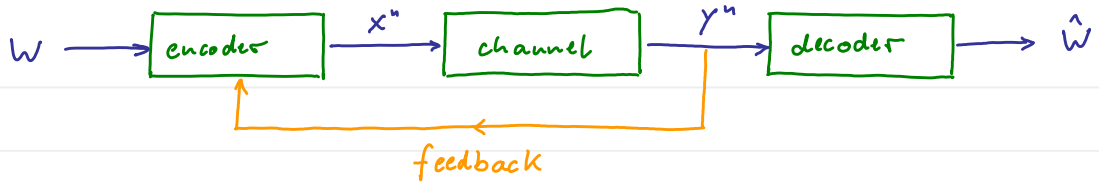

Thm.: $\boxed{C_{FB} = C}$

proof: evidently $C_{FB} \geq C$, so we need to show $C_{FB} \leq C$.

Let W be uniformly distributed over all input messages.

Similar to the proof of the converse part without feedback:

$$nR = H(W) = \underbrace{H(W|Y^n)}_{(i)} + \underbrace{I(W;Y^n)}_{(ii)}$$

(i) $H(W|Y^n) \leq H(W|\hat{W})$ data processing inequality

$\leq 1 + p_e^{(n)} nR$ Fano's inequality $(h(p_e) + p_e \log_2 |\mathcal{X}| \geq H(X|Y))$

(ii) $I(W;Y^n) = H(Y^n) - H(Y^n|W)$

$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, W)$

$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, W, X_i) \quad | \quad X_i = f_i(W, Y^{i-1})$

$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \quad | \quad Y_i$ depends on (Y^{i-1}, W) only via X_i

$\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) \quad |$ subadditivity

$= \sum_{i=1}^{n} I(X_i;Y_i) \leq nC$

$\Rightarrow nR \leq 1 + p_e^{(n)} nR + nC \Rightarrow R(1 - p_e^{(n)}) \leq \frac{1}{n} + C$

$\Rightarrow R \leq C$ via $n \to \infty$

□

remarks: in practice, however, feedback can help / simplify.

example: for the binary erasure channel resend the bit until it has not been erased → average nr. of channels used:

$$(1-p)\underbrace{\sum_{n=1}^{\infty} n p^{n-1}}_{(1-p)^{-2}} = \frac{1}{(1-p)}$$

→ $(1-p)$ is achievable rate with feedback.

But we know also that $C_{FB} = C = (1-p)$

note: • the capacity is in this case easily achieved with zero error

• without feedback codes coming close to capacity are far more complicated & the error is non-zero

- another resource which doesn't change capacity is "shared randomness" between sender & receiver.

IV. 10. Source-channel separation

Question: what if the messages to be transmitted are not uniformly distributed?

- one possibility is to separate source coding (data compression) & channel coding
- a more general approach would be to combine them. Such codes are called source-channel codes.
- the following shows that we don't loose anything, if we separate the two:

Thm.: (source-channel coding theorem)

Consider a discrete memoryless channel with $C := \max_{p(x)} I(X;Y)$.

(i) Let $\{V_i\}_{i \in \mathbb{N}}$ be a stochastic process which satisfies the AEP w.r.t. its entropy rate $H(\{V_i\})$ (e.g. an i.i.d source or, more generally, a stationary ergodic stochastic process). If $H(\{V_i\}) < C$, there is a source-channel code which allows transmission s.t. $\text{prob}(\hat{V}^n \neq V^n) \to 0$ as $n \to \infty$.

(ii) For any stationary stochastic process, if $H(\{V_i\}) > C$, then

$\exists \delta > 0 \ \forall n \in \mathbb{N} \ \forall$ source-channel codes : $\text{prob}(\hat{V}^n \neq V^n) > \delta$.

proof: similar to what we did before → exercise.

V. Error correcting codes / coding theory

Note: "random coding" (as in the proof of Shannon's noisy channel coding thm.) is completely useless for actual information transmission. We need something more concrete & more efficient ...

Example: "[7,4] Hamming code"

Let $x \in \{0,1\}^4$ be a message which we want to protect against errors.

Define $g := \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$ and $G := \begin{pmatrix} \mathbb{1}_4 \\ g \end{pmatrix} \in \mathbb{Z}_2^{7 \times 4}$.

"Encode" the message into $y := Gx \in \mathbb{Z}_2^7$ with addition mod 2.

Claim: the image of any vector $x' \in \mathbb{Z}_2^4$ with $x' \neq x$ differs from y in at least three bits, i.e. $|\{i \mid (G(x-x'))_i \neq 0\}| \geq 3$.

Consequence: if an arbitrary single bit in y is corrupted, we can correct for it.

proven by inspection: $G(\Delta x)$ has at least 3 non-zero components if $\Delta x \neq 0$.

V.1. Basic definitions

Def.: Let $\mathcal{X}$ be a finite alphabet, $0 \in \mathcal{X}$ and $x, x' \in \mathcal{X}^n$.

- $d(x) := \left| \{ i \in \{1, \dots, n\} \mid x_i \neq 0 \} \right|$ "Hamming weight"
- $d(x-x') := \left| \{ i \mid x_i \neq x'_i \} \right|$ "Hamming distance"
- $\{ x' \in \mathcal{X}^n \mid d(x-x') \leq r \}$ "Hamming ball" of radius r around x

remark: $(x, x') \mapsto d(x-x')$ is a metric on $\mathcal{X}^n$.

Def.: An "error correcting code" C of length $n \in \mathbb{N}$ over an alphabet $\mathcal{X}$ is a subset $C \subseteq \mathcal{X}^n$ whose elements are called "codewords".

remarks:
- we will often associate an "encoding map" $E : \{1, \dots, |C|\} \to C \subseteq \mathcal{X}^n$ with the error correcting code (= code in the following)
- the above codes are also called "block codes" with "block length" n,
- the code is called q-ary (binary) if $|\mathcal{X}| = q$ ($|\mathcal{X}| = 2$)

Def.: Let $C \subseteq \mathcal{X}^n$.

- $R(C) := \dfrac{\log |C|}{\log |\mathcal{X}^n|}$ is called the "rate" of the code.
- $d(C) := \min\limits_{\substack{c, c' \in C \\ c \neq c'}} d(c-c')$ is called its "distance", and $\dfrac{d(C)}{n}$ "relative distance".

remarks:
- $R(C)$ ~ fraction of non-redundant info in the codewords of C.
- the $[7,4]$ Hamming code has distance 3 & rate $R(C) = \frac{4}{7}$
- a corrupted message x' is said to have k errors w.r.t. its uncorrupted version x if $d(x-x') = k$.

Note: A code with distance d allows to correct

(i) $\left\lfloor \frac{(d-1)}{2} \right\rfloor$ errors,

(ii) $(d-1)$ symbol erasures.

proof: just choose the codeword closest in Hamming distance. □

Def.: Let $\mathcal{C} = \{C_i\}_{i \in \mathbb{N}}$ be a sequence of codes with lengths n_i so that $n_{i+1} > n_i$. $\mathcal{C}$ is called "asymptotically good" if

$\liminf_i R(C_i)$ and $\liminf_i \left(\frac{d(C_i)}{n_i} \right)$ are both strictly positive.