INSTRUCTIONS: The exam has six questions labeled 1 through 6. The exam is worth 100 points in total.

The exam is two hours in length.

To receive full credit you must explain your answers.

Write all answers on the exam. You may use the backs of pages if necessary.

Pre-approved calculator (Casio 991 or equivalent) is permitted.
1. Let \( n \in \mathbb{N} \). The binary repetition code of length \( n \) has the codewords \( 00 \cdots 0 \) and \( 11 \cdots 1 \). Show that it is a \([n, 1, n]\) linear code.

2. Let \( C \) be the binary code with all words of length 4 over \( F = \mathbb{F}_2 = \{0, 1\} \) as its codewords. Let \( C' \) be the single parity check code obtained by adding an extra bit at the end of each codeword in \( C \) to make the total number of 1s even. Show that \( C' \) is a linear \([5, 4, 2]\)-code.

3. Let \( C \) be the binary linear code of length 5 with codewords 

\[
00000, 01101, 10110, 11011.
\]

Show that \( C \) is 1-error correcting and 2-error detecting.

Solutions:

1. Let \( C = \{00 \cdots 0, 11 \cdots 1\} \) where each codeword is of length \( n \). Then \( C \) is obviously a subspace of \( \mathbb{F}_2^n \) as sums of any two elements are in \( C \). A generator matrix of \( C \) is 

\[
G = [1, 1, \ldots, 1],
\]

and the minimum distance is \( d = n \). Hence \( C \) is a binary linear code \([n, 1, n]\).

2. Let \( C \) be a binary code of length 4. So 

\[
C = \{x = [x_1, x_2, x_3, x_4] | x_i \in \mathbb{F}_2 \quad \text{for all } i \}.
\]

Then \( |C| = 2^4 \) and the minimum distance is \( d = 1 \), so \( C \) is a \((4, 16, 1)\) code. Also \( \dim_{\mathbb{F}_2} C = \log_2 2^4 = 4 \). \( C \) is linear as it is a subspace of \( \mathbb{F}_2^4 \). In fact, for any \( x, y \in C \), 

\[
x + y \in C.
\]

Hence \( C \) is a binary linear \([4, 4, 1]\) code.

Now let 

\[
C' = \{y = xx_5 = [x_1, x_2, x_3, x_4, x_5] | [x_1, x_2, x_3, x_4] \in C, \sum_{i=1}^{5} x_i \equiv 0 \pmod{2} \}.
\]

So \( C' \) is a single parity check code over \( \mathbb{F}_2 \). Obviously, \( cC' \) is a subspace of \( \mathbb{F}_2^5 \). In fact, for any \( y, y' \), 

\[
y + y' = xx_5 + x'x_5' \quad \text{and} \quad \sum_{i=1}^{5} (x_i + x'_i) = 0 \in \mathbb{F}_2.
\]

Also \( \dim_{\mathbb{F}_2} C' = 4 \) and the minimum distance is \( d = 2 \) as the fifth bit is added to check parity. Hence \( C' \) is a binary linear \([5, 4, 2]\) code.

3. We see that the minimum distance \( d \) of \( C \) is \( d = 3 \). Hence \( C \) is \( \lfloor 3 - 1 \rfloor = 1 \) error correcting, and \( 3 - 1 = 2 \) error detecting.
2. [5 pts each= 15 pts] The ISBN-10 coding system is a code $C$ over $F = \mathbb{F}_{11} = \{0, 1, 2, \cdots, 9, 10\}$. An ISBN-10 codeword $\mathbf{x} = x_1 x_2 \cdots x_{10}$ is ten bit in length, and satisfies $\sum_{i=1}^{10} i x_i \equiv 0 \pmod{11}$. Suppose that $C$ has a parity check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{bmatrix}$$

1. Show that $C$ is a linear $[10, 8]$ code over $F = \mathbb{F}_{11}$.
2. Show that $C$ is a 1-error correcting code.
3. What is the missing bit to make $0257x41326$ an ISBN code?

Solutions:

1. Each codeword has length 10. A parity check matrix $H$ has a standard form

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{bmatrix}$$

$\rightarrow r_1 \rightarrow r_1 + r_2 \begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 0 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{bmatrix}$

$\rightarrow r_1 \times (-1), r_2 \times (-1) \begin{bmatrix} 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \\ 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$

$\rightarrow r_2 \rightarrow r_2 - 2r_1 \begin{bmatrix} 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 0 & 1 \end{bmatrix}$

Hence $\text{rank}(H) = 2 = 10 - k$ so that $k = 8$. This implies that $C$ is a linear $[10, 8]$ code.

A generator matrix $G$ is given by

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 8 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 3 & 7 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 4 & 6 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 6 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 8 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

which has $k = 8$. 
2. We see from the standard form of \( H \) that \( d \geq 3 \). In fact, any 2 columns of \( H \) are linearly independent but there are 3 columns that become linearly dependent, e.g., the last three columns. So \( d = 3 \).

3. 

\[
1 \cdot 0 + 2 \cdot 2 + 3 \cdot 5 + 4 \cdot 7 + 5 \cdot x + 6 \cdot 4 + 7 \cdot 1 + 8 \cdot 3 + 9 \cdot 2 + 10 \cdot 6
\]

\[
= 4 + 15 + 28 + 5x + 24 + 7 + 24 + 18 + 60 = 5x + 4 = 0 \in \mathbb{F}_{11}
\]

Hence \( x = -4 \cdot 5^{-1} = -4 \cdot 9 = -36 = -3 = 8 \in \mathbb{F}_{11} \).
3. [10 pts] A codeword of the code $C = \{01010, 10101\}$ is transmitted through a BSC with crossover probability $p < 1/2$, and MDD decoder $D$ is applied to the received word. Give a formula for the decoding error probability $P_{err}$ of $D$. (No need to do the actual calculation for $P_{err}$.)

**Solution:** For $c \in C$, we have

$$P_{err}(c) = \sum_{y : D(y) \neq c} \text{Prob}(y \text{ received } | c \text{ transmitted})$$

and

$$P_{err} = \max_{c \in C} P_{err}(c).$$

In our case, we have

$$P_{err}(c) = \sum_{y : D(y) \neq c} (1 - p)^5 \left( \frac{p}{1 - p} \right)^{d(y, c)}$$

where $d(y, c)$ is the Hamming distance.
4. [5pts=20pts] If $B$ is a matrix, $B^T$ stands for its transpose, and $I_r$ denotes the $r \times r$ identity matrix. Let $C$ be a linear $[n, k, d]$ code over $F = \mathbb{F}_q$.

1. What is a generator matrix of $C$?

2. What is a parity check matrix of $C$?

3. Show that $C$ is self-dual if and only if $k = n/2$. (A linear code $C$ is said to be self-dual if $C^\perp = C$.)

4. If $C$ has a generator matrix $G$ with standard form $[I_k A]$ where $A$ is a $k \times (n - k)$ matrix, show that a parity check matrix $H$ has standard form $[-A^T I_{n-k}]$.

**Solutions:**

1. A $k \times n$ matrix $G$ whose rows form a basis of $C$ is called a generator matrix of $C$. However, $G$ is not unique as a basis is not unique.

2. First we define the dual code $C^\perp$ of $C$.

$$C^\perp = \{ x \in F^n \mid x \cdot y = 0 \quad \text{for all} \quad y \in C \}.$$  

$C^\perp$ is clearly a subspace of $F^n$, so it is a linear $[n, n - k]$ code by the fundamental theorem of Linear Algebra.

A parity check matrix $H$ of $C$ is defined to be a generator matrix of the dual code $C^\perp$.

Alternatively, a parity check matrix for $C$ is an $(n - k) \times n$ matrix $H$ such that for all words $x \in F^n$, $x \in C$ if and only if $Hx^t = 0$.

3. If $C$ is an $[n, k]$ code, then $C^\perp$ is an $[n, n - k]$ code as noted above. Then $C = C^\perp$ if and only if $k = n - k$ if and only if $k = n/2$.

4. A generator matrix $G$ of an $[n, k, d]$ code is a $k \times n$ matrix. A parity check matrix is an $(n - k) \times n$ matrix. Suppose that a generator matrix is given by

$$G = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(k) \end{bmatrix}$$

where $r(1), \ldots, r(k)$ are $k$ linearly independent rows. Then

$$HG^T = H[r(1)^T, r(2)^T, \ldots, r(k)^T] = [Hr(1)^T, Hr(2)^T, \ldots, Hr(k)^T] = 0$$
so that
\[ Hr(i)^T = 0 \quad \iff \quad r(i) \in \text{Ker}(H) \]
for every \( i = 1, 2, \ldots, k \). If \( G = [I_k \ A] \) and \( H = [B \ I_{n-k}] \), then \( H^T = \begin{bmatrix} B^T \\ I_{n-k} \end{bmatrix} \) and

\[ 0 = GH^T = [I_k \ A] \begin{bmatrix} B^T \\ I_{n-k} \end{bmatrix} = B^T + A \]

Hence \( B^T = -A \) and \( B = -A^T \).
5. [5 pts each=25 pts] Let $C$ be a binary linear $[6, 3, 3]$ code with parity check matrix

$$H = \begin{bmatrix}
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 1
\end{bmatrix}. $$

- What is the syndrome of a word?
- Show that two words have the same syndrome if and only if they lie in the same coset of $C$.
- Find coset leaders.
- Find syndromes.
- Decode $y = 110001$.

**Solutions:**

1. A syndrome of a word $y \in \mathbb{F}_2^n$ is $s = s(y) = Hy^t$.

2. Let $y_1, y_2 \in \mathbb{F}_2^n$. Then

$$y_1 + C = y_2 + C \iff y_1 - y_2 \in C \iff H(y_1 - y_2)^T = 0 \iff s(y_1) = s(y_2).$$

3. We know that $C$ has $2^3 = 8$ codewords, and $|\mathbb{F}_2^6| = 2^6 = 64$, so there are $64/8 = 8$ cosets and hence 8 coset leaders. Now

$$C = \{000000, 100110, 010101, 001001, 110011, 100111, 011100, 111010 \}.$$  

Obvious coset leaders (corresponding to 1-errors) are

$$000000, 100000, 010000, 001000, 000100, 000010, 000001$$

There is one more coset leader corresponding to 2-errors, that is, 100001.

4. We calculate the syndromes corresponding to coset leaders:

$$s(000000) = 000, s(100000) = 100, s(010000) = 010, s(001000) = 001$$

$$s(000100) = 110, s(000010) = 101, s(000001) = 011, s(100001) = 111.$$

Note that except for the first one, syndromes are the transposes of the columns of $H$.

5. $H(110001)^t = 101$, which is the binary representation of 5. So the corresponding coset leader is $e = 000010$ and

$$110001 - 000010 = 110011 \in C.$$
6. [5pts each=15 pts]

1. Let $H$ be a parity check matrix for a linear $[n,k]$ code $C$. Prove that the minimum distance of $C$ is $d$ if and only if any set of $d-1$ columns of $H$ are linearly independent but some set of $d$ columns are linearly dependent.

2. Find the minimum weight of the linear code over $F = \mathbb{F}_3$ with parity check matrix

$$H = \begin{bmatrix} 1 & 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 & 2 \end{bmatrix}.$$

3. Let $C$ be a $[n,k,d]$ linear code over $F = \mathbb{F}_q$. Prove the Singleton bound:

$$d \leq n - k + 1.$$

Solutions:

1. Let $H = [h_1, \cdots, h_n]$ be an $(n-k) \times n$ parity check matrix. Then

$$Hc^T = \sum_i c_i h_i$$

so if $c$ is a word of weight $w(c) = w$, then there exists indexes $i_1, \cdots, i_w$ such that $c_{i_1}, c_{i_2}, \cdots, c_{i_w}$ are nonzero. Then

$$Hc^T = c_{i_1} h_{i_1} + \cdots + c_{i_w} h_{i_w} = 0$$

if and only if $\{h_{i_1}, \cdots, h_{i_w}\}$ are linearly dependent. Hence $d = w(c)$.

2. $d = 4$ as $d-1 = 3$ columns are linearly independent, while any 4 columns are linearly dependent.

3. Since $C$ has minimum distance $d$, any set of $d-1$ columns of a parity check matrix $H$ are linearly independent, so $\text{rank}(H) \geq d-1$, and $\text{rank}(H) = n-k$. So

$$n - k \geq d - 1 \iff d \leq n - k + 1.$$