Lecture 4: BCH Codes, Code Concatenation

Topics in Error-Correcting Codes (Fall 2022)
University of Toronto
Swastik Kopparty
Scribe: Anatoly Zavyalov and Gal Gross

1 BCH Codes

We first focus on **BCH codes**, which meet the volume-packing (Hamming) bound with a constant distance d and $|\Sigma| = 2$ as $n \to \infty$.

Throughout, we let $n = 2^t$ for some positive integer t, and fix $|\Sigma| = 2$. We consider $\mathbb{F}_{2^t} = \{\alpha_1, \dots \alpha_n\}$, where $\mathbb{F}_{2^t}$ denotes the field consisting of 2^t elements. We will also consider $\mathbb{F}_2^t$, which denotes the vector space over $\mathbb{F}_2$ of dimension t.

Fact 0. *There exists an $\mathbb{F}_2$-linear bijection $\psi \colon \mathbb{F}_{2^t} \to \mathbb{F}_2^t$ between $\mathbb{F}_{2^t}$ and $\mathbb{F}_2^t$.*

We give two equivalent definitions of BCH codes. When defining the code, a constant parameter $k \leq 2^t - 2$ of the BCH code is given.

Definition 1 (First definition of BCH codes). *A* ***BCH code*** *$C \subseteq \mathbb{F}_2^n$ is*

$$C = \{f \colon \mathbb{F}_{2^t} \to \mathbb{F}_2 \mid f \textit{ is the evaluation of a polynomial } p(x) \subseteq \mathbb{F}_{2^t}(x) \textit{ of degree} \leq 2^t - k - 2\}.$$

Definition 2 (Second definition of BCH codes). *A* ***BCH code*** *$C \subseteq \mathbb{F}_2^n$ is*

$$C = \left\{ f \colon \mathbb{F}_{2^t} \to \mathbb{F}_2 \mid \sum_{\alpha \in \mathbb{F}_{2^t}} f(\alpha) \cdot v_\alpha = 0 \right\},$$

where

$$v_\alpha = \begin{pmatrix} \psi(1) \\ \psi(\alpha) \\ \psi(\alpha^2) \\ \vdots \\ \psi(\alpha^k) \end{pmatrix} \in \mathbb{F}_2^{t(k+1)}.$$

Note that v_α has length $t(k+1)$ because $\psi(1)$ and each $\psi(\alpha^i)$ have length t.

Claim 3. *A BCH code C has distance $\geq k + 2$.*

Proof. We prove this claim with both definitions of the BCH code.

Using the first definition of BCH codes (**Definition ??**), if $f, g \in C$ are distinct codewords, the number of agreements (that is, $|\{x \in \mathbb{F}_{2^t} \mid f(x) = g(x)\}|$) of f and g is $\leq 2^t - k - 2$ (as otherwise

$f = g$, a contradiction). Hence, the distance (the number of disagreements) between f and g is $\geq k+2$.

Using the second definition of BCH codes (**Definition ??**), let $S = \{i \in \{1, \ldots, n\} \mid f(\alpha_i) = 1\}$ be the set of indices where f is equal to 1. As C is a linear code, it suffices to show that $|S| \geq k+2$. For $f \in C$, we have

$$\begin{bmatrix} | & | & & | \\ v_{\alpha_1} & v_{\alpha_2} & \cdots & v_{\alpha_{2^t}} \\ | & | & & | \end{bmatrix} \begin{bmatrix} | \\ f(\alpha) \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

From the definition of S,

$$\begin{aligned} \sum_{\alpha \in \mathbb{F}_{2^t}} f(\alpha) v_\alpha = 0 &\implies \sum_{\alpha \in S} v_\alpha = 0 \\ &\implies \sum_{\alpha \in S} \psi(\alpha^i) = 0 \quad \text{for all } i \in \{0, 1, \ldots, k\} && \text{by definition of } v_\alpha \\ &\implies \sum_{\alpha \in S} \alpha^i = 0 \quad \text{for all } i \in \{0, 1, \ldots, k\} && \text{by linearity of } \psi \end{aligned}$$

We can equivalently express this as below, where each column of the matrix contains powers of all elements $\alpha \in S$:

$$k+1 \text{ rows} \left\{ \underbrace{\begin{bmatrix} & & 1 & \\ & & \alpha & \\ \cdots & \cdots & \alpha^2 & \cdots \\ & & \vdots & \\ & & \alpha^k & \end{bmatrix}}_{|S| \text{ columns}} \right. \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

If $|S| \leq k+1$, then, as in the previous lecture, we can let $\alpha_{i_1}, \ldots, \alpha_{i_{|S|}}, \alpha_{i_{|S|+1}}, \ldots, \alpha_{i_{k+1}} \in \mathbb{F}_{2^t}$ be such that $\alpha_{i_1}, \ldots \alpha_{i_{|S|}} \in S$ are distinct, and $\alpha_{i_{|S|+1}}, \ldots, \alpha_{i_{k+1}} \in S$ are distinct from each other, but not from $\alpha_{i_1}, \ldots \alpha_{i_{|S|}}$, where $i_1, ..., i_{k+1} \in \{1, \ldots, n\}$. But then the Vandermonde matrix

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ \alpha_{i_1} & \alpha_{i_2} & \alpha_{i_3} & \cdots & \alpha_{i_{k+1}} \\ \alpha_{i_1}^2 & \alpha_{i_2}^2 & \alpha_{i_3}^2 & \cdots & \alpha_{i_{k+1}}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{i_1}^k & \alpha_{i_2}^k & \alpha_{i_3}^k & \cdots & \alpha_{i_{k+1}}^k \end{bmatrix}$$

has a determinant of zero (as the $\alpha_{i_1}, \ldots, \alpha_{i_{k+1}}$ are not distinct), which is a contradiction. Hence, we must have $|S| \geq k+2$, as desired.

□

Observation 4. If we set $k = 1$, notice that we get the Hamming code, as the second definition of BCH codes (**Definition ??**) gives the generator matrix of the Hamming code with a fixed vector $\psi(1)$ at the top of each column.

2 Size of a BCH Code

Let's determine a lower bound on the size of the BCH code C with a constant distance d. However, we first make a simplification to the definition of a BCH code. We know that $f \in C$ if and only if

$$t(k+1) \text{ rows} \left\{ \begin{bmatrix} \psi(1) & \psi(1) & \cdots & \psi(1) \\ \psi(\alpha_1) & \psi(\alpha_2) & \cdots & \psi(\alpha_n) \\ \vdots & \vdots & \ddots & \vdots \\ \psi(\alpha_1^k) & \psi(\alpha_2^k) & \cdots & \psi(\alpha_n^k) \end{bmatrix} \right. \begin{bmatrix} | \\ f(\alpha) \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{1}$$

We see that as $\psi(1) \in \mathbb{F}_2^t$ and it is nonzero (as ψ is linear), each of the first t rows of the parity-check matrix above consist of all 1s or all 0s. So, we can replace the first t rows with a single row of all 1's and end up with the same set C, as all rows of zeroes impose no conditions on f, and all rows of ones impose the same condition $\sum_{\alpha \in \mathbb{F}_{2^t}} f(\alpha) = 0$. Hence, (**??**) is equivalent to

$$tk+1 \text{ rows} \left\{ \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \psi(\alpha_1) & \psi(\alpha_2) & \cdots & \psi(\alpha_n) \\ \vdots & \vdots & \ddots & \vdots \\ \psi(\alpha_1^k) & \psi(\alpha_2^k) & \cdots & \psi(\alpha_n^k) \end{bmatrix} \right. \begin{bmatrix} | \\ f(\alpha) \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2}$$

So, $\dim(C) \geq 2^t - tk - 1$, implying that $|C| \geq 2^{2^t - tk - 1} = 2^{n - k\log_2(n) - 1} = \frac{1}{2}\frac{2^n}{n^k}$. If the distance of the code is $d = k+2$, this gives a code of size $\Omega\left(\frac{2^n}{n^{d-2}}\right)$.

However, we can do even better! In fact, this very same code goes well beyond the Gilbert-Varshamov bound of $\Omega\left(\frac{2^n}{n^{d-1}}\right)$, and meets the Hamming bound $\omega\left(\frac{2^n}{n^{\left(\frac{d-1}{2}\right)}}\right)$. To show this, we first touch on the algebra of $\mathbb{F}_{2^t}$.

Fact 5. *Given any $\alpha, \beta \in \mathbb{F}_{p^t}$ where p is prime, we have $(\alpha+\beta)^p = \alpha^p + \beta^p$. In particular, given $\alpha, \beta \in \mathbb{F}_{2^t}$, we have $(\alpha+\beta)^2 = \alpha^2 + \beta^2$, i.e. squaring is a linear operation.*[1]

The above fact holds because any field $\mathbb{F}_{p^m}$ where p is prime has characteristic p, hence $\mathbb{F}_{2^t}$ has characteristic 2. It also implies that $\psi(\alpha^2) \in \operatorname{span}(\psi(\alpha))$ for any $\alpha \in \mathbb{F}_{2^t}$.

Claim 6. *Suppose that $\sum\limits_{\alpha \in \mathbb{F}_{2^t}} f(\alpha)\psi(\alpha) = 0$. Then $\sum\limits_{\alpha \in \mathbb{F}_{2^t}} f(\alpha)\psi(\alpha^2) = 0$.*

Proof. We have

$$\begin{aligned} \sum_{\alpha \in \mathbb{F}_{2^t}} f(\alpha)\psi(\alpha) = 0 &\implies \sum_{\alpha \in \mathbb{F}_{2^t}} \psi(f(\alpha)\cdot\alpha) = 0 \quad \text{by linearity of } \psi\text{, viewing } \mathbb{F}_2 \text{ as a subset of } \mathbb{F}_{2^t} \\ &\implies \psi\left(\sum_{\alpha \in \mathbb{F}_{2^t}} f(\alpha)\alpha\right) = 0 \end{aligned}$$

[1] This is an example of where the "freshman's dream" is fulfilled!

$$\begin{aligned}
&\Longrightarrow \sum_{\alpha\in\mathbb{F}_{2^t}} f(\alpha)\alpha = 0 \\
&\Longrightarrow \left(\sum_{\alpha\in\mathbb{F}_{2^t}} f(\alpha)\alpha\right)^2 = 0 \\
&\Longrightarrow \sum_{\alpha\in\mathbb{F}_{2^t}} f(\alpha)\alpha^2 = 0 \qquad\qquad \text{by } \textbf{Fact ??}\text{, and } f(\alpha)^2 = f(\alpha) \text{ since } f(\alpha)\in\mathbb{F}_2 \\
&\Longrightarrow \psi\left(\sum_{\alpha\in\mathbb{F}_{2^t}} f(\alpha)\alpha^2\right) = 0 \\
&\Longrightarrow \sum_{\alpha\in\mathbb{F}_{2^t}} f(a)\psi(\alpha^2) = 0,
\end{aligned}$$

as desired.

□

Theorem 7. *A BCH code C with distance $d \in O(1)$ meets the Hamming bound $\omega\left(\frac{2^n}{n^{\left(\frac{d-1}{2}\right)}}\right)$.*

Proof. The first thing we prove is that the rank of the parity-check matrix in (**??**) is not $tk+1$, it's actually half of that! By **Fact ??**, $\psi(\alpha^{2i}) \in \operatorname{span}(\psi(\alpha^i))$ for all $\alpha \in \mathbb{F}_{2^t}$, so we can keep only the odd-numbered rows in the parity-check matrix in (**??**), giving

$$C = \left\{ f : \mathbb{F}_{2^t} \to \mathbb{F}_2 | \sum_{\alpha\in\mathbb{F}_{2^t}} f(\alpha)\widetilde{v}_\alpha = 0 \right\},$$

where, setting $\ell = \left\lceil \frac{k}{2} \right\rceil$,

$$\widetilde{v}_\alpha = \begin{pmatrix} 1 \\ \psi(\alpha) \\ \psi(\alpha^3) \\ \vdots \\ \psi(\alpha^{2\ell-1}) \end{pmatrix}.$$

Pushing this through the calculation done at the start of this section, when d is constant, we get a size of $\omega\left(\frac{2^n}{n^{\left(\frac{d-1}{2}\right)}}\right)$, which is the Hamming bound.

□

For an even k, we get $|C| \geq 2^{2^t - t\frac{k}{2} - 1} = \frac{1}{2}\frac{2^n}{n^{\frac{k}{2}}} \in \Omega\left(\frac{2^n}{n^{\frac{d}{2}-1}}\right)$.

If $d \in \Theta(1)$, then the BCH codes will match the Hamming bound up to a factor in $\Theta(1)$. If we consider d to be non-constant, we would get similar bounds when $d \in O(\log_2(n))$, but when $d \in \Theta(n)$, we would get a very low lower bound, which isn't that interesting.

3 BCH codes over bigger alphabets

In most practical applications, an alphabet Σ of size 2 often suffices, but, for instance, packets sent over the internet have *huge* alphabet sizes. What happens if we consider differently sized alphabets, for instance, if $|\Sigma| = 3$?

Let $n = 3^t$ and consider $\mathbb{F}_{3^t}$ and $\mathbb{F}_3^t$. Similarly to **Fact ??**, there exists an $\mathbb{F}_3$-linear bijection $\psi : \mathbb{F}_{3^t} \to \mathbb{F}_3^t$ between $\mathbb{F}_{3^t}$ and $\mathbb{F}_3^t$, and we can define a BCH code as in **Definition ??**:

$$C = \left\{ f : \mathbb{F}_{3^t} \to \mathbb{F}_3 \mid \sum_{\alpha \in \mathbb{F}_{3^t}} f(\alpha) v_\alpha = 0 \right\}, \tag{3}$$

where

$$v_\alpha = \begin{pmatrix} \psi(1) \\ \psi(\alpha) \\ \psi(\alpha^2) \\ \vdots \\ \psi(\alpha^k) \end{pmatrix}.$$

As before, C has distance $\geq k+2$, and by **Fact ??**, cubing is a linear operation in $\mathbb{F}_{3^t}$. So, we can eliminate every third row similarly to v_α as in the proof of **Theorem ??** (as $\psi(\alpha^{3i}) \in \operatorname{span}(\psi(\alpha^i))$ for all $\alpha \in \mathbb{F}_{3^t}$) to get $\tilde{v}_\alpha \in \mathbb{F}_3^{\approx t \cdot (\frac{2k}{3})}$ (the size of $\tilde{v}_\alpha$ will vary depending on whether k is divisible by 3). This gives

$$|C| \gtrsim 3^{3^t - \frac{2k}{3}t} = \frac{3^n}{n^{\frac{2k}{3}}} = \frac{3^n}{n^{\frac{2d}{3} + O(1)}},$$

which does *not* show tightness of the Hamming bound. There are ways to make tiny improvements to the BCH code's size, but those improvements are nowhere close to changing the $\frac{2}{3}$ factor into $\frac{1}{2}$. For different values of t, one could compute the ranks of the parity-check matrices in (**??**), but, they would still (sadly) give codes of size close to $\frac{3^n}{n^{\frac{2d}{3}+O(1)}}$.

4 Asymptotically Good Codes

For a long while coding theorists struggled with the question of whether there exist codes with constant rate and constant relative distance. The other parameters are allowed the vary, leading us to the following definition.

Definition 8. *Let $(\Sigma_n)_{n=1}^\infty$ be a sequence of finite alphabets, and $(C_n)_{n=1}^\infty$ a sequence of codes such that $C_n \subseteq (\Sigma_n)^n$ for each $n \in \mathbb{N}$. We say the sequence (C_n) is* asymptotically good *if there exist $R = \Omega(1)$ and $\delta = \Omega(1)$ such that for all sufficiently large n*

$$\mathit{rate}(C_n) \geq R \text{ and } \Delta(C_n) \geq \delta n.$$

It is easy to see that Reed-Solomon codes are asymptotically good for $|\Sigma_n| \geq n$, and the question is whether there exists asymptotically good codes with $|\Sigma_n| = O(1)$.

This question was answered in the affirmative by Forney in the 1970's, though it took a while for the community to appreciate the significance of his technique. Forney's key insight was to use *code concatenation.*

Code concatenation requires two codes, and their encoding schemes

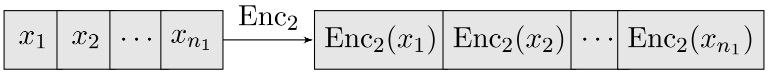
$$C_1 \subseteq \Sigma_1^{n_1} \qquad \mathrm{Enc}_1 : \Sigma_1^{k_1} \to C_1;$$
$$C_2 \subseteq \Sigma_2^{n_2} \qquad \mathrm{Enc}_2 : \Sigma_2^{k_2} \to C_2$$

satisfying the condition

$$|\Sigma_1| = |\Sigma_2|^{k_2}\,.$$

(Remember from Lecture 1 that for an arbitrary code $k = \log|C| \,/\, \log|\Sigma|$ is the parameter such that $|\Sigma|^k = |C|$.)

The condition allows us to fix some bijection $\Sigma_1 \to \Sigma_2^k$ which identifies the two sets. We can therefore regard each *symbol* of Σ_1 as a string of symbols in Σ_2^k; applying Enc_2 to this string we obtain a codeword in C_2, which is string of n_2 symbols. Repeating this process for each symbol in a *codeword* $x = x_1x_2\ldots x_{n_1} \in C_1$, we obtain a string s of codewords in C_2 of length $|s| = n_1n_2$ (see figure below.)



If we regard the resulting string s as a code-word in its own right, what are the parameters of the resulting 'concatenation code' C? Starting with a word in $\Sigma_1^{k_1} \equiv \Sigma_2^{k_1k_2}$, we apply $\mathrm{Enc} = \mathrm{Enc}_2 \circ \mathrm{Enc}_1$ to obtain a word in $\Sigma_2^{n_1n_2}$.

Moreover, if $\Delta(C_1) = d_1$ and $\Delta(C_2) = d_2$, then we are guaranteed that $\Delta(C) \geq d_1d_2$. Indeed, given two words $x, y \in C_1$, we know they differ in at least d_1 symbols, but applying Enc_2 to two different symbols we obtain strings that differ in at least d_2 symbols, so that $\mathrm{Enc}(x), \mathrm{Enc}(y)$ differ in at least d_1d_2 symbols.

code	alphabet	size	length	distance
C_1	Σ_1	k_1	n_1	d_1
C_2	Σ_2	k_2	n_2	d_2
C	Σ_2	k_1k_2	n_1n_2	d_1d_2

Code concatenation allows us to construct asymptotically good codes. Our plan (to be carried out in detail next class) is as follows.

- For C_1 we take the Reed-Solomon code over $\mathbb{F}_2^{2^t}$ of length $n = 2^t$ and distance $d_1 = n_1/2$ (i.e., we consider the set of degree $n_1/2$ polynomials).
- For C_2 we brute-force search for a binary linear code with size $k_2 = t$, length 10^3t, and distance $d_2 = 10^2t$.

Since $1 - H(10^{-1}) > 10^{-3}$, the parameters of C_2 satisfy the GV bound. Thus, the existence of the code C_2 is guaranteed by our greedy algorithm, which achieves the GV bound. This algorithm runs in time $2^{O(t)}$, and since $t = \log n$ we see that we can explicitly construct asymptotically good codes in time $poly(n)$.