Problem 1.

a) For all $x \in \mathcal{X}$, since $P(x^*) \geq P(x)$, then $\log(\frac{1}{P(x)}) \geq \log(\frac{1}{P(x^*)})$. Hence,

$$H(X) = \sum_{x \in \mathcal{X}} P(x) \log(\frac{1}{P(x)}) \geq (\log(\frac{1}{P(x^*)})) \sum_{x \in \mathcal{X}} P(x) = \log(\frac{1}{P(x^*)}).$$

b) As we have seen in class, we define

$$Z = \begin{cases} 0, & \hat{X} = X \\ 1, & \hat{X} \neq X \end{cases}$$

Then, $H(X, Z|Y) = H(X|Y) + H(Z|X, Y) = H(Z|Y) + H(X|Z, Y)$.
Moreover, $H(Z|X, Y) \leq H(Z|X, g(Y) = \hat{X}) = 0$ and $H(Z|Y) \leq H(P_e)$. Therefore,

$$H(X|Y) \leq H(P_e) + H(X|Z, Y) = H(P_e) + P_e H(X|Z = 1, Y) \leq H(P_e) + P_e \log(|\mathcal{X}| - 1).$$

c) Assume that $\hat{x} = g(y)$ for some observation y. This means that $P(\hat{x}|y) \geq P(x|y)$ for all $x \in \mathcal{X}$. According to part (a), $H(X|Y = y) \geq \log(\frac{1}{P(\hat{x}|y)})$. Combining these, we obtain

$$P(\hat{x}|y) \geq e^{-H(X|Y=y)}.$$

On the other hand, $P_e = P\{\hat{X} \neq X\} = 1 - P\{\hat{X} = X\}$. So,

$$\begin{aligned} P_e &= 1 - \sum_{y \in \mathbf{Y}} P(Y = y) P(\hat{x}|y) \\ &\leq 1 - \sum_{y \in \mathcal{Y}} P(Y = y) e^{-H(X|Y=y)} \\ &\leq 1 - e^{-\sum_{y \in \mathcal{Y}} P(Y=y) H(X|Y=y)} \\ &= 1 - e^{H(X|Y)}. \end{aligned}$$

where we used the hint in the last inequality.

Problem 2.

(a)

$$\begin{aligned} I(X; YZ) &= I(X; Z) + I(X; Y|Z) = I(X; Y|Z) \\ &= I(X; Y|Z = 1) \Pr\{Z = 1\} + I(X; Y|Z = 2) \Pr\{Z = 2\} \\ &= pI(X; Y^1) + (1 - p) I(X; Y^2) \end{aligned}$$

(b)

$$\begin{aligned} \max_{p(x)} I(X; YZ) &= \max_{p(x)} pI(X; Y_1) + (1 - p) I(X; Y_2) \leq p \max_{p(x)} I(X; Y_1) + (1 - p) \max_{p(x)} I(X; Y_2) \\ &= pC_1 + (1 - p) C_2 \end{aligned}$$

If both terms are positive, we have equality if the maximizing input distribution is the same for both terms. In our case, at any δ, the BSC has the uniform distribution that achieves capacity. So, we need to have that the Z-channel also has the uniform distribution as capacity achieving distribution, which happens only in degenerated cases: if $\epsilon = 0$ or $\epsilon = 1$. (For details, see Homework 7).

Also, we can have equality if one (or both) of the terms are 0. This happens in four cases: $p = 1$, $p = 0$, $\epsilon = 1$ or $\delta = 0.5$.

(c) *Encoding*: We design two different codes, one for $\mathcal{C}_1$ another for $\mathcal{C}_2$ using the corresponding capacity achieving distribution (as seen in class). The length of the first code is $(n(p-\varepsilon))$ the length of the second is $(n(1-p-\epsilon))$. Take those X_is for which $Z_i = 1$ together and treat them as one block, then choose their values according to the first code. Similarly, for the block that consist of X_is for which $Z_i = 2$, we use the second code.

If the block is larger than code length, set the leftover to 0. If it is shorter, declare error.

Decoding: The decoding is done similarly. First arrange the output into two blocks based on Z_i, ignore the outputs of the padded 0s, and do the decoding according to the corresponding code (as seen in class).

For ε-typical Z sequences this code achieves $(p-\varepsilon)C_1 + (1-p-\varepsilon)C_2$ rate. ε can be arbitrarily small, and for any ε the probability that the Z sequence is not typical goes to zero, so with sufficiently large n we can reach $pC_1 + (1-p)C_2$.

Note: One can show that $I(X;Y|Z)$ is a valid upper-bound for this case also, so $pC_1 + (1-p)C_2$ is in fact the capacity of this non-casual channel.

PROBLEM 3.

$$
\begin{aligned}
\log(p(y_1...y_n|x_1...x_n)) &= \log(\prod_{i=1}^{n} p(y_i|x_i)) \\
&= \log(\prod_{x\in\mathcal{X},y\in\mathcal{Y}} p(y|x)^{N(x,y)}) \\
&= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \log(p(y|x)^{N(x,y)}) \\
&\leq \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \log(p(y|x)^{n(1-\epsilon)p(y|x)}) \\
&= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} n(1-\epsilon)p(x,y)\log(p(y|x)) \\
&= -n(1-\epsilon)H(Y|X)
\end{aligned}
$$

$\Rightarrow p(y_1...y_n)|(x_1...x_n)) \leq 2^{-n(1-\epsilon)H(Y|X)}$

By similar steps, we find $\log(p(y_1...y_n)|(x_1...x_n)) \geq 2^{-n(1+\epsilon)H(Y|X)}$. The cardinality of the typical set is then upper bounded as:

$$
\begin{aligned}
1 &\geq \sum_{y\in A^{\epsilon,n}_{p_{Y|X}}} p(y_1,...,y_n|x_1,...x_n) \\
&\geq \sum_{y\in A^{\epsilon,n}_{p_{Y|X}}} 2^{-n(1+\epsilon)H(Y|X)} \\
&= \|A^{\epsilon,n}_{p_{Y|X}}\| 2^{-n(1+\epsilon)H(Y|X)}
\end{aligned}
$$

$\Rightarrow \|A^{\epsilon,n}_{p_{Y|X}}\| \leq 2^{n(1+\epsilon)H(Y|X)}$.

Problem 4.

1. Let $P_{e,0}$ and $P_{e,1}$ denote the conditional error probabilities given that the input 0 and 1 are sent, respectively. Then, we have

$$\begin{aligned} P_{e,0} &= \sum_{y\in\mathcal{Y}} P(y|0)\mathbf{1}\{y : \frac{P(y|1)}{P(y|0)} \geq 1\} \\ &\leq \sum_{y\in\mathcal{Y}} P(y|0)\sqrt{P(y|1)/P(y|0)} = Z(P) \\ P_{e,1} &= \sum_{y\in\mathcal{Y}} p(y|1)\mathbf{1}\{y : \frac{P(y|0)}{P(y|1)} \geq 1\} \\ &\leq \sum_{y\in\mathcal{Y}} P(y|1)\sqrt{P(y|0)/P(y|1)} = Z(P) \end{aligned}$$

where $\mathbf{1}\{.\}$ is the indicator function.

Hence the average error probability P_e is given by

$$P_e = Pr(X=0)P_{e,0} + Pr(X=1)P_{e,1} = Z(P).$$

2. The function $Z(P)$ is a concave function of the channel transition probabilities, i.e., given any collection of B-DMCs, $P_j : \mathcal{X} \to \mathcal{Y}, j \in \mathcal{J}$, and a probability distribution Q on $\mathcal{J}$, if we define $P : \mathcal{X} \to \mathcal{Y}$ as the channel $P(y|x) = \sum_{j\in\mathcal{J}} Q(j)P_j(y|x)$, then,

$$\sum_{j\in\mathcal{J}} Q(j)Z(P_j) \leq Z(P).$$

To show this, we start using the hint

$$\begin{aligned} Z(P) &= \sum_y \sqrt{P(y|0)P(y|1)} \\ &= -1 + \frac{1}{2}\sum_y \left[\sum_x \sqrt{P(y|x)}\right]^2 \end{aligned}$$

Then, we apply Minkowsky's inequality to get

$$\begin{aligned} Z(P) &\geq -1 + \frac{1}{2}\sum_y \sum_{j\in\mathcal{J}} Q(j)\left[\sum_x \sqrt{P_j(y|x)}\right]^2 \\ &= \sum_{j\in\mathcal{J}} Q(j)Z(P_j). \end{aligned}$$