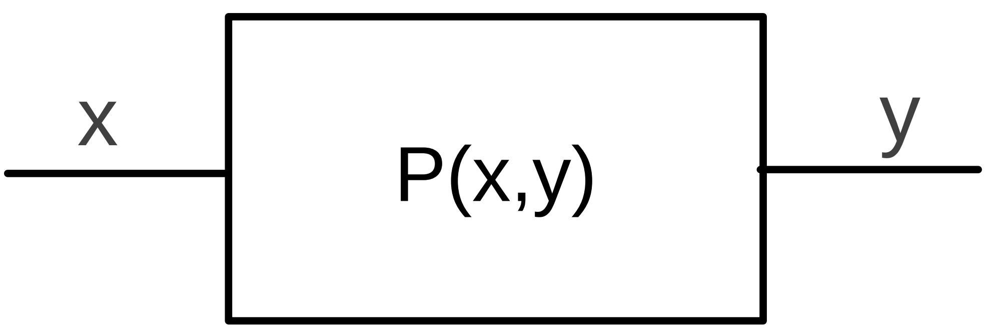
Mutual information and channel coding

S-72.340

Channel probabilities

- Information theory exploits a type of channel model which consists of probabilities related to the input and output symbols of the channel
- These probabilities may be derived using modulation theoretical studies<



$$P(x,y) = P(y|x) \cdot P(x) \tag{4.1.1}$$

$P(x)$ a priori probability of input symbol

$P(y|x)$ transition probability from input to output.

Mutual information

- Received symbols give information about the transmitted symbols.
- In reasoning about the input symbol the probability structure of the channel should be exploited. Here the concept of mutual information is useful.
- Assume the symbols x are one of the set $\{a_\ell\}$. (ℓ=1,...,L) The event $y=b_{\ell'}$ (ℓ'=1,...,L') give information of the event $x=a_\ell$

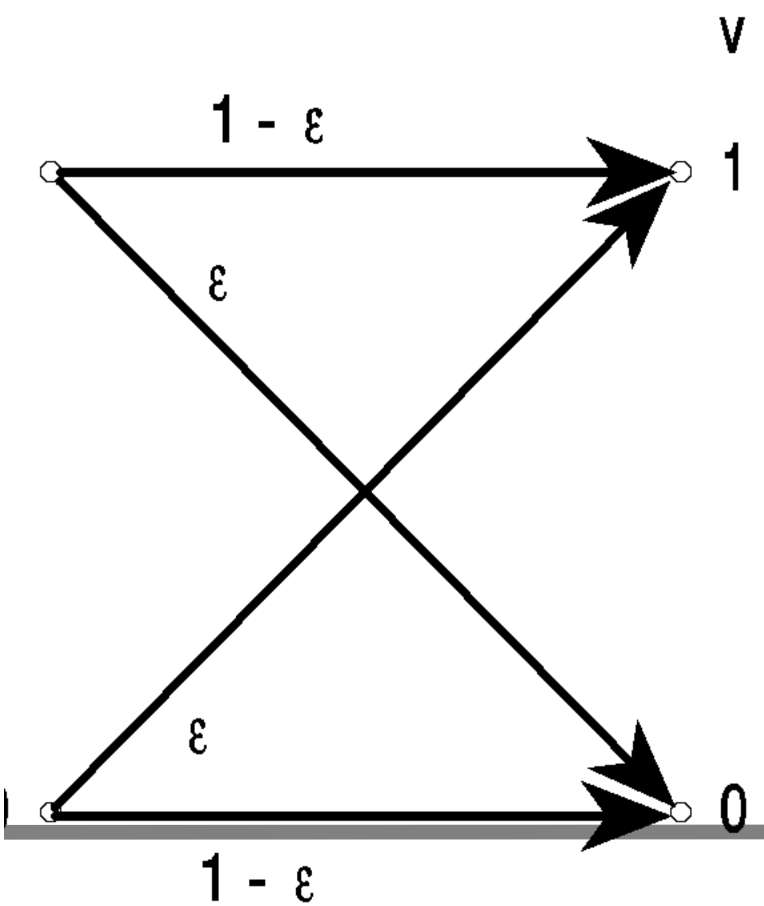
$$I(a_\ell , b_{\ell'}) = \log[P(x=a_\ell|y=b_{\ell'})/P(x=a_\ell)] \qquad (4.1.2)$$

- Here P(x|y) is the a posteriori probability. It can be computed using the Bayes rule:

$$P(x|y) = P(y|x)\cdot P(x)/P(y) \qquad (4.1.4)$$

Binary symmetric channel

- Example 4.1. A sagittal diagram is used to describe the binary symmetric channel (BSC).



The channel has u as input and v as output. Both input and output symbols are binary (0 or 1). An error probability ε is assumed.

BSC comes naturally in an FSK modem. (V.21 or V.23)

Computation of mutual information for BSC

$P(0|1) = \varepsilon$

$P(1|1) = 1- \varepsilon$

- Assume $P(0) = P(1) = 0.5$. Assume $\varepsilon = 1/65536 = 2^{-16}$. Then

  $I(1,1) = I(0,0) = \log(2\cdot(1- \varepsilon))$ » 1,0 bits

  $I(1,0) = I(0,1) = \log(2\cdot\varepsilon) = -15$ bits. (4.1.5)

- Without errors about 1 bit of information is conveyed. If an error happens much negative information is received.

Average mutual information

- Mutual information is a random variable. Find its average

$$I(U,V) = \sum_{u} \sum_{v} P(u,v) \cdot \log[P(u|v)/P(u)] \qquad (4.1.6)$$

- For P(u|v) = 1 and assuming same alphabet for input and output I(U,V=H(U).
- Average mutual information can be expressed as difference of two entropy functions:
- $I(U,V) = -\Sigma\,\Sigma\, P(u,v) \cdot \log[P(u)] - \{-\Sigma\,\Sigma\, P(u,v) \cdot \log[P(u|v)]\}$
  $= H(U) - H(U|V), \qquad (4.1.7)$
- The average mutual information is seen equal to the entropy of the source from which the average equivocation H(U|V) of the observation is subtracted.

Properties of average mutual information

- I(U,V) is non-negative:

  $$- I(U,V) = \log(e)\cdot\Sigma\,\Sigma P(u,v)\cdot\ln[P(u)/P(u|v)]$$
  $$\leq \log(e)\cdot\Sigma\,\Sigma P(u,v)\cdot[P(u)/P(u|v) - 1] \qquad (4.1.8)$$
  $$= \log(e)\cdot[\Sigma\,\Sigma P(u)\cdot P(v) - \Sigma\,\Sigma P(u,v)]$$
  $$\leq 0,$$

- Equality applies here whenever U and V are statistically independent.

Channel capacity

- Capacity of a channel may be defined as the maximal value of the average mutual information with respect to the choice of the probabilities of the source alphabet.
- We assume that the channel is discrete and memoryless. Denote: $P(a_\ell) = Q_\ell$, $P(b_{\ell'}|a_\ell) = P(\ell'|\ell)$. Define

$$C = \max_{Q_\ell} \left\{ \sum_{\ell',\ell} Q_\ell \cdot P(\ell'|\ell) \cdot \log\left[P(\ell'|\ell) / \left(\sum_{\ell''} Q_{\ell''} \cdot P(\ell'|\ell'') \right) \right] \right\} \quad (4.1.9)$$

BSC capacity

- Set the binary alphabet 0,1 equally probable. For error probability ε

  $$C = 1 - \mathcal{H}(\varepsilon), \qquad (4.1.10)$$

- The value of the capacity is 1 for $\varepsilon = 0$ or 1 and 0 for $\varepsilon = 0$.

4.2. Converse of the coding theorem

- It is easier to prove the converse: For the information rate higher than the channel capacity the error rate has a positive lower bound, in fact it is very high.
- A discrete source produces symbols $[u_1, u_2, \dots u_K]$

$$H_K(U) = H(\mathbf{U}_K)/K = -(1/K)\cdot \sum_{\mathbf{u}} p(\mathbf{u})\cdot \log[p(\mathbf{u})] \qquad (4.2.1)$$

- For a stationary source this entropy decreases monotonously and has the limit $H_\infty(U)$, kun $K \to \infty$.
- Now define the average probability of error for a sequence of length K

$$\langle P_e \rangle = (1/K)\cdot \sum_{k=1}^{K} P_{e,k} \qquad (4.2.2)$$

Proof of the converse

- Assume that the input and output alphabet a_m of the channel are the same: The probability that the input and output differ

$$P_e = \sum_{u=a_m}^{M} \sum_{v \neq u} P(u,v). \qquad (4.2.3)$$

- Then the following inequality is true:

$$P_e \cdot \log(M-1) + \mathcal{H}(P_e) > H(U|V). \qquad (4.2.4)$$

- This sets a lower bound to the error probability P_e which is determined by the equivocation $H(U|V)$ of the channel. The lower bound may be found by solving the nonlinear inequality.

Proof continued

- Write out the equivalocation

$$H(U|V) = \sum_{v}\sum_{u\neq v} P(u,v)\cdot\log(1/P(u|v)) + \sum_{v,u=v} P(u,v)\cdot\log(1/P(u|v)) \qquad (4.2.5)$$

- Subtract from this the left side of eq. (4.2.4).

$$H(U|V) - P_e\cdot\log(M-1) - \mathcal{H}(P_e) =$$

$$\sum_{v}\sum_{u\neq v} P(u,v)\cdot\log\{P_e/[(M-1)\cdot P(u|v)]\} + \sum_{v,\, u=v} P(u,v)\cdot\log[(1-P_e)/P(u|v)]$$

- Now apply the logarithmic inequality

$$\leq \log(e)\cdot \ \{\sum_{v,\, u\neq v} P(u,v)\cdot[P_e/((M-1)\cdot P(u|v))-1] +$$

$$+ \sum_{v,\, u=v} P(u,v)\cdot[(1-P_e)/P(u|v) -1]\}$$

Proof continued #2

$$= \log(e)\cdot\{[P_e/(M-1)]\cdot \sum_{v}\sum_{u\neq v} P(v) - \sum_{v}\sum_{u\neq v} P(u,v) +$$

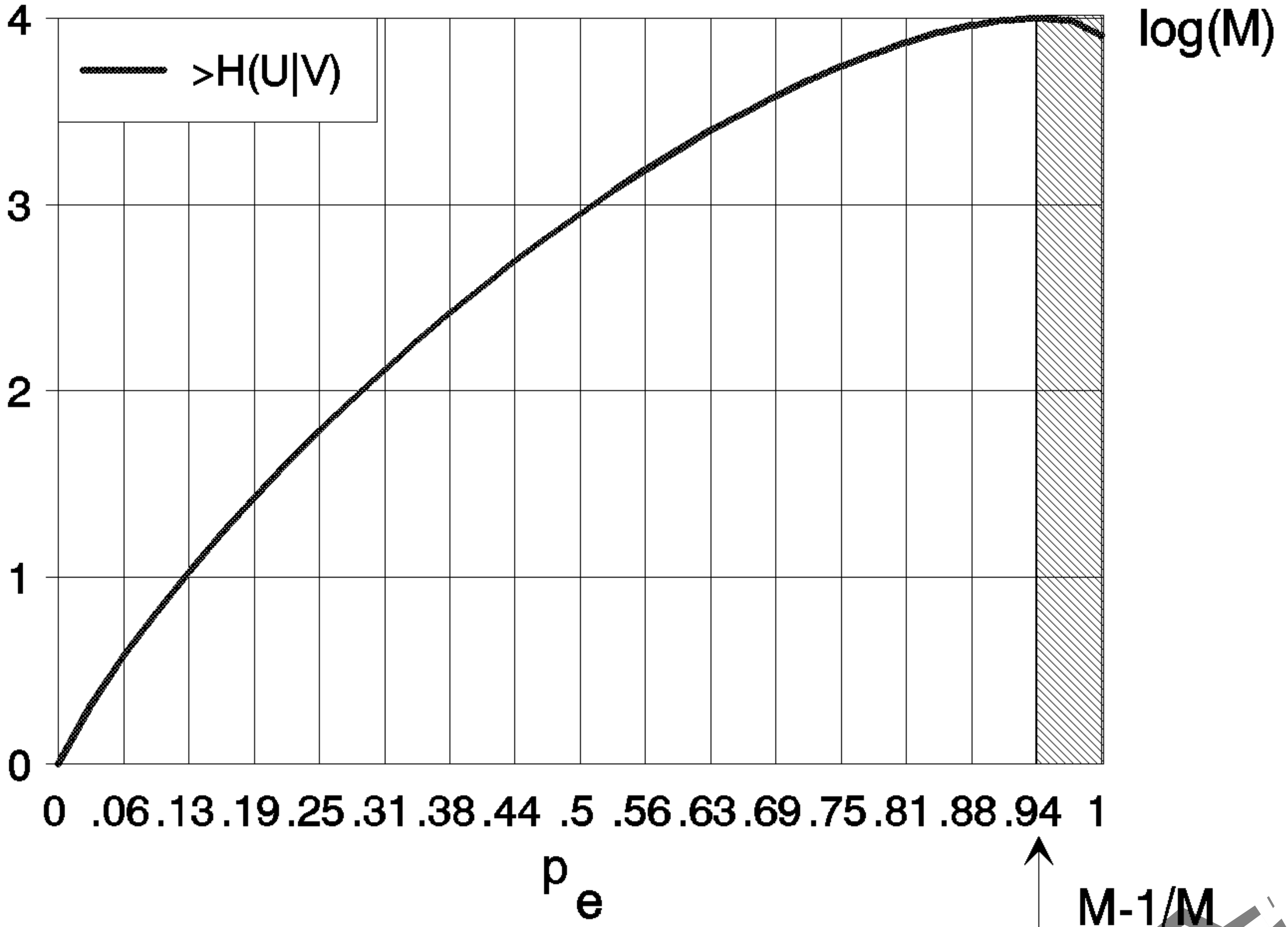
$$+ (1-P_e)\cdot \sum_{v} P(v) - \sum_{v,\, u=v} P(u,v)\}$$

$$= \log(e)\cdot[P_e - P_e + (1-P_e) - (1-P_e)]$$

$$= 0, \qquad (4.2.6)$$

► QED.

Interpretation of equivocation upper bound



Upper bound of equivocation for sequences

- We have to prove the following:

$$\langle P_e \rangle \cdot \log(M-1) + \mathcal{H}(\langle Pe \rangle) > H(\mathbf{U}_K|\mathbf{V}_K)/K, \qquad (4.2.7)$$

- where P_e is defined in (4.1.13). Setting $\mathbf{U}_K = \{U_1 \times U_2 \times \ldots \times U_K\}$

$$H(\mathbf{U}_K|\mathbf{V}_K) = H(U_1|\mathbf{V}_K) + H(U_2|U_1 \times \mathbf{V}_K) + \ldots + H(U_K|U_1 \times U_2 \times \ldots \times U_{K-1} \times \mathbf{V}_K)$$

$$\leq \sum_{k=1}^{K} H(U_k|V_k). \qquad (4.2.8)$$

- Entropy grows when conditioning is reduced.

Proof for sequences continued

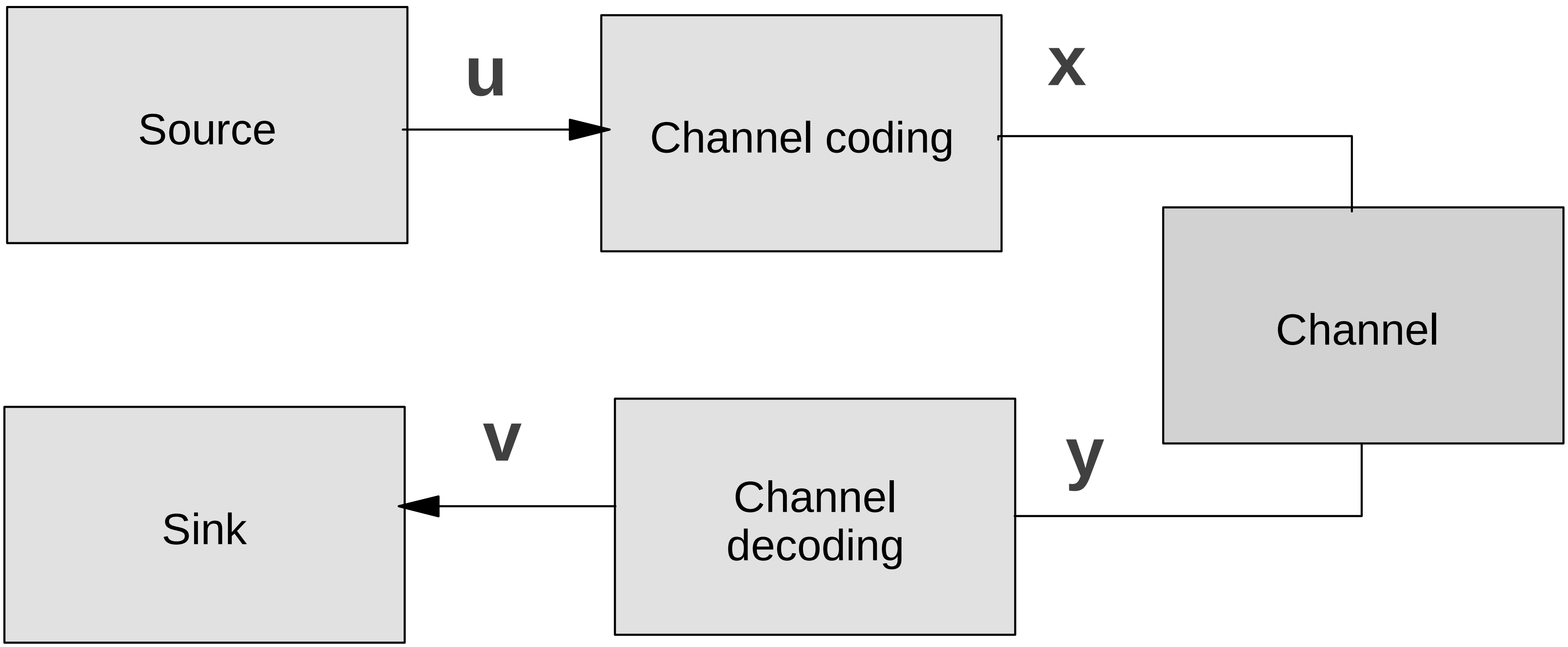
- Use inequality (4.2.4) for the right side of (4.2.8) divided by K:

$$\sum H(U_k|V_k)/K < \sum_{k=1}^{K} [P_{e,k}\cdot\log(M-1) + \mathcal{H}(P_{e,k})]/K$$

$$\leq \langle P_e \rangle\cdot\log(M-1) + \left[\sum_{k=1}^{K} \mathcal{H}(P_{e,k})\right]/K \qquad (4.2.9)$$

where the right side follows from convexity of $\mathcal{H}(.)$.

Model for a digital communication system



Data processing theorem

- The channel model indicates following probabilistic properties of the source sequence **u** = $[u_1, u_2, ..., u_K]$ within the definition space $U_K \times X_N \times Y_N \times V_K$.
- The output sequence of the channel **y** = $[y_1, y_2, ..., y_N]$ is inde- pendent of the source **u** when the channel input **x** = $[x_1, x_2, ..., x_N]$ is given.
- The input sequence of the sink **v** = $[v_1, v_2, ..., v_K]$ is independent of the sequences **u** or **x** when the channel output **y** is given. These properties mean simply that there are no other connections between the quantities in the digital channel model than those indicated in the picture,
- The channel is modeled with the mutual information I(**x**,y) while the source and sink are modeled with H(U) and H(V).

Data processing theorem #2

- Following inequality applies

  $I(\mathbf{U}_K;\mathbf{V}_K) \leq I(\mathbf{X}_N;\mathbf{Y}_N)$ (4.2.10)

- Proof: We start from the definitions

  $I(\mathbf{U}_K \times \mathbf{X}_N;\mathbf{Y}_N) = I(\mathbf{U}_K;\mathbf{Y}_N) + I(\mathbf{X}_N;\mathbf{Y}_N|\mathbf{U}_K)$ (4.2.11)

  $= I(\mathbf{X}_N;\mathbf{Y}_N) + I(\mathbf{U}_K;\mathbf{Y}_N|\mathbf{X}_N)$

- Because $I(\mathbf{U}_K;\mathbf{Y}_N|\mathbf{X}_N) = 0$,

  $I(\mathbf{U}_K;\mathbf{Y}_N) = I(\mathbf{X}_N;\mathbf{Y}_N) - I(\mathbf{X}_N;\mathbf{Y}_N|\mathbf{U}_K) \leq I(\mathbf{X}_N;\mathbf{Y}_N)$ (4.2.12)

- Similarly study $I(\mathbf{U}_K;\mathbf{Y}_N \times \mathbf{V}_K)$ and exploit the fact that $I(\mathbf{U}_K;\mathbf{V}_K|\mathbf{Y}_N) = 0$ to lead to (4.2.10).
- The conclusion drawn from the data processing theorem is that any data processing on the transmission path decreases the mutual information and thus deteriorates performance.

Converse of the channel coding theorem

- Putting together (4.2.7) ja (4.2.10) we get

$$\langle P_e \rangle \cdot \log(M-1) + H(\langle P_e \rangle) > \mathcal{H}(\mathbf{U}_K|\mathbf{V}_K)/K$$

$$= H_K(U) - I(\mathbf{U}_K;\mathbf{V}_K)/K \qquad (4.2.13)$$

$$> H_K(U) - I(\mathbf{X}_N;\mathbf{Y}_N)/K,$$

Here $H_K(U) = H(\mathbf{U}_K)/K$.

- For a discrete memoryless channel the capacity per symbol is defined by $I(\mathbf{X}_N;\mathbf{Y}_N) < N \cdot C$, so that

$$\langle P_e \rangle \cdot \log(M-1) + \mathcal{H}(\langle P_e \rangle) > H_K(U) - (N/K) \cdot C \qquad (4.2.14)$$

- Letting $K \to \infty$ we obtain the converse of the coding theorem:

$$\langle P_e \rangle \cdot \log(M-1) + \mathcal{H}(\langle P_e \rangle) > H_\infty(U) - (\tau_s/\tau_c) \cdot C, \qquad (4.2.15)$$

where τ_s is the duration of the source symbol and τ_c that of the channel symbol.

Comments on converse of coding theorem

- For source rate higher than the capacity of the channel the error probability has a positive lower bound.
- In fact the error probability in such case is very high.