Chap. 1 Introduction

□ Fundamental Results from Information Theory

- Let X be a (discrete) random variable with probability distribution

$$p(x) = \Pr(X = x)$$

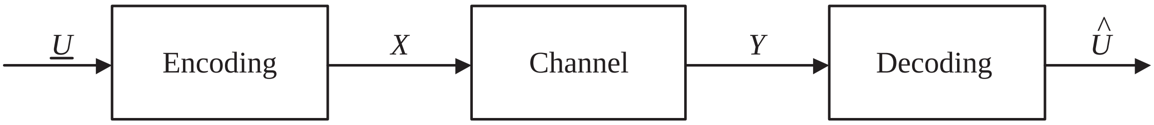
- The *entropy* of X is defined by

$$\begin{aligned} H(X) &= -\sum_{x} p(x) \log p(x) \\ &= E\left(\frac{1}{\log p(X)}\right) \end{aligned}$$

Note that $H(X)$ can be thought of as a measure of the following things about X:

- the amount of information provided by an observation of X;
- the *uncertainty* about X;
- the randomness of X

- (*Discrete Memoryless*) Channel (DMC)



  - The DMC is completely described by the channel transition probability

$$q(y|x) = \mathsf{Pr}(Y = y|X = x)$$

where X is an channel input and Y is its corresponding channel output.

  - Using the total probability theorem, we get

$$p(y) = \sum_x p(x)q(y|x).$$

  - In other words, if Y is discrete and and takes on a finite number of values,

$$P_Y = Q\, P_X$$

where P_Y is a column vector of length m, P_X is a column vector of length n and Q is the channel transition matrix whose (y, x)-entry is given by

$$Q_{y,x} = q(y|x).$$

Example: For the BSC with crossover probability ϵ,

$$Q = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}.$$

- The *conditional entropy* of X, given Y, is defined by

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y).$$

  where $p(x|y)$ is the conditional probability of x given y, and $p(x,y)$ is the joint probability of x and y.
  Note that $H(X|Y)$ represents the amount of uncertainty remaining about X after Y has been observed.

- The *mutual information* between X and Y is defined by

$$I(X:Y) = H(X) - H(X|Y)$$

- The *channel capacity* of a DMC Channel is defined by

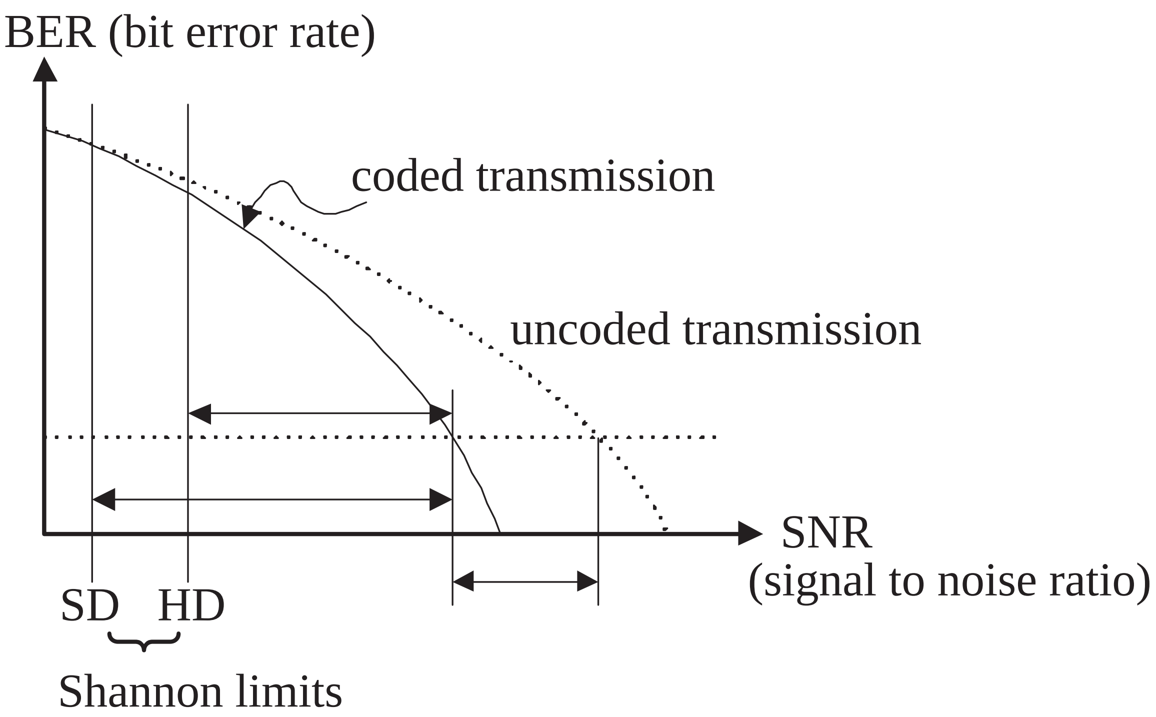
$$C = \sup_{\{p(x)\}} I(X:Y)$$

  where the supremum is taken over all inputs X with input distribution $p(x)$.

Theorem 1 (Channel coding Theorem) *If the code rate $R < C$, then it is possible to transmit information at arbitrarily low error probability.*

□ Classical Approaches

- The achievable coding gain by "classical methods" is far from Shannon's promises



- Hard to achieve with "constructable" codes !!

□ Nonclassical Approaches

- R.G. Gallager (1963)
  Low-density parity-check codes

$\downarrow$

- G.D. Forney (1966)
  Concatenated codes

$\downarrow$

- R.M. Tanner (1981),
  A recursive approach to
  low complexity codes

$\downarrow$

- G. Berrou, A. Glavieux, P. Thitimajshima,
  Near-Shannon limit error-correcting coding
  and decoding: Turbo codes

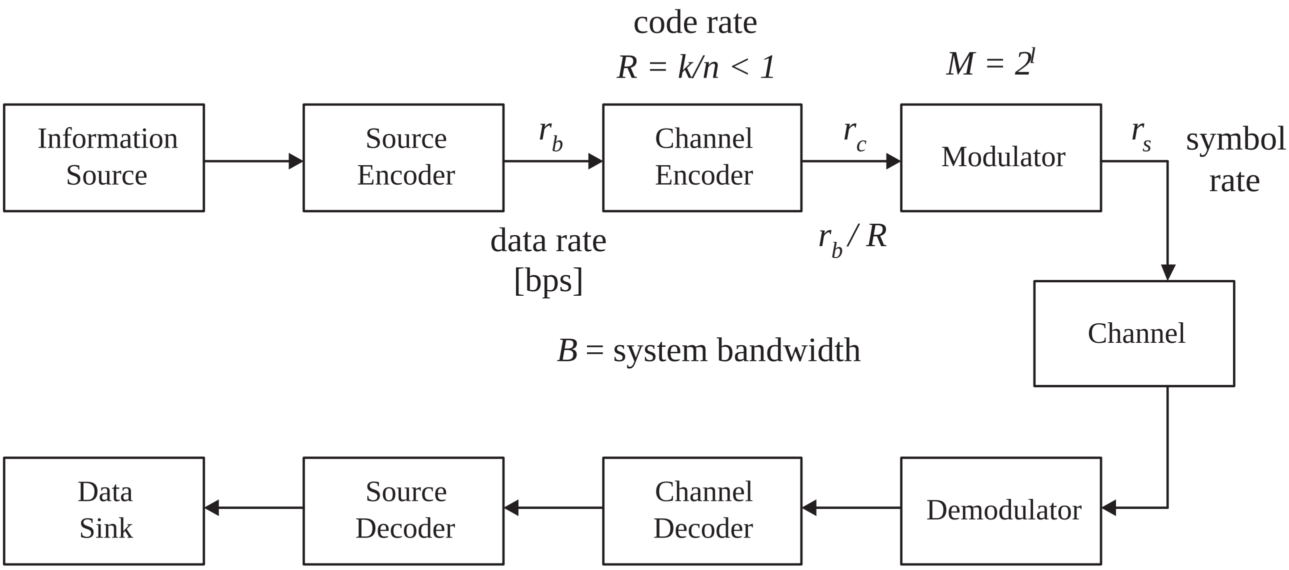
$\downarrow$

- D.J.C. Mackay and R.M. Neal (1995),
  Good codes based on very sparse matrices

- Serially concatenated turbo codes

——— Factor Graphs ———

□ Digital Communication System



- System parameters
  - r_b = data rate [bps]
  - r_c = channel data rate [bps] $= r_b/R$
  - r_s = symbol rate $= 1/T$, where T = signaling interval
  - minimum signal bandwidth $= r_s/2 = r_c/2l = r_b/2Rl$ [Hz]

- *Spectral efficiency*

$$\begin{aligned}\eta &= r_b/B \quad \text{[bits/sec/Hz]} \\ &= r_s l R/B.\end{aligned}$$

Therefore,

$$\eta_{\max} = 2lR \quad \text{[bits/sec/Hz]}$$

since min. bandwidth $= r_s/2 = 1/2T$.

(Note that $W = 1/T$ is quite often assumed.)

- The *bit error rate* (BER) or *bit error probability* measures the reliability of information transmission in digital communication systems.

- The *power efficiency* is captured by the required bit energy to one-sided noise power spectral density ratio, E_b/N_o, to achieve a specified BER.

- The *Signal-to-noise ratio* (SNR) is given by

$$\frac{S}{N} = \frac{E_s/T}{WN_o} = \frac{lRE_b/T}{(1/2T)N_o} = 2lR\frac{E_b}{N_o}$$

  where S is the signal power and N is the noise power within the signal bandwidth.

- The *channel capacity* for an AWGN Channel is given by

$$C = B\log_2\left(1 + \frac{S}{N}\right) \qquad \text{[bits/sec]}.$$

- For error-free transmission,

$$\begin{aligned} \eta_{\max} &\leq \frac{C}{B} \\ &= \log_2\left(1 + 2lR\frac{E_b}{N_o}\right) \\ &= \log_2\left(1 + \eta_{\max}\frac{E_b}{N_o}\right) \end{aligned}$$

  since $2lR = \eta_{max}$.

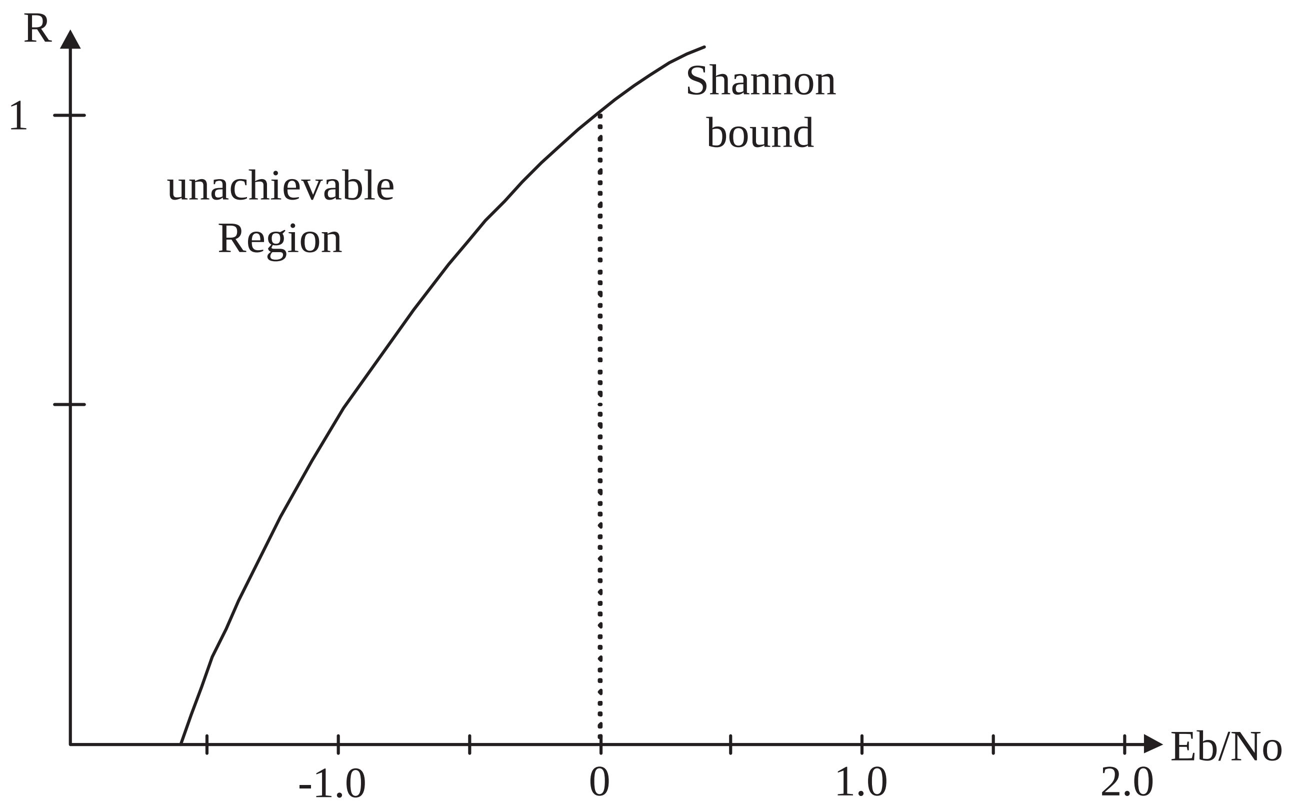
- Therefore, the minimum required SNR for error-free transmission is

$$\frac{E_b}{N_o} \geq \frac{2^{\eta_{\max}} - 1}{\eta_{\max}} \xrightarrow[]{B\to\infty \quad (\text{or } \eta_{\max}\to 0)} \ln 2 = -1.59 \text{ [dB]}$$

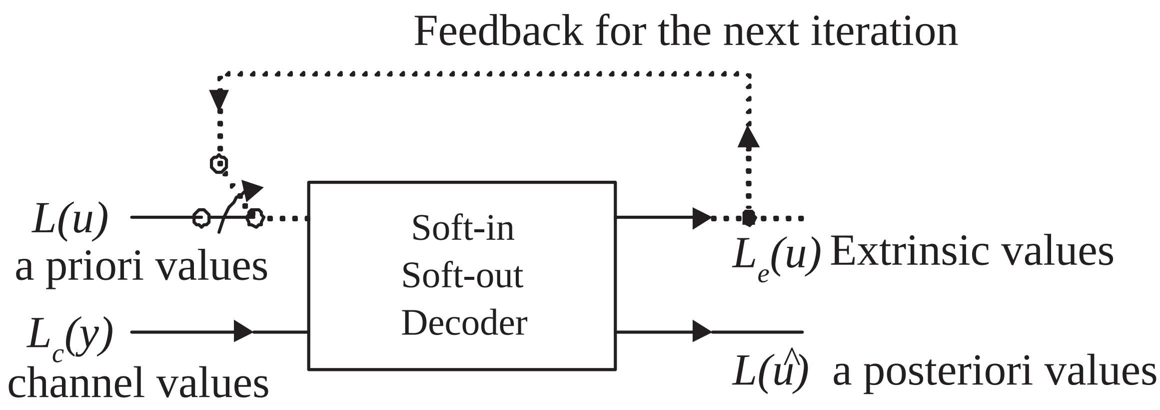
  if the bandwidth is not limited.

- Binary case $(l = 1)$
  - $\eta_{\max} = 2lR = 2R$
  - For error-free transmission,

$$\frac{E_b}{N_o} \geq \frac{2^{2R} - 1}{2R}.$$



□ Iterative Decoding Procedure for Soft-Input/Soft-Output Decoders



- Systematic $[N, k]$ code
  - information bits $\quad u_1 u_2 u_3 \cdots u_k$
  - coded bits $\quad c_1\, c_2\, c_3 \cdots c_k \,\vdots\, c_{k+1} \cdots c_N = \underline{c}$
  - transmitted symbol $\quad x_1 x_2 x_3 \cdots x_k \,\vdots\, x_{k+1} \cdots x_N = \underline{x}$
  - matched filter output $\quad y_1\, y_2\, y_3 \cdots y_k \,\vdots\, y_{k+1} \cdots y_N = \underline{y}$

- The *log-likelihood ratio* (LLR) of x_i, conditioned on $\mathbf{y}$, is defined as

$$L(x_i|\mathbf{y}) \triangleq \log \frac{P(x_i = +1|\mathbf{y})}{P(x_i = -1|\mathbf{y})}.$$

Then

$$\begin{aligned} L(x_i|\mathbf{y}) &= \log \frac{P(\mathbf{y}|x_i = +1)}{P(\mathbf{y_i}|x_i = -1)} \cdot \frac{P(x_i = +1)}{P(x_i = -1)} \\ &= \log \frac{P(x_i = +1)}{P(x_i = -1)} + \log \frac{P(y_i|x_i = +1)}{P(y_i|x_i = -1)} \\ &\quad + \log \frac{P(\tilde{\mathbf{y}}_i|x_i = +1, y_i)}{P(\tilde{\mathbf{y}}_i|x_i = -1, y_i)} \end{aligned}$$

where $\tilde{\mathbf{y}}_i = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_N)$.

- Bit-to-symbol mapping for BPSK:

$$0 \longleftrightarrow +1 \qquad\qquad 1 \longleftrightarrow -1$$

- Let $L(x_i)$ be the *a priori LLR* of the data bit x_i, defined by

$$L(x_i) \triangleq \log \frac{P(x_i = +1)}{P(x_i = -1)} \quad (x_i = u_i, \;\; \text{for} \;\; 1 \le i \le K)$$

and $L_c(y)$ the channel measurement made at the detector, defined by

$$L_c(y) \triangleq \log \frac{P(y \mid x = +1)}{P(y \mid x = -1)}.$$

- For a Gaussian/fading channel,

$$\begin{aligned} L_c(y) &= \log \frac{\exp\left(-\frac{E_s}{N_o}(y-a)^2\right)}{\exp\left(-\frac{E_s}{N_o}(y+a)^2\right)} \\ &= 4a \cdot E_s/N_o \cdot y \end{aligned}$$

Note:

1) Gaussian case:

$$a = 1 \;\; \text{and} \;\; E_s/N_o = \frac{1}{2\sigma^2}.$$

2) $a =$ fading amplitude

3) BSC:

$$L_c(y) = y \log \frac{1-\epsilon}{\epsilon} \qquad \text{for} \;\; y \in \{+1, -1\}.$$

where ϵ is the cross-over probability.

4) $L_c(y) =$ reliability value of the channel

- *Extrinsic LLR* (values):

$$L_e(x_i) \triangleq \log \frac{P(\tilde{\mathbf{y}}_i \mid x_i = +1, y_i)}{P(\tilde{\mathbf{y}}_i \mid x_i = -1, y_i)} = \log \frac{P(\tilde{\mathbf{y}}_i \mid x_i = +1)}{P(\tilde{\mathbf{y}}_i \mid x_i = -1)}$$

assuming that $\tilde{\mathbf{y}}_i$ is independent of y_i.

- The *a posteriori LLR* is the LLR (soft output) of the decoder, defined by

$$L(\hat{x}_i) \triangleq L(x_i \mid \mathbf{y})$$

Then

$$\boxed{L(\hat{x}_i) = L_c(y_i) + L(x_i) + L_e(x_i)}$$

Remark on the a priori LLR:

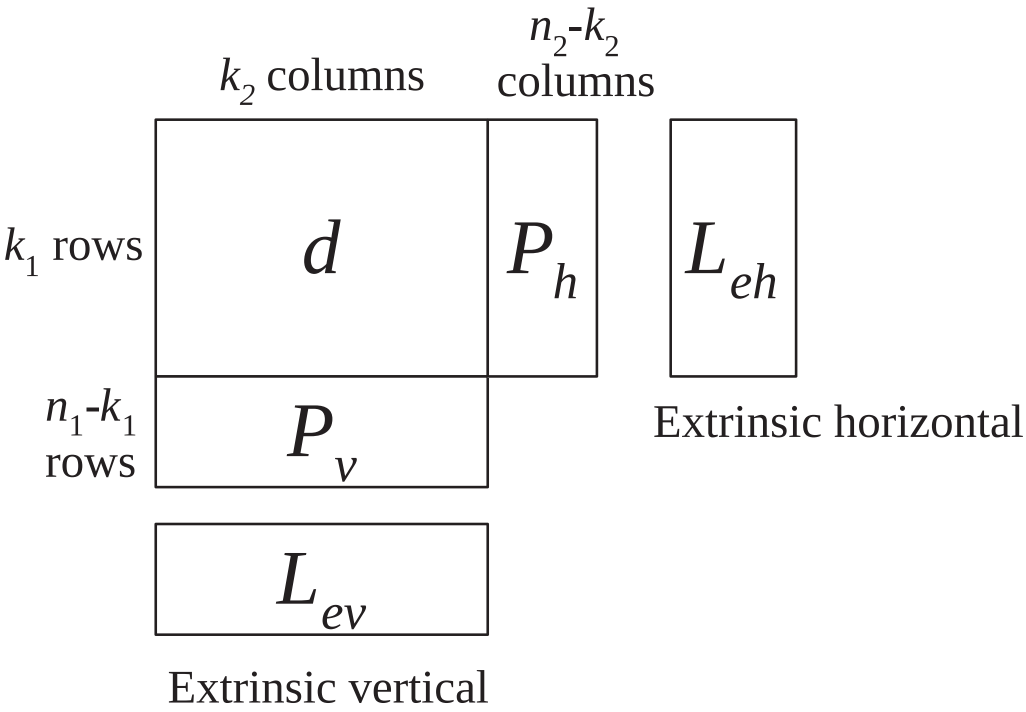
1) For the first decoding iteration, we set

$$L^{(1)}(x_i) = 0.$$

2) For $r \geq 2$, the a priori LLR at the rth iteration is set to

$$L^{(r)}(x_i) = L_e^{(r-1)}(x_i).$$

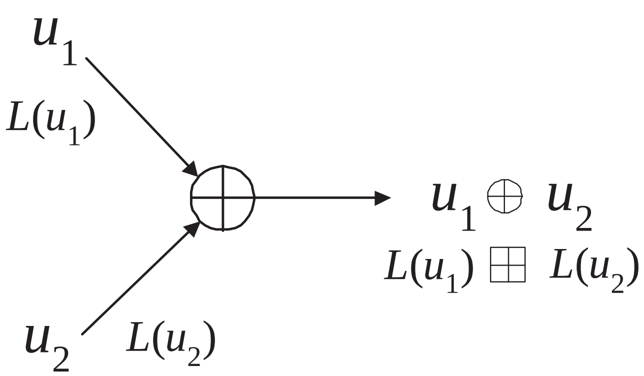
Example: Two-dimensional product code



Procedure of iterative decoding

1) Set the a priori information $L(d) = 0$.
2) Decode horizontally and obtain $L_{eh}(d)$:
$$L_{eh}(d) = L(\hat{d}) - L_c(y) - L(d).$$
3) Set $L(d) = L_{eh}(d)$.
4) Decode vertically and obtain $L_{ev}(d)$:
$$L_{ev}(d) = L(\hat{d}) - L_c(x) - L(d).$$
5) Set $L(d) = L_{ev}(d)$.
6) If there have been enough iterations to yield a reliable decision, goto step 7; otherwise, go to step 2.
7) The soft output is
$$L(\hat{d}) = L_c(y) + L_{eh}(d) + L_{ev}(d).$$

□ Log-Likelihood Algebra



- Bit-to-symbol mapping for BPSK: $0 \leftrightarrow +1, \quad 1 \leftrightarrow -1$
- Relation between the LLR $L(u)$ and the probability $P(u=0)$:

  Since
  $$L(u) = \log \frac{P(u=0)}{P(u=1)}$$
  and
  $$P(u=0) + P(u=1) = 1,$$
  we have
  $$P(u=0) = \frac{e^{L(u)}}{1+e^{L(u)}}.$$

- Compute the following probabilities:

$$\begin{aligned}
P(u_1 \oplus u_2 = 0) &= P(u_1=0)\cdot P(u_2=0) \\
&\quad + (1-P(u_1=0))\cdot(1-P(u_2=0)) \\
&= \frac{1+e^{L_1+L_2}}{(1+e^{L_1})(1+e^{L_2})}
\end{aligned}$$

$$P(u_1 \oplus u_2 = 1) = \frac{e^{L_1}+e^{L_2}}{(1+e^{L_1})(1+e^{L_2})}$$

- Defining

$$
\begin{aligned}
L(u_1) \boxplus L(u_2) &\triangleq L(u_1 \oplus u_2) \\
&= \log \frac{1 + e^{L(u_1)+L(u_2)}}{e^{L(u_1)} + e^{L(u_2)}},
\end{aligned}
$$

we get the following **tanh** rule:

$$
\boxed{\tanh\tfrac{L}{2} = \tanh\tfrac{L_1}{2} \cdot \tanh\tfrac{L_2}{2}}
$$

Exercise: Derive the tanh rule.

Hint:

$$
\tanh \frac{L}{2} = \frac{e^L - 1}{e^L + 1}.
$$

Remark on the LLR Algebra:

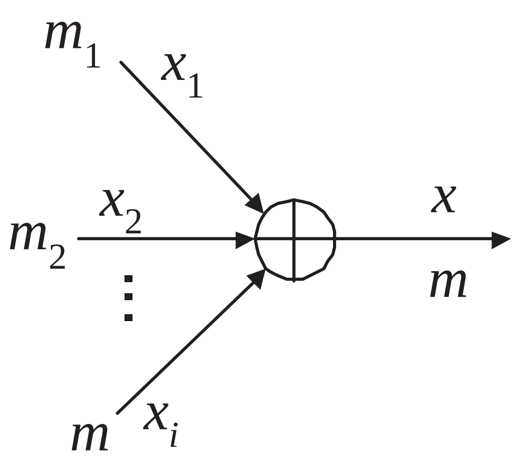
1) LLR algebra with special values:

$$
\begin{aligned}
L(u) \boxplus \infty &= L(u), \\
L(u) \boxplus (-\infty) &= -L(u), \\
L(u) \boxplus 0 &= 0
\end{aligned}
$$

2) Approximation:

$$
L(u_1 \oplus u_2) \approx \operatorname{sign}(L(u_1)) \operatorname{sign}(L(u_2)) \min(|L(u_1)|, |L(u_2)|)
$$

3) General case:



$$\tanh\frac{m}{2} = \prod_{j=1}^{J} \tanh\frac{m_j}{2}$$

4) Approximation in general case:

$$\sum_{j=1} \boxplus\, L(u_j) = 2\tanh^{-1}\left(\prod_{j=1}^{J} \tanh\left(\frac{L(u_j)}{2}\right)\right)$$

$$\approx \left(\prod_{j=1}^{J} \mathrm{sign}\,(L(u_j))\right) \min_{j=1,\cdots,J} |\, L(u_j)\, |$$

<u>Exercise</u>: Show the properties 1), 2) and 4) in the above.

Example: Consider a two-dimensional product code:

u_{11}	u_{12}	p_1^h
u_{21}	u_{22}	p_2^h
p_1^v	p_2^v	

+	+	+
+	-	-
+	-	

Code values

+0.5	+1.5	+1.0
+4.0	+1.0	−1.5
+2.0	−2.5	

Received values $L_c(y)$

+1.0	+0.5	+0.5
−1.0	−1.5	+1.0

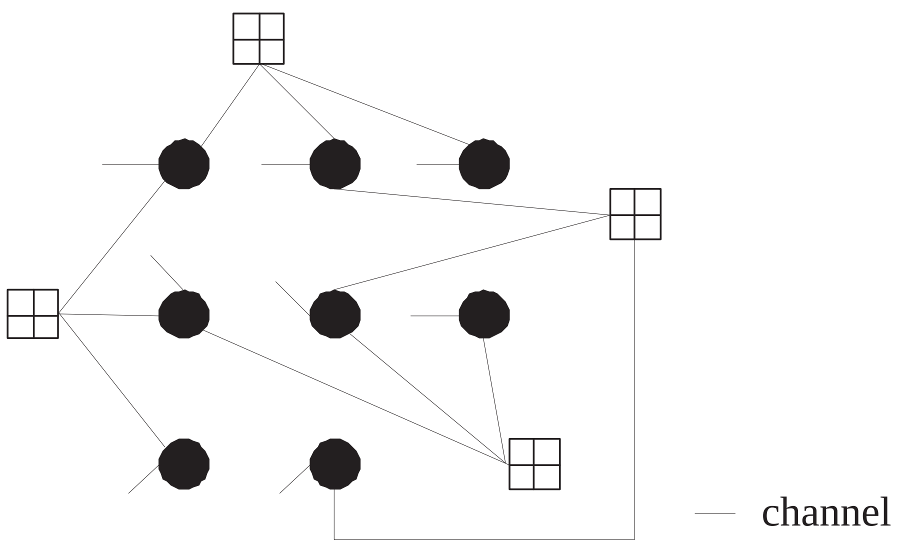
Extrinsic information L_e^h
after first horizontal decoding

+2.0	+0.5
+1.5	−2.0
+1.5	−0.5

Extrinsic information L_e^v
after first vertical decoding

+3.5	+2.5
+4.5	−2.5

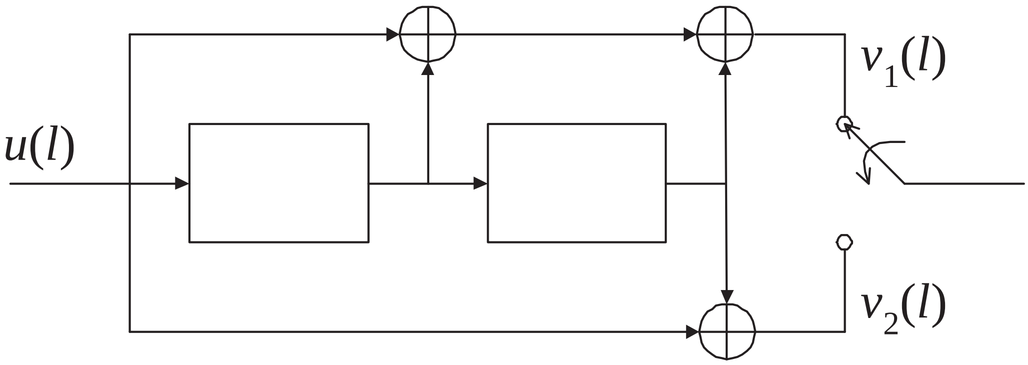
Soft output after the first
horizontal and vertical decoding



Factor graph

☐ Review of Convolutional Codes

Example: Consider the convolutional encoder



- Encoding Procedure:

$$
\begin{array}{llll}
u(l): & u(0) & u(1) & u(2) \;\cdots \\
v_1(l): & v_1(0) & v_1(1) & v_1(2) \cdots \\
 & \downarrow\nearrow & \downarrow\nearrow & \downarrow \\
v_2(l): & v_2(0) & v_2(1) & v_2(2) \cdots
\end{array}
$$

- *Time-domain representation* :

$$v_1(l) = u(l) + u(1-1) + u(l-2)$$
$$v_2(l) = u(l) + u(l-2)$$

- *Impulse response* due to $\mathbf{u} = (1000\cdots)$

$$\mathbf{v}_1 : \; 1\;1\;1\;0\;0\;\;\cdots \rightarrow g_1(l)$$
$$\mathbf{v}_2 : \; 1\;0\;1\;0\;0\;\;\cdots \rightarrow g_2(l)$$

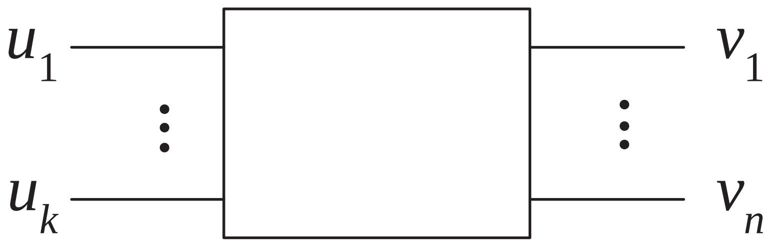
Then

$$
\begin{aligned}
v_j(l) &= \sum_{m=0}^{2} u(l-m) g_j(m) \\
&= u(l) * g_j(l)
\end{aligned}
$$

In general, the jth output for a rate k/n convolutional code

$$v_j(l) = \sum_{i=1}^{k} u_i(l) * g_{ij}(l)$$

where $u_i(l)$ is the ith input and $g_{ij}(l)$ is jth impulse response due to ith input.



□ Power Series Representation

- Let

$$\begin{aligned} u(l) \longleftrightarrow U(D) &\triangleq \sum_{l=0}^{\infty} u(l) D^l \\ v_j(l) \longleftrightarrow V_j(D) &\triangleq \sum_{l=0}^{\infty} v(l) D^l \\ &= U(D) g_j(D) \end{aligned}$$

- The power series of the output in the example is given by

$$\begin{aligned} \mathbf{V}(D) &= [V_1(D) \quad V_2(D)] \\ &= U(D) [g_1(D) \quad g_2(D)] \\ &= U(D) G(D) \end{aligned}$$

where $G(D)$ is the *transfer function matrix* defined by

$$G(D) = [1 + D + D^2 \quad 1 + D^2]$$

- In general,

$$\begin{array}{ccc} \mathbf{V}(D) & = & \mathbf{U}(D) \quad G(D) \\ 1 \times n & & 1 \times k \quad k \times n \end{array}$$

where $G(D)$ is called the *transfer function matrix* or *generator matrix*.

☐ Structure of Convolutional Codes

- FIR (Finite Impulse Response):
  $g_{ij}(D)$ is a polynomial $\forall\, i, j$

  IIR (Infinit Impulse Response)
  $g_{ij}(D)$ is a rational function for some i, j

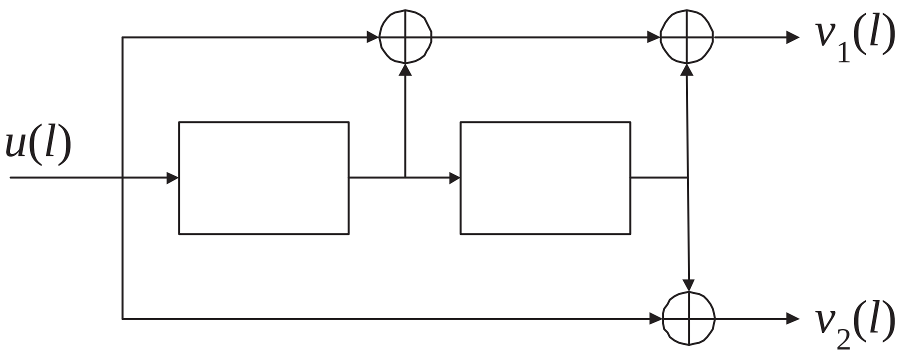
- Systematic CC
  $G(D) = \big[I : P(D)\big]$

  Nonsystematic CC

Example:

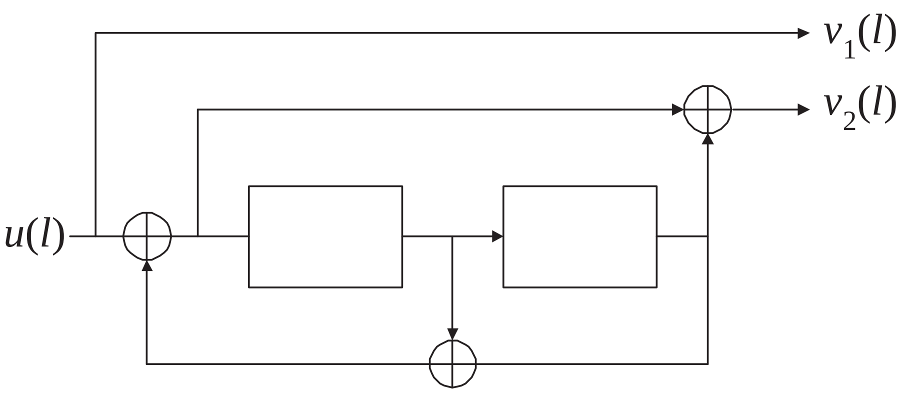
- FIR and nonsystematic

$$G(D) = [1 + D + D^2 \quad 1 + D^2]$$



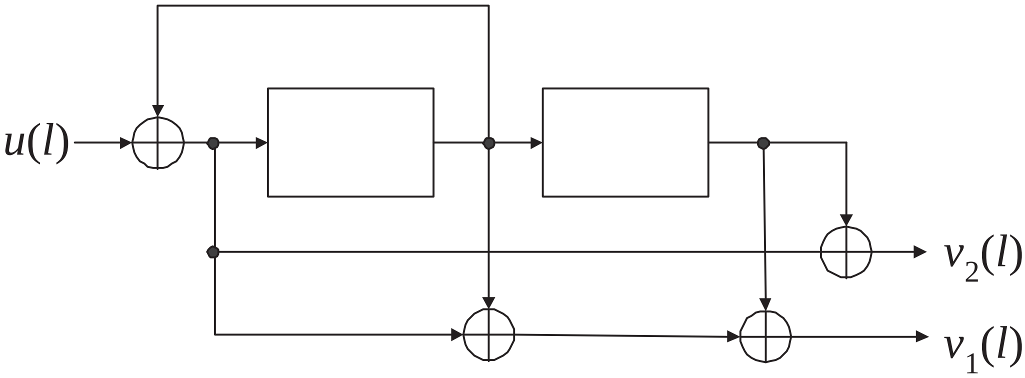
- IIR and systematic or recursive systematic code (RSC)

$$G(D) = \left[1 \quad \frac{1 + D^2}{1 + D + D^2}\right]$$



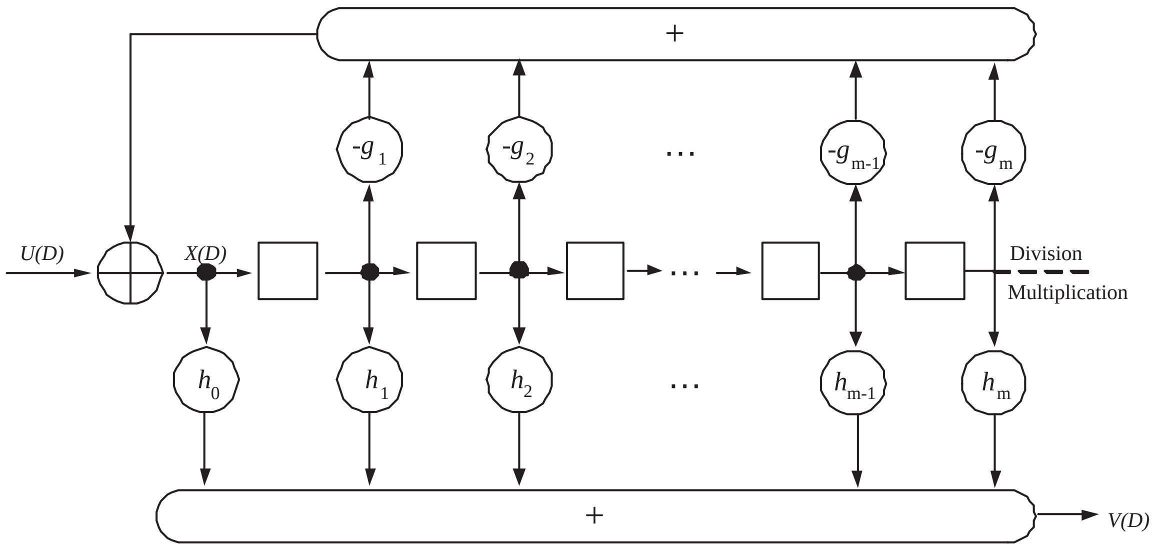
- IIR and nonsystematic

$$G(D) = \left[\frac{1+D+D^2}{1+D} \quad \frac{1+D^2}{1+D} \right] = \left[\frac{1+D+D^2}{1+D} \quad 1+D \right]$$



□ Rational Transfer Function $T(D) = h(D)/g(D)$:

- *Controller Canonical Form* or *Fibonacci Configuation*



$$
\begin{aligned}
X(D) &= U(D) - g_1 D X(D) - \cdots - g_m D^m X(D) \\
\Rightarrow \quad X(D) &= \frac{U(D)}{1 + g_1 D + \cdots + g_m D^m} \\
&= \frac{U(D)}{g(D)}.
\end{aligned}
$$

$$
\begin{aligned}
V(D) &= h_0 X(D) + h_1 D X(D) + \cdots + h_m D^m X(D) \\
&= (h_0 + h_1 D + \cdots + h_m D^m) X(D) \\
&= h(D) X(D)
\end{aligned}
$$

$$
\Rightarrow \quad V(D) = \underbrace{\frac{h(D)}{g(D)}}_{\text{transfer function}} U(D)
$$

- *Observer Canonical Form* or *Galois Configuation*

  Rewriting

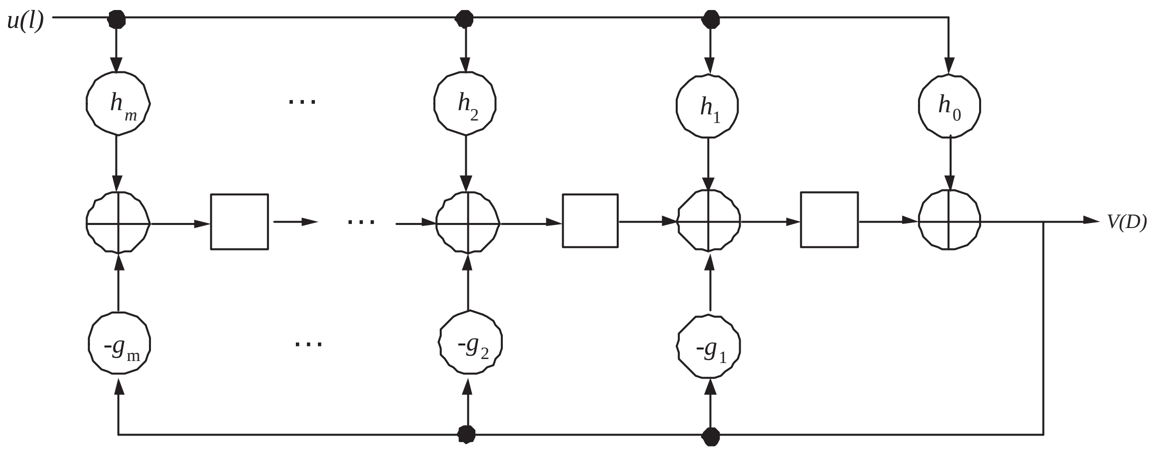
$$V(D) = \frac{h(D)}{g(D)} U(D),$$

  we get

$$(1 + g_1 D + \cdots + g_m D^m) V(D) = U(D)(h_0 + h_1 D + \cdots + h_m D^m).$$

  That is,

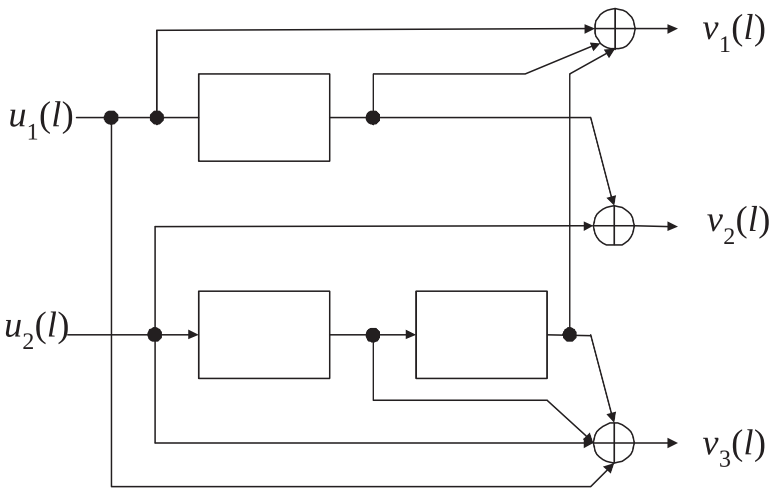
$$\begin{aligned} V(D) &= U(D)(h_0 + h_1 D + \cdots + h_m D^m) \\ &\quad + V(D)(g_1 D + g_2 D^2 + \cdots + g_m D^m). \end{aligned}$$



Major Concepts in Convolutional Codes

- catastrophic error propagation
- state diagram
- trellis diagram
- weight distribution
- Viterbi decoder
- performance analysis, etc.

Example: Nonsystematic $\rightarrow$ Systematic (p.55, Fig 3.8)



The transfer function matrix is

$$G(D) = \begin{bmatrix} 1+D & D & 1 \\ D^2 & 1 & 1+D+D^2 \end{bmatrix}.$$

Choose

$$T(D) = \begin{bmatrix} 1+D & D \\ D^2 & 1 \end{bmatrix}.$$

Note that

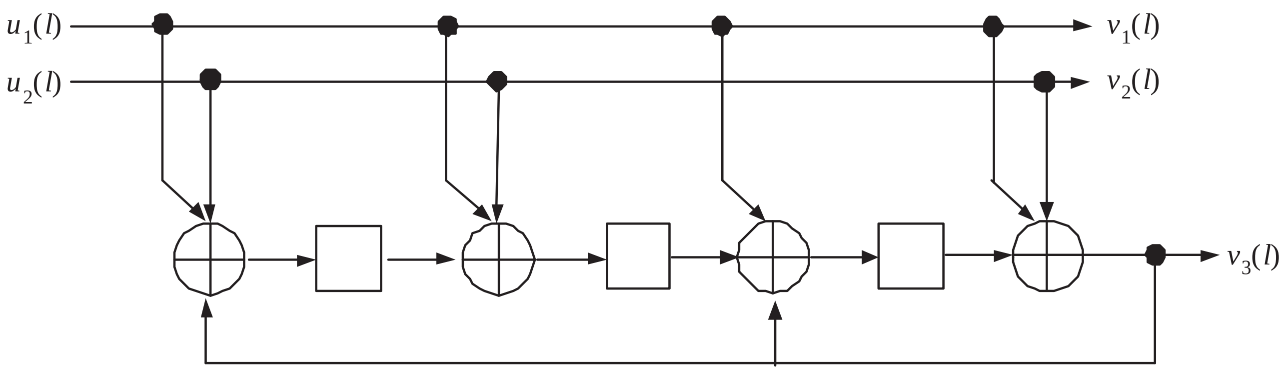
$$\det T(D) \neq 0$$

and

$$T^{-1}(D) = \frac{1}{1+D+D^3}\begin{bmatrix} 1 & D \\ D^2 & 1+D \end{bmatrix}.$$

Then the transfer function matrix can be transformed into

$$\begin{aligned} G_{\mathsf{sys}}(D) &= T^{-1}(D)G(D) \\ &= \begin{bmatrix} 1 & 0 & \frac{1+D+D^2+D^3}{1+D+D^3} \\ 0 & 1 & \frac{1+D^2+D^3}{1+D+D^3} \end{bmatrix}. \end{aligned}$$



$\square$