Tutorial Overview

- **Channel capacity**
- **Convolutional codes**
  - the MAP algorithm
- **Turbo codes**
  - Standard binary turbo codes: UMTS and cdma2000
  - Duobinary CRSC turbo codes: DVB-RCS and 802.16
- **LDPC codes**
  - Tanner graphs and the message passing algorithm
  - Standard binary LDPC codes: DVB-S2
- **Bit interleaved coded modulation (BICM)**
  - Combining high-order modulation with a binary capacity approaching code.
- **EXIT chart analysis of turbo codes**
Software to Accompany Tutorial

- Iterative Solution's Coded Modulation Library (CML) is a library for simulating and analyzing coded modulation.
- Available for free at the Iterative Solutions website:
  - www.iterativesolutions.com
- Runs in matlab, but uses c-mex for efficiency.
- Supported features:
  - Simulation of BICM
    - Turbo, LDPC, or convolutional codes.
    - PSK, QAM, FSK modulation.
    - BICM-ID: Iterative demodulation and decoding.
  - Generation of ergodic capacity curves (BICM/CM constraints).
  - Information outage probability in block fading.
  - Calculation of throughput of hybrid-ARQ.
- Implemented standards:
  - Binary turbo codes: UMTS/3GPP, cdma2000/3GPP2.
  - Duobinary turbo codes: DVB-RCS, wimax/802.16.
  - LDPC codes: DVB-S2.

Noisy Channel Coding Theorem

- Every channel has associated with it a capacity $C$.
  - Measured in bits per channel use (modulated symbol).
- The channel capacity is an upper bound on information rate $r$.
  - There exists a code of rate $r < C$ that achieves reliable communications.
    - Reliable means an arbitrarily small error probability.
Computing Channel Capacity

The capacity is the **mutual information** between the channel’s input X and output Y maximized over all possible input distributions:

\[ C = \max_{p(x)} \{ I(X;Y) \} \]

\[ = \max_{p(x)} \left\{ \int \int p(x,y) \log_2 \left( \frac{p(x,y)}{p(x)p(y)} \right) dx dy \right\} \]

Capacity of AWGN with Unconstrained Input

Consider an AWGN channel with 1-dimensional input:

- \( y = x + n \)
- where n is Gaussian with variance \( N_0/2 \)
- x is a signal with average energy (variance) \( E_s \)

The capacity in this channel is:

\[ C = \max_{p(x)} \{ I(X;Y) \} = \frac{1}{2} \log_2 \left( \frac{2E_s}{N_0} + 1 \right) = \frac{1}{2} \log_2 \left( \frac{2rE_b}{N_0} + 1 \right) \]

- where \( E_b \) is the energy per (information) bit.

This capacity is achieved by a Gaussian input x.
- This is not a practical modulation.
If we only consider antipodal (BPSK) modulation, then

\[ X = \pm \sqrt{\mathbb{E}} \]

and the capacity is:

\[
C = \max_{p(x)} \{ I(X;Y) \} \\
= I(X;Y)_{p(x),p=1/2} \\
= H(Y) - H(N) \\
= \int p(y) \log_2 p(y) dy - \frac{1}{2} \log_2 (\pi e N_0) \\
\text{maximized when two signals are equally likely}
\]

This term must be integrated numerically with

\[
p_x(y) = p_x(y) y p_x(y) = \int_{-\infty}^\infty p_x(\lambda) p_x(y-\lambda) d\lambda
\]

Capacity of AWGN w/ 1-D Signaling

It is theoretically impossible to operate in this region.

It is theoretically possible to operate in this region.
Power Efficiency of Standard Binary Channel Codes

Spectral Efficiency

- Code Rate &
- Shannon Capacity Bound

- BPSK Capacity Bound

- Eb/No in dB

<table>
<thead>
<tr>
<th>Code</th>
<th>Year</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turbo Code</td>
<td>1993</td>
<td>Mariner</td>
</tr>
<tr>
<td>LDPC Code</td>
<td>2001</td>
<td>Chung, Forney, Richardson, Urbanke</td>
</tr>
<tr>
<td>Voyager</td>
<td>1977</td>
<td>Galileo: BVD</td>
</tr>
<tr>
<td>Galileo: LGA</td>
<td>1996</td>
<td>Galileo: BVD</td>
</tr>
<tr>
<td>Iridium</td>
<td>1998</td>
<td>Galileo: BVD</td>
</tr>
<tr>
<td>Pioneer 1968-72</td>
<td>1988</td>
<td>Odenwalder, Convolutional Codes 1976</td>
</tr>
<tr>
<td>Mariner 1969</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odenwalder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convolutional Codes</td>
<td>1976</td>
<td></td>
</tr>
</tbody>
</table>

Arbitrarily low BER, $P_e = 10^{-6}$

Binary Convolutional Codes

- A convolutional encoder comprises:
  - k input streams
  - We assume k=1 throughout this tutorial.
  - n output streams
  - m delay elements arranged in a shift register.
  - Combinatorial logic (OR gates).
    - Each of the n outputs depends on some modulo-2 combination of the k current inputs and the m previous inputs in storage.
- The constraint length is the maximum number of past and present input bits that each output bit can depend on.
  - $K = m + 1$
State Diagrams

- A convolutional encoder is a finite state machine, and can be represented in terms of a state diagram.

![State Diagram](image)

- Input data bit
- Corresponding output code bits
- $2^m = 4$ total states
- Since $k=1$, 2 branches enter and 2 branches leave each state

Trellis Diagram

- Although a state diagram is a helpful tool to understand the operation of the encoder, it does not show how the states change over time for a particular input sequence.

- A **trellis** is an expansion of the state diagram which explicitly shows the passage of time.
  - All the possible states are shown for each instant of time.
  - Time is indicated by a movement to the right.
  - The input data bits and output code bits are represented by a unique path through the trellis.
Recursive Systematic Convolutional (RSC) Codes

- An **RSC** encoder is constructed from a standard convolutional encoder by feeding back one of the outputs.
- An RSC code is **systematic**.
  - The input bits appear directly in the output.
- An RSC encoder is an **Infinite Impulse Response** (IIR) Filter.
  - An arbitrary input will cause a “good” (high weight) output with high probability.
  - Some inputs will cause “bad” (low weight) outputs.
With an RSC code, the output labels are the same. However, input labels are changed so that each state has an input “0” and an input “1.”

Messages labeling transitions that start from $S_1$ and $S_2$ are complemented.

Trellis Diagram of RSC Code

$m = 2$ tail bits no longer all-zeros must be calculated by the encoder.
Convolutional Codewords

- Consider the trellis section at time t.
  - Let S(t) be the encoder state at time t.
  - When there are four states, S(t) ∈ \{S_0, S_1, S_2, S_3\}
- Let u(t) be the message bit at time t.
  - The encoder state S(t) depends on u(t) and S(t-1)
- Depending on its initial state S(t-1) and the final state S(t), the encoder will generate an n-bit long word
  - \( x(t) = (x_1, x_2, \ldots, x_n) \)
- The word is transmitted over a channel during time t, and the received signal is:
  - \( y(t) = (y_1, y_2, \ldots, y_n) \)
  - For BPSK, each \( y = (2x-1) + n \)
- If there are L input data bits plus m tail bits, the overall transmitted codeword is:
  - \( x = [x(1), x(2), \ldots, x(L), \ldots, x(L+m)] \)
- And the received codeword is:
  - \( y = [y(1), y(2), \ldots, y(L), \ldots, y(L+m)] \)

MAP Decoding

- The goal of the maximum a posteriori (MAP) decoder is to determine \( P( u(t)=1 \mid y ) \) and \( P( u(t)=0 \mid y ) \) for each t.
  - The probability of each message bit, given the entire received codeword.
- These two probabilities are conveniently expressed as a log-likelihood ratio:
  \[
  \lambda(t) = \log \frac{P[u(t)=1 \mid y]}{P[u(t)=0 \mid y]}
  \]
Determining Message Bit Probabilities from the Branch Probabilities

- Let \( p_{i,j}(t) \) be the probability that the encoder made a transition from \( S_i \) to \( S_j \) at time \( t \), given the entire received codeword.
  - \( p_{i,j}(t) = P( S_i(t-1) \rightarrow S_j(t) \mid y ) \)
  - where \( S_j(t) \) means that \( S(t)=S_j \)
- For each \( t \),
  \[
  \sum_{S_i \rightarrow S_j} P(S_i(t-1) \rightarrow S_j(t) \mid y) = 1
  \]
- The probability that \( u(t) = 1 \) is
  \[
  P(u(t) = 1 \mid y) = \sum_{S_i \rightarrow S_j, u=1} P(S_i(t-1) \rightarrow S_j(t) \mid y)
  \]
- Likewise
  \[
  P(u(t) = 0 \mid y) = \sum_{S_i \rightarrow S_j, u=0} P(S_i(t-1) \rightarrow S_j(t) \mid y)
  \]

Determining the Branch Probabilities

- Let \( \gamma_{i,j}(t) \) = Probability of transition from state \( S_i \) to state \( S_j \) at time \( t \), given just the received word \( y(t) \)
  - \( \gamma_{i,j}(t) = P( S_i(t-1) \rightarrow S_j(t) \mid y(t) ) \)
- Let \( \alpha_i(t-1) \) = Probability of starting at state \( S_i \) at time \( t \), given all symbols received prior to time \( t \).
  - \( \alpha_i(t-1) = P( S_i(t-1) \mid y(1), y(2), \ldots, y(t-1) ) \)
- \( \beta_j(t) \) = Probability of ending at state \( S_j \) at time \( t \), given all symbols received after time \( t \).
  - \( \beta_j(t) = P( S_j(t) \mid y(t+1), \ldots, y(L+m) ) \)
- Then the branch probability is:
  - \( p_{i,j}(t) = \alpha_i(t-1) \gamma_{i,j}(t) \beta_j(t) \)
Computing $\alpha$

- $\alpha$ can be computed recursively.
- Prob. of path going through $S_i(t-1)$ and terminating at $S_j(t)$, given $y(1)\ldots y(t)$ is:
  - $\alpha_{i}(t-1) \gamma_{i,j}(t)$
- Prob. of being in state $S_j(t)$, given $y(1)\ldots y(t)$ is found by adding the probabilities of the two paths terminating at state $S_j(t)$.
- For example,
  - $\alpha_3(t)=\alpha_1(t-1) \gamma_{1,3}(t) + \alpha_3(t-1) \gamma_{3,3}(t)$
- The values of $\alpha$ can be computed for every state in the trellis by “sweeping” through the trellis in the forward direction.

Computing $\beta$

- Likewise, $\beta$ is computed recursively.
- Prob. of path going through $S_i(t+1)$ and terminating at $S_j(t)$, given $y(t+1), \ldots, y(L+m)$
  - $\beta_{j}(t+1) \gamma_{i,j}(t+1)$
- Prob. of being in state $S_i(t)$, given $y(t+1), \ldots, y(L+m)$ is found by adding the probabilities of the two paths starting at state $S_i(t)$.
- For example,
  - $\beta_3(t)=\beta_2(t+1) \gamma_{1,2}(t+1) + \beta_3(t+1) \gamma_{3,3}(t+1)$
- The values of $\beta$ can be computed for every state in the trellis by “sweeping” through the trellis in the reverse direction.
Computing $\gamma$

- Every branch in the trellis is labeled with:
  - $\gamma_{i,j}(t) = P(S_i(t-1) \rightarrow S_j(t) | y(t))$
- Let $x_{i,j} = (x_1, x_2, \ldots, x_n)$ be the word generated by the encoder when transitioning from $S_i$ to $S_j$.
  - $\gamma_{i,j}(t) = P(x_{i,j} | y(t))$
- From Bayes rule,
  - $\gamma_{i,j}(t) = P(x_{i,j} | y(t)) = P(y(t) | x_{i,j}) P(x_{i,j}) / P(y(t))$
  - $P(y(t))$
    - Is not strictly needed because will be the same value for the numerator and denominator of the LLR $\lambda(t)$.
    - Instead of computing directly, can be found indirectly as a normalization factor (chosen for numerical stability)
- $P(x_{i,j})$
  - Initially found assuming that code bits are equally likely.
  - In a turbo code, this is provided to the decoder as “a priori” information.

Computing $P(y(t) | x_{i,j})$

- If BPSK modulation is used over an AWGN channel, the probability of code bit $y$ given $x$ is conditionally Gaussian:
  $$P(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y-m_x)^2}{2\sigma^2} \right\}$$
  $$m_x = \sqrt{E}_{y|x}(2x-1)$$
  $$\sigma^2 = \frac{N_0}{2}$$
  - In Rayleigh fading, multiply $m_x$ by $a$, the fading amplitude.
- The conditional probability of the word $y(t)$
  $$P(y | x) = \prod_{i=1}^{n} p(y_i | x_i)$$
Overview of MAP algorithm

- Label every branch of the trellis with $\gamma_{i,j}(t)$.
- Sweep through trellis in forward-direction to compute $\alpha_i(t)$ at every node in the trellis.
- Sweep through trellis in reverse-direction to compute $\beta_j(t)$ at every node in the trellis.
- Compute the LLR of the message bit at each trellis section:

$$\lambda(t) = \log \frac{P[u(t) = 1 | y]}{P[u(t) = 0 | y]}$$

$$= \log \frac{\sum \alpha_i(t-1) \gamma_{i,j}(t) \beta_j(t)}{\sum \alpha_i(t-1) \gamma_{i,j}(t) \beta_j(t)}$$

- MAP algorithm also called the “forward-backward” algorithm (Forney).

Log Domain Decoding

- The MAP algorithm can be simplified by performing in the log domain.
  - exponential terms (e.g. used to compute $\gamma$) disappear.
  - multiplications become additions.
  - Addition can be approximated with maximization.
- Redefine all quantities:
  - $\gamma_{i,j}(t) = \log P(S_i(t-1) \rightarrow S_j(t) | y(t))$
  - $\alpha_i(t-1) = \log P(S_i(t-1) | y(1), y(2), ..., y(t-1))$
  - $\beta_j(t) = \log P(S_j(t) | y(t+1), ..., y(L+m))$
- Details of the log-domain implementation will be presented later…
Parallel Concatenated Codes with Nonuniform Interleaving

- A stronger code can be created by encoding in parallel.
- A nonuniform interleaver scrambles the ordering of bits at the input of the second encoder.
  - Uses a pseudo-random interleaving pattern.
- It is very unlikely that both encoders produce low weight code words.
- MUX increases code rate from 1/3 to 1/2.

Random Coding Interpretation of Turbo Codes

- **Random codes** achieve the best performance.
  - Shannon showed that as $n \to \infty$, random codes achieve channel capacity.
- However, random codes are not feasible.
  - The code must contain enough structure so that decoding can be realized with actual hardware.
- **Coding dilemma:**
  - “All codes are good, except those that we can think of.”
- With turbo codes:
  - The nonuniform interleaver adds apparent randomness to the code.
  - Yet, they contain enough structure so that decoding is feasible.
Comparison of a Turbo Code and a Convolutional Code

First consider a K=12 convolutional code.
- \( d_{\text{min}} = 18 \)
- \( \beta_d = 187 \) (output weight of all \( d_{\text{min}} \) paths)

Now consider the original turbo code.
- Same complexity as the K=12 convolutional code
- Constraint length 5 RSC encoders
- \( k = 65,536 \) bit interleaver
- Minimum distance \( d_{\text{min}} = 6 \)
- \( a_d = 3 \) minimum distance code words
- Minimum distance code words have average information weight of only \( f_d = 2 \)

\[
\frac{E_b}{N_0} \text{ in dB}
\]

Comparison of Minimum-distance Asymptotes

Convolutional code:

- \( d_{\text{min}} = 18 \)
- \( c_{d_{\text{min}}} = 187 \)

- \( P_b \approx \left( \frac{18}{18} \right) \left( \frac{E_b}{N_0} \right) \)

Turbo code:

- \( d_{\text{min}} = 6 \)
- \( c_{d_{\text{min}}} \approx \frac{a_{d_{\text{min}}} W_{d_{\text{min}}}}{k} \approx \frac{3.2}{65536} \)

- \( P_b \approx \left( 9.2 \times 10^{-5} \right) \left( \frac{E_b}{N_0} \right) \)
The Turbo-Principle

- Turbo codes get their name because the decoder uses feedback, like a turbo engine.

Performance as a Function of Number of Iterations

- $K = 5$ - constraint length
- $r = 1/2$ - code rate
- $L = 65,536$ - interleaver size
- number data bits
- Log-MAP algorithm
Summary of Performance Factors and Tradeoffs

- **Latency vs. performance**
  - Frame (interleaver) size $L$

- **Complexity vs. performance**
  - Decoding algorithm
  - Number of iterations
  - Encoder constraint length $K$

- **Spectral efficiency vs. performance**
  - Overall code rate $r$

- **Other factors**
  - Interleaver design
  - Puncture pattern
  - Trellis termination

Tradeoff: BER Performance versus Frame Size (Latency)

- $K = 5$
- Rate $r = 1/2$
- 18 decoder iterations
- AWGN Channel

![Graph showing BER performance versus frame size (latency)]
Characteristics of Turbo Codes

- Turbo codes have extraordinary performance at low SNR.
  - Very close to the Shannon limit.
  - Due to a low multiplicity of low weight code words.
- However, turbo codes have a BER “floor”.
  - This is due to their low minimum distance.
- Performance improves for larger block sizes.
  - Larger block sizes mean more latency (delay).
  - However, larger block sizes are not more complex to decode.
  - The BER floor is lower for larger frame/interleaver sizes
- The complexity of a constraint length $K_{TC}$ turbo code is the same as a $K = K_{CC}$ convolutional code, where:
  - $K_{CC} = 2 + K_{TC} + \log_2(\text{number decoder iterations})$

UMTS Turbo Encoder

- From 3GPP TS 25 212 v6.6.0, Release 6 (2005-09)
  - UMTS Multiplexing and channel coding
- Data is segmented into blocks of $L$ bits.
  - where $40 \leq L \leq 5114$
UMTS Interleaver: Inserting Data into Matrix

- Data is fed row-wise into a R by C matrix.
  - \( R = 5, 10, \) or 20.
  - \( 8 \leq C \leq 256 \)
  - If \( L < RC \) then matrix is padded with dummy characters.

In the CML, the UMTS interleaver is created by the function \texttt{CreateUMTSInterleaver}.
Interleaving and Deinterleaving are implemented by \texttt{Interleave} and \texttt{Deinterleave}.

UMTS Interleaver: Intra-Row Permutations

- Data is permuted within each row.
  - Permutation rules are rather complicated.
  - See spec for details.
UMTS Interleaver:
Inter-Row Permutations

- Rows are permuted.
  - If \( R = 5 \) or 10, the matrix is reflected about the middle row.
  - For \( R = 20 \) the rule is more complicated and depends on \( L \).
    - See spec for \( R = 20 \) case.

```
X8 X1 X4 X3 X7 X5 X6 X2  
X16 X9 X14 X13 X15 X11 X12 X10 
X24 X17 X20 X19 X23 X21 X22 X18 
X32 X25 X30 X29 X31 X27 X28 X26 
X34 X33 X38 X37 X39 X35 X36 X40
```

UMTS Interleaver:
Reading Data From Matrix

- Data is read from matrix column-wise.

```
X_{40} X_{36} X_{32} X_{28} X_{24} X_{20} X_{16} X_{12} 
X_{18} X_{14} X_{10} X_{6} X_{2} X_{3} X_{5} X_{7} 
X_{10} X_{14} X_{18} X_{22} X_{26} X_{30} X_{34} X_{38} 
X_{2} X_{6} X_{10} X_{14} X_{18} X_{22} X_{26} X_{30} 
```

Thus:
- \( X'_{1} = X_{40} \) \( X'_{2} = X_{26} \) \( X'_{3} = X_{18} \) \( \ldots \)
- \( X'_{38} = X_{24} \) \( X'_{39} = X_{16} \) \( X'_{40} = X_{8} \)
UMTS Constituent RSC Encoder

- Upper and lower encoders are identical:
  - Feedforward generator is 15 in octal.
  - Feedback generator is 13 in octal.

Trellis Termination

- After the $L^{th}$ input bit, a 3 bit tail is calculated.
  - The tail bit equals the fed back bit.
  - This guarantees that the registers get filled with zeros.
- Each encoder has its own tail.
  - The tail bits and their parity bits are transmitted at the end.
Output Stream Format

The format of the output steam is:

\[ X_1 Z_1 Z'_1 X_2 Z_2 Z'_2 \ldots X_L Z_L Z'_L X_{L+1} Z_{L+1} Z'_{L+1} \]

- \( X_i \) and \( Z_i \) are data bits
- \( Z'_i \) are parity bits
- \( Z'_{L+1} \) are parity bits for the upper encoder
- \( Z'_{L+2} \) are parity bits for the lower encoder

Total number of coded bits = 3L + 12

Code rate: \( r = \frac{L}{3L + 12} \approx \frac{1}{3} \)

Channel Model and LLRs

Channel gain: \( a \)
- Rayleigh random variable if Rayleigh fading
- \( a = 1 \) if AWGN channel

Noise
- variance is: \( \sigma^2 = \frac{1}{2r \left( \frac{E_b}{N_0} \right)^2} = \frac{3}{2 \left( \frac{E_b}{N_0} \right)} \)
SISO-MAP Decoding Block

This block is implemented in the CML by the `SisoDecode` function.

- **Inputs:**
  - $\lambda_{u,i}$ LLR’s of the data bits. This comes from the other decoder $r$.
  - $\lambda_{c,i}$ LLR’s of the code bits. This comes from the channel observations $r$.

- **Two output streams:**
  - $\lambda_{u,o}$ LLR’s of the data bits. Passed to the other decoder.
  - $\lambda_{c,o}$ LLR’s of the code bits. Not used by the other decoder.

---

Turbo Decoding Architecture

- **Initialization and timing:**
  - Upper $\lambda_{u,i}$ input is initialized to all zeros.
  - Upper decoder executes first, then lower decoder.
Performance as a Function of Number of Iterations

Log-MAP Algorithm: Overview

- Log-MAP algorithm is MAP implemented in log-domain.
  - Multiplications become additions.
  - Additions become special "max*" operator (Jacobi logarithm)
- Log-MAP is similar to the Viterbi algorithm.
  - Except "max" is replaced by "max*" in the ACS operation.
- Processing:
  - Sweep through the trellis in **forward** direction using modified Viterbi algorithm.
  - Sweep through the trellis in **backward** direction using modified Viterbi algorithm.
  - Determine LLR for each trellis section.
  - Determine output extrinsic info for each trellis section.
The max* operator

- max* must implement the following operation:

\[ z = \max(x, y) + \ln(1 + \exp[-|y - x|]) \]
\[ = \max(x, y) + f_\epsilon(|y - x|) \]
\[ = \max^*(x, y) \]

- Ways to accomplish this:
  - C-function calls or large look-up-table. log-MAP
  - (Piecewise) linear approximation.
  - Rough correction value.

\[ z = \max(x, y) + \begin{cases} 0 & \text{if } |y - x| > 1.5 \\ 0.5 & \text{if } |y - x| \leq 1.5 \end{cases} \]

- Max operator.

The Correction Function

The dec_type option in SisoDecode
-0 For linear-log-MAP (DEFAULT)
-1 For max-log-MAP algorithm
-2 For Constant-log-MAP algorithm
-3 For log-MAP, correction factor from small nonuniform table and interpolation
-4 For log-MAP, correction factor uses C function calls
The Trellis for UMTS

- Dotted line = data 0
- Solid line = data 1
- Note that each node has one each of data 0 and 1 entering and leaving it.
- The branch from node $S_i$ to $S_j$ has metric $\gamma_{ij}$

$$\gamma_{ij} = X_k(i,j)\lambda_{ij}^d + X_k(i,j)\lambda_{ij}^c + Z_k(i,j)\lambda_{ij}^z$$

- Data bit associated with branch $S_i \rightarrow S_j$
- The two code bits labeling with branch $S_i \rightarrow S_j$

Forward Recursion

- A new metric must be calculated for each node in the trellis using:

$$\alpha_j = \max^*\{(\alpha_{i_1} + \gamma_{ij}), (\alpha_{i_2} + \gamma_{ij})\}$$

- where $i_1$ and $i_2$ are the two states connected to $j$.
- Start from the beginning of the trellis (i.e. the left edge).
- Initialize stage 0:

$$\alpha_0 = 0$$
$$\alpha_i = -\infty \text{ for all } i \neq 0$$
Backward Recursion

A new metric must be calculated for each node in the trellis using:

$$\beta_i = \max \left\{ \left( \beta_j^{*} + \gamma_{ji} \right), \left( \beta_j^{*} + \gamma_{ji}^{*} \right) \right\}$$

where $j_1$ and $j_2$ are the two states connected to $i$.

Start from the end of the trellis (i.e. the right edge).

Initialize stage L+3:

$$\beta_0 = 0$$

$$\beta_i = -\infty \text{ for all } i \neq 0$$

Log-likelihood Ratio

The likelihood of any one branch is:

$$\alpha_i + \gamma_{ji} + \beta_j$$

The likelihood of data 1 is found by summing the likelihoods of the solid branches.

The likelihood of data 0 is found by summing the likelihoods of the dashed branches.

The log likelihood ratio (LLR) is:

$$\Lambda(X_k) = \ln \left( \frac{P[X_k = 1]}{P[X_k = 0]} \right)$$

$$= \max_{j_i \rightarrow j_{i-1}} \left\{ \alpha_i + \gamma_{ji} + \beta_j \right\}$$

$$- \max_{j_i \rightarrow j_{i-1}} \left\{ \alpha_i + \gamma_{ji}^{*} + \beta_j^{*} \right\}$$
Memory Issues

- A naïve solution:
  - Calculate $\alpha$’s for entire trellis (forward sweep), and store.
  - Calculate $\beta$’s for the entire trellis (backward sweep), and store.
  - At the $k$th stage of the trellis, compute $\lambda$ by combining $\gamma$’s with stored $\alpha$’s and $\beta$’s.

- A better approach:
  - Calculate $\beta$’s for the entire trellis and store.
  - Calculate $\alpha$’s for the $k$th stage of the trellis, and immediately compute $\lambda$ by combining $\gamma$’s with these $\alpha$’s and stored $\beta$’s.
  - Use the $\alpha$’s for the $k$th stage to compute $\alpha$’s for state $k+1$.

- Normalization:
  - In log-domain, $\alpha$’s can be normalized by subtracting a common term from all $\alpha$’s at the same stage.
  - Can normalize relative to $\alpha_0$, which eliminates the need to store $\alpha_0$.
  - Same for the $\beta$’s.

Sliding Window Algorithm

- Can use a sliding window to compute $\beta$’s
  - Windows need some overlap due to uncertainty in terminating state.

  ![Diagram of sliding window algorithm](image-url)
Extrinsic Information

- The extrinsic information is found by subtracting the corresponding input from the LLR output, i.e.
  \[ \lambda^{u,i}_{\text{lower}} = \lambda^{u,o}_{\text{upper}} - \lambda^{u,i}_{\text{upper}} \]
  \[ \lambda^{u,i}_{\text{upper}} = \lambda^{u,o}_{\text{lower}} - \lambda^{u,i}_{\text{lower}} \]

- It is necessary to subtract the information that is already available at the other decoder in order to prevent "positive feedback".
- The extrinsic information is the amount of new information gained by the current decoder step.

Performance Comparison

Graph showing the BER of 640 bit turbo code with max-log-MAP, constant-log-MAP, and log-MAP decoding algorithms. The graph compares performance in AWGN and fading conditions with 10 decoder iterations.
cdma2000 uses a rate $\frac{1}{2}$ constituent encoder.

- Overall turbo code rate can be 1/5, 1/4, 1/3, or 1/2.
- Fixed interleaver lengths:
  - 378, 570, 762, 1146, 1530, 2398, 3066, 4602, 6138, 9210, 12282, or 20730

![Diagram of Turbo Encoder](diagram)

The performance of cdma2000 turbo code in AWGN with interleaver length 1530 is shown in the following graph.
CRSC codes use the concept of tailbiting.
- Sequence is encode so that initial state is same as final state.

Advantage and disadvantages
- No need for tail bits.
- Need to encode twice.
- Complicates decoder.

Duobinary codes
- Duobinary codes are defined over GF(4).
  - two bits taken in per clock cycle.
  - Output is systematic and rate 2/4.
- Hardware benefits
  - Half as many states in trellis.
  - Smaller loss due to max-log-MAP decoding.
Digital Video Broadcasting – Return Channel via Satellite.
- Consumer-grade Internet service over satellite.
- 144 kbps to 2 Mbps satellite uplink.
- Uses same antenna as downlink.
- QPSK modulation.

DVB-RCS uses a pair of duobinary CRSC codes.

Key parameters:
- input of $N = k/2$ couples
- $N = \{48, 64, 212, 220, 228, 424, 432, 440, 752, 848, 856, 864\}$
- $r = \{1/3, 2/5, 1/2, 2/3, 3/4, 4/5, 6/7\}$


DVB-RCS: Influence of Decoding Algorithm

- Rate $r=\frac{1}{2}$
- Length $N=212$
- 8 iterations.
- AWGN.
DVB-RCS: Influence of Block Length

- Rate ⅓
- max-log-MAP
- 8 iterations
- AWGN

DVB-RCS: Influence of Code Rate

- N=212
- max-log-MAP
- 8 iterations
- AWGN
802.16 (WiMax)

- The standard specifies an optional convolutional turbo code (CTC) for operation in the 2-11 GHz range.
- Uses same duobinary CRSC encoder as DVB-RCS, though without output W.
- Modulation: BPSK, QPSK, 16-QAM, 64-QAM, 256-QAM.
- Key parameters:
  - Input message size 8 to 256 bytes long.
  - \( r = \{1/2, 2/3, 3/4, 5/6, 7/8\} \)

Prelude to LDPC Codes:
Review of Linear Block Codes

- \( V_n = n \)-dimensional vector space over \( \{0, 1\} \)
- A \((n, k)\) linear block code with dataword length \( k \), codeword length \( n \) is a \( k \)-dimensional vector subspace of \( V_n \)
- A codeword \( c \) is generated by the matrix multiplication \( c = uG \), where \( u \) is the \( k \)-bit long message and \( G \) is a \( k \) by \( n \) generator matrix
- The parity check matrix \( H \) is a \( n-k \) by \( n \) matrix of ones and zeros, such that if \( c \) is a valid codeword then, \( cH^T = 0 \)
- Each row of \( H \) specifies a parity check equation. The code bits in positions where the row is one must sum (modulo-2) to zero
Low-Density Parity-Check Codes

- **Low-Density Parity-Check** (LDPC) codes are a class of linear block codes characterized by sparse parity check matrices $H$
  - $H$ has a low-density of 1's

- LDPC codes were originally invented by Robert Gallager in the early 1960's but were largely ignored until they were “rediscovered” in the mid-1990's by MacKay

- Sparseness of $H$ can yield large minimum distance $d_{\text{min}}$ and reduces decoding complexity

- Can perform within 0.0045 dB of Shannon limit

Decoding LDPC codes

- Like Turbo codes, LDPC can be decoded iteratively
  - Instead of a trellis, the decoding takes place on a **Tanner graph**
  - Messages are exchanged between the v-nodes and c-nodes
  - Edges of the graph act as *information pathways*

- Hard decision decoding
  - *Bit-flipping* algorithm

- Soft decision decoding
  - *Sum-product* algorithm
    - Also known as message passing/ belief propagation algorithm
  - *Min-sum* algorithm
    - Reduced complexity approximation to the sum-product algorithm

- In general, the per-iteration complexity of LDPC codes is less than it is for turbo codes
  - However, many more iterations may be required (max=100; avg=30)
  - Thus, overall complexity can be higher than turbo
Tanner Graphs

- A **Tanner graph** is a **bipartite** graph that describes the parity check matrix $H$.
- There are two classes of nodes:
  - **Variable-nodes**: Correspond to bits of the codeword or equivalently, to columns of the parity check matrix.
    - There are $n$ v-nodes.
  - **Check-nodes**: Correspond to parity check equations or equivalently, to rows of the parity check matrix.
    - There are $m = n - k$ c-nodes.
  - **Bipartite** means that nodes of the same type cannot be connected (e.g. a c-node cannot be connected to another c-node).
- The $i$th check node is connected to the $j$th variable node iff the $(i, j)$th element of the parity check matrix is one, i.e. if $h_{ij} = 1$.
- All of the v-nodes connected to a particular c-node must sum (modulo-2) to zero.

Example: Tanner Graph for (7,4) Hamming Code

$$
H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1
\end{bmatrix}
$$

---

Turbo and LDPC Codes
More on Tanner Graphs

- A **cycle** of length $l$ in a Tanner graph is a path of $l$ distinct edges which closes on itself.
- The **girth** of a Tanner graph is the minimum cycle length of the graph.
  - The shortest possible cycle in a Tanner graph has length 4.

![Tanner Graph Diagram]

Bit-Flipping Algorithm:
(7,4) Hamming Code

- Received code word:
  - $c_n = 1$  $c_i = 0$  $c_j = 1$  $c_k = 0$  $c_m = 0$  $c_p = 1$

- Transmitted code word:
  - $c_0 = 1$  $c_1 = 0$  $c_2 = 1$  $c_3 = 0$  $c_4 = 0$  $c_5 = 1$
Bit-Flipping Algorithm:
(7,4) Hamming Code

y₀ = 1  y₂ = 1  y₃ = 1  y₆ = 1  y₄ = 0  y₅ = 0

f₀ = 1  f₁ = 1  f₂ = 0

y₀ = 1  y₁ = 0  y₂ = 1  y₃ = 1  y₄ = 0  y₅ = 0  y₆ = 1
Generalized Bit-Flipping Algorithm

- Step 1: Compute parity-checks
  - If all checks are zero, stop decoding

- Step 2: Flip any digit contained in $T$ or more failed check equations

- Step 3: Repeat 1 to 2 until all the parity checks are zero or a maximum number of iterations are reached

- The parameter $T$ can be varied for a faster convergence

Generalized Bit Flipping: (15,7) BCH Code

Transmitted code word

$\begin{align*}
  f_0 &= 1 \\
  f_1 &= 0 \\
  f_2 &= 0 \\
  f_3 &= 0 \\
  f_4 &= 1 \\
  f_5 &= 0 \\
  f_6 &= 0 \\
  f_7 &= 1 \\
  c_0 &= 0 \\
  c_1 &= 0 \\
  c_2 &= 0 \\
  c_3 &= 0 \\
  c_4 &= 0 \\
  c_5 &= 0 \\
  c_6 &= 0 \\
  c_7 &= 0 \\
  c_8 &= 0 \\
  c_9 &= 0 \\
  c_{10} &= 0 \\
  c_{11} &= 0 \\
  c_{12} &= 0 \\
  c_{13} &= 0 \\
  c_{14} &= 0
\end{align*}$

Received code word

$\begin{align*}
  y_0 &= 0 \\
  y_1 &= 0 \\
  y_2 &= 0 \\
  y_3 &= 0 \\
  y_4 &= 1 \\
  y_5 &= 0 \\
  y_6 &= 0 \\
  y_7 &= 0 \\
  y_8 &= 0 \\
  y_9 &= 0 \\
  y_{10} &= 0 \\
  y_{11} &= 0 \\
  y_{12} &= 0 \\
  y_{13} &= 0 \\
  y_{14} &= 1
\end{align*}$
Generalized Bit Flipping: 
(15,7) BCH Code
Sum-Product Algorithm:

Notation

- $Q_0 = P(c_i = 0 | y, S_i)$, $Q_1 = P(c_i = 1 | y, S_i)$
- $S_i$ = event that bits in $c$ satisfy the $d_v$ parity check equations involving $c_i$
- $q_{ij}(b)$ = extrinsic info to be passed from v-node $i$ to c-node $j$
  - Probability that $c_i = b$ given extrinsic information from check nodes and channel sample $y_i$
- $r_{ij}(b)$ = extrinsic info to be passed from c-node $j$ to v-node $l$
  - Probability of the $j^{th}$ check equation being satisfied given that $c_i = b$
- $C_i$ = $\{ j : h_{ji} = 1 \}$
  - This is the set of row location of the 1’s in the $i^{th}$ column
- $C_i = (j : h_{ji} = 1) \setminus \{ j \}$
  - The set of row locations of the 1’s in the $i^{th}$ column, excluding location $j$
- $R_j$ = $\{ i : h_{ji} = 1 \}$
  - This is the set of column location of the 1’s in the $j^{th}$ row
- $R_j = \{ i : h_{ji} = 1 \} \setminus \{ i \}$
  - The set of column locations of the 1’s in the $j^{th}$ row, excluding location $i$

Step 1: Initialize $q_{ij}(0) = 1 - p_i = 1/(1 + \exp(-2y_i/\sigma^2))$
- $q_{ij}(1) = p_i = 1/(1 + \exp(2y_i/\sigma^2))$
- $q_{ij}(b) = \text{probability that } c_i = b, \text{ given the channel sample}$

Received code word (output of AWGN)
Sum-Product Algorithm

Step 2: At each c-node, update the $r$ messages

$$r_{j}(0) = \frac{1}{2} (1 - 2q_{j}(0))$$

$$r_{j}(1) = 1 - r_{j}(0)$$

$r_{j}(b)$ is the probability that $j$'th check equation is satisfied given $c_{j} = b$

\[
\begin{array}{c}
    v_0 & \quad v_1 & \quad v_2 & \quad v_3 & \quad v_4 & \quad v_5 & \quad v_6 \\
    f_0 & \quad f_1 & \quad f_2 & \quad f_3 & \quad f_4 & \quad f_5 & \quad f_6 \\
\end{array}
\]

$q_{j}(0)$ and $q_{j}(1)$ are used for hard decision making.

\[
\begin{array}{c}
    v_0 & \quad v_1 & \quad v_2 & \quad v_3 & \quad v_4 & \quad v_5 & \quad v_6 \\
    f_0 & \quad f_1 & \quad f_2 & \quad f_3 & \quad f_4 & \quad f_5 & \quad f_6 \\
\end{array}
\]

\[
\begin{array}{c}
    v_0 & \quad v_1 & \quad v_2 & \quad v_3 & \quad v_4 & \quad v_5 & \quad v_6 \\
    f_0 & \quad f_1 & \quad f_2 & \quad f_3 & \quad f_4 & \quad f_5 & \quad f_6 \\
\end{array}
\]

Make hard decision

$$Q_{c} = \begin{cases} 
1 & \text{if } Q_{c}(1) \geq 0.5 \\
0 & \text{otherwise}
\end{cases}$$
Halting Criteria

- After each iteration, halt if:
  \[ \hat{c} \mathbf{H}^T = \mathbf{0} \]
- This is effective, because the probability of an undetectable decoding error is negligible
- Otherwise, halt once the maximum number of iterations is reached
- If the Tanner graph contains no cycles, then \( Q_i \) converges to the true APP value as the number of iterations tends to infinity

Sum-Product Algorithm in Log Domain

- The sum-product algorithm in probability domain has two shortcomings
  - Numerically unstable
  - Too many multiplications
- A log domain version is often used for practical purposes
- \( Q_i = \log \left( \frac{P(c_i = 0 | y, S_i)}{P(c_i = 1 | y, S_i)} \right) \) LLR of the \( i \)th code bit (ultimate goal of algorithm)
- \( q_{ij} = \log \left( \frac{q_{ij}(0)}{q_{ij}(1)} \right) \) extrinsic info to be passed from v-node \( i \) to c-node \( j \)
- \( r_{ij} = \log \left( \frac{r_{ij}(0)}{r_{ij}(1)} \right) \) extrinsic info to be passed from c-node \( j \) to v-node \( i \)
Sum-Product Decoder
(in Log-Domain)

- Initialize:
  \[ q_{ij} = \lambda_i = 2y_i/\sigma^2 = \text{channel LLR value} \]
- Loop over all i,j for which \( h_{ij} = 1 \)
  - At each c-node, update the r messages:
    \[
    r_{ji} = \left( \prod_{j \in B_{ji}} \alpha_{r_j} \right) \phi \left( \sum_{j \in B_{ji}} \beta_{r_j} \right)
    \]
  - At each v-node update the q message and Q LLR:
    \[
    Q_i = \lambda_i + \sum_{j \in \mathcal{C}_i} r_{ji}
    \]
    \[
    q_{ij} = Q_i - r_{ji}
    \]
  - Make hard decision:
    \[
    \hat{c}_i = \begin{cases} 
      1 & \text{if } Q_i < 0 \\
      0 & \text{otherwise}
    \end{cases}
    \]

Sum-Product Algorithm:
Notation

- \( \alpha_{ij} = \text{sign} \left( q_{ij} \right) \)
- \( \beta_{ij} = | q_{ij} | \)
- \( \phi(x) = -\log \tanh(x/2) = \log \left( \frac{e^x+1}{e^x-1} \right) = \phi^{-1}(x) \)
Min-Sum Algorithm

- Note that:
  \[ \phi \left( \sum_{j} \phi(\beta_j) \right) = \phi \left( \min_j \beta_j \right) = \min_j \beta_j \]

- So we can replace the \( r \) message update formula with
  \[ r_{ji} = \prod_{\ell \in K_{ji}} \alpha_{\ell j} \left( \min_{\ell \in K_{ji}} \beta_{\ell j} \right) \]

- This greatly reduces complexity, since now we don’t have to worry about computing the nonlinear \( \phi \) function.

- Note that since \( \alpha \) is just the sign of \( q \), \( \prod \alpha \) can be implemented by using XOR operations.

---

BER of Different Decoding Algorithms

- Code #1:
  - MacKay’s construction 2A
  - AWGN channel
  - BPSK modulation

![BER graph](image-url)
Extrinsic-information Scaling

- As with max-log-MAP decoding of turbo codes, min-sum decoding of LDPC codes produces an extrinsic information estimate which is biased.
  - In particular, $r_{ji}$ is overly optimistic.

- A significant performance improvement can be achieved by multiplying $r_{ji}$ by a constant $\kappa$, where $\kappa<1$.
  
  $$r'_{ji} = \kappa \left( \prod_{i \in R_{ji}} \alpha_{i,j} \right) \min_{i \in R_{ji}} \beta_{r_{i'}}$$


  - Experimentation shows that $\kappa=0.9$ gives best performance.

---

### BER of Different Decoding Algorithms

- Code #1: MacKay’s construction 2A
- AWGN channel
- BPSK modulation

- Min-sum
- Min-sum w/ extrinsic info scaling
  Scale factor $\kappa=0.9$

- Sum-product

- Eb/No in dB vs. BER
Regular vs. Irregular LDPC codes

- An LDPC code is **regular** if the rows and columns of $H$ have uniform weight, i.e., all rows have the same number of ones ($d_v$) and all columns have the same number of ones ($d_c$)
  - The codes of Gallager and MacKay were regular (or as close as possible)
  - Although regular codes had impressive performance, they are still about 1 dB from capacity and generally perform worse than turbo codes
- An LDPC code is **irregular** if the rows and columns have non-uniform weight
  - Irregular LDPC codes tend to outperform turbo codes for block lengths of about $n>10^5$

- The degree distribution pair $(\lambda, \rho)$ for a LDPC code is defined as
  $$\lambda_i(x) = \sum_{j=0}^{\infty} \lambda_j x^{i-j}$$
  $$\rho_i(x) = \sum_{j=0}^{\infty} \rho_j x^{i-j}$$
- $\lambda_i, \rho_i$ represent the fraction of edges emanating from variable (check) nodes of degree $i$

Constructing Regular LDPC Codes: MacKay, 1996

- Around 1996, Mackay and Neal described methods for constructing sparse $H$ matrices
  - The idea is to randomly generate a $M \times N$ matrix $H$ with weight $d_v$ columns and weight $d_c$ rows, subject to some constraints
  - Construction 1A: Overlap between any two columns is no greater than 1
    - This avoids length 4 cycles
  - Construction 2A: $M/2$ columns have $d_v=2$, with no overlap between any pair of columns. Remaining columns have $d_v=3$. As with 1A, the overlap between any two columns is no greater than 1
  - Construction 1B and 2B: Obtained by deleting select columns from 1A and 2A
    - Can result in a higher rate code
Constructing Irregular LDPC Codes: Luby, et. al., 1998

- Luby et. al. (1998) developed LDPC codes based on irregular LDPC Tanner graphs
- Message and check nodes have conflicting requirements
  - Message nodes benefit from having a large degree
  - LDPC codes perform better with check nodes having low degrees
- Irregular LDPC codes help balance these competing requirements
  - High degree message nodes converge to the correct value quickly
  - This increases the quality of information passed to the check nodes, which in turn helps the lower degree message nodes to converge
- Check node degree kept as uniform as possible and variable node degree is non-uniform
  - Code 14: Check node degree =14, Variable node degree =5, 6, 21, 23
- No attempt made to optimize the degree distribution for a given code rate

Density Evolution: Richardson and Urbanke, 2001

- Given an irregular Tanner graph with a maximum $d_v$ and $d_c$, what is the best degree distribution?
  - How many of the v-nodes should be degree $d$, $d_v$-1, $d_v$-2,... nodes?
  - How many of the c-nodes should be degree $d$, $d_c$-1,... nodes?
- Question answered using Density Evolution
  - Process of tracking the evolution of the message distribution during belief propagation
- For any LDPC code, there is a “worst case” channel parameter called the threshold such that the message distribution during belief propagation evolves in such a way that the probability of error converges to zero as the number of iterations tends to infinity
- Density evolution is used to find the degree distribution pair $(\lambda, \rho)$ that maximizes this threshold
Density Evolution:
Richardson and Urbanke, 2001

- Step 1: Fix a maximum number of iterations
- Step 2: For an initial degree distribution, find the threshold
- Step 3: Apply a small change to the degree distribution
  - If the new threshold is larger, fix this as the current distribution
- Repeat Steps 2-3

Richardson and Urbanke identify a rate $\frac{1}{2}$ code with degree distribution pair which is 0.06 dB away from capacity
Chung et al., use density evolution to design a rate $\frac{1}{2}$ code which is 0.0045 dB away from capacity

More on Code Construction

- LDPC codes, especially irregular codes exhibit error floors at high SNRs
- The error floor is influenced by $d_{\text{min}}$
  - Directly designing codes for large $d_{\text{min}}$ is not computationally feasible
- Removing short cycles indirectly increases $d_{\text{min}}$ (*girth conditioning*)
  - Not all short cycles cause error floors
- *Trapping sets* and *Stopping sets* have a more direct influence on the error floor
- Error floors can be mitigated by increasing the size of minimum stopping sets
- Trapping sets can be mitigated using *averaged belief propagation decoding*
  - Milenkovic, “Algorithmic and combinatorial analysis of trapping sets in structured LDPC codes”, *in Proc. Intl. Conf. on Wireless Ntw., Communications and Mobile computing*, 2005
- LDPC codes based on *projective geometry* reported to have very low error floors
A linear block code is encoded by performing the matrix multiplication $c = uG$.

A common method for finding $G$ from $H$ is to first make the code systematic by adding rows and exchanging columns to get the $H$ matrix in the form $H = [P^T I]$.
- Then $G = [I P]$.
- However, the result of the row reduction is a non-sparse $P$ matrix.
- The multiplication $c = [u uP]$ is therefore very complex.

As an example, for a (10000, 5000) code, $P$ is 5000 by 5000.
- Assuming the density of 1’s in $P$ is 0.5, then $0.5 \times (5000)^2$ additions are required per codeword.

This is especially problematic since we are interested in large $n (>10^5)$.

An often used approach is to use the all-zero codeword in simulations.

Richardson and Urbanke show that even for large $n$, the encoding complexity can be (almost) linear function of $n$.


Using only row and column permutations, $H$ is converted to an approximately lower triangular matrix.
- Since only permutations are used, $H$ is still sparse.
- The resulting encoding complexity in almost linear as a function of $n$.

An alternative involving a sparse-matrix multiply followed by differential encoding has been proposed by Ryan, Yang, & Li.
Encoding LDPC Codes

- Let $H = [H_1 \ H_2]$ where $H_1$ is sparse and
  
  \[
  H_2 = \begin{bmatrix}
  1 & 1 & \cdots & 1 \\
  1 & 1 & \cdots & 1 \\
  1 & \cdots & 1 \\
  1 & 1 & \cdots & 1
  \end{bmatrix}
  \quad \text{and} \quad
  H_2^T = \begin{bmatrix}
  1 & 1 & \cdots & 1 \\
  1 & 1 & \cdots & 1 \\
  1 & \cdots & 1 \\
  \cdots & \cdots & \cdots & \cdots \\
  1 & 1 & \cdots & 1
  \end{bmatrix}
  \]

- Then a systematic code can be generated with $G = [I \ H_1 \ H_2^T]$.
- It turns out that $H_2^T$ is the generator matrix for an accumulate-code (differential encoder), and thus the encoder structure is simply:

\[
\begin{align*}
  &u \\
  \downarrow \text{Multiply} \quad \text{by} \quad H_2^T \\
  \downarrow \oplus \\
  \downarrow \text{D} \\
  \rightarrow uH_1^T H_2^T
\end{align*}
\]

- Similar to Jin & McEliece’s Irregular Repeat Accumulate (IRA) codes.
  - Thus termed “Extended IRA Codes”

Performance Comparison

- We now compare the performance of the maximum-length UMTS turbo code against four LDPC code designs.
- Code parameters
  - All codes are rate $\frac{1}{3}$
  - The LDPC codes are length $(n,k) = (15000, 5000)$
    - Up to 100 iterations of log-domain sum-product decoding
    - Code parameters are given on next slide
  - The turbo code has length $(n,k) = (15354, 5114)$
    - Up to 16 iterations of log-MAP decoding
- BPSK modulation
- AWGN and fully-interleaved Rayleigh fading
- Enough trials run to log 40 frame errors
  - Sometimes fewer trials were run for the last point (highest SNR).
LDPC Code Parameters

- **Code 1: MacKay’s regular construction 2A**

- **Code 2: Richardson & Urbanke irregular construction**

- **Code 3: Improved irregular construction**
  - Idea is to avoid small stopping sets

- **Code 4: Extended IRA code**

LDPC Degree Distributions

- The distribution of row-weights, or check-node degrees, is as follows:

<table>
<thead>
<tr>
<th>i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10000</td>
<td>4999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>5458</td>
<td>5000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9987</td>
<td>4542</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

  Code number:
  1 = MacKay construction 2A
  2 = Richardson & Urbanke
  3 = Jones, Wesel, & Tian
  4 = Ryan’s Extended-IRA

- The distribution of column-weights, or variable-node degrees, is:

<table>
<thead>
<tr>
<th>i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>8282</td>
<td>9045</td>
<td>9999</td>
</tr>
<tr>
<td>3</td>
<td>10000</td>
<td>2238</td>
<td>2267</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2584</td>
<td>569</td>
<td>5000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>206</td>
<td>1941</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1689</td>
<td>1178</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FER for DVB-S2 LDPC Code
Normal Frames in BPSK/AWGN

FER for DVB-S2 LDPC Code
Short Frames in BPSK/AWGN
M-ary Complex Modulation

- \( \mu = \log_2 M \) bits are mapped to the symbol \( x_k \), which is chosen from the set \( S = \{ x_1, x_2, \ldots, x_M \} \)
  - The symbol is multidimensional.
  - 2-D Examples: QPSK, M-PSK, QAM, APSK, HEX
  - M-D Example: FSK, block space-time codes (BSTC)
- The signal \( y = hx_k + n \) is received
  - \( h \) is a complex fading coefficient.
  - More generally (BSTC), \( Y = HX + N \)
- Modulation implementation in the ISCML
  - The complex signal set \( S \) is created with the CreateConstellation function.
  - Modulation is performed using the Modulate function.

Log-likelihood of Received Symbols

- Let \( p(x_k|y) \) denote the probability that signal \( x_k \in S \) was transmitted given that \( y \) was received.
- Let \( f(x_k|y) = K \ p(x_k|y) \), where \( K \) is any multiplicative term that is constant for all \( x_k \).
- When all symbols are equally likely, \( f(x_k|y) \propto f(y|x_k) \)
- For each signal in \( S \), the receiver computes \( f(y|x_k) \)
  - This function depends on the modulation, channel, and receiver.
  - Implemented by the Demod2D and DemodFSK functions, which actually computes \( \log f(y|x_k) \).
- Assuming that all symbols are equally likely, the most likely symbol \( x_k \) is found by making a hard decision on \( f(y|x_k) \) or \( \log f(y|x_k) \).
Example: QAM over AWGN.

Let \( y = x + n \), where \( n \) is complex i.i.d. \( N(0,N_0/2) \) and the average energy per symbol is \( E[|x|^2] = E_s \)

\[
p(y|x_k) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{|y-x_k|^2}{2\sigma^2} \right\}
\]

\[
f(y|x_k) = \exp \left\{ -\frac{|y-x_k|^2}{2\sigma^2} \right\}
\]

\[
\log f(y|x_k) = -\frac{|y-x_k|^2}{2\sigma^2}
\]

\[
= -\frac{E_s |y-x_k|^2}{N_o}
\]

Log-Likelihood of Symbol \( x_k \)

The log-likelihood of symbol \( x_k \) is found by:

\[
\Lambda_k = \log p(x_k | y)
\]

\[
= \log \frac{p(x_k | y)}{\sum_{x_n \in S} p(x_k | y)}
\]

\[
= \log \frac{f(y | x_k)}{\sum_{x_n \in S} f(y | x_m)}
\]

\[
= \log f(y | x_k) - \log \sum_{x_n \in S} f(y | x_m)
\]

\[
= \log f(y | x_k) - \log \sum_{x_n \in S} \exp\left\{ \log f(y | x_m) \right\}
\]

\[
= \log f(y | x_k) - \max_{x_n \in S} \left\{ \log f(y | x_m) \right\}
\]

6/7/2006
The max* function

\[
\text{max}^* (x, y) = \log(\exp(x) + \exp(y)) = \max(x, y) + \log(1 + \exp(-|y - x|)) = \max(x, y) + f_c(|y - x|)
\]

\[
f_c(z) = \log(1 + \exp(-z))
\]

Capacity of Coded Modulation (CM)

- Suppose we want to compute capacity of M-ary modulation
  - In each case, the input distribution is constrained, so there is no need to maximize over \(p(x)\)
  - The capacity is merely the mutual information between channel input and output.
- The mutual information can be measured as the following expectation:

\[
C = I(X; Y) = E_{x,y} \left[ \log M + \log p(x_k | y) \right] \text{ nats}
\]
Monte Carlo Calculation of the Capacity of Coded Modulation (CM)

- The mutual information can be measured as the following expectation:
  \[ C = I(X; Y) = E_{x_k} \left[ \log M + \log p(x_k | y) \right] \text{ nats} \]
  \[ = \log M + E_{x_k} \left[ \Lambda_k \right] \text{ nats} \]
  \[ = \log_2 M + \frac{E_{x_k} \left[ \Lambda_k \right]}{\log(2)} \text{ bits} \]
  \[ = \mu + \frac{E_{x_k} \left[ \Lambda_k \right]}{\log(2)} \text{ bits} \]

- This expectation can be obtained through Monte Carlo simulation.

Simulation Block Diagram

This function is computed by the CML function Demod2D

This function is computed by the CML function Capacity

Modulator:
Pick \( x_k \) at random from S

Receiver:
Compute \( \log f(y|x_k) \) for every \( x_k \in S \)

Calculate:
\[ \Lambda_k = \log f(y | x_k) \]
\[ - \max_{x_{k+1}} \left[ \log f(y | x_{k+1}) \right] \]

Noise Generator

Benefits of Monte Carlo approach:
- Allows high dimensional signals to be studied.
- Can determine performance in fading.
- Can study influence of receiver design.

After running many trials, calculate:
\[ C = \mu + \frac{E[\Lambda_k]}{\log(2)} \]
Capacity of 2-D modulation in AWGN

Capacity of M-ary Noncoherent FSK in AWGN


**BICM**

- Coded modulation (CM) is required to attain the aforementioned capacity.
  - Channel coding and modulation handled jointly.
  - e.g. trellis coded modulation (Ungerboeck); coset codes (Forney)
- Most off-the-shelf capacity approaching codes are binary.
- A pragmatic system would use a binary code followed by a bitwise interleaver and an M-ary modulator.
  - Bit Interleaved Coded Modulation (BICM); Caire 1998.
Transforming Symbol Log-Likehoods Into Bit LLRs

- Like the CM receiver, the BICM receiver calculates \( \log f(y|x_k) \) for each signal in \( S \).
- Furthermore, the BICM receiver needs to calculate the log-likelihood ratio of each code bit:

\[
\lambda_n = \log \frac{p(c_n = 1 | y)}{p(c_n = 0 | y)} = \log \frac{\sum_{x_k \in S_1} p(x_k | y) p[x_k]}{\sum_{x_k \in S_0} p(x_k | y) p[x_k]}
\]

\[
= \max_{x_k \in S_1} \log f(y|x_k) - \max_{x_k \in S_0} \log f(y|x_k)
\]

- where \( S_1 \) represents the set of symbols whose \( n \)-th bit is a 1.
- and \( S_0 \) is the set of symbols whose \( n \)-th bit is a 0.

BICM Capacity

- BICM transforms the channel into \( \mu \) parallel binary channels, and the capacity of the \( n \)-th channel is:

\[
C_n = E_{c,n} \left[ \log(2) + \log p(c_n | y) \right] \text{ nats}
\]

\[
= \log(2) + E_{c,n} \left[ \log \frac{p(c_n = 1 | y)}{p(c_n = 0 | y) + p(c_n = 1 | y)} \right] \text{ nats}
\]

\[
= \log(2) - E_{c,n} \left[ \log \frac{p(c_n = 0 | y) + p(c_n = 1 | y)}{p(c_n | y)} \right] \text{ nats}
\]

\[
= \log(2) - E_{c,n} \left[ \log \left( \exp \log \frac{p(c_n = 0 | y)}{p(c_n | y)} + \exp \log \frac{p(c_n = 1 | y)}{p(c_n | y)} \right) \right] \text{ nats}
\]

\[
= \log(2) - E_{c,n} \left[ \max \left\{ \log \frac{p(c_n = 0 | y)}{p(c_n | y)}, \log \frac{p(c_n = 1 | y)}{p(c_n | y)} \right\} \right] \text{ nats}
\]

\[
= \log(2) - E_{c,n} \left[ \max \left\{ 0, (-1)^c \lambda_n \right\} \right] \text{ nats}
\]
BICM Capacity (Continued)

Since capacity over parallel channels adds, the capacity of BICM is:

\[
C = \sum_{k=1}^{\mu} C_k
\]

\[
= \sum_{k=1}^{\mu} \left( \log(2) - E_{c,k} \left[ \max \left\{ 0, (-1)^{c_k} \lambda_k \right\} \right] \right) \text{ nats}
\]

\[
= \mu \log(2) - \sum_{k=1}^{\mu} E_{c,k} \left[ \max \left\{ 0, (-1)^{c_k} \lambda_k \right\} \right] \text{ nats}
\]

\[
= \mu - \frac{1}{\log(2)} \sum_{k=1}^{\mu} E_{c,k} \left[ \max \left\{ 0, (-1)^{c_k} \lambda_k \right\} \right] \text{ bits}
\]

BICM Capacity

As with CM, this can be computed using a Monte Carlo integration.

Modulator:
Pick \( x_k \) at random from \( S \)

Receiver:
Compute \( p(y|x_k) \) for every \( x_k \in S \)

\[ \sum_{x_k \in S} p(y|x_k) \]

\[ \lambda_{\text{in}} = \log \left( \sum_{x_k \in S} p(y|x_k) \right) \]

For each bit, calculate:

\[ \sum_{x_k \in S} p(y|x_k) \]

For each symbol, calculate:

\[ \Lambda = -\sum_{k=1}^{\mu} E_{c,k} \left[ \max \left\{ 0, (-1)^{c_k} \lambda_k \right\} \right] \text{ bits} \]

Unlike CM, the capacity of BICM depends on how bits are mapped to symbols

\[ C = \mu + \frac{E[\Lambda]}{\log(2)} \]
The conventional BICM receiver assumes that all bits in a symbol are equally likely:

$$\lambda_n = \log \frac{\sum_{x \in S_n^{(1)}} p(x | y)}{\sum_{x \in S_n^{(1)}} p(x | y)} = \log \frac{\sum_{y \in S_n^{(0)}} p(y | x) p(x | c_n = 1)}{\sum_{y \in S_n^{(0)}} p(y | x) p(x | c_n = 0)}$$

However, if the receiver has estimates of the bit probabilities, it can use this to weight the symbol likelihoods.

$$\lambda_n = \log \frac{\sum_{x \in S_n^{(1)}} p(y | x) p(x | c_n = 1)}{\sum_{x \in S_n^{(1)}} p(y | x) p(x | c_n = 0)}$$

This information is obtained from decoder feedback.

- Bit Interleaved Coded Modulation with Iterative Demodulation
- Li and Ritcey 1999.
Now consider a receiver that has a priori information about the code bits (from a soft output decoder).

Assume the following:
- The a priori information is in LLR form.
- The a priori LLR’s are Gaussian distributed.
- The LLR’s have mutual information $I_v$

Then the mutual information $I_z$ at the output of the receiver can be measured through Monte Carlo Integration.
- $I_z$ vs. $I_v$ is the Mutual Information Transfer Characteristic.
- ten Brink 1999.

Generating Random a Priori Input

There is a one-to-one correspondence between the mutual information and the variance of the Gaussian distributed a priori input.
Mutual Information Characteristic

EXIT Chart

16-QAM
AWGN
6.8 dB
adding curve for a FEC code
makes this an extrinsic information transfer (EXIT) chart
EXIT Chart Analysis of Turbo Codes

- PCCC (turbo) codes can be analyzed with an EXIT chart by plotting the mutual information transfer characteristics of the two decoders.
Conclusions

- It is now possible to closely approach the Shannon limit by using turbo and LDPC codes.
- Binary capacity approaching codes can be combined with higher order modulation using the BICM principle.
- These code are making their way into standards
  - Binary turbo: UMTS, cdma2000
  - Duobinary turbo: DVB-RCS, 802.16
  - LDPC: DVB-S2 standard.
- Software for simulating turbo and LDPC codes can be found at www.iterativesolutions.com