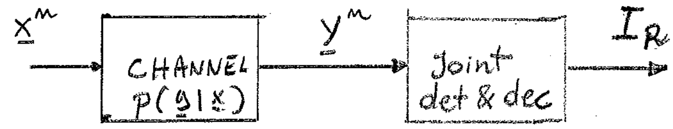
Information Theoretic Analysis for Detectors DECEMBER 2012

Advantages:
- it shows the potentiality of a detection algorithm when coupled with a good code
- it can reveal gains that cannot be noticed with BER test simulations
- ... and many others given by the information theory

MAIN DRAWBACK: it cannot be performed with all the detectors

ACHIEVABLE INFORMATION RATE



Let's suppose
- $\underline{x}^n$ be the vector of transmitted symbols.
- $\underline{y}^n$ a sufficient statistic
- $p(\underline{y}^n|\underline{x}^n)$ the probability density function (PDF) of the channel (well known as CHANNEL LAW)

If we use a receiver who knows the channel law $p(\underline{y}^n|\underline{x}^n)$, the highest rate I_R that we can achieve is the information rate (under suitable hypothesis):

$$I_R = \lim_{n \to +\infty} \frac{1}{n} I(\underline{x}^n; \underline{y}^n) = \lim_{n \to +\infty} \frac{1}{n} E\left\{ \log_2\left(\frac{p(\underline{y}^n|\underline{x}^n)}{p(\underline{y}^n)} \right) \right\} \quad (1)$$

Often I_R is hard to be computed in closed form.
The problem can be solved by means of the Monte Carlo method. This technique is known as ARNOLD & LOELIGER algorithm

ARNOLD & LOELIGER, (2006)

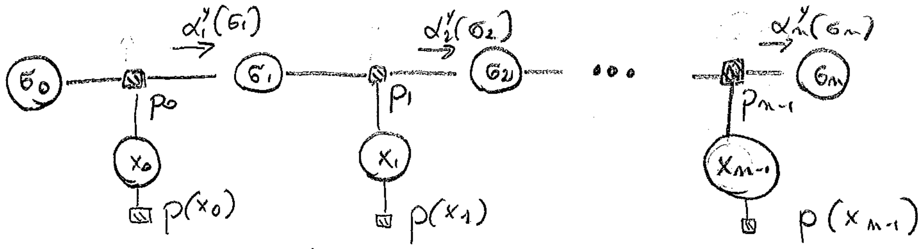
By means of the WEAK CENTRAL LIMIT THEOREM, eq. (1) can be computed by sampling two long sequeces $\tilde{y}^n$, $\tilde{x}^n$ and computing

$$\lim_{n \to +\infty} \frac{1}{n} \log\left(\frac{p(\tilde{\underline{y}}^n|\underline{x}^n)}{p(\tilde{\underline{y}}^n)} \right) \xrightarrow[\text{probability}]{\text{in}} \lim_{n \to +\infty} \frac{1}{n} E\left\{ \log\left(\frac{p(\underline{y}^n|\underline{x}^n)}{p(\underline{y}^n)} \right) \right\}$$

How can we compute $p(\tilde{\underline{y}}^n)$ and $p(\tilde{\underline{y}}^n|\tilde{\underline{x}}^n)$?
Let's use the Factor graphs (FG)!
Defining σ_n the state of the channel we have:

$$p(\underline{y}^n, \underline{x}^n, \underline{\sigma}^n) \propto p(\underline{y}^n|\underline{x}^n, \underline{\sigma}^n)\, p(\underline{\sigma}^n, \underline{x}^n)$$

$$= \prod_n p(y_n|x_n, \sigma_n)\, p(\sigma_{n+1}|\sigma_n, x_n)\, p(x_n) \quad (1)$$



$$p_i = p(y_i \mid x_i, \sigma_i)\, p(\sigma_{i+1} \mid \sigma_i, x_i)$$

It can be noticed that the FG is the same of the BCJR!
If we run the forward recursion on $\tilde{y}^n$

$$\alpha^y_{i+1}(\sigma_{i+1}) = \sum_{x_i, \sigma_i} \alpha^y_i(\sigma_i)\, p(\tilde{y}_i \mid x_i, \sigma_i)\, p(\sigma_{i+1} \mid \sigma_i, x_i)\, p(x_i),$$

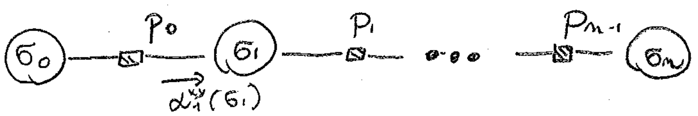
the last term is $\alpha^y_n(\sigma_n)$ who reads $p(\underline{\tilde{y}}^n, \sigma_n)$.
Thus,

$$\boxed{p(\underline{\tilde{y}}^n) = \sum_{\sigma_n} \alpha^y_n(\sigma_n)}$$

straightforwadly, we can derive $p(\underline{\tilde{y}}^n \mid \underline{\tilde{x}}^n)$:

$$p(\underline{y}^n, \underline{\sigma}^n \mid \underline{x}^n) \propto p(\underline{y}^n \mid \underline{x}^n, \underline{\sigma}^n)\, p(\underline{\sigma}^n \mid \underline{x}^n)$$

$$= \prod_k p(y_k \mid x_k, \sigma_k)\, p(\sigma_{k+1} \mid \sigma_k, x_k)$$



$$\alpha^{xy}_{i+1}(\sigma_{i+1}) = \sum_{\sigma_i} \alpha^{xy}_i(\sigma_i)\, p(\tilde{y}_i \mid \tilde{x}_i, \sigma_i)\, p(\sigma_{i+1} \mid \sigma_i, x_i)$$

$$\boxed{p(\underline{\tilde{y}}^n \mid \underline{\tilde{x}}^n) = \sum_{\sigma_n} \alpha^{xy}_n(\sigma_n)}$$

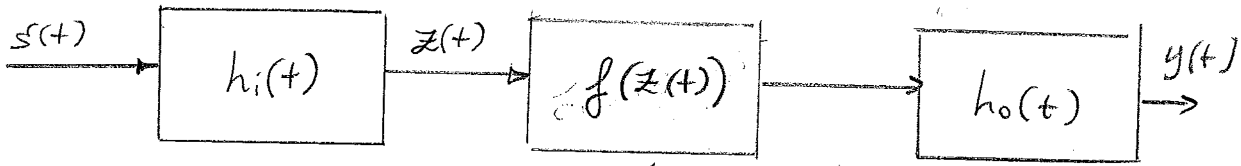
Summary:
- sample two long sequeses $\tilde{x}^n$ and $\tilde{y}^n$
- run the BCJR forward recursion on $\tilde{y}^n$ to compute $\alpha^y_n(\sigma_n) \rightarrow$ carry out $p(\underline{\tilde{y}}^n)$
- run again on $\tilde{x}^n$ and $\tilde{y}^n$ $\rightarrow p(\underline{\tilde{y}}^n \mid \underline{\tilde{x}}^n)$
- $I_R \simeq -\frac{1}{n} \log_2 p(\underline{\tilde{y}}^n) + \frac{1}{n} \log_2 p(\underline{\tilde{y}}^n \mid \underline{\tilde{x}}^n)$

ACHIEVABLE INFORMATION RATE with MISMATCHED DETECTORS

Sometimes running the optimal receiver is prohibitive!

- the complexity can be too high (e.g. for ISI channels is $\mathcal{O}(M^{L+1})$ where M is the alphabet cardinality, and L the memory)
- the actual channel law is not known

e.g. SATELLITE CHANNEL



where $f(z(t)) = \mu(|z(t)|)\, e^{j\left(\phi(\angle z(t)) + \angle z(t)\right)}$

Thus the receiver will use a mismatched channel law $q(\underline{y}|\underline{x}) \neq p(\underline{y}|\underline{x})$

Thm The output of the Arnold-Loeliger algorithm using a channel $p(\underline{y}|\underline{x})$ and a receiver based on the law $q(\underline{y}|\underline{x})$ is a LOWER BOUND on the actual information rate!

Namely

$$\boxed{I_q(\underline{x}^n; \underline{y}^n) = E\left\{ \log_2 \frac{q(\underline{y}^n|\underline{x}^n)}{q(\underline{y}^n)} \right\} \leq I(\underline{x}^n; \underline{y}^n)}$$

proof

$$I(\underline{x}^n; \underline{y}^n) - I_q(\underline{x}^n; \underline{y}^n) = E_p\left\{ \log_2\left(\frac{p(\underline{y}^n|\underline{x}^n)}{p(\underline{y}^n)} \cdot \frac{q(\underline{y}^n)}{q(\underline{y}^n|\underline{x}^n)} \right)\right\}$$

$$= E_p\left\{ \log_2\left(\frac{p(\underline{y}^n, \underline{x}^n)}{p(\underline{y}^n)p(\underline{x}^n)} \cdot \frac{q(\underline{y}^n)}{q(\underline{y}^n|\underline{x}^n)} \right)\right\} = (*)$$

$$\frac{q(\underline{y}^n|\underline{x}^n)\, p(\underline{x}^n)\, p(\underline{y}^n)}{q(\underline{y}^n)} = q(\underline{x}^n|\underline{y}^n)\, p(\underline{y}^n) \rightarrow$$ is still a PDF!

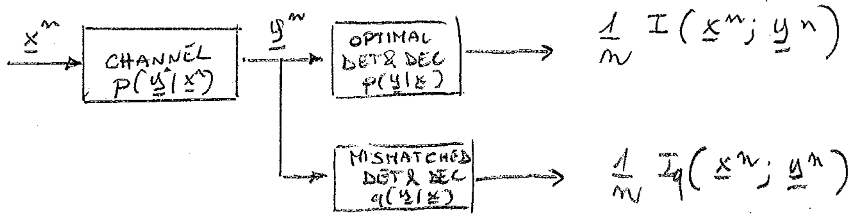
$$(*) = D\left(p(\underline{y}^n, \underline{x}^n) \,\|\, q(\underline{x}^n|\underline{y}^n)\, p(\underline{y}^n) \right) \geq 0$$

Thm The lower bound $\frac{1}{n} I_q(\underline{x}^n; \underline{y}^n)$ is the achievable information rate by a receiver based on the mismatched channel law $q(\underline{y}^n; \underline{x}^n)$

proof at a glance in few words we can define a new kind of typical set based on the mismatched entropies

$$-E_p\{\log q(\underline{y}^n)\}, \quad -E_p\{\log q(\underline{y}^n | \underline{x}^n)\}$$

Using Shannon's random codes it can be proven that the rate $\lim_{n \to +\infty} \frac{1}{n} I_q(\underline{x}^n; \underline{y}^n)$ is achievable by means of joint detection and decoding.



It is worth to point out that the algorithm carries out the information rate only for JOINT DETECTION & DECODING!
This means that the rates can be achieved with iterative detection & decoding. The detector must be based on the channel law $q(\underline{y}^n | \underline{x}^n)$.

NOTES
- Even with a BCJR on $q(\underline{y}^n | \underline{x}^n)$ and a good code it is not guaranteed that the rate can be approached, since iterative is not joint
- We cannot measure the information rate of suboptimal detectors that do not have a mismatched channel law (e.g. Tikhonov synch.).
  In other words the MISMATCHED DETECTOR must be optimal for an auxiliary channel $q(\underline{y}^n | \underline{x}^n)$

BCJR ONCE RATE (or PRAGMATIC CAPACITY)

$$\underline{b} \rightarrow \boxed{\text{MOD}} \xrightarrow{\underline{x}} \boxed{\text{CH.}} \xrightarrow{\underline{y}} \boxed{\text{DET}} \xrightarrow{\underline{l}}$$

$$l_i = \ln\left(\frac{p(\underline{y} \mid b_i = 0)}{p(\underline{y} \mid b_i = 1)}\right)$$

log Likelihood Ratio (LLR)

Often iterative det & dec can be prohibitively due to the high complexity of the detector. Thus we could be interested in the rate achievable by a detector without any information from the code.

Namely $I(b_i; l_i)$, well known as BCJR once rate.

Thm the $I(b_i; l_i)$ is always less than or equal to $\frac{1}{m \log_2 M} I(\underline{x}^n; \underline{y}^n)$

proof the l_i is a sufficient statistics for $\underline{y}^n$, when determining b_i.

$$I(b_i; l_i) = I(b_i; \underline{y}^n).$$

$$I(\underline{x}^n; \underline{y}^n) = I(\underline{b}^n; \underline{y}^n) = \sum_{i=0}^{m \log_2 M - 1} I(b_i; \underline{y}^n \mid b^{i-1}) \geq m \log_2 M \, I(b_i; \underline{y}^n)$$

$$= m \log_2 M \, I(b_i; l_i)$$

$$\Rightarrow \quad I(b_i; l_i) \leq \frac{1}{m \log_2 M} I(\underline{x}^n; \underline{y}^n) \qquad \boxtimes$$

How can it be computed?

Let's define $f(x) = \log_2(1 + e^{-x}) \Rightarrow$
$f(l_i) = -\log_2 p(b_i = 0 \mid \underline{y})$
$f(-l_i) = -\log_2 p(b_i = 1 \mid \underline{y})$

Thus $f(l_i(1 - 2b_i)) = -\log_2 p(b_i \mid \underline{y})$

$$I(b_i; l_i) = 1 - E\{ f(l_i(1 - 2b_i)) \}$$

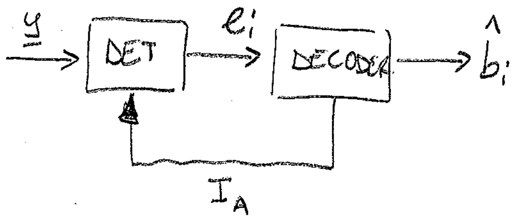
$$\overset{MC}{\simeq} 1 - \frac{1}{m} \sum_{i=0}^{m-1} \log_2\left(1 + e^{-\tilde{l}_i(1 - 2\tilde{b}_i)}\right) \qquad (2)$$

Summary:
- sample two "long" sequences $\tilde{b}_i$, $\tilde{l}_i$
- run (2)

EXIT CHART (Extrinsic Information chart)

The once rate can be extended to study on a quality level how the detector works when it has a priori knowledge of the bit.

e.g. iterative det & dec



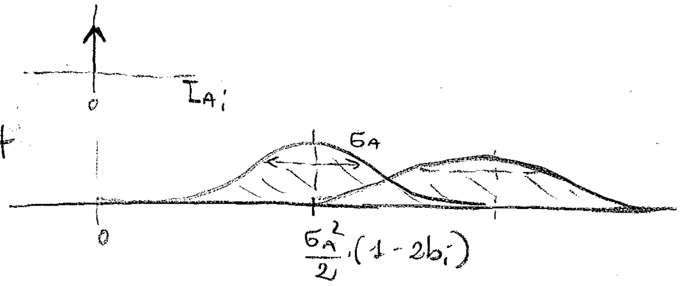
The apriori information is built as a Gaussian process

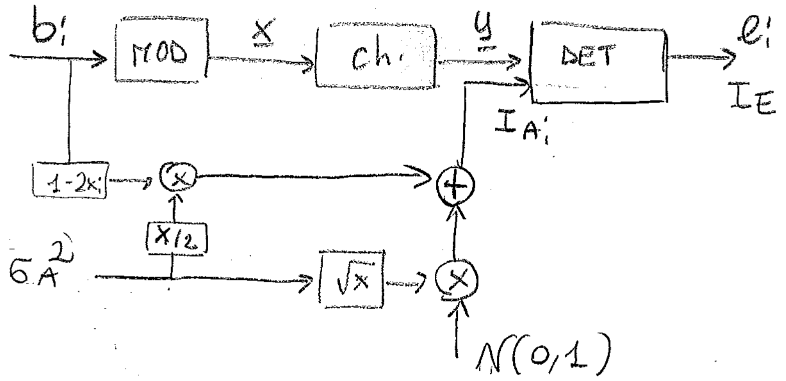
$$I_{A_i} = \frac{\sigma_A^2}{2}(1-2b_i) + N(0, \sigma_A^2)$$

where σ_A^2 is the "level" of information

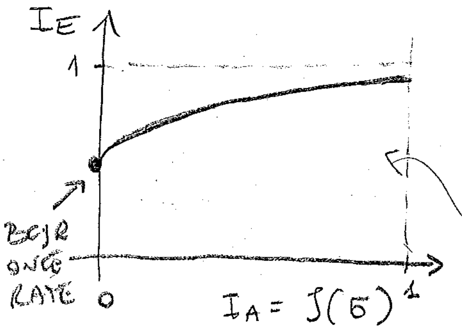
$\sigma_A^2 \to 0$ $\quad I_{A_i} = 0$

$\sigma_A^2 \to +\infty$ $\quad I_{A_i}$ guess right





The extrinsic information is computed with (2)!



where $J(\sigma) = 1$ if $\sigma \to +\infty$

$J(\sigma) = 0$ if $\sigma = 0$

EXIT CHART OF THE DETECTOR

it is useful when coupled with a code! EXIT for a code can be carried out and showed in the same figure. This technique gives an idea on how iterative det & dec. works