for C_{in}. One can do this because the Varshamov bound (Theorem 4.2.1) states that there exists a linear code which lies on the GV bound. This will take $q^{O(kn)}$ time. Using $k = rn$ (or $n = O(k)$), we get $q^{O(kn)} = q^{O(k^2)} = N^{O(\log N)}$, which is upper bounded by $(nN)^{O(\log(nN))}$, a quasi-polynomial time bound.

- The second option is to construct C_{in} in $q^{O(n)}$ time and thus use $(nN)^{O(1)}$ time overall. See Exercise 9.4.1 for one way to construct codes on the GV bound in time $q^{O(n)}$.

Thus,

Theorem 9.2.1. *We can construct a code that achieves the Zyablov bound in polynomial time.*

In particular, we can construct explicit asymptotically good code in polynomial time, which answers Question 9.2.1 in the affirmative.

A somewhat unsatisfactory aspect of this construction (in the proof of Theorem 9.2.1) is that one needs a brute force search for a suitable inner code (which led to the polynomial construction time). A natural followup question is

Question 9.2.2. *Does there exist a strongly explicit asymptotically good code?*

9.3 Strongly Explicit Construction

We will now consider what is known as the *Justesen code*. The main insight in these codes is that if we are only interested in asymptotically good codes, then the arguments in the previous section would go through even if (i) we pick different inner codes for each of the N outer codeword positions and (ii) most (but not necessarily all) inner code lie on the GV bound. It turns out that constructing an "ensemble" of codes such that most of the them lie on the GV bound is much easier than constructing a single code on the GV bound. For example, the ensemble of all linear codes have this property– this is exactly what Varshamov proved. However, it turns out that we need this ensemble of inner codes to be a smaller one than the set of all linear codes.

Justesen code is concatenated code with multiple, *different* linear inner codes. Specifically, it is composed of an $(N, K, D)_{q^k}$ outer code C_{out} and different inner codes $C_{\text{in}}^i : 1 \le i \le N$. Formally, the concatenation of these codes, denoted by $C_{\text{out}} \circ \left(C_{\text{in}}^1, \ldots, C_{\text{in}}^N\right)$, is defined as follows: given a message $\mathbf{m} \in \left[q^k\right]^K$, let the outer codeword be denoted by $(c_1, \ldots, c_N) \stackrel{\text{def}}{=} C_{\text{out}}(\mathbf{m})$. Then $C_{\text{out}} \circ \left(C_{\text{in}}^1, \ldots, C_{\text{in}}^N\right)(\mathbf{m}) = \left(C_{\text{in}}^1(c_1), C_{\text{in}}^2(c_2), \ldots, C_{\text{in}}^n(c_N)\right)$.

We will need the following result.

Theorem 9.3.1. *Let $\varepsilon > 0$. There exists an ensemble of inner codes $C_{\text{in}}^1, C_{\text{in}}^2, \ldots, C_{\text{in}}^N$ of rate $\frac{1}{2}$, where $N = q^k - 1$, such that for at least $(1-\varepsilon)N$ values of i, C_{in}^i has relative distance $\ge H_q^{-1}\left(\frac{1}{2} - \varepsilon\right)$.*

In fact, this ensemble is the following: for $\alpha \in \mathbb{F}^*_{q^k}$, the inner code $C^\alpha_{\text{in}} : \mathbb{F}^k_q \to \mathbb{F}^{2k}_q$ is defined as $C^\alpha_{\text{in}}(x) = (x, \alpha x)$. This ensemble is called the *Wozencraft ensemble*. We claim that C^α_{in} for every $\alpha \in \mathbb{F}^*_{q^k}$ is linear and is strongly explicit. (The proof if left as an exercise.)

9.3.1 Justesen code

For the Justesen code, the outer code C_{out} is a Reed-Solomon code evaluated over $\mathbb{F}^*_{q^k}$ of rate R, $0 < R < 1$. The outer code C_{out} has relative distance $\delta_{\text{out}} = 1 - R$ and block length of $N = q^k - 1$. The set of inner codes is the Wozencraft ensemble $\{C^\alpha_{\text{in}}\}_{\alpha \in \mathbb{F}^*_{q^k}}$ from Theorem 9.3.1. So the Justesen code is the concatenated code $C^* \stackrel{\text{def}}{=} C_{\text{out}} \circ (C^1_{\text{in}}, C^2_{\text{in}}, \ldots, C^N_{\text{in}})$ with the rate $\frac{R}{2}$. The following proposition estimates the distance of C^*.

Proposition 9.3.2. *Let $\varepsilon > 0$. C^* has relative distance at least $(1 - R - \varepsilon) \cdot H_q^{-1}\left(\frac{1}{2} - \varepsilon\right)$*

Proof. Consider $\mathbf{m}^1 \neq \mathbf{m}^2 \in (\mathbb{F}_{q^k})^K$. By the distance of the outer code $|S| \geq (1 - R)N$, where

$$S = \left\{ i | C_{\text{out}}(\mathbf{m}^1)_i \neq C_{\text{out}}(\mathbf{m}^2)_i \right\}.$$

Call the ith inner code *good* if C^i_{in} has distance at least $d \stackrel{\text{def}}{=} H_q^{-1}\left(\frac{1}{2} - \varepsilon\right) \cdot 2k$. Otherwise, the inner code is considered bad. Note that by Theorem 9.3.1, there are at most εN bad inner codes. Let S_g be the set of all good inner codes in S, while S_b is the set of all bad inner codes in S. Since $S_b \leq \varepsilon N$,

$$|S_g| = |S| - |S_b| \geq (1 - R - \varepsilon)N. \tag{9.3}$$

For each good $i \in S$, by definition we have

$$\Delta\left(C^i_{\text{in}}\left(C_{\text{out}}\left(\mathbf{m}^1\right)_i\right), C^i_{\text{in}}\left(C_{\text{out}}\left(\mathbf{m}^2\right)_i\right)\right) \geq d. \tag{9.4}$$

Finally, from (9.3) and (9.4), we obtain that the distance of C^* is at least

$$(1 - R - \varepsilon) \cdot Nd = (1 - R - \varepsilon) H_q^{-1}\left(\frac{1}{2} - \varepsilon\right) N \cdot 2k,$$

as desired. □

Since the Reed-Solomon codes as well as the Wozencraft ensemble are strongly explicit, the above result implies the following:

Corollary 9.3.3. *The concatenated code C^* from Proposition 9.3.2 is an asymptotically good code and is strongly explicit.*

Thus, we have now satisfactorily answered Question 9.2.2 modulo Theorem 9.3.1, which we prove next.

Proof of Theorem 9.3.1. Fix $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in \mathbb{F}_q^{2k} \setminus \{\mathbf{0}\}$. Note that this implies that $\mathbf{y}_1 = \mathbf{0}$ and $\mathbf{y}_2 = \mathbf{0}$ are not possible. We claim that $\mathbf{y} \in C_{\text{in}}^{\alpha}$ for at most one $\alpha \in \mathbf{F}_{2^k}^*$. The proof is by a simple case analysis. First, note that if $\mathbf{y} \in C_{\text{in}}^{\alpha}$, then it has to be the case that $\mathbf{y}_2 = \alpha \cdot \mathbf{y}_1$.

- Case 1: $\mathbf{y}_1 \neq \mathbf{0}$ and $\mathbf{y}_2 \neq \mathbf{0}$, then $\mathbf{y} \in C_{\text{in}}^{\alpha}$, where $\alpha = \frac{\mathbf{y}_2}{\mathbf{y}_1}$.
- Case 2: $\mathbf{y}_1 \neq \mathbf{0}$ and $\mathbf{y}_2 = \mathbf{0}$, then $\mathbf{y} \notin C_{\text{in}}^{\alpha}$ for every $\alpha \in \mathbb{F}_{2^k}^*$ (as $\alpha \mathbf{y}_1 \neq \mathbf{0}$ since product of two elements in $\mathbb{F}_{2^k}^*$ also belongs to $\mathbb{F}_{2^k}^*$).
- Case 3: $\mathbf{y}_1 = \mathbf{0}$ and $\mathbf{y}_2 \neq \mathbf{0}$, then $\mathbf{y} \notin C_{\text{in}}^{\alpha}$ for every $\alpha \in \mathbb{F}_{2^k}^*$ (as $\alpha \mathbf{y}_1 = \mathbf{0}$).

Now assume that $wt(\mathbf{y}) < H_q^{-1}(1-\varepsilon)n$. Note that if $\mathbf{y} \in C_{\text{in}}^{\alpha}$, then C_{in}^{α} is "bad"(i.e. has relative distance $< H_q^{-1}\left(\frac{1}{2} - \varepsilon\right)$). Since $\mathbf{y} \in C_{\text{in}}^{\alpha}$ for at most one value of α, the total number of bad codes is at most

$$\begin{aligned}
\left|\left\{\mathbf{y} | wt(\mathbf{y}) < H_q^{-1}\left(\frac{1}{2} - \varepsilon\right) \cdot 2k\right\}\right| &\leq Vol_q\left(H_q^{-1}\left(\frac{1}{2} - \varepsilon\right) \cdot 2k, 2k\right) \\
&\leq q^{H_q(H_q^{-1}(\frac{1}{2}-\varepsilon)) \cdot 2k} \qquad (9.5) \\
&= q^{(\frac{1}{2}-\varepsilon) \cdot 2k} \\
&= \frac{q^k}{q^{2\varepsilon k}} \\
&< \varepsilon(q^k - 1) \qquad (9.6) \\
&= \varepsilon N. \qquad (9.7)
\end{aligned}$$

In the above, (9.5) follows from our good old upper bound on the volume of a Hamming ball (Proposition 3.3.1) while (9.6) is true for large enough k. Thus for at least $(1-\varepsilon)N$ values of α, C_{in}^{α} has relative distance at least $H_q^{-1}\left(\frac{1}{2} - \varepsilon\right)$, as desired. □

By concatenating an outer code of distance D and an inner code of distance d, we can obtain a code of distance at least $\geq Dd$ (Theorem 9.1.1). Dd is called the concatenated code's *design distance*. For asymptotically good codes, we have obtained polynomial time construction of such codes (Theorem 9.2.1, as well as strongly explicit construction of such codes (Corollary 9.3.3). Further, since these codes were linear, we also get polynomial time encoding. However, the following natural question about decoding still remains unanswered.

Question 9.3.1. *Can we decode concatenated codes up to half their design distance in polynomial time?*

9.4 Exercises

Exercise 9.4.1. In Section 4.2.1, we saw that the Gilbert construction can compute an $(n,k)_q$ code in time $q^{O(n)}$. Now the Varshamov construction (Section 4.2.2) is a randomized construction and it is natural to ask how quickly can we compute an $[n,k]_q$ code that meets the GV bound. In this exercise, we will see that this can also be done in $q^{O(n)}$ deterministic time, though the deterministic algorithm is not that straight-forward anymore.

1. (*A warmup*) Argue that Varshamov's proof gives a $q^{O(kn)}$ time algorithm that constructs an $[n,k]_q$ code on the GV bound. (Thus, the goal of this exercise is to "shave" off a factor of k from the exponent.)

2. A $k \times n$ *Toeplitz Matrix* $A = \{A_{i,j}\}_{i=1,\,j=1}^{k\,,\,n}$ satisfies the property that $A_{i,j} = A_{i-1,j-1}$. In other words, any diagonal has the same value. For example, the following is a 4×6 Toeplitz matrix:
$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 1 & 2 & 3 & 4 & 5 \\ 8 & 7 & 1 & 2 & 3 & 4 \\ 9 & 8 & 7 & 1 & 2 & 3 \end{pmatrix}$$
A *random* $k \times n$ Toeplitz matrix $T \in \mathbb{F}_q^{k\times n}$ is chosen by picking the entries in the first row and column uniformly (and independently) at random.

   Prove the following claim: For any non-zero $\mathbf{m} \in \mathbb{F}_q^k$, the vector $\mathbf{m} \cdot T$ is uniformly distributed over $\mathbb{F}_q^n$, that is for every $\mathbf{y} \in \mathbb{F}_q^n$, $\Pr\left[\mathbf{m} \cdot T = \mathbf{y}\right] = q^{-n}$.

   (*Hint:* Write down the expression for the value at each of the n positions in the vector $\mathbf{m} \cdot T$ in terms of the values in the first row and column of T. Think of the values in the first row and column as variables. Then divide these variables into two sets (this "division" will depend on $\mathbf{m}$) say S and $\overline{S}$. Then argue the following: for every fixed $\mathbf{y} \in \mathbb{F}_q^n$ and for every fixed assignment to variables in S, there is a unique assignment to variables in $\overline{S}$ such that $\mathbf{m}T = \mathbf{y}$.)

3. Briefly argue why the claim in part (b) implies that a random code defined by picking its generator matrix as a random Toeplitz matrix with high probability lies on the GV bound.

4. Conclude that an $[n,k]_q$ code on the GV bound can be constructed in time $2^{O(k+n)}$.

9.5 Bibliographic Notes

Code concatenation was first proposed by Forney[12].

Justesen codes were constructed by Justesen [34]. In his paper, Justesen attributes the Wozencraft ensemble to Wozencraft.