Algorithms in the Real World (15-853), Fall 06
Assignment #6 **Due: 29 Nov 06**

You can look up material on the web and books, but you cannot look up solutions to the given problems. You can work in groups, but must write up the answers individually, and must write down your collaborators' names. Note that there are six problems on two pages.

Problem 1: Linear Codes Primer (5pt)
Show that for a linear binary code, the minimum distance is d if and only if the minimum weight of a non-zero codeword is d. (The *weight* of a codeword is just the number of 1's in the codeword.)

Problem 2: Binary versus q-ary Codes (15pt)
In class we showed a lower bound on the number of code bits required to fix s bit-errors for a binary linear code with with k message bits. In particular by arguing that every codeword requires a ball of radius $(d-1)/2$ around it we derived the formula

$$n \geq k + \log(1 + \sum_{i=1}^{s} \binom{n}{i}) \ .$$

1. Generalize this to q-ary codes.
2. Give an upper bound on the the number of code bits that must be sufficient for some code to correct s bit-errors. (Hint: the argument is similar, but the bound is not tight).

Problem 3: Ever Wonder about ISBN? (15pt)
The ISBN is a 10-digit codeword such as 0-471-06259-6. The first digit indicates the language (0 for English), the next group specifies the publisher (471 for Wiley); the next group forms the book number assigned by the publisher. The final "digit" is chosen to make the entire number $x_1 \cdots x_{10}$ satisfy the single check equation: $\sum_{i=1}^{10}(ix_i) = 0 \pmod{11}$.
Note that the first 9 digits lie between 0 and 9, whereas the last "digit" can take any value between 0 and 10, where the value 10 is represented by the letter X.

A. What are the parameters of this code in the $(n, k, d)_q$ notation (i.e., what are n, k, d and q).

B. Calculate the check digit for the message 0-13-200809.

C. It is easy to see that the ISBN code can detect any single digit error. Show that the code can detect the transposition of any two digits (not necessarily consecutive).

D. The sixth digit in the code 0-13-28?796-X was smudged. Find the missing digit.

Problem 4: Encoding and Decoding LDPC codes (15pt)
An LDPC code is given by a bipartite graph $B = (U, V, E)$, with $|U| = n$ and $|V| = n - k$. The codewords are defined as bit vectors $x \in \{0, 1\}^n$ such that the neighboring bits of any vertex $v \in V$ on the right have even parity. Assume that the graph is regular: that is, each node on the left has the same degree δ_U, and each node on the right has the same degree δ_V.
Given such a bipartite graph, explain how to obtain the message in $O(n)$ time, once you have corrected the errors and obtained a valid codeword. Also, given a message, explain how to encode a message in quadratic time, i.e. $O(n^2)$. You are allowed to do $O(n^3)$ preprocessing.

Problem 5: Conditional Probabilities (10pt)
Given the following conditional probabilities for a two state Markov Chain what factor would one save by using the conditional entropy instead of the unconditional entropy?

$$\begin{array}{cc} p(w|w) = .95 & p(b|w) = .05 \\ p(w|b) = .2 & p(b|b) = .8 \end{array}$$

Problem 6: Bounds on Prefix Codes (20pt)

A. Prove the first part of the Kraft-McMillan inequality for Prefix Codes. In particular show that for any prefix code C,

$$\sum_{(s,w)\in C} 2^{-l(w)} \leq 1.$$

B. Prove that if you have $n = 2^k$ codewords in a prefix code and that if one of them is shorter than k bits, then at least two must be longer than k bits.