Statistical Information Theory Coursework

Hamming Codes

Belfiore Asia, *CID*: 02129867

MSc Advanced Computing
Department of Computing, Imperial College London

24 November 2024

Task 2c

1) Decoding Accuracy for Noisy Channels

The decoding accuracy of a Hamming Code with m parity bits for any codeword t sent through a binary symmetric channel with noise level p is the probability of correctly decoding the string. Since the Hamming Code can detect and correct up to 1 bit error in each codeword, the probability of a codeword being correctly decoded can be expressed as the probability (for that string) of having at most one bit error:

$$Accuracy_t = p(at\ most\ 1\ bit\ error) = p(no\ error) + p(1\ bit\ error)$$

For any number of parity bits m, any codeword t resulting from the Hamming encoding with m parity bits has length $n = 2^m - 1$ bits. If the channel is completely noisy, i.e. the probability of flipping each bit in the codeword is equal to the probability of not flipping, $p = 1 - p = 0.5$. Then we have:

$$p(0\ bit\ error) = (1-p)^n = 0.5^n$$

i.e. the probability of having 0 bit error $(1-p)$ for every (n) bit in the string; similarly:

$$p(1\ bit\ error) = np(1-p)^{n-1} = n(0.5)(0.5)^{n-1} = n(0.5)^n$$

i.e. the probability of having no error $(1-p)$ in $n-1$ bits in the string, and one-bit error p anywhere in the string (n possible combinations). Thus, the probability of having at most one bit of error in any string decoded is:

$$p(at\ most\ 1\ bit\ error) = (1-p)^n + np(1-p)^{n-1} = (0.5)^n + n(0.5)^n = (0.5)^n(n+1)$$

Since $n = 2^m - 1$ we can substitute it in the formula above to get:

$$Accuracy_t = (0.5)^{2^m-1}(2^m - 1 + 1) = 2^m \left(\frac{1}{2}\right)^{2^m-1} = \frac{2^m}{2^{2^m-1}}$$

For different parity bits m, the accuracy of a Hamming Code for a channel with noise level $p = 0.5$ is:

a) m=2, gives $n = 2^m - 1 = 3$ and $k = n - m = 1$, i.e. Hamming(3,1) with accuracy:

$$Accuracy = \frac{2^2}{2^{2^2-1}} = \frac{4}{2^3} = \frac{4}{8} = \frac{1}{2} = 0.5$$

b) m=3, gives $n = 2^m - 1 = 7$ and $k = n - m = 4$, i.e. Hamming(7,4) with accuracy:

$$Accuracy = \frac{2^3}{2^{2^3-1}} = \frac{8}{2^7} = \frac{8}{128} \approx 0.0625$$

a) m=4, gives $n = 2^m - 1 = 15$ and $k = n - m = 11$, i.e. Hamming(15,11) with accuracy:

$$Accuracy = \frac{2^4}{2^{2^4-1}} = \frac{16}{2^{15}} = \frac{16}{32768} = \frac{1}{2048} \approx 0.000488$$

It follows that the accuracy of Hamming Codes for noisy channels is fully dependent on the length of the codeword to be decoded (and thus on the number of parity bits used by the code). As the number of parity bits added to the source string grows, the decoding accuracy rapidly vanishes to 0 (for $m > 2$). This makes intuitive sense since every bit has a 50% probability of being flipped, thus having more bits means having a higher chance of 2 or more bits being flipped. In particular, for any $m > 2$ the probability of block error for $p = 0.5$, is greater than:

$$p(block\ error) = 1 - p(at\ most\ 1\ bit\ error) > 1 - \frac{2^m}{2^{2^m-1}} = 1 - 0.5 = 0.5$$

meaning that a Hamming Code with any $m > 2$ will have a higher probability of *not detecting* (and thus not correctly decoding) any error in a codeword than of correctly decoding it. Figure 1 below shows the validity of the above calculations, comparing them to actual accuracies of simulated Haming encodings and decodings (with different parity bits) on 1000 random strings (for a symmetric channel with $p = 0.5$).

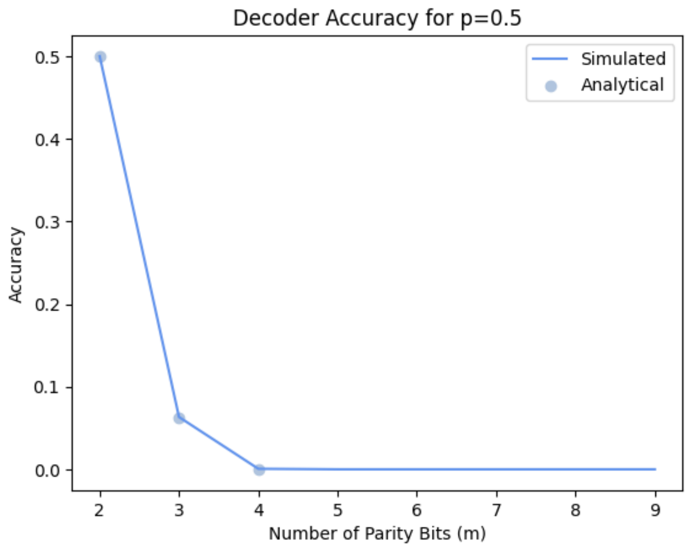


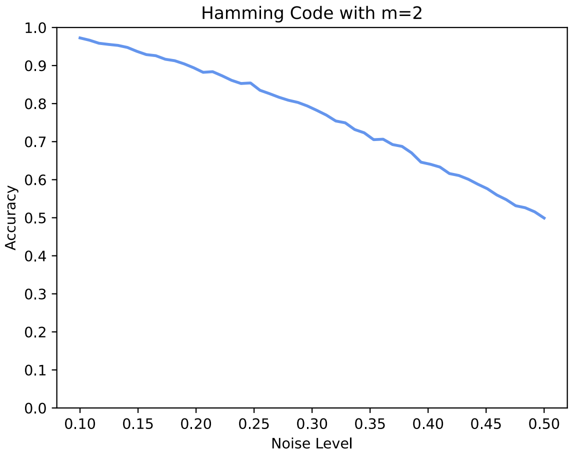
Figure 1: Hamming Code accuracies for different numbers of parity bits (m) for a completely noisy channel. The solid line represents the computer simulated accuracies, the dots represent the above manually calculated accuracies for $m = 2,3,4$.

2) Noise Level and Decoder Accuracy

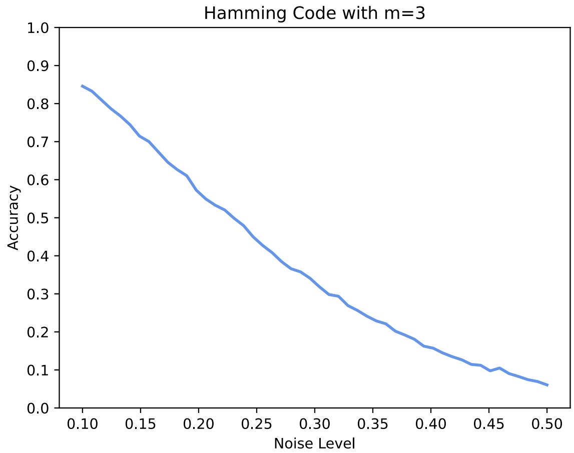
The accuracy of a decoder can also be determined by simulating the transmission of a large number of strings (eg. 1000, as done in Figure 1) through a noisy channel and checking how many of these are correctly decoded over the overall number of transmitted strings:

$$Accuracy = \frac{\#\ correctly\ decoded\ strings}{\#\ number\ of\ transmitted\ strings}$$

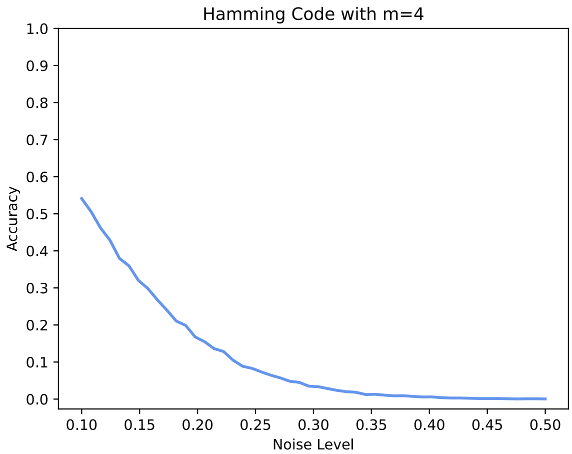
For different numbers of parity bits ($m = 2,3,4$), the accuracy of the resulting Hamming Codes are shown in Figure 2 (respectively in Figure 2a, 2b and 2c) for different noise levels between 0 and 0.5, simulated for 10000 random strings:



a) Hamming(3,1) accuracy with different noise levels.



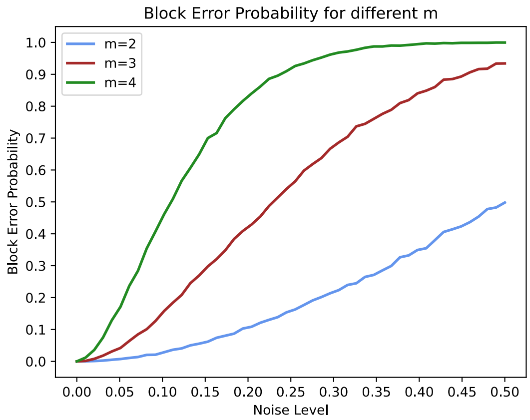
b) Hamming(7,4) accuracy with different noise levels.



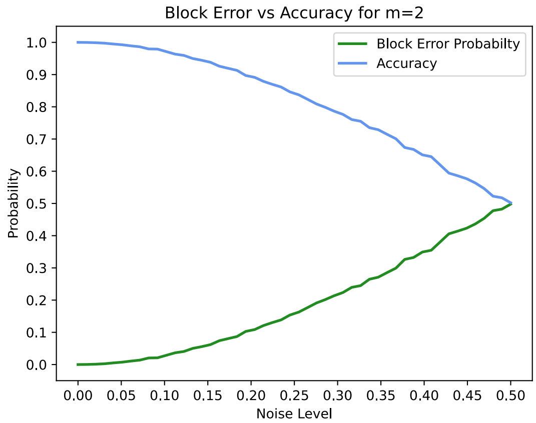
c) Hamming(15,11) accuracy with different noise levels.

Figure 2: Hamming Code accuracies for different Noise Levels and number of parity bits (m), simulated for 10000 random codewords.

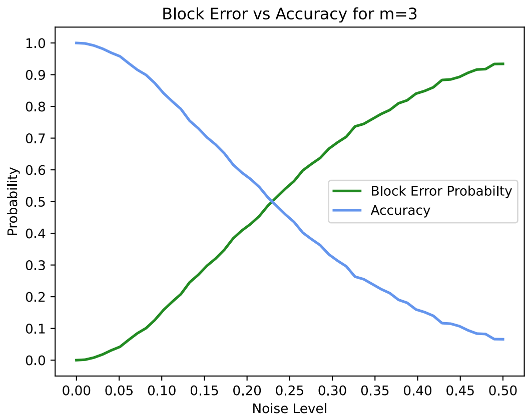
This confirms the above results in Section 1 and further expands the observations made for completely noisy channels to channels with noise levels different from 0.5 (but greater than 0). As the number of parity (and data) bits increases, the accuracy of the Hamming Code quickly drops, even for low noise levels. This is a direct consequence of the fact that the accuracy of the decoder is inversely proportional to the number of bits sent through the channel: the more information is sent (higher bit content), the less accurate the decoding will be because the probability of block error (i.e. getting more than one-bit error) grows exponentially with the length of the codeword, thus quickly exceeding the amount of bit error that the Hamming Code can detect and correct. This means that any Hamming Code is a perfect decoder only in situations where each codeword has at most 1 bit error, but becomes quickly inefficient when the transmission channels are noisy. Then the code's accuracy becomes quickly unsatisfactory as $m \geq 3$, even for noise levels close to 0, because the probability of getting more bit errors (≥ 2) than the amount that is detectable and correctable by the code itself (1 bit) grows alongside m (Figure 3).



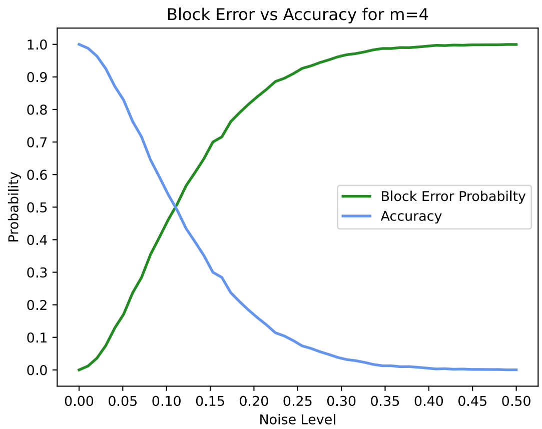
a) Probability of Block Error for different noise levels and parity bits.



b) Rates of Block Error vs Accuracy for m=2



c) Rates of Block Error vs Accuracy for m=3



d) Rates of Block Error vs Accuracy for m=4

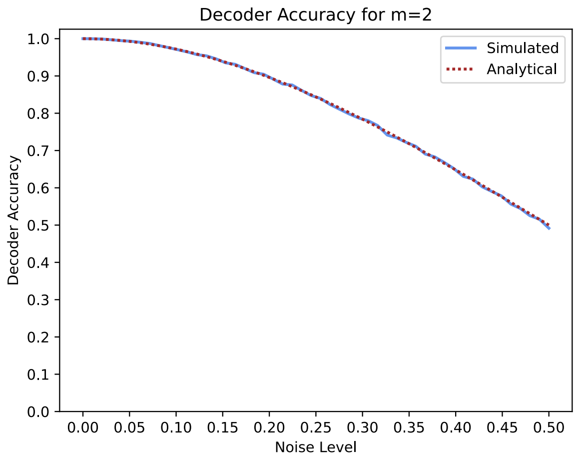
Figure 3: Probability of Block Error for different noise levels and parity bits (a) and comparison of the Code Accuracy and Probability of having more than 1 bit error with different parity bits (b),(c),(d). Simulated for 10000 random codewords and with different noise levels.

3) Analytical and Simulated Accuracy*

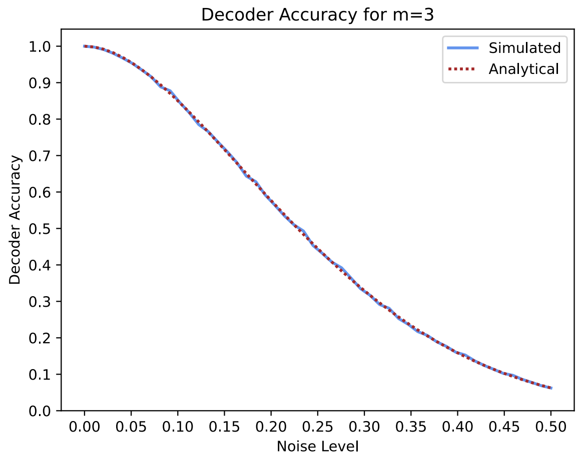
As further proof of the validity of the derivations in Section 1, Figure 4 shows the relationship between the simulated accuracies and the analytically calculated ones with different parity bits and noise levels. To manually calculate the code accuracy for channels with noise levels $p \neq 0.5$ I used the general formula from Section 1:

$$p(at\ most\ 1\ bit\ error) = (1-p)^n + np(1-p)^{n-1}$$

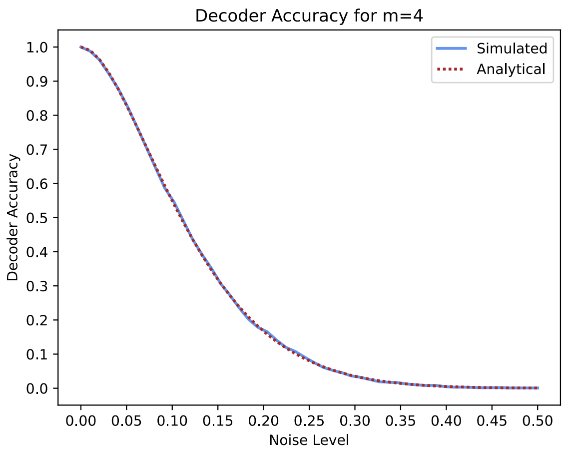
which, as shown below, correctly approximates the actual accuracy of the code.



a) Hamming(3,1) accuracy with different noise levels.



b) Hamming(7,4) accuracy with different noise levels.



c) Hamming(15,11) accuracy with different noise levels.

Figure 4: Comparing of Analytical and simulated Hamming Code accuracies for different numbers of parity bits (m) for various noise levels. The blue solid line represents the computer simulated accuracies for 10000 random strings, the dotted red line represent the analytically calculated accuracies with the above formula derived in Section 1.