Name: ____

DEPARTMENT OF EECS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.02 Spring 2012

Quiz I

March 13, 2012

"×" your section	Section	Time	Recitation Instructor	TA
☐	1	10-11	Vincent Chan	Jared Monnin
☐	2	11-12	Vincent Chan	Anirudh Sivaraman
☐	3	12-1	Sidhant Misra	Sungwon Chung
☐	4	1-2	Sidhant Misra	Omid Aryan / Nathan Lachenmyer
☐	5	2-3	Katrina LaCurts	Sunghyun Park
☐	6	3-4	Katrina LaCurts	Muyiwa Ogunnika

Please read and follow these instructions:

0. Please write your name in the space above and × your section.
1. There are **15 questions** (some with multiple parts) and **10 pages** in this quiz booklet.
2. Answer each question according to the instructions given, within **120 minutes**.
3. **Please answer legibly. Explain your answers, especially when we ask you to.** If you find a question ambiguous, write down your assumptions. Show your work for partial credit.
4. Use the empty sides of this booklet if you need scratch space. *If you use the blank sides for answers, make sure to say so!*

One two-sided "crib sheet" and a calculator allowed. No other aids.

PLEASE NOTE: SOME STUDENTS WILL TAKE THE MAKE-UP QUIZ TOMORROW AT 9.30 AM. PLEASE DON'T DISCUSS THIS QUIZ WITH ANYONE IN THE CLASS, UNLESS YOU'RE SURE THEY HAVE TAKEN IT WITH YOU TODAY.

Do not write in the boxes below

1-3(x/21)	4 (x/15)	5-7 (x/18)	8-10 (x/11)	11-12 (x/11)	13-15 (x/24)	Total (x/100)

I Green Eggs and Hamming

By writing *Green Eggs and Ham*, Dr. Seuss won a $50 bet with his publisher because he used only 50 distinct English words in the entire book of 778 words. The probability of occurrences of the most common words in the book is given in the table below, in decreasing order of popularity:

Rank	Word	Probability of occurrence of word in book
1	not	10.7%
2	I	9.1%
3	them	7.8%
4	a	7.6%
5	like	5.7%
6	in	5.1%
7	do	4.6%
8	you	4.4%
9–50	(*all other words*)	45.0%

1. [4 × 2 = 8 points]: I pick a secret word from the book.

The Bofa tells you that the secret word is one of the 8 most common words in the book.
Yertle tells you it is *not* the word "*not*".
The Zlock tells you it is three letters long.

Express your answers to the following in $\log_2(\frac{100}{\cdot})$ form, which will be convenient; you don't need to give the actual numerical value. (The 100 is because the probabilities in the table are shown as percentages.)

How many bits of information about the secret word have you learned from:

The Bofa alone? ________________________

Yertle alone? ________________________

The Bofa and the Zlock together? ________________________

All of them together? ________________________

2. [2+2+2+1 = 7 points]: The Cat in the Hat compresses *Green Eggs and Ham* with the LZW compression method described in 6.02 (codewords from 0 to 255 are initialized to the corresponding ASCII characters, which includes all the letters of the alphabet and the space character). The book begins with these lines:

I_am_Sam
I_am_Sam
Sam_I_am

We have replaced each space with an underscore (_) for clarity, and eliminated punctuation marks.

A. What are the strings corresponding to codewords 256 and 257 in the string table?

B. When compressed, the sequence of codewords starts with the codeword 73, which is the ASCII value of *I*. The initial few codewords in this sequence will all be ≤ 255, and then one codeword > 255 will appear. What **string** does that codeword correspond to?

C. Cat finds that codeword 700 corresponds to the string *I_do_not_l*. This string comes from the sentence *I_do_not_like_them_with_a_mouse* in the book. What are the first two letters of the codeword numbered 701 in the string table?

D. Thanks to a stuck keyboard (or because Cat is an ABBA fan), the phrase *IdoIdoIdoIdoIdo* shows up at the input to the LZW compressor. The decompressor gets a codeword, already in its string table, and finds that it corresponds to the string *Ido*. This codeword is followed immediately by a new codeword **not** in its string table. What string should the decompressor return for this new codeword?

3. [2+2+2 = 6 points]: The Lorax decides to compress *Green Eggs and Ham* using Huffman coding, treating each **word** as a distinct symbol, ignoring spaces and punctuation marks. He finds that the expected code length of the Huffman code is 4.92 bits. The average length of a word in this book is 3.14 English letters. Assume that in uncompressed form, each English letter requires 8 bits (ASCII encoding). Recall that the book has 788 total words (and 50 distinct ones).

A. What is the uncompressed (ASCII-encoded) length of the book? Show your calculations.

B. What is the expected length of the Huffman-coded version of the book? Show your calculations.

C. The words "*if*" and "*they*" are the two least popular words in the book. In the Huffman-coded format of the book, what is the Hamming distance between their codewords?

4. [9+6=15 points]: The Lorax now applies Huffman coding to all of Dr. Seuss's works. He treats each word as a distinct symbol. There are n distinct words in all. Curiously, he finds that **the most popular word (symbol) is represented by the codeword** 0 **in the Huffman encoding.**

Symbol i occurs with probability p_i; $p_1 \geq p_2 \geq p_3 \ldots \geq p_n$. Its length in the Huffman code tree is ℓ_i.

A. Given the conditions above, circle **True** or **False** for these statements. *To receive credit, explain your answers below; if True, a correctness argument, and if False, a counter-example.*

(a) **True / False** $p_1 \geq 1/3$.

(b) **True / False** $p_1 \geq \sum_{i=3}^{n} p_i$.

B. The Grinch removes the most-popular symbol (whose probability is p_1) and implements Huffman coding over the remaining symbols, retaining the same probabilities proportionally; i.e., the probability of symbol i (where $i > 1$) is now $\frac{p_i}{1-p_1}$. What is the expected code length of the Grinch's code tree, in terms of $L = \sum_{i=1}^{n} p_i \ell_i$ (the *expected code length of the* **original** *code tree*) and p_1? Explain your answer.

II *Errare humanum est* (To err is human)

5. [6 points]: A binary symmetric channel has bit-flip probability ε. Suppose we take a stream of S bits in which zeroes and ones occur with equal probability, divide it into blocks of k bits each, and apply an $(n, k, 3)$ code to each block. To correct bit errors, we decode each block of n bits at the receiver using maximum-likelihood decoding. Calculate the probability that the stream of S bits is decoded correctly, assuming that any n-bit block with two or more errors will not be decoded correctly. Show your work.

6. [2+3 = 5 points]: For an (n, k, d) linear block code, fill in the blanks below in terms of one of more of n, k, d.

A. The maximum number of bit errors that can always be **detected** is _______________.

B. $2^{n-k} \geq 1 + n + \binom{n}{2} + \ldots + \binom{n}{t}$, where $t =$ _________________________________.

7. [1+2+4 = 7 points]: Consider the following variant of the rectangular parity code with r rows and c columns. Each row has a row parity bit. Each column has a column parity bit. In addition, **each codeword has an overall parity bit**, to ensure that the number of ones in each codeword is even.

A. Code rate = ______________________________. Hamming distance of code = ____________.

B. Ben Bitdiddle takes each codeword of this code and *removes* the first row parity bit from each codeword. Note that the overall parity bit is calculated by **including** the row parity bit he eliminated. Ben then transmits these codewords over a noisy channel. What is the largest number of bit errors in a codeword that this new code can always **correct**? Explain.

III The Matrix Reloaded

Neo receives a 7-bit string, $D_1D_2D_3D_4P_1P_2P_3$ from Morpheus, sent using a code, $\mathcal{C}$, with parity equations

$$\begin{aligned} P_1 &= D_1 + D_2 + D_3 \\ P_2 &= D_1 + D_2 + D_4 \\ P_3 &= D_1 + D_3 + D_4 \end{aligned}$$

8. [4 points]: Write down the generator matrix, G, for $\mathcal{C}$.

9. [3 points]: Write down the parity check matrix, H, for $\mathcal{C}$.

10. [4 points]: If Neo receives 1000010 and does maximum-likelihood decoding on it, what would his estimate of the data transmission $D_1D_2D_3D_4$ from Morpheus be? For your convenience, the syndrome s_i corresponding to data bit D_i being wrong are given below, for $i = 1, 2, 3, 4$:

$$s_1 = (111)^T, s_2 = (110)^T, s_3 = (101)^T, s_4 = (011)^T.$$

On Trinity's advice, Morpheus decides to augment each codeword in $\mathcal{C}$ with an **overall parity bit**, so that each codeword has an even number of ones. Call the resulting code $\mathcal{C}^+$.

11. [6 points]: Write down the generator matrix, G^+, of code $\mathcal{C}^+$. Express your answer as a concatenation (or stacking) of G (the generator for code $\mathcal{C}$) and another matrix (which you should specify). Explain your answer.

12. [5 points]: Morpheus would like to use a code that corrects all patterns of 2 or fewer bit errors in each codeword, by adding an appropriate number of parity bits to the data bits $D_1D_2D_3D_4$. He comes up with a code, $\mathcal{C}^{++}$, which adds 5 parity bits to the data bits to produce the required codewords. Explain whether or not $\mathcal{C}^{++}$ will meet Neo's error correction goal.

IV Convolutionally yours

Dona Ferentes is debugging a Viterbi decoder for her client, The TD Company, which is building a wireless network to send gifts from mobile phones. She picks a rate-$1/2$ code with constraint length 4, no puncturing. Parity stream p_0 has the generator $g_0 = 1110$. Parity stream p_1 has the generator $g_1 = 1xyz$, but she needs your help determining x, y, z, as well as some other things about the code. In these questions, each state is labeled with the most-recent bit on the left and the least-recent bit on the right.

13. [4+7+4 = 15 points]: These questions are about the state transitions and generators.

A. From state 010, the possible **next states** are __________ and __________.

From state 010, the possible **predecessor states** are __________ and __________.

B. Given the following facts, find g_1, the generator for parity stream p_1. g_1 has the form $1xyz$, with the standard convention that the left-most bit of the generator multiplies the most-recent input bit.

Starting at state 011, receiving a 0 produces $p_1 = 0$.
Starting at state 110, receiving a 0 produces $p_1 = 1$.
Starting at state 111, receiving a 1 produces $p_1 = 1$.

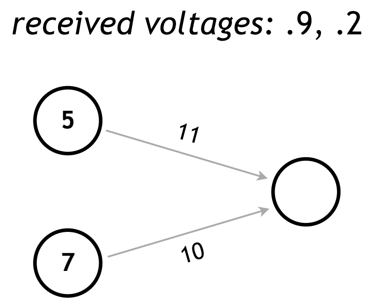
C. Dona has just completed the forward pass through the trellis and has figured out the path metrics for all the end states. Suppose the state with smallest path metric is 110. The traceback from this state looks as follows:

$$000 \leftarrow 100 \leftarrow 010 \leftarrow 001 \leftarrow 100 \leftarrow 110$$

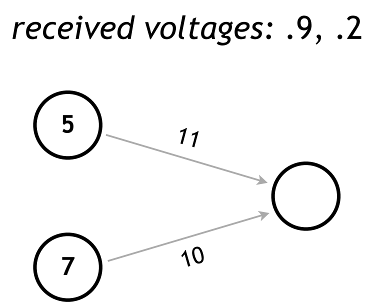
What is the most likely transmitted message? Explain your answer, and if there is not enough information to produce a unique answer, say why.

14. [8 points]: During the decoding process, Dona observes the voltage pair $(0.9, 0.2)$ volts for the parity bits p_0p_1, where the sender transmits 1.0 volts for a "1" and 0.0 volts for a "0". The threshold voltage at the decoder is 0.5 volts. In the portion of the trellis shown below, each edge shows the expected parity bits p_0p_1. The number in each circle is the path metric of that state.

A. With **hard-decision decoding**, give the branch metric near each edge and the path metric inside the circle.



B. Timmy Dan (founder of TD Corp.) suggests that Dona use **soft-decision decoding** using the **squared Euclidean distance** metric. Give the branch metric near each edge and the path metric inside the circle.



15. [1 free point!]: The real purpose behind Dona Ferentes decoding convolutionally is some awful wordplay with Virgil's classical Latin. What does *Timeo Danaos et dona ferentes* mean?

(Circle ALL that apply.)

A. Timmy Dan and Dona are friends.

B. It's time to dance with Dona Ferentes.

C. I fear the Greeks, even those bearing gifts.

D. I fear the Greeks, especially those bearing debt.

E. You *#@$*@!#s. This is the last straw; I'm reporting you to the Dean. If I'd wanted to learn this, I'd have gone to that school up the Charles!

FIN