ENTROPY IS NOT NEGATIVE

$$H(x) = \mathbb{E}\left[\log_2\left(\frac{1}{P(x)}\right)\right] = \sum_{x \in \mathcal{A}_X} \underbrace{P(x)}_{\geq 0} \cdot \overbrace{\log_2\left(\underbrace{\frac{1}{P(x)}}_{\geq 0}\right)}^{\geq 0} \geq 0$$

ENTROPY CHAIN RULE

$$H(x, Y) = H(x|Y) + H(Y)$$

PROOF

$$H(X,Y) = \mathbb{E}\left[\log_2\left(\frac{1}{P(x,Y)}\right)\right] = \mathbb{E}\left[\log_2\left(\frac{1}{P(x|Y)\cdot P(Y)}\right)\right] = \mathbb{E}\left[\log_2\left(\frac{1}{P(x|Y)}\right) + \log_2\left(\frac{1}{P(Y)}\right)\right] =$$

$$= H(x|Y) + H(Y) \quad \square$$

FANO'S INEQUALITY

$$H(x|Y) \leq H(Pe) + Pe \log_2(M-1), \quad M = \mathrm{CARD}(\mathcal{A}_x)$$

LET BE:

$$E = \begin{cases} 0 & \text{IF } \hat{X} = X \\ 1 & \text{, OTHERWISE} \end{cases} \qquad \text{SO } E \sim \mathcal{B}(Pe)$$

PROOF

$$H(E, x|Y) = \begin{cases} \underbrace{H(E|x,Y)}_{\text{IT'S 0 SINCE } \begin{matrix} X \text{ KNOWN} \\ Y \text{ KNOWN} \end{matrix} \Rightarrow E \text{ KNOWN}} + H(x|Y) \\ \\ H(x|E,Y) + H(E|Y) \end{cases}$$

$$\Rightarrow H(x|Y) = H(x|E,Y) + \underbrace{H(E|Y)}_{\leq H(E) = H(Pe)}$$

$$= Pe\, \underbrace{H(x|Y, E=1)}_{\leq \log_2(M-1) \text{ IT'S THE LARGEST ENTROPY}} + (1-Pe)\underbrace{H(x|Y, E=0)}_{\text{IT'S 0 SINCE } X = \hat{X} = Y}$$

$$\Rightarrow H(x|Y) \leq H(Pe) + Pe \log_2(M-1) \quad \square$$

ACHILLE CANNAVALE

DATA PROCESSING INEQUALITY

LET $X \to Y \to Z$ BE A MARKOV'S CHAIN, X, Y, Z R.V.s

$$\boxed{I(X;Z) \leq I(X;Y)}$$

PROOF

$$I(X;Y,Z) = \begin{cases} I(X;Y|Z) + I(X;Z) \\ \underbrace{I(X;Z|Y)}_{0 \text{ SINCE } X \text{ AND } Z|Y \text{ ARE INDEPENDENT}} + I(X;Y) \end{cases}$$

$$\Rightarrow I(X;Y) = \underbrace{I(X;Y|Z)}_{\geq 0} + I(X|Z) \Rightarrow I(X;Z) \leq I(X;Y) \quad \square$$

STATIONARY PROCESS

$\{X_i\}_{i \in \mathbb{Z}}$ IS A STATIONARY PROCESS IFF:

$$\mathbb{P}((X_n, \dots, X_m) \in A) = \mathbb{P}((X_{n-\Delta}, \dots, X_{m-\Delta}) \in A)$$

$\forall n \in \mathbb{Z}, \ \forall m \geq n, \ m \in \mathbb{Z}, \ \forall A, \ \forall \Delta \in \mathbb{N}$

ENTROPY RATE

IF $\exists \lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n)$,

THEN $H_\infty = \lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n)$ $\left[\text{bit}/\text{SYMBOL} \right]$

KRAFT'S INEQUALITY

1. C IS AN INSTANTANEOUS D-ARY SOURCE CODE WITH CODEWORD LENGTH $\{l_i\}_{i=1}^{m}$, $m < \infty$ $\Rightarrow$ $\sum_{i=1}^{m} D^{-l_i} \leq 1$

2. IF $\{l_i\}_{i=1}^{m}$, $m < \infty$ ARE SUCH THAT $l_i \in \mathbb{N}, \forall i$ AND $\sum_{i=1}^{m} D^{-l_i} \leq 1$ $\Rightarrow$ $\exists$ A D-ARY INSTANTANEOUS CODEWORD WITH CODEWORD LENGTH: $\{l_i\}_{i=1}^{m}$

PROOF 1

Every D-ary code with maximum codeword l_{MAX} can be represented with a d-ary tree.

If the code is prefix and a node is used for a codeword, then all the descending nodes cannot be used, otherwise the corresponding codewords would violate the prefix condition.

$\Longrightarrow$ A codeword of length l_i eliminates $D^{l_{MAX}-l_i}$ nodes at depth l_{MAX}

$$\underbrace{\sum_{i=1}^{m} D^{l_{MAX}-l_i}}_{\text{* ELIMINATED TERMINAL NODES}} \leq \underbrace{D^{l_{MAX}}}_{\text{* TERMINAL NODES}} \Rightarrow \sum_{i=1}^{m} D^{-l_i} \leq 1 \quad \square$$

PROOF 2

IF $l_i \in \mathbb{N}$ AND $\sum_{i=1}^{m} D^{-l_i}$ $\Rightarrow$

1) Build a d-ary tree of depth $l_{MAX} = MAX\{l_1, \dots l_M\}$

2) Place a codeword on a node of depth l_i

3) Remove all descending nodes

4) Iterate until there is no l_i available

5) Assign a label to every branch of the tree

6) Assign a codeword to every node by reading the labels from the root to the node

$\square$

STRONG LAW LARGE NUMBER

$\{x_i\}_{i\in\mathbb{N}}$ i.i.d. r.v.

$\mathbb{E}[|x_i|] < \infty \quad \forall i \in \mathbb{N}$

LET $\bar{x}_m = \frac{1}{m}\sum_{i=1}^{m} x_i$, THEN $\bar{x}_m \xrightarrow[m\to\infty]{} \mathbb{E}[x_i]$ ALMOST SURELY AND IN MEAN SQUARE.

$\mathbb{E}[x_i] = \mu$

PROOF (WEAK LAW): $VAR(x_i) = \sigma^2 < \infty$, CONV. IN $\mathbb{P}$

$$\forall \varepsilon > 0 \quad \mathbb{P}(|\bar{x}_m - \mu| > \varepsilon) \xrightarrow[m\to\infty]{} 0$$

$$\mathbb{P}(|\bar{x}_m - \mu| > \varepsilon) = \mathbb{P}(|\bar{x}_m - \mu|^2 > \varepsilon^2) \underbrace{\leq}_{\text{MARKOV'S INEQUALITY}} \frac{\mathbb{E}[(\bar{x}_m - \mu)^2]}{\varepsilon} = \frac{VAR(\bar{x}_m)}{\varepsilon} = *$$

$$\mathbb{E}[\bar{x}_m] = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} x_i\right] = $$
$$= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[x_i] = \mu$$

$$VAR(\bar{x}_m) = VAR\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) = \frac{1}{m}\left(\sum_{i=1}^{m}\underbrace{VAR(x_i)}_{\sigma^2} + \sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}\underbrace{VAR(x_i, x_j)}_{0 \Leftarrow \text{ i.i.d}}\right)$$

$$* = \frac{\sigma^2}{m\varepsilon} \xrightarrow[m\to\infty]{} 0 \quad \square$$

ASYMPTOTIC EQUIPARTITION PROPERTY

$\{x_i\}_{i\in\mathbb{N}}$ i.i.d

$H(X_i) = H(X) < \infty$

$\frac{1}{m}\log_2\left(\frac{1}{P(x_1,\dots,x_m)}\right) \xrightarrow[m\to\infty]{} H(X)$ ALMOST SURE AND IN MEAN SQUARE.

PROOF

$$\frac{1}{m}\log_2\left(\frac{1}{P(x_1,\dots,x_m)}\right) = \frac{1}{m}\sum_{i=1}^{m}\underbrace{\log_2\left(\frac{1}{P(x_i)}\right)}_{Y_i}$$

IF IT SATISFIES STRONG LAW LARGE NUMBER $\Rightarrow \xrightarrow[m\to\infty]{} \mathbb{E}[Y_i] = \mathbb{E}\left[\log_2\left(\frac{1}{P(x_i)}\right)\right] = H(X)$

- i.i.d? YES, CAUSE Y_i IS A FUNCTION OF x_i THAT IS i.i.d.
- $\mathbb{E}[|Y_i|] < \infty$? YES, CAUSE $\log_2(\cdot)$ IS POSITIVE. $\square$

TYPICAL SET

LET $\{X_i\}_{i\in\mathbb{N}}$ iid , $H(X)<\infty$

$$A_\varepsilon^{(m)} = \left\{ \underline{x} = (x_1,\dots,x_m) \in \mathcal{A}_X^m : \left| \frac{1}{m} \log_2\left(\frac{1}{P(\underline{x})}\right) - H(X) \right| < \varepsilon \right\}$$

PROP. 1

$$\mathbb{P}\left((x_1,\dots,x_m) \in A_\varepsilon^{(m)} \right) > 1-\varepsilon \quad \forall \varepsilon > 0 \text{, FOR } m \text{ SUFF. LARGE.}$$

PROOF. 1

SINCE $\mathbb{P}\left((x_1,\dots,x_m) \in A_\varepsilon^{m} \right) \xrightarrow[m\to\infty]{} 1 \Rightarrow \mathbb{P}\left((x_1,\dots,x_m) \in A_\varepsilon^{m} \right) > 1-\varepsilon$

$\forall \varepsilon > 0$, FOR m SUFF. LARGE

PROP. 2

$$\forall \varepsilon > 0 \quad \mathrm{CARD}\left(A_\varepsilon^{(m)}\right) \leq 2^{m(H(X)+\varepsilon)}$$

PROOF 2

FROM THE DEF. OF TYPICAL SET:

$$-\frac{1}{m} \log_2\left(P(\underline{x})\right) - H(\underline{X}) < \varepsilon \Rightarrow P(\underline{x}) \geq 2^{-m(H(X)+\varepsilon)}$$

$$-\left(-\frac{1}{m} \log_2\left(P(\underline{x})\right) - H(\underline{x})\right) < \varepsilon \rightarrow P(\underline{x}) \leq 2^{-m(H(X)-\varepsilon)}$$

NOW WE CAN WRITE: $1 = \sum_{x\in\mathcal{A}_X^m} P(\underline{x}) \geq \sum_{x\in A_\varepsilon^{(m)}} P(\underline{x}) \geq \sum_{x\in A_\varepsilon^{(m)}} 2^{-m(H(\underline{X})+\varepsilon)} =$

$= \mathrm{CARD}\left(A_\varepsilon^{(m)}\right) \cdot 2^{-m(H(\underline{X})+\varepsilon)} \Rightarrow 1 \geq \mathrm{CARD}\left(A_\varepsilon^{(m)}\right) \cdot 2^{-m(H(\underline{X})+\varepsilon)} \Rightarrow$

$\Rightarrow 2^{m(H(\underline{X})+\varepsilon)} \geq \mathrm{CARD}\left(A_\varepsilon^{(m)}\right)$

PROP. 3

$$\forall \varepsilon > 0 \quad \mathrm{CARD}\left(A_\varepsilon^{(m)}\right) > (1-\varepsilon) \cdot 2^{m(H(X)-\varepsilon)} \text{, FOR } m \text{ SUFF. LARGE}$$

PROOF. 3

FOR m SUFF. LARGE $\mathbb{P}\left(\underline{x} \in A_\varepsilon^{(m)}\right) > 1-\varepsilon$, SO:

$$1-\varepsilon < P\left(\underline{x} \in A_\varepsilon^{(m)}\right) \overset{②}{\leq} \sum_{x\in A_\varepsilon^{(m)}} 2^{-m(H(\underline{X})-\varepsilon)} = \mathrm{CARD}\left(A_\varepsilon^{(m)}\right) \cdot 2^{-m(H(\underline{X})-\varepsilon)} \Rightarrow$$

$$\Rightarrow \mathrm{CARD}\left(A_\varepsilon^{m}\right) > (1-\varepsilon)\, 2^{m(H(\underline{X})-\varepsilon)}, \quad m \longrightarrow \infty$$

ACHILLE CANNAVALE

SOURCE CODE

A SOURCE CODE FOR A R.V. $\mathcal{X}$ IS AN APPLICATION:

$$C : \mathcal{A}_x \longrightarrow \mathbb{D}^*$$

ALPHABET OF $\mathcal{X}$

SET OF ALL FINITE LENGTH STRING OF SYMBOLS TAKEN FROM THE D-ARY ALPHABET:

$$\mathbb{D} = \{1, \dots, D-1\}$$

NON-SINGULAR

A CODE IS NON-SINGULAR IF C IS INJECTIVE:

$$\text{IF } \forall x_1 \neq x_2 \Rightarrow C(x_1) \neq C(x_2)$$

EXTENDED CODE

THE EXTENSION C^* OF A CODE C IS THE APPLICATION:

$$C^* : \mathcal{A}_x^* \longrightarrow \mathbb{D}^* = C^*(x_1, \dots x_n) =$$

$$= C(x_1)\, C(x_2) \dots C(x_n)$$

SET OF ALL FINITE-LENGTH SEQUENCES OF SYMBOLS TAKEN FROM $\mathcal{A}_x$.

UNIQUELY DECODABLE

A CODE C IS UNIQUELY DECODABLE IF C^* IS NON-SINGULAR.

PREFIX CODE

C IS A PREFIX CODE IF IT SATISFIES THE PREFIX CONDITION:

NO CODEWORD IS PREFIX OF ANY CODEWORD

ACHILLE CANNAVALE

PROVE THAT EXPECTED LENGTH OF ANY PREFIX CODE IS ALWAYS GREATER THAN ENTROPY

C IS A D-ARY PREFIX CODE FOR x R.V.

$$L = \mathbb{E}[\ell(x)] \geq H(x)$$

= IFF P(x) IS D-ADIC

PROOF

$$L - H(x) = \mathbb{E}[\ell(x)] - \mathbb{E}\left[\log_D\left(\frac{1}{P(x)}\right)\right] = \mathbb{E}\left[\ell(x) - \log_D\left(\frac{1}{P(x)}\right)\right] =$$

$$= \mathbb{E}\left[-\log_D\left(D^{-\ell(x)}\right) - \log_D\left(\frac{1}{P(x)}\right)\right] = \mathbb{E}\left[\log_D\left(\frac{P(x)}{D^{-\ell(x)}}\right)\right] =$$

$$= \mathbb{E}\left[\log_D\left(\frac{P(x)}{\underbrace{\frac{D^{-\ell(x)}}{\sum_{y \in A_x} D^{-\ell(y)}}}_{q(x)}}\right)\right] + \mathbb{E}\left[\log_D\left(\frac{1}{\sum_{y \in A_x} D^{-\ell(y)}}\right)\right] =$$

IT'S JUST A NUMBER

$$= \mathbb{E}_p\left[\log_D\left(\frac{P(x)}{q(x)}\right)\right] + \log_D\left(\frac{1}{\sum_{y \in A_x} D^{-\ell(y)}}\right) =$$

$$= \underbrace{D_p(P(x) \,//\, q(x))}_{\geq 0} + \underbrace{\log_D\left(\frac{1}{\sum_{y \in A_x} D^{-\ell(y)}}\right)}_{\geq 0}$$

PREFIX $\Rightarrow$ KRAFT $\Rightarrow \leq 1$

$$\Rightarrow L - H(x) \geq 0 \quad \square$$

BOUNDS TO THE AVERAGE LENGTH

C IS AN OPTIMAL CODE $\Rightarrow \quad H(x) \leq L^* \leq H(x) + 1$

PROOF

LET C BE THE SHANNON CODE:

$$\Rightarrow \ell(x) = \left\lceil \log_D\left(\frac{1}{P(x)}\right) \right\rceil \quad \text{SINCE } a \leq \lceil a \rceil \leq a+1$$

$$\Rightarrow \log_D\left(\frac{1}{P(x)}\right) \leq \ell(x) \leq \log_D\left(\frac{1}{P(x)}\right) + 1$$

$$\xrightarrow{\mathbb{E}[\cdot]} \Rightarrow \quad H(x) \leq L \leq H(x) + 1 \quad \square$$

ACHILLE CANNAVALE

CHANNEL CODE

AN (M, m) CHANNEL CODE IS THE MAPPING:

$$\underline{x}_m : \{1, \dots, M\} \to \mathcal{A}_x^m$$

$$\{\underline{x}_m(1), \dots, \underline{x}_m(M)\} = \text{CODEBOOK}$$

$\underline{x}_m(M) \to$ CODEWORD

DECODING RULE

A DECODING RULE IS THE MAPPING:

$$g : \mathcal{A}_y^m \longrightarrow \{1, \dots, M\}$$

$$\hat{W} = g(\underline{y}_m) = \text{ESTIMATE OF } W$$

RATE

THE RATE R OF A (M, m) CHANNEL CODE IS:

$$R = \frac{\log_2(M)}{m} \left[\text{bits} / \text{CHANNEL USES} \right]$$

ACHIEVABLE RATE

THE RATE R IS ACHIEVABLE IF $\exists$ A SEQUENCE OF $(\lceil 2^{mR} \rceil, m)$ CHANNEL CODES:

$$P_m^{MAX}(e) \xrightarrow[m \to \infty]{} 0$$

ACHILLE CANNAVALE

CHANNEL CODING THEOREM

1. $R < C \Rightarrow \exists$ A SEQUENCE OF $(\lceil 2^{mR} \rceil, m)$ CHANNEL CODES: $P_m^{MAX}(e) \xrightarrow[m \to \infty]{} 0$

2. IF FOR A SEQUENCE OF $(\lceil 2^{mR} \rceil, m)$ CHANNEL CODES: $P_m^{MAX}(e) \xrightarrow[m \to \infty]{} 0 \Rightarrow R \leq C$

PROOF

LET $M = \lceil 2^{mR} \rceil \approx 2^{mR}$, THEN

1) RANDOM CODING

THE (M, m) CHANNEL CODE IS GENERATED AS FOLLOWS:

$$\begin{pmatrix} x_1(1) & \dots & x_m(1) \\ x_1(2) & \dots & x_m(2) \\ \vdots & & \\ x_1(M) & \dots & x_m(M) \end{pmatrix} = \begin{pmatrix} \underline{x}_m(1) \\ \underline{x}_m(2) \\ \vdots \\ \underline{x}_m(M) \end{pmatrix} = \text{CODEBOOK}$$

THE ENTRIES OF THIS MATRIX ARE i.i.d. r.v. DRAWN FROM THE DISTRIBUTION $p^*(x) = \underset{p(x)}{\text{ARGMAX}}\ I(x; Y)$

THE CODE IS SHARED WITH THE DESTINATION.

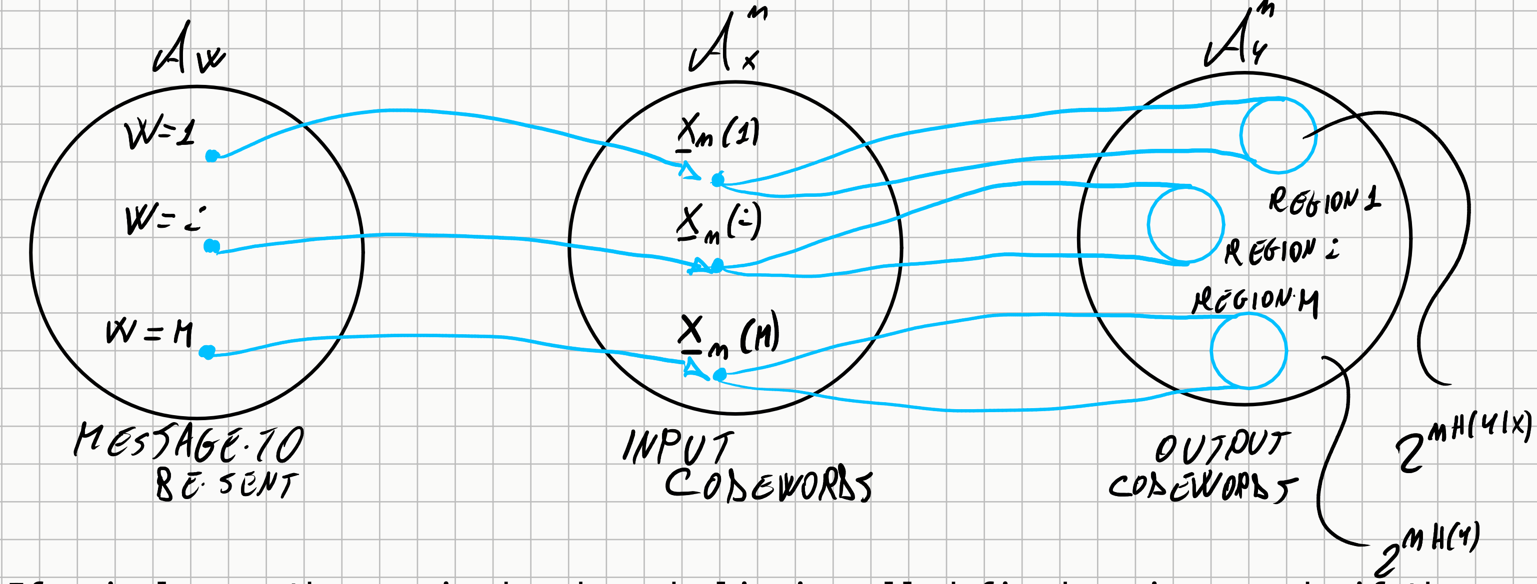
2) JOINTLY TYPICAL DECODING RULE

$\hat{W} = g(\underline{Y}_m) = i \Leftrightarrow \underline{X}_m(i)$ (THE CODEWORD ASSOCIATED TO $W = i$) IS THE ONLY SEQUENCE JOINTLY TYPICAL WITH $\underline{Y}_m$, i.e.

$$\begin{cases} (\underline{X}_m(i), \underline{Y}_m) \in A_\varepsilon^{(m)} \\ (\underline{X}_m(j), \underline{Y}_m) \notin A_\varepsilon^{(m)} \quad \forall j \neq i \end{cases}$$

ACHILLE CANNAVALE

3) SPHERE



If n is large, the received codewords lie in well defined regions, and, if these regions do not overlap, there is no error in decoding; this is possible only if the rate is sufficiently small.

(R SMALL $\Rightarrow$ M SMALL $\Rightarrow$ FEW REGIONS)

BUT:
- THE TOTAL NUMBER OF TYPICAL (RECEIVED) SEQUENCES IS $2^{nH(Y)}$
- $\forall\, \underline{X}_n$ TYPICAL, THERE ARE $2^{nH(Y|X)}$ $\underline{Y}_n$ TYPICAL.
- THE NUMBER OF REGIONS IS M.

IN ORDER TO HAVE NON-OVERLAPPED REGIONS, WE MUST HAVE:

$$\underbrace{M}_{\#\text{ OF REGIONS}}\ \underbrace{2^{nH(Y|X)}}_{\#\text{ OF CODEWORDS IN EACH REGION}} \leq \underbrace{2^{nH(Y)}}_{\text{TOTAL }\#\text{ OF CODEWORDS.}}$$

SINCE $M = 2^{nR}$

$$2^{nR}\, 2^{nH(Y|X)} \leq 2^{nH(Y)}$$

$$2^{nR} \leq 2^{n(H(Y)-H(Y|X))} = 2^{nI(X;Y)} = 2^{nC}$$

SINCE $P^*(x) = \underset{P(x)}{\mathrm{ARGMAX}}\ I(X;Y)$

$\Rightarrow R \leq C$ $\square$

CAPACITY OF A GAUSSIAN CHANNEL WITH POWER CONSTRAINT

$$C = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

AND CAN BE ACHEIVED WITH: $x \sim \mathcal{N}(0,P)$

PROOF

$$I(x;y) = h(y) - \underbrace{h(y|x)}_{} =$$

$$\hookrightarrow h(x+z|x) = h(z|x) \overset{z \text{ AND } x \text{ ARE INDEP.}}{=} h(z) = \frac{1}{2}\log_2(2\pi e N)$$

$\hookrightarrow z \sim \mathcal{N}(0,N)$

$$= h(y) - \frac{1}{2}\log_2(2\pi e N)$$

$$VAR(y) = VAR(x+z) \overset{x,z \text{ INDEP}}{=} VAR(x) + VAR(z) = \underbrace{\mathbb{E}[x^2]}_{\leq P} - \underbrace{(\mathbb{E}[x])^2}_{\geq 0} + N \leq P+N$$

$$\Rightarrow h(y) \leq \frac{1}{2}\log_2(2\pi e(P+N))$$

"=" IF $y \sim \mathcal{N}(0,(P+N))$

IF $x \sim \mathcal{N}(0,P)$

$$\Rightarrow I(x;y) = h(y) - \frac{1}{2}\log_2(2\pi e N) \leq \frac{1}{2}\log_2(2\pi e(P+N)) - \frac{1}{2}\log_2(2\pi e N)$$

$$= \frac{1}{2}\log_2\left(\frac{P+N}{N}\right) = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

$$\Rightarrow C = \max_{P_x \leq P}\{I(x;y)\} = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

FOR $x \sim \mathcal{N}(0,P)$ $\square$

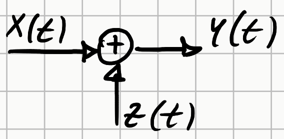
ACHILLE CANNAVALE

BAND-LIMITED GAUSSIAN CHANELL

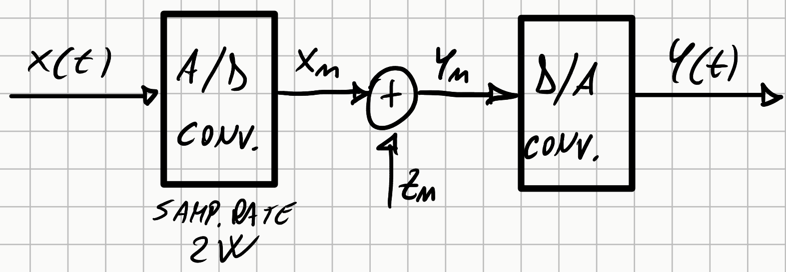
$$C = W \log_2\left(1+\frac{P}{N_0 W}\right) \left[bit/s\right]$$

PROOF

THE CHANNEL:



HAS THE SAME CAPACITY AS



$$C = \frac{1}{2}\log_2\left(1+\frac{P}{N_0 W}\right) \left[bit/ch.\ us.\right]$$

SINCE IT IS USED 2W TIMES PER SECOND WE HAVE:

$$C = \underbrace{2W}_{\left[CH\ USE/s\right]} \cdot \underbrace{\frac{1}{2}\log_2\left(1+\frac{P}{N_0 W}\right)}_{\left[bit/CH.\ USE\right]} \Longrightarrow$$

$$\Longrightarrow W \cdot \log_2\left(1+\frac{P}{N_0 W}\right) \left[bit/s\right] \quad \square$$