PROBLEM 1.

(a)
$$\begin{aligned}I(X;Y) &= \sum_{x,y} p_{XY}(xy) \log \frac{p_{XY}(xy)}{p_X(x)p_Y(y)} \\ &= \sum_{xy} p_X(x)W(y|x) \log \frac{W(y|x)}{p_Y(y)} \\ &= \sum_x p_X(x) D(W_{Y|X=x} \| p_Y).\end{aligned}$$

(b) For any distribution q_Y on $\mathcal{Y}$

$$\begin{aligned}\sum_x p_X(x) D(W_{Y|X=x} \| q_Y) - I(X;Y) &= \sum_{x,y} p_{XY}(x,y) \log \frac{p_Y(y)}{q_Y(y)} \\ &= \sum_y p_Y(y) \log \frac{p_Y(y)}{q_Y(y)} \\ &= D(p_Y \| q_Y) \\ &\geq 0.\end{aligned}$$

We obtain the desired result as a special case when $q_Y = W_{Y|X=x_0}$.

(c) Using (b), and noting that $D(W_{Y|X=x} \| W_{Y|X=x_0}) \leq C_1 b(x)$, we find $I(X;Y) \leq C_1 \sum_x p_X(x) b(x) = C_1 E[b(X)]$.

(d) Noting that $D(\cdot\|\cdot) \geq 0$, we see that each term in the right hand side of (a) is a lower bound to $I(X;Y)$.

(e) For the given distribution note that $p_Y = \delta W_{Y|X=x_1} + (1-\delta) W_{Y|X=x_0}$. Using (d) we can lower bound $I(X;Y)$ by $\delta D(W_{Y|X=x_1} \| p_Y)$. Noting that $E[b(X)] = \delta b(x_1)$ the result follows.

(f) By (c) we see that $\sup_{p_X} I(X;Y)/E[b(X)] \leq C_1$. Now choose x_1 be the x that achieves the maximum that defines C_1, so that $C_1 = D(W_{Y|X=x_1} \| W_{Y|X=x_0})/b(x_1)$. Using (d) with $\delta \to 0$ we see that the $\sup_{p_X} I(X;Y)/E[b(X)] \geq C_1$.

PROBLEM 2. (a) The constraints that define $\mathcal{Y}$ fix k of the coordinates of y^n, allowing $n-k$ coordinates to be free. Thus $|\mathcal{Y}(f^n, s^n)| = 2^{n-k}$.

(b) For each $y^n \in \mathcal{Y}$ the probability that $\text{read}(y^n) \neq w$ is $1 - 2^{-nR}$. Since these events are independent the probability of $\text{read}(y^n) \neq w$ for all $y^n \in \mathcal{Y}$ is $(1-2^{-nR})^{|\mathcal{Y}|} = (1-2^{-nR})^{2^{n-k}}$.

(c) Using (b) and upper bounding $1 - 2^{-nR}$ by $\exp(-2^{-nR})$ we see that the probability in (b) is upper bounded by $\exp(-2^{n-k}2^{-nR})$. Noting that $k = qn$ the result follows.

(d) Given $R < 1-p$, fix q_0 such that $p < q_0 < 1-R$. Let A be the event that for all $y^n \in \mathcal{Y}(F^n, S^n)$, $\text{read}(y^n) \neq w$, and let B be the event at that $K/n < q_0$. We then have

$$\Pr(A) \leq \Pr(A \cap B) + \Pr(B^c) \leq \Pr(A|B) + \Pr(B^c).$$

By the law of large numbers $K/n \to p$ as n gets large. Thus $\Pr(B^c) \to 0$ since $q_0 > p$. Moreover, by (c), $\Pr(A|B) \leq \exp\big(-2^{n(1-R-q_0)}\big)$ which also approaches 0 as n gets large since $q_0 < 1-R$. Consequently $\Pr(A) \to 0$ as n gets large.

(e) Given the randomly constructed read() as above, define $\text{write}(w_n, f^n, s^n)$ as follows: if there is a $y^n \in \mathcal{Y}(f^n, s^n)$ with $\text{read}(y^n) = w$, set $\text{write}() = y^n$, otherwise randomly choose write(). Note that in the first case $\hat{w}_n = w_n$. Thus $\Pr(\hat{W}_n \neq W_n)$ is upper bounded by the probability we found in (d), which can be made less than ϵ by choosing n large enough.

(f) No. Even if f^n we revealed to *both* the reader and writer there are only $n-K$ memory locations that they can use to store data. For $R > 1-p$, by the law of large numbers $nR \leq n-K$ is a small probability event, so there is a small probaiblity that nR bits of data can be stored in $n-K$ locations.

Problem 3. (a) Blocklength of enc is $2n$. Also, enc encodes $k_1 + k_2$ bits of information to $2n$ channel symbols, so $R = (k_1+k_2)/2n = (R_1+R_2)/2$.

(b) $w_H(x) = w_H(x_1) + w_H(x_1+x_2)$. By the triangle inequality $w_H(x_1) + w_H(x_1+x_2) \geq w_H(x_1+x_1+x_2) = w_H(x_2)$.

(c) If $x_2 = 0$, we clearly have $w_H(x) = 2w_H(x_1)$. Otherwise, by (b) we have $w_H(x) \geq w_H(x_2)$. In either case the claim $w_H(x) \geq 2w_H(x_1)\mathbb{1}(x_2 = 0) + w_H(x_2)\mathbb{1}(x_2 \neq 0)$ holds.

(d) Recall that for linear encoders the minimum distance is equal to the minimum weight. Note that the codewords of enc are of the form x above with x_i a codeword of enc_i for $i = 1, 2$. A non-zero codeword x of enc must that either $x_1 \neq 0$ or $x_2 \neq 0$. Thus by (c), we see that the minimum weight codeword of enc has weight at least $\min\{2d_1, d_2\}$, and thus $d \geq \min\{2d_1, d_2\}$. Moreover, with x_1 a minimum weight codeword of enc_1 and x_2 a minimum weight codeword of enc_2, observe that both $[x_1, x_1]$ and $[0, x_2]$ are non-zero codewords of enc, thus $d \leq \min\{2d_1, d_2\}$.

(e) The encoder that corresponds to generator M_i takes one bit and repeats it 2^i times. Thus it is of rate $1/2^i$ and has minimum distance 2^i. Consequently, n_i, R_i and d_i satisfy: $n_{i+1} = 2n_i$, $R_{i+1} = (R_i + 2^{-i})/2$ and $d_{i+1} = \min\{2d_i, 2^i\}$, starting with $n_1 = 2$, $R_1 = 1$, $d_1 = 1$. We thus see that $n_i = 2^i$, $d_i = 2^{i-1}$, and $R_i = (i+1)/2^i$.

Problem 4. (a) Suppose a scheme that achieves (R, D) with $R < R(D)$. The same scheme must achieve R and $E[\log d(X^n, Y^n)] \leq \log D$. Since $\log d(X^n, Y^n)$ is an additive distortion, we know from the standard converse that $R \geq R(D)$. Hence, it is a contradiction and $R \geq R(D)$ must hold.

(b) Let $\tilde{R}(D) := \inf_{p_{y|x}: E[\log d(X,Y)] \leq D} I(X;Y)$. We know $\tilde{R}(D)$ is convex and $R(D) = \tilde{R}(\log(D))$. Hence, $R(\lambda D_1 + (1-\lambda)D_2) = \tilde{R}(\log(\lambda D_1 + (1-\lambda)D_2)) \overset{(*)}{\leq} \tilde{R}(\lambda \log(D_1) + \log((1-\lambda)D_2)) \overset{(**)}{\leq} \lambda R(D_1) + (1-\lambda)R(D_2)$. $(*)$ follows from concavity of $\log(.)$ and due to the fact that $\tilde{R}(.)$ is non-increasing. $(**)$ follows from convexity of $\tilde{R}(D)$.

(c) Since $\sum_i \mathbb{1}(x_i = x, y_i = y) \leq n p_{XY}(x,y)(1+\epsilon)$ for every ϵ-typical (x^n, y^n) and $d(x,y) \geq 1$ for all (x,y), $d(x^n, y^n) = \prod_i d(x_i, y_i)^{\frac{1}{n}} \leq \prod_{x,y} d(x,y)^{p_{XY}(x,y)(1+\epsilon)} = \exp\big(E[\log d(X,Y)](1+\epsilon)\big) = D^{1+\epsilon}$ for every ϵ-typical (x^n, y^n).

(d)
$$\begin{aligned} E[d(X^n,Y^n)] &= E[d(X^n,Y^n)|(X^n,Y^n) \text{ is not } \epsilon\text{-typical}]\Pr((X^n,Y^n) \text{ is not } \epsilon\text{-typical}]) \\ &\quad + E[d(X^n,Y^n)|(X^n,Y^n) \text{ is } \epsilon\text{-typical}]\Pr((X^n,Y^n) \text{ is } \epsilon\text{-typical}]) \\ &\leq d_{\max}\Pr((X^n,Y^n) \text{ is not } \epsilon\text{-typical}]) + E[d(X^n,Y^n)|(X^n,Y^n) \text{ is } \epsilon\text{-typical}] \end{aligned}$$

From the course, we know $\epsilon' := d_{\max}\Pr((X^n,Y^n) \text{ is not } \epsilon\text{-typical}]) \to 0$ if $R > R(D)$. Part (c) implies $E[d(X^n,Y^n)|(X^n,Y^n) \text{ is } \epsilon\text{-typical}] \leq D^{1+\epsilon}$. Hence, $E[d(X^n,Y^n)] \leq \epsilon' + D^{1+\epsilon}$.