Problem 8.1:

(a) $X \to Y \to \tilde{Y} = g(Y)$ form a Markov chain. Hence by the Data Processing Theorem,

$$I(X;Y) \geq I(X;\tilde{Y}) \quad \forall\, p(x) \qquad (*)$$

Let $\tilde{p}(x)$ be the input distribution that achieves $\tilde{C} = \max_{p(x)} I(X;\tilde{Y})$, then

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &\geq I(X;Y)|_{p(x)=\tilde{p}(x)} \\ &\geq I(X;\tilde{Y})|_{p(x)=\tilde{p}(x)} \qquad \text{by } (*) \\ &= \tilde{C}. \end{aligned}$$

(b) We have no *strict* decrease in capacity if $C = \tilde{C}$. This holds iff

$$X \to \tilde{Y} = g(Y) \to Y$$

form a Markov chain (e.g., X and Y are independent or $g(\cdot)$ is invertible).

Problem 8.2:

(a)

Y_1	Y_2	Y_3	$\Pr\{W = a_1 \mid Y_1, Y_2, Y_3\}$	$\Pr\{W = a_2 \mid Y_1, Y_2, Y_3\}$	$\beta(Y_1Y_2Y_3)$
0	0	0	$(1-\epsilon)^3/[(1-\epsilon)^3+\epsilon^3]$	$\epsilon^3/[(1-\epsilon)^3+\epsilon^3]$	a_1
0	0	1	$\epsilon(1-\epsilon)^2/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	$\epsilon^2(1-\epsilon)/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	a_1
0	1	0	$\epsilon(1-\epsilon)^2/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	$\epsilon^2(1-\epsilon)/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	a_1
1	0	0	$\epsilon(1-\epsilon)^2/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	$\epsilon^2(1-\epsilon)/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	a_1
0	1	1	$\epsilon^2(1-\epsilon)/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	$\epsilon(1-\epsilon)^2/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	a_2
1	0	1	$\epsilon^2(1-\epsilon)/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	$\epsilon(1-\epsilon)^2/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	a_2
1	1	0	$\epsilon^2(1-\epsilon)/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	$\epsilon(1-\epsilon)^2/[\epsilon(1-\epsilon)^2+\epsilon^2(1-\epsilon)]$	a_2
1	1	1	$\epsilon^3/[\epsilon^3+(1-\epsilon)^3]$	$(1-\epsilon)^3/[\epsilon^3+(1-\epsilon)^3]$	a_2

(b) Maximum likelihood (ML) decoding minimizes the probability of error when messages are equally likely. For a ML decoder:

$$\beta(Y_1Y_2Y_3) = \begin{cases} a_1 & \text{if } \Pr\{W = a_1 \mid Y_1, Y_2, Y_3\} \geq \Pr\{W = a_2 \mid Y_1, Y_2, Y_3\} \\ a_2 & \text{if } \Pr\{W = a_1 \mid Y_1, Y_2, Y_3\} < \Pr\{W = a_2 \mid Y_1, Y_2, Y_3\} \end{cases}$$

From (a) we can see that this is what the decoder does when $\epsilon < 1/2$, i.e., $\epsilon < 1-\epsilon$.

(c) $\Pr\{\hat{W} \neq W\} = \binom{3}{2}\epsilon^2(1-\epsilon) + \binom{3}{3}\epsilon^3 = \epsilon^2(3-2\epsilon)$.

(d) Consider the ratio

$$\frac{\Pr\{W = a_1 \mid \mathbf{Y}^{2n+1}\}}{\Pr\{W = a_2 \mid \mathbf{Y}^{2n+1}\}} = \frac{\epsilon^{\sum Y_i}(1-\epsilon)^{2n+1-\sum Y_i}}{\epsilon^{2n+1-\sum Y_i}(1-\epsilon)^{\sum Y_i}} = \left(\frac{1-\epsilon}{\epsilon}\right)^{2n+1-2\sum Y_i}.$$

This is < 1 if $\sum_{i=1}^{2n+1} Y_i > n$; and it is > 1 if $\sum_{i=1}^{2n+1} Y_i \leq n$. Thus the optimal decoder will be

$$\hat{W} = \beta(\mathbf{Y}^{2n+1}) = \begin{cases} a_1 & \text{if } \sum Y_i \leq n \\ a_2 & \text{if } \sum Y_i > n \end{cases}.$$

$\Pr\{\text{error}\} = \Pr\{\hat{W} \neq a_1 | W = a_1\} = \Pr\{\frac{1}{2n+1}\sum_{i=1}^{2n+1} Z_i > \frac{1}{2}\}$, where $\{Z_i\}$ is the i.i.d. noise process. Let $\delta = 1/2 - \epsilon > 0$. By the WLLN,

$$\Pr\{\frac{1}{2n+1}\sum_{i=1}^{2n+1} Z_i > \epsilon + \delta\} \to 0 \qquad \text{as } n \to \infty.$$

Thus $\Pr\{\text{error}\} \to 0$ as $n \to \infty$.
Note that in this problem, the rate, R, in bits/channel use goes to zero also. Before Shannon published his paper, it was thought that the only way for $\Pr\{\text{error}\} \to 0$ is to have $R \to 0$. But Shannon showed that it is possible to have $\Pr\{\text{error}\} \to 0$ while maintaining a fixed R.

Problem 8.3: $\mathcal{X} = \{0,1\}$, $P(Z=0) = P(Z=a) = \frac{1}{2}$, so $H(Z) = 1$ (if $a \neq 0$). $Y = X + Z$ (ordinary addition). The capacity of the channel depends on the value of a; there are four cases:

Case (i) $a = 0$: Here $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and the channel reduces to a BSC with crossover probability 0 (or $Y = X$; Noiseless Channel).
So $C = 1 - h_b(0) = 1$ bit/channel use.

Case (ii) $a = +1$: Here $\mathcal{Y} = \{0,1,2\}$

$$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

Hence the channel reduces to a BEC with parameter $\alpha = \frac{1}{2}$.
So $C = 1 - \alpha = 1 - 1/2 = \frac{1}{2}$ bits/channel use.

Case (iii) $a = -1$: Here $\mathcal{Y} = \{-1,0,1\}$

$$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

Here also, we get a BEC with parameter $\alpha = \frac{1}{2}$; so $C = \frac{1}{2}$ bits/channel use.

Case (iv) $a \neq 0$ and $a \neq \pm 1$: Here $\mathcal{Y} = \{0, 1, a, a+1\}$

$$p(y|x) = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

The channel is symmetric, so $p(x) = \frac{1}{2}\ \forall x$ achieves capacity and

$$C = \log_2 4 - 1 = 1 \quad \text{bit/channel use.}$$

Problem 8.4: $C = 1 - h_b(p)$. Now,

$$\begin{aligned}
I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) \\
&= H(Y^n) - H(X_1 \oplus Z_1, \ldots, X_n \oplus Z_n | X_1, \ldots X_n) \\
&= H(Y^n) - H(Z_1, \ldots, Z_n) \quad \text{(since } X_i \text{ and } Z_i \text{ are independent)} \\
&= H(Y^n) - \sum_{i=1}^{n} H(Z_i|Z^{i-1}) \quad \text{(chain rule)} \\
&\geq H(Y^n) - \sum_{i=1}^{n} H(Z_i) \quad \text{(since conditioning reduces entropy)} \\
&= H(Y^n) - nh_b(p) \quad (Z_i \text{ has constant marginal pmf } (p, 1-p)).
\end{aligned}$$

But $Y^n \in \mathcal{Y}^n = \{0,1\}^n$; thus for any X^n

$$H(Y^n) \leq \log_2 |\mathcal{Y}^n| = \log_2(2^n) = n,$$

with equality iff $Y^n \sim \text{Bern}(1/2)$ (iid uniform); this occurs iff $X^n \sim \text{Bern}(1/2)$ (verify it). Hence, $\max_{p(x^n)} H(Y^n) = H(Y^n)|_{X^n \sim \text{Bern}(1/2)} = n$.
Therefore,

$$\begin{aligned}
\max_{p(x^n)} I(X^n; Y^n) &= \max_{p(x^n)} H(Y^n) - H(Z^n) \\
&\geq \max_{p(x^n)} H(Y^n) - nh_b(p) \\
&= n - nh_b(p) = n(1 - h_b(p)) \\
&= nC.
\end{aligned}$$

Thus, a channel with memory has higher capacity than an "equivalent" memoryless channel.

Problem 8.5:
The DMC is described by $Y = X \oplus Z \pmod{11}$, where $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, \ldots, 10\}$, X and Z are independent ($X \perp Z$), and $\Pr\{Z = 1\} = \Pr\{Z = 2\} = \Pr\{Z = 3\} = 1/3$.

(a) $C = \max_{p(x)} I(X;Y) = \max_{p(x)}(H(Y) - H(Y|X))$. But $H(Y|X) = H(X \oplus Z|X) = H(Z|X) = H(Z) = \log_2(3)$, since $X \perp Z$. Furthermore, since the channel is symmetric, capacity is achieved when X is uniform. This yields a uniform Y with $\max_{p(x)} H(Y) = \log_2(11)$. Hence,

$$C = \log_2(11) - \log_2(3) = 1.87 \quad \text{bits/channel use.}$$

(b) From (a), $p^*(x) = 1/11$ for all $x \in \{0, 1, 2, \ldots, 10\}$.

Problem 8.6:
We assume that the two channels are independent. So Y_1 and Y_2 are conditionally independent given X_1 and X_2. Let $p^*(x_1)$ and $p^*(x_2)$ be the distributions which achieve C_1 and C_2, respectively. The capacity of the parallel channel is given by

$$C = \max_{p(x_1, x_2)} I(X_1, X_2; Y_1, Y_2).$$

Note that

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) \\
&\overset{(a)}{=} H(Y_1, Y_2) - [H(Y_1|X_1) + H(Y_2|X_2)] \\
&\overset{(b)}{\leq} H(Y_1) + H(Y_2) - [H(Y_1|X_1) + H(Y_2|X_2)] \\
&= I(X_1; Y_1) + I(X_2; Y_2) \\
&\leq C_1 + C_2.
\end{aligned}$$

Here, (a) is by the independence of the two channels and (b) is by the independence bound on (Y_1, Y_2). Note that inequality (b) is achieved if X_1 and X_2 are independent.
Thus $C = C_1 + C_2$ and the distribution which achieves capacity is

$$p^*(x_1, x_2) = p^*(x_1)p^*(x_2).$$

Problem 8.7:

(a) $C = \log_2 26 = 4.70$ bits/channel use.

(b)

$$p(y|x) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix}.$$

This channel is (weakly) symmetric, so

$$\begin{aligned} C &= \log_2 |\mathcal{Y}| - H(Y|X = x) \\ &= \log_2 26 - 1 \\ &= \log_2 13 = 3.70 \text{ bits/channel use.} \end{aligned}$$

(c) $M = 13, \mathcal{C} = \{a, c, e, g, i, k, m, o, q, s, u, w, y\}$.

$$\begin{aligned} \beta(a) &= \beta(b) = 1 \\ \beta(c) &= \beta(d) = 2 \\ \beta(e) &= \beta(f) = 3 \\ \beta(g) &= \beta(h) = 4 \\ \beta(i) &= \beta(j) = 5 \\ \beta(k) &= \beta(l) = 6 \\ \beta(m) &= \beta(n) = 7 \\ \beta(o) &= \beta(p) = 8 \\ \beta(q) &= \beta(r) = 9 \\ \beta(s) &= \beta(t) = 10 \\ \beta(u) &= \beta(v) = 11 \\ \beta(w) &= \beta(x) = 12 \\ \beta(y) &= \beta(z) = 13 \end{aligned}$$

Here, $\lambda_{max}(\mathcal{C}) = 0$ and the rate is $R = \log_2 13 = C$ bits/channel use. This is a special DMC in which a small blocklength code can achieve capacity.

Problem 8.8:
Here, we have a cascade of n independent BSCs with identical crossover probability p. The cascaded channel has two inputs and two outputs. It is straightforward to show that the cascaded channel is a BSC. To determine its crossover probability, proceed as follows.

$$\begin{aligned} p(0|0) &= \Pr\{\#\text{ of errors in the } n \text{ component channels is even}\} \\ &= \sum_{\substack{k=0 \\ k \text{ even}}}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{\substack{k=0 \\ k \text{ even}}}^{n} F_k, \end{aligned}$$

where F_k is defined as $F_k \triangleq \binom{n}{k} p^k (1-p)^{n-k}$. Now

$$\begin{aligned}
2p(0|0) - 1 &= 2 \sum_{\substack{k=0 \\ k \text{ even}}}^{n} F_k - \sum_{k=0}^{n} F_k \\
&= \sum_{\substack{k=0 \\ k \text{ even}}}^{n} F_k + \sum_{\substack{k=0 \\ k \text{ even}}}^{n} F_k - \sum_{\substack{k=0 \\ k \text{ even}}}^{n} F_k - \sum_{\substack{k=0 \\ k \text{ odd}}}^{n} F_k \\
&= \sum_{\substack{k=0 \\ k \text{ even}}}^{n} F_k - \sum_{\substack{k=0 \\ k \text{ odd}}}^{n} F_k \\
&= \sum_{k=0}^{n} (-1)^k F_k \\
&= \sum_{k=0}^{n} \binom{n}{k} (-p)^k (1-p)^{n-k} \\
&= (1-2p)^n.
\end{aligned}$$

The last equality is due to the binomial theorem. Thus

$$\begin{aligned}
p(0|0) &= \frac{1}{2}[(1-2p)^n + 1] \\
p(1|0) &= 1 - p(0|0) = \frac{1}{2}[1 - (1-2p)^n]
\end{aligned}$$

Therefore, $C_n = h_b(\frac{1}{2}[(1-2p)^n + 1])$. For any $p \in (0,1)$, as $n \to \infty$, $C_n \to 0$.

Problem 8.9:
The capacity of the Z channel can be calculated using elementary calculus. Let

$$p(x) = \begin{cases} p & \text{if } x = 0 \\ 1-p & \text{if } x = 1 \end{cases}.$$

Then

$$p(y) = \begin{cases} p + (1-p)/2 & \text{if } y = 0 \\ (1-p)/2 & \text{if } y = 1 \end{cases}.$$

Thus

$$\begin{aligned}
H(Y) &= \frac{1+p}{2} \log_2 \frac{2}{1+p} + \frac{1-p}{2} \log_2 \frac{2}{1-p} = h_b(\frac{1+p}{2}) \\
H(Y|X) &= p \underbrace{H(Y|X=0)}_{=0} + (1-p) \underbrace{H(Y|X=1)}_{=1} = 1 - p.
\end{aligned}$$

Now

$$I(X;Y) = H(Y) - H(Y|X) = h_b(\frac{1+p}{2}) - (1-p).$$

To find the capacity, we maximize $I(X;Y)$ w.r.t. p. To do this, we take the derivative and set it to zero:

$$\frac{dI(X;Y)}{dp} = \frac{1}{2} h_b'(\frac{1+p}{2}) + 1 = 0.$$

This implies that

$$h_b'(\frac{1+p}{2}) = \log_2 \left[\frac{(1-p)/2}{(1+p)/2} \right] = -2.$$

This in turn implies that $p = 3/5$. Note that this is the maximum value of $I(X;Y)$ instead of the minimum because of the concavity of $I(X;Y)$ w.r.t. $p(x)$. Hence,

$$C = h_b(4/5) - 2/5 = \log_2 5 - 2 = 0.322 \text{ bits/channel use.}$$

Problem 8.10:
Using the same arguments as in the direct part of the channel coding theorem:

$$\lambda_{ave} = \Pr\{\beta(\mathbf{Y}) \neq i\} = \underbrace{\Pr\{(\mathbf{X}_i, \mathbf{Y}) \notin A_\epsilon^{(n)}\}}_{\star} + \underbrace{\sum_{\substack{j=1 \\ j \neq i}}^{n} \Pr\{(\mathbf{X}_i, \mathbf{Y}) \in A_\epsilon^{(n)}\}}_{\#}.$$

The first part $\star \to 0$ as $n \to \infty$. Now

$$\Pr\{(\mathbf{X}_i, \mathbf{Y}) \in A_\epsilon^{(n)}\} \leq 2^{-n[I(X;Y)-3\epsilon]},$$

where $I(X;Y) = h_b(3/4) - 1/2$ [uniform input distribution]. Thus $\# \to 0$ if $R < R_{max} = \frac{3}{4}(2 - \log_2 3) = 0.311$. Note that $R_{max} < C = 0.322$.

Problem B:

(a)

$$\begin{aligned} Q = [p(y|x)] &= \begin{bmatrix} p(0|1) & p(1|1) & p(2|1) & p(3|1) \\ p(0|2) & p(1|2) & p(2|2) & p(3|2) \end{bmatrix} \\ &= \begin{bmatrix} \epsilon & 1-2\epsilon & \epsilon & 0 \\ 0 & \epsilon & 1-2\epsilon & \epsilon \end{bmatrix}. \end{aligned}$$

(b) Q is quasi-symmetric since it can be partitioned (along its columns) into two symmetric sub-arrays. Thus $p^* = [1/2, 1/2]$ is the optimal input distribution that achieves C. Also, every row in Q is a permutation of every other row, so $H(Y|X) = H(\epsilon, 1-2\epsilon, \epsilon) = -2\epsilon \log \epsilon - (1 - 2\epsilon)\log(1-2\epsilon)$.
So $C = I(X;Y)|_{p^*=[1/2,1/2]}$. Hence,

$$C = H(Y)|_{p^*=[1/2,1/2]} - H(Y|X).$$

If X is distributed by $[1/2, 1/2]$, then Y is distributed by,

$$p(y) = \begin{cases} \epsilon/2 & \text{if } y \in \{0,3\} \\ \frac{1}{2}(1-\epsilon) & \text{if } y \in \{1,2\} \end{cases}$$

So,

$$\begin{aligned} H(Y) &= -\epsilon \log\left(\frac{\epsilon}{2}\right) - (1-\epsilon)\log\left(\frac{1-\epsilon}{2}\right) \\ &= 1 + h_b(\epsilon) \end{aligned}$$

Thus, $C = 1 + h_b(\epsilon) + 2\epsilon\log(\epsilon) + (1-2\epsilon)\log(1-2\epsilon)$ bits/channel use.

(c) $H(X,Y) = H(X) + H(Y|X) = 1 - 2\epsilon\log\epsilon - (1-2\epsilon)\log(1-2\epsilon)$ bits. Also, $H(X,Y) = H(Y) + H(X|Y)$. Thus,

$$\begin{aligned} H(X|Y) &= H(X,Y) - H(Y) \\ &= 1 - 2\epsilon\log\epsilon - (1-2\epsilon)\log(1-2\epsilon) - 1 - h_b(\epsilon) \\ &= -2\epsilon\log\epsilon - (1-2\epsilon)\log(1-2\epsilon) + \epsilon\log\epsilon + (1+\epsilon)\log(1-\epsilon) \\ &= -\epsilon\log\epsilon - (1-2\epsilon)\log(1-2\epsilon) + (1-\epsilon)\log(1-\epsilon). \end{aligned}$$

Problem C:

(a) The DMC is described by: $Y = 3X + (-1)^X Z$, where $\mathcal{X} = \{0, 1\}$, $\Pr\{Z = 0\} = \Pr\{Z = 1\} = \Pr\{Z = 2\} = \epsilon$, and $X \perp Z$. Clearly, $\mathcal{Y} = \{0, 1, 2, 3\}$ and

$$\begin{aligned} Q = [p(y|x)] &= \left[Pr\left\{ Z = \frac{y - 3x}{(-1)^x} \right\} \right] \\ &= \begin{bmatrix} p(0|0) & p(1|0) & p(2|0) & p(3|0) \\ p(0|1) & p(1|1) & p(2|1) & p(3|1) \end{bmatrix} \\ &= \begin{bmatrix} \epsilon & \epsilon & \epsilon & 1-3\epsilon \\ 1-3\epsilon & \epsilon & \epsilon & \epsilon \end{bmatrix}. \end{aligned}$$

(b) The channel is *quasi-symmetric*, since Q can be partitioned along its columns into two symmetric subarrays:

$$\begin{bmatrix} \epsilon & 1-3\epsilon \\ 1-3\epsilon & \epsilon \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} \epsilon & \epsilon \\ \epsilon & \epsilon \end{bmatrix}.$$

Therefore, the uniform input distribution ($Pr\{X = 0\} = Pr\{X = 1\} = 1/2$) achieves capacity, and

$$\begin{aligned} C &= \max_X I(X; Y) \\ &= H(Y)|_{Pr\{X=0\}=1/2} - H(\epsilon, \epsilon, \epsilon, 1 - 3\epsilon) \\ &= H(Y)|_{Pr\{X=0\}=1/2} + 3\epsilon \log_2(\epsilon) + (1 - 3\epsilon) \log_2(1 - 3\epsilon). \end{aligned}$$

Under the (maximizing) uniform input distribution, the output distribution becomes:

$$Pr\{Y = 0\} = Pr\{Y = 3\} = \frac{1}{2}\epsilon + \frac{1}{2}(1 - 3\epsilon) = \frac{1}{2} - \epsilon,$$

and

$$Pr\{Y = 1\} = Pr\{Y = 2\} = \epsilon.$$

Thus,

$$H(Y) = -2\epsilon log_2(\epsilon) - 2\left(\frac{1}{2} - \epsilon\right) \log_2\left(\frac{1}{2} - \epsilon\right).$$

So,

$$C = \epsilon log_2(\epsilon) - 2\left(\frac{1}{2} - \epsilon\right) \log_2\left(\frac{1}{2} - \epsilon\right) + (1 - 3\epsilon) \log_2(1 - 3\epsilon) \quad \text{bits/channel use.}$$

(c) If $\epsilon = 1/4$, then

$$C = (1/4) \log_2(1/4) - 2(1/4) \log_2(1/4) + (1/4) \log_2(1/4) = 0.$$

Thus, *zero* bits/channel use can be reliably sent over the channel. In other words, *no* reliable transmission is possible. (Note that when $\epsilon = 1/4$, the noise is uniformly distributed; so the noise entropy is maximum and the channel becomes totally random).

Please let me know if you see a mistake in this solution.