Exam (January 13, 2018) Solution

Last modified January 19, 2018

Problem 1 - Source Coding and Channel Capacity [15 points]

Part I

Since X and Y are independent, $\mathsf{H}(X|Y) = \mathsf{H}(X)$, $\mathsf{H}(Y|X) = \mathsf{H}(Y)$, and $\mathsf{H}(X,Y) = \mathsf{H}(X) + \mathsf{H}(Y)$. The entropy of X is

$$\mathsf{H}(X) = \sum_{i=1}^{8} p_i \log_2\left(\frac{1}{p_i}\right) = \log_2(8) = 3 \text{ bits.}$$

The entropy of Y is

$$\begin{aligned}
\mathsf{H}(Y) &= \sum_{k=1}^{\infty} 2^{-k} \log_2(2^k) \\
&= \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k \\
&= \sum_{k=0}^{\infty} k \left(\frac{1}{2}\right)^k \\
&= \frac{0.5}{(1-0.5)^2} \\
&= 2 \text{ bits.}
\end{aligned}$$

Finally, the joint entropy of X and Y is

$$\mathsf{H}(X,Y) = \mathsf{H}(X) + \mathsf{H}(Y) = 2 + 3 = 5 \text{ bits.}$$

Part II

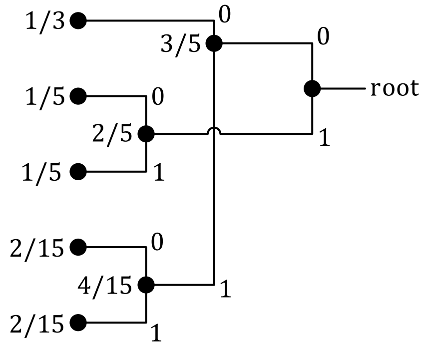
1. The source entropy is

$$\mathsf{H}(P) = \sum_{i=1}^{5} p_i \log_2\left(\frac{1}{p_i}\right) = 2.232 \text{ bits.}$$

2. One possibility is

symbol	codeword	probability
x_1	00	1/3
x_2	10	1/5
x_3	11	1/5
x_4	010	2/15
x_5	011	2/15

where the corresponding tree is depicted below.



3. The average codeword length is

$$\bar{L} = 1/3 \cdot 2 + 1/5 \cdot 2 + 1/5 \cdot 3 + 2/15 \cdot 3 + 2/15 \cdot 3 = 34/15 \approx 2.27 \text{ bits}.$$

Therefore, the efficiency is

$$\eta = \frac{\mathsf{H}(P)}{\bar{L}} = \frac{2.232}{34/15} = 98.47\%.$$

4. The average codeword length for the code from question 2, using P', is

$$1/5 \cdot 2 + 1/5 \cdot 2 + 1/5 \cdot 2 + 1/5 \cdot 3 + 1/5 \cdot 3 = 12/5.$$

If we construct an optimal Huffman code for this distribution we also end up with a code that has average codeword length of 12/5. Hence, the code obtained in question 2 is optimal for P'.

Part III

1. The entropy of the source is

$$\mathsf{H}(X) = -0.5 \log_2(0.5) - 0.5 \log_2(0.5) = 1.$$

The probability distribution of the output is

$$\begin{aligned}
P_Y(0) &= P_{Y|X}(0|0)P_X(0) + P_{Y|X}(0|1)P_X(1) \\
&= 0.5(1-\varepsilon) + 0.5\varepsilon \\
&= 0.5, \\
P_Y(1) &= P_{Y|X}(1|0)P_X(0) + P_{Y|X}(1|1)P_X(1) \\
&= 0.5\varepsilon + 0.5(1-\varepsilon) \\
&= 0.5, \\
P_Y(2) &= 0.
\end{aligned}$$

Therefore, the entropy of the output distribution is

$$\mathsf{H}(Y) = -P_Y(0)\log_2(P_Y(0)) - P_Y(1)\log_2(P_Y(1)) = 1.$$

2. The product rule gives

$$P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x).$$

Therefore, the joint probability distribution of the source and the output is

$$\begin{aligned}
P_{X,Y}(0,0) &= P_{Y|X}(0|0)P_X(0) = 0.5(1-\varepsilon) \\
P_{X,Y}(1,0) &= P_{Y|X}(0|1)P_X(1) = 0.5\varepsilon \\
P_{X,Y}(0,1) &= P_{Y|X}(1|0)P_X(0) = 0.5\varepsilon \\
P_{X,Y}(1,1) &= P_{Y|X}(1|1)P_X(1) = 0.5(1-\varepsilon) \\
P_{X,Y}(0,2) &= P_{Y|X}(2|0)P_X(0) = 0 \\
P_{X,Y}(1,2) &= P_{Y|X}(2|1)P_X(1) = 0.
\end{aligned}$$

The joint entropy is then computed as

$$\begin{aligned}
\mathsf{H}(X,Y) &= -P_{X,Y}(0,0)\log_2(P_{X,Y}(0,0)) - P_{X,Y}(0,1)\log_2(P_{X,Y}(0,1)) \\
&\quad - P_{X,Y}(1,0)\log_2(P_{X,Y}(1,0)) - P_{X,Y}(1,1)\log_2(P_{X,Y}(1,1)) \\
&= -0.5(1-\varepsilon)\log_2(0.5(1-\varepsilon)) - 0.5\varepsilon\log_2(0.5\varepsilon) - 0.5\varepsilon\log_2(0.5\varepsilon) - 0.5(1-\varepsilon)\log_2(0.5(1-\varepsilon)) \\
&= -(1-\varepsilon)\log_2(1-\varepsilon) - (1-\varepsilon)\log_2(0.5) - \varepsilon\log_2(\varepsilon) - \varepsilon\log_2(0.5) \\
&= 1 + \mathsf{H}_\mathsf{b}(\varepsilon).
\end{aligned}$$

3. The joint entropy, $\mathsf{H}(X,Y)$, can be rewritten as

$$\mathsf{H}(X,Y) = \mathsf{H}(X|Y) + \mathsf{H}(X).$$

Therefore, the mutual information can be computed using $\mathsf{H}(X)$, $\mathsf{H}(Y)$, and $\mathsf{H}(X,Y)$ as

$$\begin{aligned}\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\ &= \mathsf{H}(X) + \mathsf{H}(Y) - \mathsf{H}(X,Y) \\ &= 1 + 1 - 1 - \mathsf{H}_\mathsf{b}(\varepsilon) \\ &= 1 - \mathsf{H}_\mathsf{b}(\varepsilon).\end{aligned}$$

4. For $\varepsilon = 0$ or $\varepsilon = 1$, $\mathsf{H}_\mathsf{b}(\varepsilon) = 0$. In that case, $\mathsf{I}(X;Y)$ is maximal and the mutual information is 1 bit.

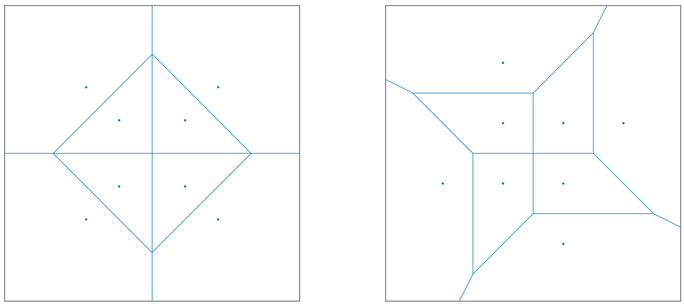
5. For $\varepsilon = 0.5$, $\mathsf{H}_\mathsf{b}(\varepsilon) = 1$. In that case, $\mathsf{I}(X;Y)$ is minimal and the mutual information is 0.

Problem 2 - Signal Constellations and Detection [15 points]

1. For Ω_1, we have 4 points with energy a^2 and 4 points with energy $(2a)^2 = 4a^2$. The average comes out to $E_s = 20a^2/8$ and therefore $a = \sqrt{2/5}$.

   For Ω_2, we have 4 points with energy $b^2 + b^2 = 2b^2$ and 4 points with energy $b^2 + 9b^2 = 10b^2$. The average comes out to $E_s = 48b^2/8 = 6b^2$ and therefore $b = \sqrt{1/6}$.

2. The maximum likelihood decision regions are depicted below.



(a) Ω_1 *(b)* Ω_2

3. In general, the nearest neighbor approximation is given by

$$P_\mathsf{s} \approx A_d Q\left(\sqrt{\frac{d_{E,\min}^2}{2N_0}}\right),$$

   where A_d is the average number of points at minimum distance $d_{E,\min}$.

   For Ω_1, we have $d_{E,\min} = a = \sqrt{2/5}$ and $A_d = 1$. Therefore

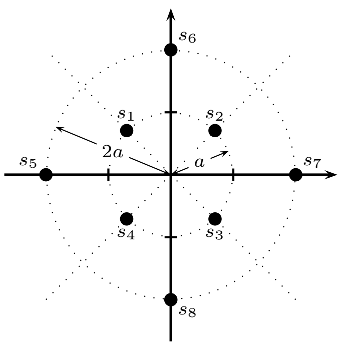
$$P_\mathsf{s} \approx Q\left(\sqrt{\frac{1}{5N_0}}\right).$$

   For Ω_2, we have $d_{E,\min} = 2b = \sqrt{2/3}$ and $A_d = (4 \cdot 1 + 4 \cdot 3)/8 = 2$. Therefore

$$P_\mathsf{s} \approx 2Q\left(\sqrt{\frac{1}{3N_0}}\right).$$

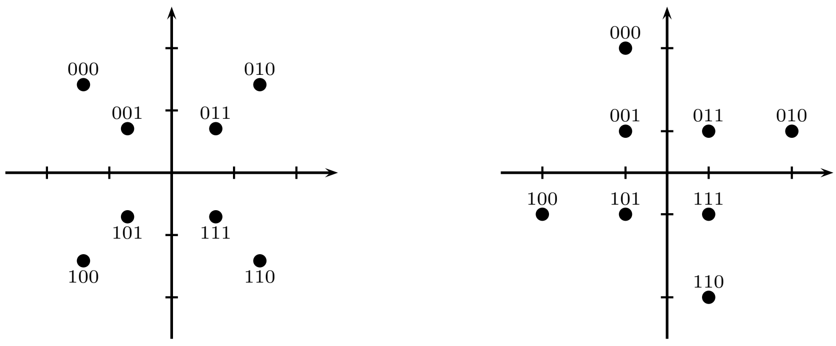
4. Ω_2 performs better because it has a larger minimum distance than Ω_1.

5. The constellation Ω_3 is depicted below.



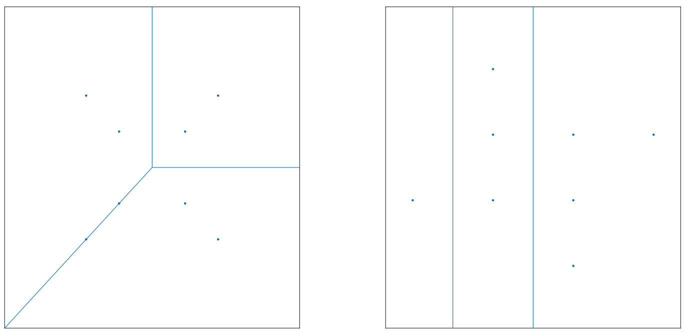
Ω_3 has $d_{E,\min} = \sqrt{2}a$, which is larger than the minimum distance of Ω_1. Therefore Ω_3 is more power efficient than Ω_1.

6. For both Ω_1 and Ω_2, a Gray mapping can be found and an example would be



(a) Ω_1 *(b)* Ω_2

7. The corresponding maximum *a posteriori* decision regions for Ω_1 and Ω_2, when $P(s_1) = P(s_2) = P(s_3) = 1/3$ and $P(s_j) = 0$ for $j = \{4, 5, 6, 7, 8\}$, are shown below.



(a) Ω_1 *(b)* Ω_2

8. The maximum likelihood decision rule can be simplified as

$$\begin{aligned}
\operatorname*{argmax}_{s\in\mathcal{X}} p(r|s) &= \operatorname*{argmax}_{s\in\mathcal{X}} \log p(r|s) \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} -|r-s|^2 \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} \left(-|r|^2 + 2\mathrm{Re}\{rs^*\} - |s|^2\right) \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} 2\mathrm{Re}\{rs^*\} - |s|^2 \qquad (1) \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} 2\mathrm{Re}\{|r||s|\exp(j\angle r - j\angle s)\} - |s|^2 \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} 2|r||s|\cos(\angle r - \angle s) - |s|^2,
\end{aligned}$$

where (1) follows from the fact that $|r|$ is constant with respect to maximization over s.

Problem 3 - Linear Block Codes and LDPC Codes [15 points]

Part I

1. The girth is the length of the shortest cycle in the Tanner graph, which is 4 in this case. It is highlighted in Fig. 1.

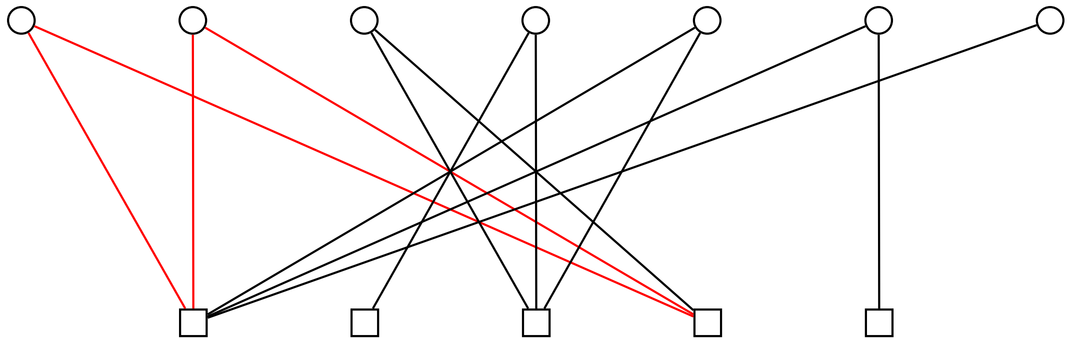


Figure 1: Tanner graph with highlighted girth.

2.

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

3. By considering the Tanner graph, we get

$$\mathsf{\Lambda}(x) = \frac{1}{7}x + \frac{6}{7}x^2 \qquad \mathsf{P}(x) = \frac{2}{5}x + \frac{2}{5}x^3 + \frac{1}{5}x^5.$$

The given code is an irregular LDPC code since the VNs and also the CNs are of different degrees.

Part II

1. If we can find a generator matrix that creates all the codewords in $\mathcal{C}_2$. We can see that the generator matrix must consist of 3 rows and pick the second, the third and the fifth codeword as the generator matrix,

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

We now need to check for each possible information sequence if it creates a valid codeword,

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \boldsymbol{G}_{\mathsf{s}} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

We note that all of those results are codewords and hence $\mathcal{C}_2$ is a linear block code.

By the choice of our generator matrix, $\boldsymbol{G}$ is already in systematic form, i.e., $\boldsymbol{G}_{\mathsf{s}} = \boldsymbol{G}$. The parity check matrix then results to

$$\boldsymbol{H}_{\mathsf{s}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

2. The error detection and error correction capabilities over the BSC depend on the minimum Hamming distance of the code, which is $d_{\mathsf{min}} = 4$. Hence, $\mathcal{C}_2$ can detect error patterns with up to 3 errors and correct all error patterns with up to 1 error.

3. The complete syndrome table consists of 31 entries. Below, we show the completed partial decoding table.

error pattern	syndrome	error pattern	syndrome
00000100	00100	00001000	01000
00010000	10000	00100000	10011
01000000	01110	10000000	11001
10000001	11000	10000010	11011

4.

$$\bar{\boldsymbol{y}}\boldsymbol{H}_{\mathsf{s}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Looking this up in the syndrome table results in the error pattern $\boldsymbol{e} = 10000010$, the codeword $\hat{\boldsymbol{c}} = 011111101$ and hence $\hat{\boldsymbol{u}} = 011$.

5. As $p > 0.5$ we need to flip each bit in order to use the same decoding table as before, i.e.,

$$(\mathbf{1} + \bar{\boldsymbol{y}})\boldsymbol{H}_{\mathsf{s}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hence, no error has been made and therefore, $\hat{\boldsymbol{u}} = 000$.

Problem 4 - Viterbi Algorithm [15 points]

1.

$$\boldsymbol{G}(\mathsf{D}) = \begin{bmatrix} \mathsf{D} & 1+\mathsf{D}+\mathsf{D}^2 & 1+\mathsf{D} \end{bmatrix}$$

2. Dividing $\boldsymbol{G}(\mathsf{D})$ by $(1+\mathsf{D})$ results in

$$\boldsymbol{G}_{\mathsf{s}}(\mathsf{D}) = \begin{bmatrix} \frac{\mathsf{D}}{1+\mathsf{D}} & \frac{1+\mathsf{D}+\mathsf{D}^2}{1+\mathsf{D}} & 1 \end{bmatrix}.$$

The corresponding block diagram is depicted in Fig. 2.

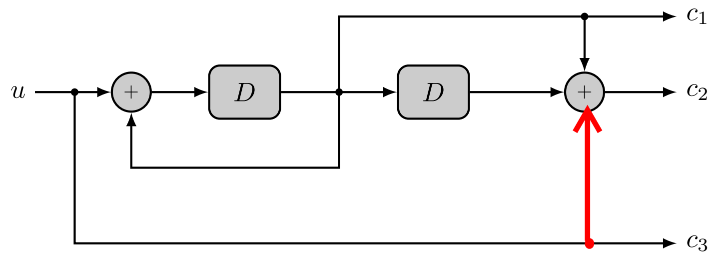


Figure 2: Encoder $\mathcal{E}_{\mathsf{s}}$

3.

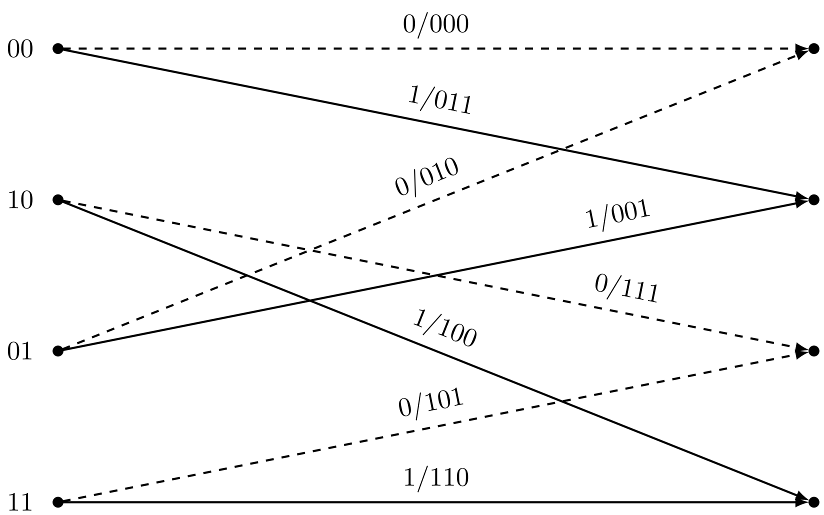


Figure 3: Full Trellis of $\mathcal{E}_1$.

4. As the channel is a BSC with $p = 0.6$, we consider the received signal $\tilde{\boldsymbol{y}} = \mathbf{1} + \bar{\boldsymbol{y}}$, where $\mathbf{1}$ is an all-one vector of appropriate size. Decoding via the Viterbi algorithm is performed in Fig. 4. Three paths have the same cumulative metric. Hence, either of them is correct. We can choose $\hat{\boldsymbol{u}} = 0100$ or $\hat{\boldsymbol{u}} = 1100$ or $\hat{\boldsymbol{u}} = 1000$.

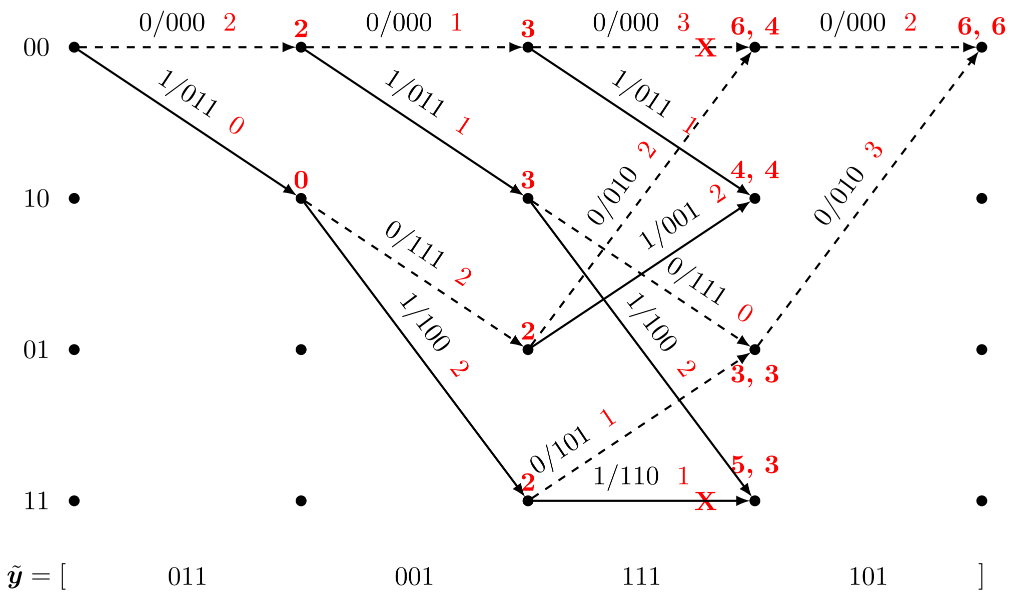


Figure 4: Viterbi algorithm.