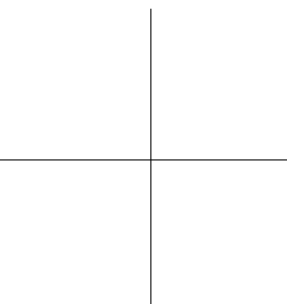
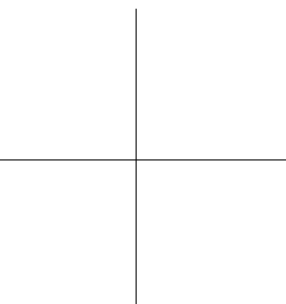
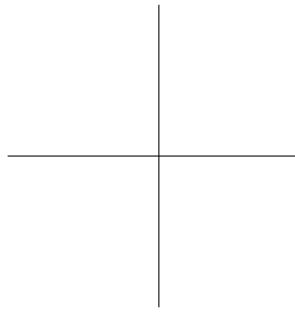
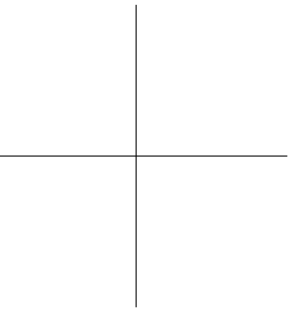




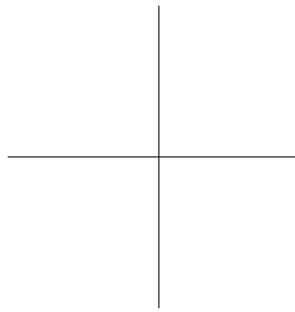
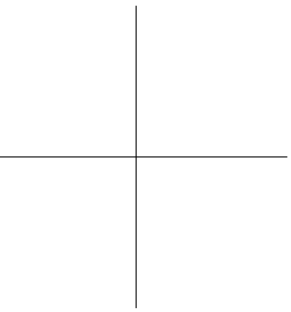
MODERN CODING THEORY - SOLUTIONS MANUAL

Preliminary version - October 28, 2008





Preliminary version – October 28, 2008



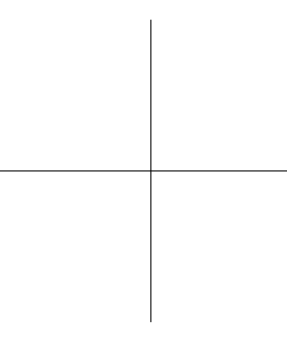
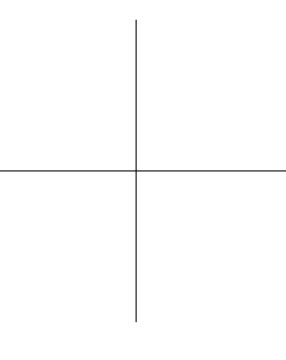


Modern Coding Theory – Solutions Manual

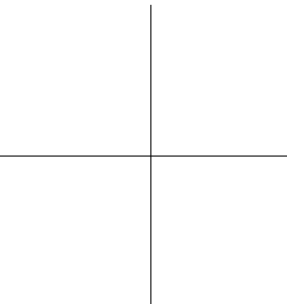
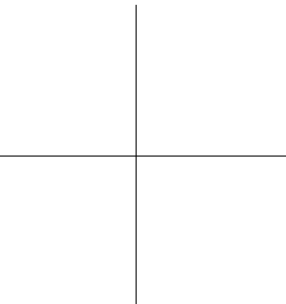
BY

T. RICHARDSON AND R. URBANKE

Cambridge University Press



Preliminary version – October 28, 2008



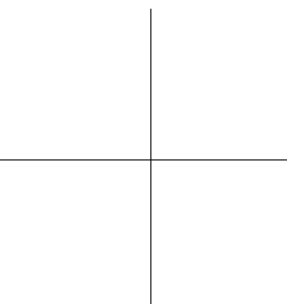
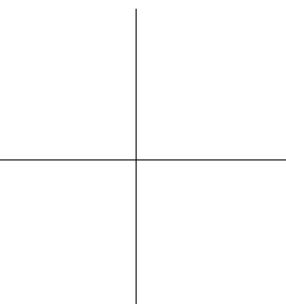
Modern Coding Theory – Solutions Manual

Copyright ©2008 by T. Richardson and R. Urbanke

All rights reserved

Library of Congress Catalog Card Number: 00-00000

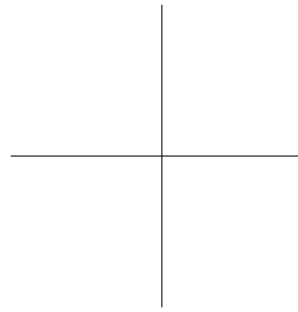
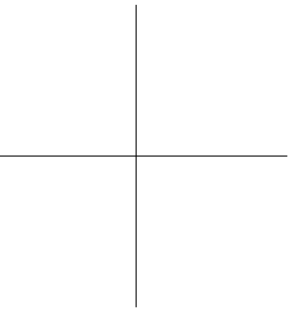
ISBN 0-000-00000-0



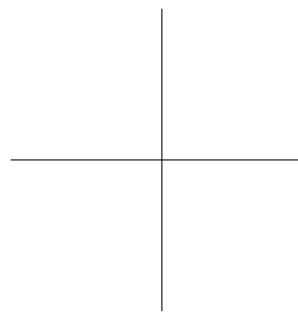
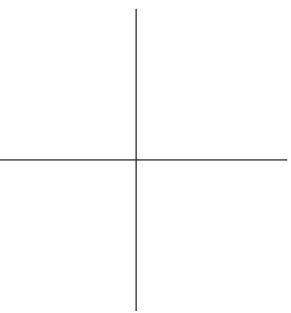
Preliminary version – October 28, 2008

CONTENTS

- 1 INTRODUCTION · page 3
- 2 FACTOR GRAPHS · page 17
- 3 BINARY ERASURE CHANNEL · page 21
- 4 BINARY MEMORYLESS SYMMETRIC CHANNELS · page 41
- 5 GENERAL CHANNELS · page 77
- 6 CONVOLUTIONAL CODES AND TURBO CODES · page 81
- 7 GENERAL ENSEMBLES · page 89
- 8 EXPANDER CODES AND THE FLIPPING ALGORITHM · page 93
- A ENCODING LOW-DENSITY PARITY-CHECK CODES · page 95
- B EFFICIENT IMPLEMENTATION OF DENSITY EVOLUTION · page 97
- C CONCENTRATION INEQUALITIES · page 101
- D FORMAL POWER SUMS · page 105



Preliminary version – October 28, 2008

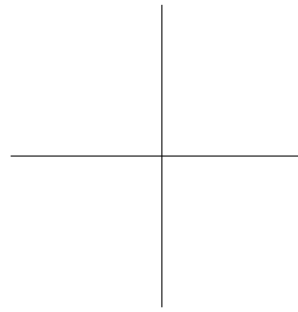
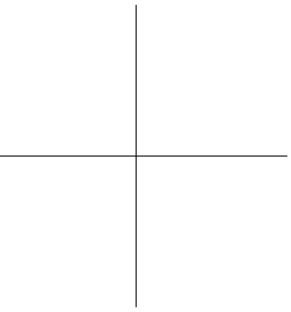




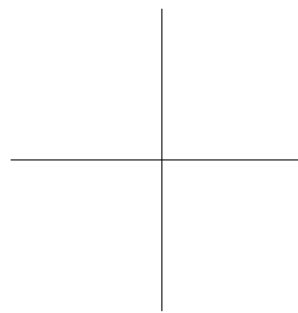
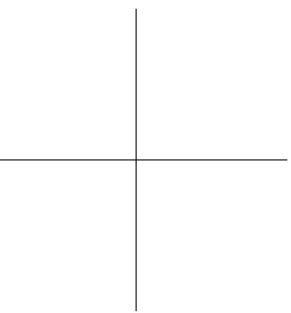
WARNING - USE AT OWN RISK

This is the solutions manual to the book `moderncodingtheory` . It contains contributions by (in alphabetical order) Abdelaziz Amraoui, Cyril Méasson, Vishwambar Rathi, and Ruediger Urbanke. These notes are work in progress and change daily. To get the current version please contact ruediger.urbanke@epfl.ch. Contributions to make this solutions manual more complete and accurate are very welcome.

If you have exercises that you are willing to share with others please send them to us.



Preliminary version – October 28, 2008



Chapter 1

INTRODUCTION

SOLUTION 1.1 (INNER PRODUCT). Parts 1, 2, and 3 are immediate. For the counterexample take $\mathbb{F} = \mathbb{F}_2$, $n = 2$, and $u = 11$. Then $\langle u, u \rangle = 0$. Recall that the set of solutions to $Gx^T = 0^T$ equals C^\perp . By explicit calculation we see that $C^\perp = \{0000, 0011, 1100, 1111\} = C$. This implies that C is self-dual. ♣ : -)

SOLUTION 1.2 (HAMMING DISTANCE). Since the distance $d(x, y)$ is defined as the sum of the distances of the components it suffices to assume that $x, y \in \mathbb{F}$. We need to check that for all $x, y, z \in \mathbb{F}$, $d(x, y) \leq d(x, z) + d(z, y)$. Since $d(\cdot, \cdot)$ is non-negative the statement is correct if $x = y$. And if $x \neq y$, so that $d(x, y) = 1$ the statement is valid as well since in this case at least one of the two terms on the right must take the value 1. ♣ : -)

SOLUTION 1.3 (EXTENDING, PUNCTURING AND SHORTENING).

Extending: results in a $(n + 1, M, d')$ code where $d' \geq d$ and $d' = d + 1$ if d is odd

Puncturing: results in a $(n - 1, M, d')$ code where $d' = d$ or $d - 1$

Shortening: results in a $(n - 1, M', d')$ code where $M' \leq M$ and $d' \geq d$

Under extending and puncturing linear codes stay linear. Under shortening a linear code stays linear if the common symbol which is chosen is the zero symbol. ♣ : -)

SOLUTION 1.4 ($u + v$ CONSTRUCTION). The non-trivial part is to show that $d = \min(2d_1, d_2)$. Let $u_1, u_2 \in C_1$ and $v_1, v_2 \in C_2$. Clearly, $\min_{u_1 \neq u_2, v_1 = v_2} d((u_1, u_1 + v_1), (u_2, u_2 + v_2)) = 2d_1$ and $\min_{u_1 = u_2, v_1 \neq v_2} d((u_1, u_1 + v_1), (u_2, u_2 + v_2)) = d_2$. Hence,

$$d = \min(2d_1, d_2, \min_{u_1 \neq u_2, v_1 \neq v_2} d((u_1, u_1 + v_1), (u_2, u_2 + v_2))).$$

Now $\min_{u_1 \neq u_2, v_1 \neq v_2} d((u_1, u_1 + v_1), (u_2, u_2 + v_2)) \geq d_1$. If $d_1 \geq d_2$, then $d = d_2$ and we are done. If, on the other hand, $d_1 \leq d_2$, then the minimization can be decomposed into two minimizations:

$$\min_{u_1 \neq u_2, v_1 \neq v_2} = \min\left(\min_{d_1 \leq d(u_1, u_2) \leq d_2}, \min_{d(u_1, u_2) \geq d_2}\right)$$

To calculate $d((u_1 + v_1), (u_2 + v_2))$, we observe that only those components contribute for which $u_{1i} + u_{2i} + v_{1i} + v_{2i} = 1$, i.e., either $u_{1i} \neq u_{2i}$ or $v_{1i} \neq v_{2i}$, but not both of them differ. So

$$\min_{d_1 \leq d(u_1, u_2) \leq d_2} d((u_1, u_1 + v_1), (u_2, u_2 + v_2)) \geq d_2$$

as $d(u_1 + v_1, u_2 + v_2) \geq d_2 - d_1$. Hence, $d = \min(2d_1, d_2)$. ♣ : -)

SOLUTION 1.5 (PROPER CODES). Fix a position i , $1 \leq i \leq n$. Since the code is proper there exists at least one codeword, call it c , in C which has a non-zero component at position i . The $|\mathbb{F}|$ distinct “multiples” of c have each a different element of \mathbb{F} at position i . Let the multiple with α at position i , $\alpha \in \mathbb{F}$, be denoted by c_α . Let us define two codewords x and y to be *equivalent* if $x = y + c_\alpha$ for some $\alpha \in \mathbb{F}$. Now note that this equivalence relationship generates equivalence classes of size $|\mathbb{F}|$ so that each codeword is in exactly one equivalence class and that each equivalence class contains exactly one codeword which has element β at position i for all $\beta \in \mathbb{F}$.

Consider now a set of positions. To keep things simple consider two positions, call them i and j . It can happen that no codeword takes on the value α at position i and β at position j , $\alpha, \beta \in \mathbb{F}$. But if there is at least one such codeword then the number of such codewords is equal to the number of codewords that take on the value 0 both at i and j . This follows from essentially the same type of argument as used above. The statement extends to any number of positions. ♣ : -)

SOLUTION 1.6 (ONE C , MANY G). Let us construct a generator matrix G for C . Start by picking any *non-zero* codeword of C and declare this codeword to constitute the first row of G . Since C has cardinality 2^k , there are $2^k - 1$ such choices. For the second row there are $2^k - 2$ remaining choices since we can pick any codeword which is not contained in the subspace spanned by the first row. For the i -th row there are by the same reasoning $2^k - 2^{i-1}$ choices. Clearly, all such choices lead to different generator matrices and any generator matrix can be constructed in this way. It follows that the total number of distinct generator matrices for a code of length n and dimension k is $\prod_{i=1}^k (2^k - 2^{i-1}) = \prod_{i=1}^k 2^{i-1} (2^{k-i+1} - 1) = 2^{\binom{k}{2}} \prod_{i=1}^k (2^i - 1)$. ♣ : -)

SOLUTION 1.7 (CONVERSION OF G INTO SYSTEMATIC FORM). By assumption G has rank k and this implies that such a submatrix A must exist. First re-order the columns of G so that the columns which make up A form the first k columns of G . Now multiply from the left by A^{-1} . This gives the desired systematic generator matrix. ♣ : -)

SOLUTION 1.8 (CONVERSION $G \leftrightarrow H$). A direct multiplication shows that

$$(I_k P) (-P^T I_{n-k})^T = -I_k P + P I_{n-k} = -P + P = 0.$$

For the $[7, 4, 3]$ Hamming code we can derive from the parity-check matrix in (1.25) the generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

♣ : -)

SOLUTION 1.9 (REED-SOLOMON CODES). To see this claim, first note that the code is linear since the evaluation map is a linear map. In more detail, if $A(x)$ and $B(x)$ are two elements of $\mathbb{F}[x]$ of degree at most $k - 1$ and if $\alpha, \beta \in \mathbb{F}$, then $C(x) = \alpha A(x) + \beta B(x)$ is also an element of $\mathbb{F}[x]$ of degree at most $k - 1$. Further, for any $x_i \in \mathbb{F}$, $\alpha A(x_i) + \beta B(x_i) = C(x_i)$. This shows that any linear combination of codewords is again a codeword.

To see that the code has dimension k first note that there are exactly $|\mathbb{F}|^k$ distinct elements $A(x)$ of $\mathbb{F}[x]$ of degree at most $k - 1$. It remains to verify that the evaluation of two distinct polynomials results in distinct codewords. Let $A(x)$ and $B(x)$ be two distinct polynomials of degree at most $k - 1$ and let $C(x) = A(x) - B(x)$. By construction $C(x)$ is a non-zero polynomial in $\mathbb{F}[x]$ of degree at most $k - 1$. If we had

$$(A(x_0), \dots, A(x_{n-1})) = (B(x_0), \dots, B(x_{n-1})),$$

then $(C(x_0), \dots, C(x_{n-1})) = 0$. But this cannot be: by the fundamental theorem of algebra the polynomial $C(x)$ can have at most $k - 1 < n$ zeros since it has degree at most $k - 1$. It follows that any non-zero codeword has weight at least $n - k + 1$. Therefore, $d(C) \geq n - k + 1$. But as discussed in Problem 1.10, the minimum distance of any code of length n and dimension k is upper bounded by $n - k + 1$. This bound is called the *Singleton bound*. We therefore conclude that $d(C) = n - k + 1$. ♣ : -)

SOLUTION 1.10 (SINGLETON BOUND). As mentioned in the hint, arrange the M codewords of length n in form of a $M \times n$ matrix and delete all but the first $\lfloor \log_{|\mathbb{F}|}(M) \rfloor$ columns. Assume first that $k = \log_{|\mathbb{F}|}(M)$ is an integer, so that $M = |\mathbb{F}|^k$. Now note that a matrix of width k over \mathbb{F} can contain at most $|\mathbb{F}|^k$ distinct rows. Therefore, at best the rows of the remaining columns are distinct and have therefore distance one. Take a pair of rows at distance one. At best, these rows can differ also in each of the chopped of $n - k$ columns, so that their total difference is upper bounded by $n - \log_{|\mathbb{F}|}(M) + 1$. If $\log_{|\mathbb{F}|}(M)$ is not an integer, then two rows in the remaining $\lfloor \log_{|\mathbb{F}|}(M) \rfloor$ columns must be the same and the minimum distance is therefore upper bounded by $n - \lfloor \log_{|\mathbb{F}|}(M) \rfloor \leq n - \log_{|\mathbb{F}|}(M) + 1$ as well. ♣ : -)

SOLUTION 1.11 (MAXIMUM DISTANCE SEPARABLE CODES). The first assertion is a simply corollary of the case of equality in the proof of the Singleton bound discussed in Problem 1.10. Arrange the $M = |\mathbb{F}|^k$ codewords in an $M \times n$ array. Take a subset $\mathcal{I} \subset [n]$ of cardinality k and delete all the columns which are not indexed by \mathcal{I} . Since we know that the code is MDS, it follows that all codewords are distinct on this subset. The claim follows since there are $|\mathbb{F}|^k$ codewords and an equal number to fill the k spots with elements from \mathbb{F} .

The dual code C^\perp has by assumption parameters $[n, n - k, d^\perp = k + 1]$. This means that any $d^\perp - 1 = k$ columns in the parity-check matrix of the dual code (which is a generator matrix of the code C) are linearly independent. In other words, they can be chosen arbitrarily and the remaining $n - k$ positions can then be filled in (in a unique way) to form a codeword of C . This means that any k positions of the code C form an information set. $\clubsuit : -)$

SOLUTION 1.12 (HAMMING BOUND). Center spheres of radius $\lfloor \frac{d-1}{2} \rfloor$ around the M codewords. Each sphere contains $\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i} (|\mathbb{F}| - 1)^i$ points. The bound now follows by noting that by definition of the minimum distance d , these spheres are disjoint and that in the total space there are $|\mathbb{F}|^n$ points. $\clubsuit : -)$

SOLUTION 1.13 (GILBERT-VARSHAMOV BOUND). Consider the following greedy code construction. Start with an arbitrary point of \mathbb{F} . Since we want to construct a code of minimum distance at least d , delete from the space the sphere of radius $d - 1$ centered on this chosen codeword. If there is a point remaining in the space, pick any such element. Again remove all points at distance $d - 1$ or less from this chosen point. Continue in this fashion until no point is left. The bound now follows by noting that at any step (any time we add a codeword) the number of points removed is at most $\sum_{i=0}^{d-1} \binom{n}{i} (|\mathbb{F}| - 1)^i$. $\clubsuit : -)$

SOLUTION 1.14 (GREEDY CODE SEARCH ALGORITHM). By construction, the minimum distance is at least d . The lower bound on the cardinality follows from the observation that M will be minimum when all the spheres of radius $d - 1$ around each codeword are disjoint. $\clubsuit : -)$

SOLUTION 1.15 (ASYMPTOTIC GILBERT-VARSHAMOV BOUND). If we start with the Gilbert-Varshamov bound discussed in Problem 1.13 and set $M = e^{\lfloor nr \rfloor}$ we get

$$e^{\lfloor nr \rfloor} = \frac{2^n}{\sum_{i=0}^{d-1} \binom{n}{i}} \stackrel{\text{Problem 1.25}}{\geq} \frac{2^n}{2^{nh_2((d-1)/n)}} = 2^{n(1-h_2((d-1)/n))}.$$

From this we conclude that

$$d/n \geq h_2^{-1} \left(1 - \frac{\lfloor nr \rfloor}{n} \right) + \frac{1}{n},$$

which implies the claim $\delta^*(r) \geq h_2^{-1}(1 - r)$. $\clubsuit : -)$

SOLUTION 1.16 (PAIRWISE INDEPENDENCE FOR GENERATOR ENSEMBLE). We have

$$\begin{aligned} \mathbb{P}\{X^{[j]} = v \mid X^{[i]} = x\} &= \mathbb{P}\{X^{[j]} = v \mid X^{[i]} = x = c + u^{[i]}G\} \\ &= \mathbb{P}\{(u^{[i]} + u^{[j]})G + x = v \mid X^{[i]} = x\} \end{aligned}$$

$$= \mathbb{P}\{(u^{[i]} + u^{[j]})G + x = v\} = 1/2^n.$$

Consider the one before last step. By assumption we pick a generator matrix G uniformly at random and (then) a shift c independently and uniformly at random from all shifts. From these two parameters we compute x . This defines the joint distribution of (G, c, x) .

Note that x has a uniform distribution as well. We get an equivalent model if we first pick x uniformly at random and (then) independent pick G uniformly at random. From these two parameters we then compute c .

If we consider this latter model, then it is clear that given x , G has a uniform distribution. ♣ : -)

SOLUTION 1.17 (MEAN AND SECOND MOMENT FOR $\mathcal{G}(n, k)$).

We start by considering $\mathbb{E}_C[A(C, w = 0)]$. Since the encoding operation is linear it follows that we always get the zero codeword if we encode the zero information word. On the other hand, if the information word is non-zero then with probability 2^{-n} the encoded word is zero. This is true since each position of the resulting codeword is a Bernoulli random variable with uniform probability. The stated formula now follows since we encode $2^{nr} - 1$ non-zero information words.

The general case $\mathbb{E}_C[A(C, w)]$ follows in a similar manner. Any time we encode one of the $2^{nr} - 1$ non-zero information words we have a chance of $\binom{n}{w} 2^{-n}$ of creating a word with Hamming weight w .

As a sanity check: if we compute $\sum_{w=0}^n \mathbb{E}_C[A(C, w)]$ we get 2^{nr} , as required.

Consider now the second moment. For $1 \leq w \leq n$, write

$$\begin{aligned} \mathbb{E}_C[A^2(C, w)] &= \mathbb{E}_C \left[\sum_{u, v \in \{0, 1\}^k} \mathbb{1}_{\{w(uG)=w\}} \mathbb{1}_{\{w(vG)=w\}} \right] \\ &= \bar{A}(w) + \mathbb{E}_C \left[\sum_{u \neq v \in \{0, 1\}^k} \mathbb{1}_{\{w(uG)=w\}} \mathbb{1}_{\{w(vG)=w\}} \right] \\ &= \bar{A}(w) + \mathbb{E}_C \left[\sum_{0 \neq u \neq v \in \{0, 1\}^k} \mathbb{1}_{\{w(uG)=w\}} \mathbb{1}_{\{w(vG)=w\}} \right]. \end{aligned}$$

The claim follows by noting that there are $(2^{nr} - 1)(2^{nr} - 2)$ distinct pairs of non-zero information vectors (u, v) , and that the probability that both u and v give rise to codewords of weight w is just the product of the probability that u gives rise to codewords of weight w and that v gives rise to codewords of weight w .

The final line (variance) follows in a straightforward manner from the previous two. ♣ : -)

SOLUTION 1.18 (MEAN AND SECOND MOMENT FOR $\mathcal{H}(n, k)$). It is shown in Problem 1.21 how to compute $\mathbb{E}_C[A(C, w)]$ for $w = 0$ as well as $w \geq 1$.

Consider now the second moment. For $1 \leq w \leq n$, write

$$\begin{aligned} \mathbb{E}_C[A^2(C, w)] &= \mathbb{E}_C \left[\sum_{u, v \in \{0,1\}^n: w(u)=w(v)=w} \mathbb{1}_{\{u \in C\}} \mathbb{1}_{\{v \in C\}} \right] \\ &= \bar{A}(w) + \mathbb{E}_C \left[\sum_{u \neq v \in \{0,1\}^n: w(u)=w(v)=w} \mathbb{1}_{\{u \in C\}} \mathbb{1}_{\{v \in C\}} \right]. \end{aligned}$$

The claim follows by noting that there are $\binom{n}{w}(\binom{n}{w} - 1)$ pairs (u, v) , $u \neq v$, where u and v are each of weight w and that for each such pair (u, v) the probability that both u and v are codewords is just the product of the probability that u is a codeword times the probability that v is a codeword. To see the last claim consider two words u and v , $u \neq v$, and both non-zero. In general these two words will share some positions in which both are 1 and each of the two will have some positions in which it takes on the value 1 but the other takes on the value 0. Finally, there are some positions in which both of them are 0. If we multiply the words with one row of H this partition shows that the two inner products have the form $\alpha + \mu$ and $\alpha + \nu$, respectively, where α corresponds to the partial sum of all positions where both take on the value 1 (this sum might be the empty sum) and μ and ν correspond to the “private” non-zero positions of each of the two words. Since the two words are distinct, and neither is the all-zero word, at least one of μ and ν is not the empty sum and either α is not the empty sum or both μ and ν correspond to non-empty sums. Since α , μ , and ν are by construction independent random variables, it follows that the event that $\alpha + \mu = 0$ is independent of the event that $\alpha + \nu$. Since the same argument applies for each row of H , the claim follows.

The expression for the variance follows in a straightforward manner from the mean and the second moment.

It remains to determine the expected number of codewords. We evaluate $\bar{A}(D)$ as given in Problem 1.21 for $D = 1$. We get

$$\bar{A}(D)|_{D=1} = 2^{-(n-k)}(1+D)^n + (1 - 2^{-(n-k)})|_{D=1} = 2^k + (1 - 2^{-(n-k)}).$$

This is slightly larger than the 2^k words that correspond to the design rate. ♣ : -)

SOLUTION 1.19 (WOLF TRELLIS). We start with the parity-check matrix

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

The codewords are

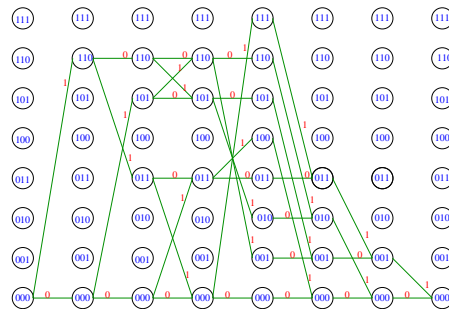


Figure 1.1: Wolf-trellis for the $(7, 4)$ Hamming code. Left to right reading of code-words from the trellis gives the bits in the sequence c_1, c_2, \dots

c_1	c_2	c_3	c_4	c_5	c_6	c_7
0	0	0	0	0	0	0
0	0	0	1	1	1	1
0	0	1	0	0	1	1
0	0	1	1	1	0	0
0	1	0	0	1	0	1
0	1	0	1	0	1	0
0	1	1	0	1	1	0
0	1	1	1	0	0	1
1	0	0	0	1	1	0
1	0	0	1	0	0	1
1	0	1	0	1	1	1
1	0	1	1	0	0	0
1	1	0	0	0	0	1
1	1	0	1	1	1	0
1	1	1	0	0	0	0
1	1	1	1	1	1	1

The associated Wolf trellis is shown in Figure 1.1.

♣ : -)

SOLUTION 1.20 (MACWILLIAMS IDENTITIES). For a code C , let us denote $k = \dim(C)$ and $k(E) = \dim(C(E))$.

1. Fix $i \in \{0, 1, \dots, n\}$. The number of codewords in C of weight i is A_i . A codeword of weight i , call it u , has $n - i$ zeros. Fix $w \in \{i, \dots, n\}$. The codeword u is counted $\binom{n-i}{n-w}$ times in the sum $\sum_{|E|=w} |C(E)|$. Therefore $\sum_{i=0}^n A_i \binom{n-i}{n-w} = \sum_{|E|=w} |C(E)|$.

2. Consider a subset E such that $|E| = w$. The all-zero codeword belongs to $C(E)$. Choose $u, v \in C(E)$, then $u + v$ has zeros outside E , i.e., $u + v \in C(E)$. In other words, $C(E)$ is a subspace of the vector space C .

Let $H(E)$ be a parity-check matrix of $C(E)$ obtained by completing a matrix H of C with $n - w$ independent vectors (having only one non-zero coordinate). We have $H(E)^T = [H^T, B^T]$ where B_E is the all-zero matrix and $B_{[n] \setminus E} = I_{n-w}$ is the $n - w$ identity matrix. The matrix $H(E)$ is block-wise triangular. Therefore row additions will transform the submatrix H without changing the rank of the set of its columns indexed by E . By performing so we can transform the matrix $H(E)$ into a block-wise diagonal matrix without changing its rank and keeping B unchanged. We finally get $\text{rank}(H(E)) = \text{rank}(H_E) + \text{rank}(I_{n-w})$, i.e., $k(E) = w - \text{rank}(H_E)$.

3. Consider a generator matrix of C built in the following special way. We start with a generator matrix for $C(E)$ and then append rows to complete it to a generator matrix of C . In a similar (dual) manner as in the previous question, we get $k = \text{rank}(G) = k(E) + \text{rank}(G_{\bar{E}})$. This result combined with the previous result shows that $w - \text{rank}(H_E) = k - \text{rank}(H_{\bar{E}})$.
4. The change of variable $v = n - u$ proves (i) because

$$\sum_{i=0}^{n-u} A_i^{\perp} \binom{n-i}{u} = \sum_{i=0}^v A_i^{\perp} \binom{n-i}{n-v} = \sum_{i=0}^n A_i^{\perp} \binom{n-i}{n-v} = \sum_{|E|=v=n-u} |C^{\perp}(E)|.$$

Next, (ii) is obtained from 2 which states that $|C^{\perp}| = 2^{|E| - \text{rank}(G_E)}$. Finally, 3 shows (iii), and 1 proves (iv).

5. Using the result from the previous question, we have

$$\underbrace{\sum_i A_i^{\perp} \binom{n-i}{u}}_{=S_u^{\perp}} = 2^{n-k-u} \underbrace{\sum_i A_i \binom{n-i}{n-u}}_{=S_u}.$$

Using the left-hand side of this equality, we have

$$\begin{aligned} \sum_u (-1)^u \binom{u}{n-j} S_u^{\perp} &= \sum_i A_i^{\perp} (-1)^i \left(\sum_u (-1)^{u-i} \binom{n-i}{u} \binom{u}{n-j} \right) \\ &= \sum_i A_i^{\perp} (-1)^i (-1)^n \underbrace{\sum_l \left((-1)^{l-i} \binom{n-i}{n-l} \binom{n-l}{n-j} \right)}_{\delta_{i,j}} \end{aligned}$$

$$(1.2) \quad = A_j^\perp (-1)^{n+j}.$$

Using the right-hand side of this equality, we have

$$(1.3) \quad \begin{aligned} \sum_u (-1)^u \binom{u}{n-j} S_u &= 2^{-k} \sum_u 2^{n-u} (-1)^u \sum_i A_i \binom{n-i}{n-u} \binom{u}{n-j} \\ &= \frac{(-1)^{n+j}}{|C|} \sum_i A_i \underbrace{\sum_u 2^{n-u} (-1)^{n+j-u} \binom{n-i}{n-u} \binom{u}{n-j}}_{=Q_j(i)}. \end{aligned}$$

We claim that $Q_j(i)$ is in fact a Krawtchouk polynomial, i.e., we claim that $Q_j(i) = P_j(i) = \sum_{l=0}^n (-1)^l \binom{i}{j-l} \binom{n-i}{l}$. If we assume this for a moment, then we see from (1.2) and (1.3) that

$$A_j^\perp = \frac{1}{|C|} \sum_{i=0}^n A_i P_j(i), \quad 0 \leq j \leq n.$$

It remains to prove our claim. To see it, observe first that

$$\begin{aligned} (1+x)^{n-i} (1-x)^i &= \sum_{j=0}^{n-i} \sum_{l=0}^i (-1)^l \binom{n-i}{j} \binom{i}{l} x^{j+l} \\ &\stackrel{j'=j+l}{=} \sum_l \sum_{j'} (-1)^l \binom{n-i}{j'-l} \binom{i}{l} x^{j'} = \sum_j P_j(i) x^j, \end{aligned}$$

and second that

$$\begin{aligned} (1-x)^n \left(1 + \frac{2x}{1-x}\right)^{n-i} &= \sum_{m=0}^{n-i} \binom{n-i}{m} 2^m x^m (1-x)^{n-m} \\ &= \sum_m \sum_k \binom{n-i}{m} \binom{n-m}{k} 2^m (-1)^k x^{m+k} \\ &\stackrel{j=k+m}{=} \sum_m \sum_j \binom{n-i}{m} \binom{n-m}{j-m} 2^m (-1)^{j-m} x^j \\ &\stackrel{u=n-m}{=} \sum_u \sum_j \binom{n-i}{n-u} \binom{u}{j-n+u} 2^{n-u} (-1)^{n-u+j} x^j \\ &= \sum_j Q_j(i) x^j. \end{aligned}$$

Finally, we use the identity $(1+x)^{n-i} (1-x)^i = (1-x)^n \left(1 + \frac{2x}{1-x}\right)^{n-i}$ to get $Q_j(i) = P_j(i)$.

6. The $[7, 4, 3]$ Hamming code has 1 all-zero codeword, 7 codewords of weight 3, 7 codewords of weight 4, and 1 all-one codeword, i.e., $A_0 = A_7 = 1$, $A_3 = A_4 = 7$, and $A_1 = A_2 = A_5 = A_6 = 0$. We can use the previous result to obtain that its dual, the $[7, 3, 4]$ Simplex code, has $A_0^\perp = 1$, $A_4^\perp = 7$, and $A_j^\perp = 0$ for $j \neq 0, 4$.

♣

SOLUTION 1.21 (UPPER BOUND ON ERROR PROBABILITY VIA WEIGHT DISTRIBUTION). By definition of the ensemble $\mathcal{H}(n, k)$, every code in the ensemble consists of the solution space of the set of equations $Hx^T = 0^T$ for some random parity-check matrix H . Hence, every code C in $\mathcal{H}(n, k)$ contains the all-zero word exactly once. It follows that $\mathbb{E}[A_0(H)] = 1$.

The chance that a fixed non-zero word fulfills a random parity-check constraint is equal to one-half. Since the rows of the parity-check matrix H are chosen independently and there are by definition exactly $n - k$ rows, it follows that the probability that a fixed non-zero word fulfills all parity-check equations equals $2^{-(n-k)}$. Since there are $\binom{n}{w}$ words of weight w we conclude that $\mathbb{E}[A_w(H)] = \binom{n}{w} 2^{-(n-k)}$.

We get

$$\begin{aligned} \bar{A}(D) &= 1 + 2^{-(n-k)} \sum_{w=1}^n \binom{n}{w} D^w \\ &= 1 + 2^{-(n-k)} ((1 + D)^n - 1) = 2^{-(n-k)} (1 + D)^n + (1 - 2^{-(n-k)}). \end{aligned}$$

Now consider the sequence of inequalities. Because of symmetry, the conditional error probability does not depend on the transmitted codeword. We are therefore free to assume that $X = 0$, i.e., that the transmitted codeword is the all-zero one. The first inequality is the union bound in which we upper bound the probability of a union of events by the sum of the individual probabilities. Since we are dealing with the Gaussian case for which the likelihoods are monotonically decreasing functions of the distance, instead of looking at $p(Y|x) \geq p(Y|0)$ we can compare $|x - Y|^2 \leq |Y|^2$. From the problem description we know that the all-zero word gets mapped to the vector of length n containing one in each component. If we consider any other codeword x of Hamming weight $w(x)$ then the indicated mapping, maps x into a vector that has Euclidean distance $d_E = 2\sqrt{w(x)}$ from the all-one vector. In Gaussian noise of variance σ^2 , the probability that we confuse two points that have Euclidean distance d_E is equal to $Q(d_E/(2\sigma))$. This explains the term $Q(\sqrt{w(x)}/\sigma)$. Now we use the standard upper bound $Q(x) \leq \frac{1}{2}e^{-x^2/2}$, $x \geq 0$.
♣ :-)

SOLUTION 1.22 (SUFFICIENT STATISTIC). Observe that equivalently we can show that $p_{X,Y}(x, y)$ can be brought into the form $a(x, z)b(y)$.

First we show that if $Z = f(Y)$ and if Z constitutes a sufficient statistic for X given Y then $p_{X,Y}(x, y)$ can be brought into the form $a(x, z)b(y)$. Write $p_{X,Y}(x, y) = \sum_z p_{X,Y,Z}(x, y, z) = \sum_z p_{X,Z}(x, z)p_{Y|Z}(y|z)$, where in the second step we have used the fact that $X \rightarrow Z \rightarrow Y$. Since Z is a function of Y we can write this further as $p_{X,Y}(x, y) = p_{X,Z}(x, f(y))p_{Y|Z}(y|f(y))$, which has the desired form $a(x, z)b(y)$.

On the other hand, assume that $p_{X,Y}(x, y)$ can be brought into the form $a(x, z)b(y)$ where $z = f(y)$. Write

$$p_{X,Y,Z}(x, y, z) = p_{X,Y}(x, y)\delta(z = f(y)) = a(x, z)b(y)\delta(z = f(y)).$$

If we marginalize out Y we get

$$p_{X,Z}(x, z) = a(x, z) \sum_y b(y)\delta(z = f(y)),$$

and if we marginalize out X we get

$$p_{Y,Z}(y, z) = b(y)\delta(z = f(y)) \sum_x a(x, z).$$

If we marginalize out X and Y we get

$$p_Z(z) = \left(\sum_x a(x, z)\right)\left(\sum_y b(y)\delta(z = f(y))\right).$$

Therefore,

$$\begin{aligned} p_{X,Y,Z}(x, y, z) &= a(x, z)b(y)\delta(z = f(y)) \\ &= \frac{p_{X,Z}(x, z)}{\sum_y b(y)\delta(z = f(y))} \frac{p_{Y,Z}(y, z)}{\sum_x a(x, z)} \\ &= p_{X,Z}(x, z)p_{Y|Z}(y|z). \quad \clubsuit : -) \end{aligned}$$

SOLUTION 1.23 (MORE ON SUFFICIENT STATISTIC). We know that (i) $X \rightarrow Y \rightarrow Z$, and that (ii) $H(X|Y) = H(X|Z)$. We have

$$\begin{aligned} H(Y|X, Z) &= H(X, Y, Z) - H(X, Z) \\ &\stackrel{(i)}{=} H(X, Y) + H(Z|X, Y) - H(X, Z) \\ &= H(X, Y) + H(Z|Y) - H(X, Z) \\ &= H(X, Y) + H(Z, Y) - H(Y) - H(X, Z) \\ &= H(Y) + H(X|Y) + H(Z, Y) - H(Y) - H(Z) - H(X|Z) \\ &\stackrel{(ii)}{=} H(Y|Z). \end{aligned}$$

Therefore, $I(X; Y|Z) = H(Y|Z) - H(Y|X, Z) = 0$, which implies that $X \rightarrow Z \rightarrow Y$.

♣

SOLUTION 1.24 (BOUND ON BINOMIALS). The first step is just the definition of $\binom{m}{k}$. Now we write

$$\frac{k!}{m(m-1)\dots(m-k+1)} = \frac{k!}{m^k \prod_{i=1}^{k-1} \frac{m-i}{m}} = \frac{k!}{m^k} e^{-\sum_{i=1}^{k-1} \ln(1-i/m)} \leq \frac{k!}{m^k} e^{k^2/m}.$$

The last step needs some justification. We need to show that $-\sum_{i=1}^{k-1} \ln(1-i/m) \leq k^2/m$. Let $z = k/m$, $z \in [0, 1)$. Note that $-\log(1-x)$ is an increasing function in x for $x \in [0, 1)$. We can therefore bound the sum by an integral:

$$\begin{aligned} -\sum_{i=1}^{k-1} \ln(1-i/m) &\leq -m \int_0^z \log(1-x) dx \\ &= m(z + (1-z)\ln(1-z)) \leq mz^2 = k^2/m. \end{aligned}$$

In the one before the last step we have used the fact that $(1-z)\ln(1-z) \leq -z(1-z)$. This is true since $-\ln(1-z) = \sum_{i \geq 1} \frac{z^i}{i}$, which implies $\ln(1-z) \leq -z$. ♣ :-)

SOLUTION 1.25 (BOUND ON SUM OF BINOMIALS). Define the polynomial $p(x) = \sum_{k=0}^n \binom{n}{k} x^k = (1+x)^n$. For any $0 < x \leq 1$

$$\sum_{k=0}^m \binom{n}{k} \leq \left(\sum_{k=0}^n \binom{n}{k} x^k \right) / x^m = p(x) / x^m.$$

We are free to choose x in the range $(0, 1]$ to get a tight bound. We choose $x = m/(n-m)$. Since by assumption $m \leq n/2$ we have indeed $x \in (0, 1]$. We get

$$\sum_{k=0}^m \binom{n}{k} \leq (n/(n-m))^n ((n-m)/m)^m = \left(\frac{n-m}{n} \right)^{-(n-m)} (m/n)^{-m} = 2^{nh_2(m/n)}.$$

♣ :-)

SOLUTION 1.26 (CHAIN RULE).

$$\begin{aligned} H(X, Y) &= -\sum_{x,y} p_{X,Y}(x, y) \log(p_X(x) p_{Y|X}(y|x)) \\ &= -\sum_{x,y} p_{X,Y}(x, y) \log p_X(x) - \sum_{x,y} p_{X,Y}(x, y) \log p_{Y|X}(y|x) \\ &= -\sum_x p_X(x) \log p_X(x) - \sum_x p_X(x) \sum_y p_{Y|X}(y|x) \log p_{Y|X}(y|x) \\ &= H(X) + \sum_x p_X(x) H(Y|X=x) = H(X) + H(Y|X). \end{aligned} \quad \clubsuit :-)$$

SOLUTION 1.27 (NON-NEGATIVITY OF MUTUAL INFORMATION).

$$\begin{aligned}
 -I(X; Y) &= \sum_{x,y} p_{X,Y}(x, y) \log \frac{p_X(x)p_Y(y)}{p_{X,Y}(x, y)} \\
 &\stackrel{(1.61)}{\leq} \log \left(\sum_{x,y} p_{X,Y}(x, y) \frac{p_X(x)p_Y(y)}{p_{X,Y}(x, y)} \right) \\
 &= 0. \quad \clubsuit : -)
 \end{aligned}$$

SOLUTION 1.28 (FANO'S INEQUALITY). As indicated in the hint,

$$H(E, X | Y) = H(X | Y) + H(E | X, Y) = H(E | Y) + H(X | E, Y).$$

Note that E is a function of X and Y so that $H(E | X, Y) = 0$. Further, $H(X | E, Y) \leq \mathbb{P} \log(|\mathcal{X}| - 1)$ since $H(X | E = 0, Y) = 0$, i.e., if there is no error then there is no uncertainty about X , but if $E = 1$, then X can take on at most $|\mathcal{X}| - 1$ values and the conditional probability distribution on these values is at worst the uniform one. Finally, $H(E | Y) \leq H(E) = H(\mathbb{P})$, where the first step is due to the fact that conditioning can only decrease entropy and the last step follows since E is binary. $\clubsuit : -)$

SOLUTION 1.29 (THE CAPACITY OF THE BSC REDERIVED). We have

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y | X) = H(Y) - \sum_x p_X(x) H(Y | X = x) \\
 &= H(Y) - \sum_x p_X(x) h_2(\epsilon) = H(Y) - h(p) \leq 1 - h_2(\epsilon).
 \end{aligned}$$

We get equality if we choose $p_X(x)$ to be uniform. $\clubsuit : -)$

SOLUTION 1.30 (DESCARTE'S RULES OF SIGNS). Consider the case $r = 0$. By the fundamental theorem of algebra we can represent $p(x)$ in the factorized form

$$p(x) = \alpha x^\beta \prod_{i=\beta}^d (x - \xi_i),$$

where the ξ_i are the non-zero (complex-valued) roots. Note that the first and last non-zero coefficient of $p(x)$ are p_β and p_d , respectively, and they have representation equal to

$$p_\beta = \alpha (-1)^{d-\beta+1} \prod_{i=\beta}^d \xi_i, \quad p_d = \alpha.$$

Now recall that because of our assumption, the non-zero roots of $p(x)$ are either negative or they appear as complex-conjugate pairs. From this and the above representation it follows that p_β and p_d have the same sign. A moments thought then

shows that the number of sign changes must be even. This proves the claim for the case $r = 0$, and this case will serve as our anchor.

Assume now that the claim is true for some $r \geq 0$ and consider a polynomial $p(x)$ with $r+1$ positive roots. Write $p(x)$ as $p(x) = (x-\xi)\tilde{p}(x)$, where ξ is positive. By assumption $\tilde{p}(x)$ has r positive roots and $r+2k$ sign changes for some $k \geq 0$. Explicitly, we have

$$p_i = \tilde{p}_{i-1} - \xi\tilde{p}_i, \quad i = 0, \dots, d+1,$$

if we define $p_{-1} = p_{d+1} = 0$. Consider now the polynomial $p(x)$. If we multiply the i -th coefficient of $p(x)$ with the positive factor ξ^{i-1} , then clearly the number of sign changes remains unaltered. Therefore, $p(x)$ has the same number of sign changes as $\sum_i (\xi^{i-1}\tilde{p}_{i-1} - \xi^i\tilde{p}_i) x^i$, which in turn has the same number of sign changes as

$$\sum_i (\xi^i\tilde{p}_i - \xi^{i-1}\tilde{p}_{i-1}) x^i = \tilde{p}(\xi x)(1-x).$$

We claim that for a general polynomial $p(x)$ the number of sign changes of $p(x)(1-x)$ always differs by an odd number from the number of sign changes of $p(x)$. This will settle the claim. We use again induction, this time on the degree of $p(x)$. The claim is trivially fulfilled for any non-zero degree-zero polynomial. Let us also check the case $d = 1$. Without loss of generality we can eliminate multiples of x . Assume therefore first that $p(x) = p_0 + p_1x$ with $p_0 \neq 0$, so that $p(x)(1-x) = p_0 + (p_1 - p_0)x - p_1x^2$. Since we are only interested in sign *changes* we can assume without loss of generality that $p_0 > 0$. An explicit check then shows that the claim is true for both $p_1 > 1$ as well as $p_1 < 0$. Assume therefore that the claim is true for any polynomial of degree up to d , where d is some natural number. Consider a polynomial of degree $d+1$. It has the form $\sum_{i=0}^{d+1} p_i x^i$, where $p_{d-1} \neq 0$.

Assume first that $p_d p_{d+1} > 0$, i.e., they have the same sign. Without loss of generality assume that they are both positive. In this case, if we eliminate the highest term of $p(x)$ and call it $\tilde{p}(x)$, then $p(x)$ and $\tilde{p}(x)$ have the same number of sign changes. If we compare $\tilde{p}(x)(1-x)$ with $p(x)(1-x)$ then the second has the final two terms $p_{d+1} - p_d$ and $-p_{d+1}$, whereas, $\tilde{p}(x)(1-x)$ has the final term $-p_d$. In both cases the final term is negative. If $p_{d+1} - p_d < 0$ then the number of sign changes is the same in both cases, if $p_{d+1} - p_d > 0$ then it differs by two. In either case the claim is fulfilled since by induction on the degree the statement is fulfilled for $\tilde{p}(x)(1-x)$. The case $p_d p_{d+1} > 0$ can be treated in a similar way. ♣ :-)