To Iterate or Feedback: How to Achieve Universally Good Performance on Channels with Memory?

Krishna Narayanan

Department of Electrical Engineering

Texas A&M University

College Station, TX, USA

Acknowledgement: Nitin Nangare and Kapil Bhattad

Definition of Channel with Memory

- Memoryless Channel: $p(y_1, \ldots, y_N | x_1, \ldots, x_N) = \prod_{i=1}^{N} p(y_i | x_i)$
- Channel with memory: The above condition is not true
- In this talk, focus is on the case when there is a Markov structure

Coding for Memoryless Channels

- Consider the following Memoryless Channels

Additive White Noise: $y_k = x_k + n_k \qquad x_k \in \{\pm 1\}$

BSC: $y_k = x_k \oplus e_k$

BEC: $y_k = x_k$ OR $y_k =$ Erasure

Fading with CSI at Rx: $y_k = \alpha_k \, x_k + n_k, \quad \alpha_k$ is known at the receiver

- Code design for such channels is well understood
  - Low Density Parity Check (LDPC) codes perform close to capacity
  - Tools to analyze and design good code ensembles exist

Universal LDPC Codes for Memoryless channels

- LDPC codes perform uniformly well on several memoryless channels (S.Y. Chung, Ph.D. thesis, MIT)

j	k	rate	BEC	BSC	Laplace	AWGN
3	4	0.25	0.353	0.349	0.348	0.346
3	5	0.4	0.482	0.489	0.485	0.480
3	6	0.5	0.571	0.584	0.577	0.571
3	9	2/3	0.717	0.739	0.728	0.720
3	12	0.75	0.790	0.814	0.801	0.793
4	10	0.6	0.692	0.689	0.684	0.683

- A single good LDPC code exists for the class of AWN channels

$$y_k = x_k + n_k$$

Examples of Channels with memory

- ISI channel: $y_k = \sum_{i=-l_2}^{l_1} h_i \, x_{k-i} + n_k \qquad x_k \in \mathcal{X}$

- Additive correlated noise: $y_k = x_k \oplus w_k$

- MIMO Channels: $\boldsymbol{Y}_k = \boldsymbol{H} \, \boldsymbol{x}_k + \boldsymbol{n}_k$

- Fading channels without CSI: $y_k = \alpha_k \, x_k + n_k$

- We want good codes for these channels (preferably universal)

Capacity of ISI Channels with Input Constraints

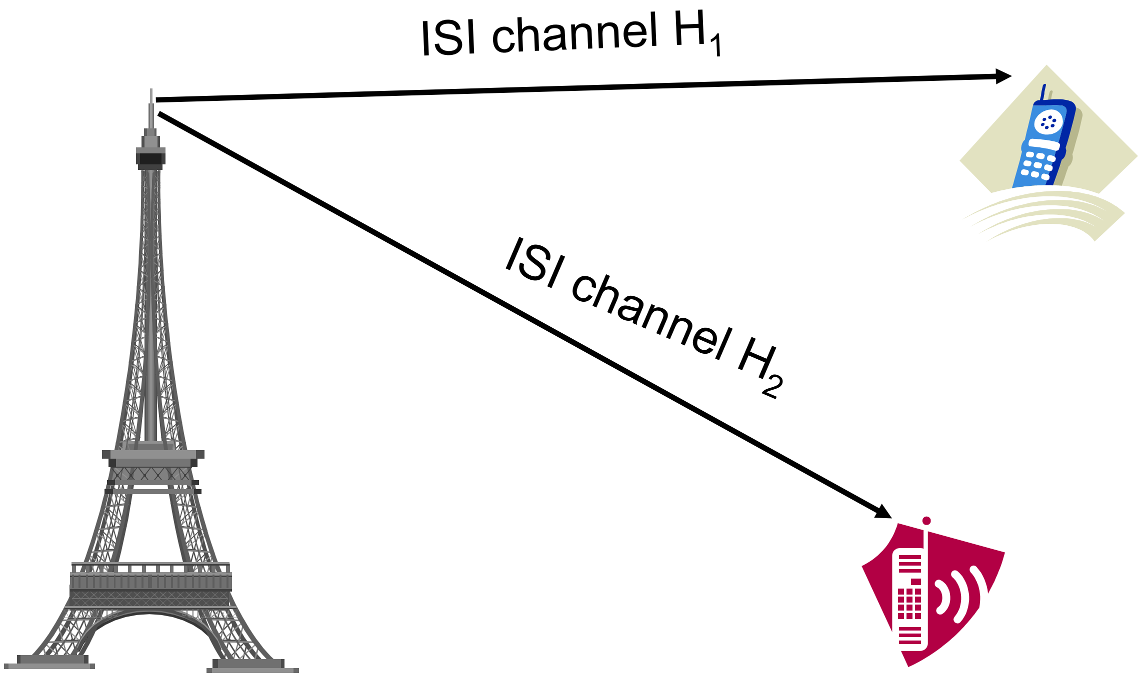
$$y_k = \sum_{i=-l_2}^{l_1} h_i \, x_{k-i} + n_k \qquad x_k \in \{\pm 1\}$$

- Computing capacity with input constraints is quite difficult
- Good upper and lower bounds are available
- $C_{i.i.d}$ – is a lower bound assuming x_k's are i.i.d
- Better bounds – assume x_k's are the output of a Markov chain
- When channel is not known at transmitter, $C_{i.i.d}$ is a meaningful measure of performance

What is the universality we are looking for?

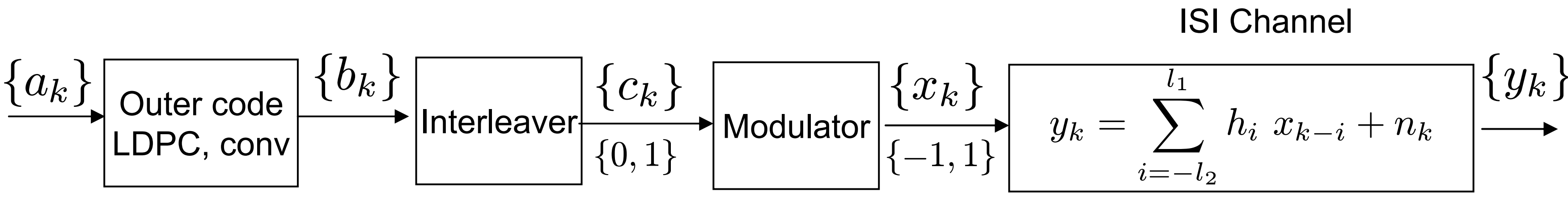
- Set of ISI channels $\{H_1, H_2, \ldots.\}$, $C_{iid}(H_i) = R$
- We want a coding scheme of rate R (or close to R) and a receiver such that for any H_i, reliable communication is possible
- Note that the channel energy is not normalized
- In the MIMO case, for every realization of **H** for which $C_{i.i.d}(\mathbf{H}) > R$, we want to the code to work
- With some distribution on H, we talk of Outage capacity

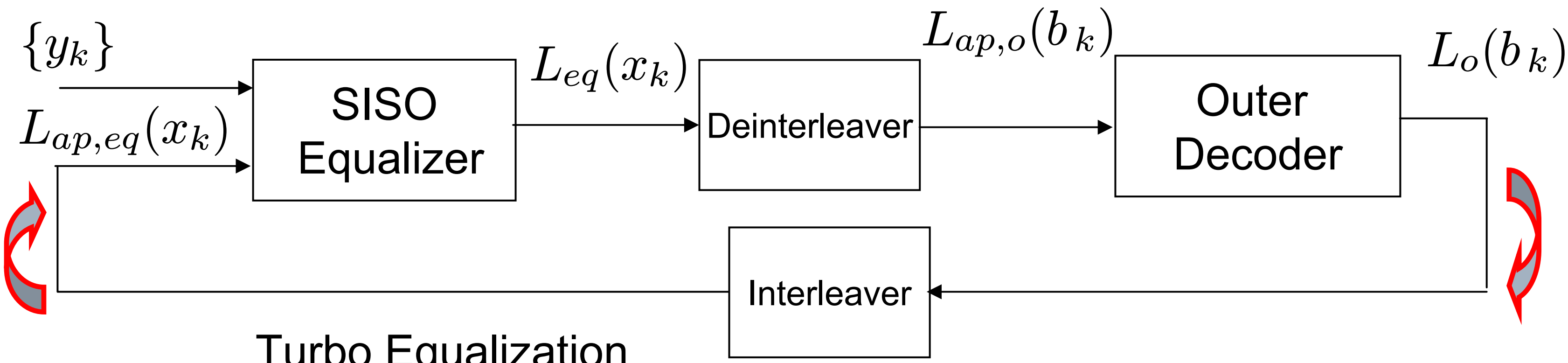
A Possible Application



- Note that the problem is not to optimize sum rate
- In this sense, it is different from conventional broadcast problem

How to deal with the memory? - Equalization





Turbo Equalization

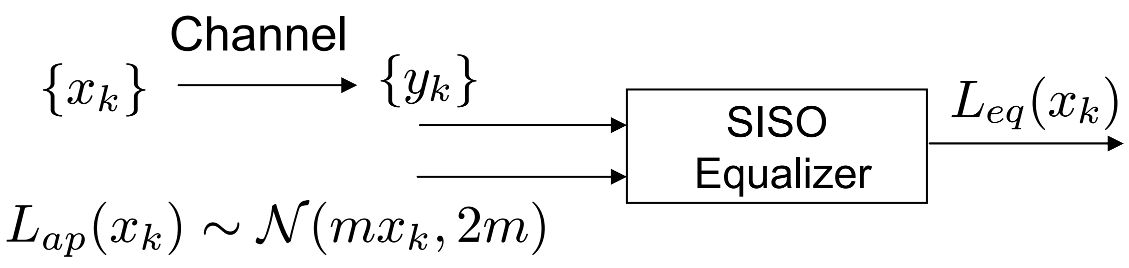
- Computationally efficient way to approximate the MAP decoder
- Code design with turbo equalization [Varnica & Kavcic '02]

Main point of the talk

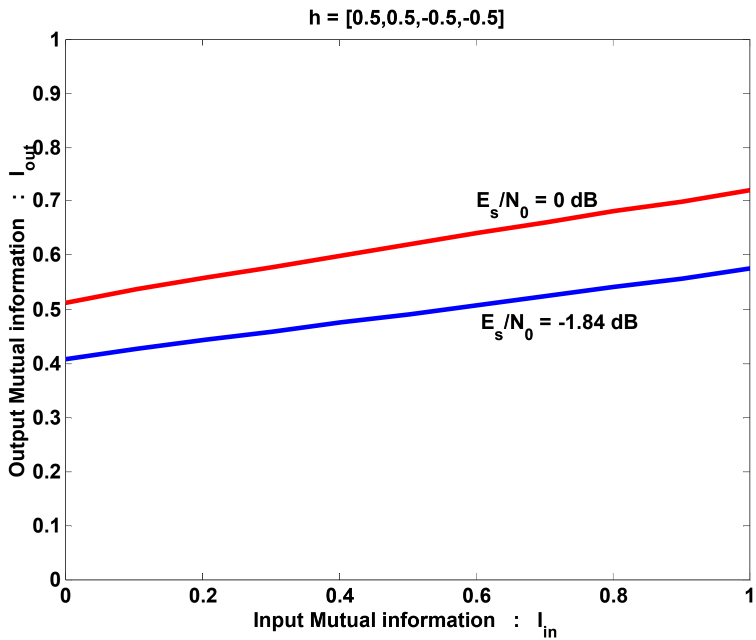
- Can Turbo equalization provide near capacity performance?
  - Yes!
  - But optimal codes are different for different channel realizations
  - Not a computationally efficient solution for near-capacity performance
- Coded Decision Feedback Equalization
  - Non-iterative structure that can provide close to capacity performance
  - Leads to universal codes and receiver structures
  - Complexity is lower than that of turbo equalization
- This principle applies to other signal processing functions as well

Extrinsic Information Transfer (EXIT) Curves

- Proposed by Ten Brink, Similar ideas by some other researchers also



F_{eq} - Plot of $I_{in} = I(L_{ap}; X)$ vs $I_{out} = I(L_{eq}; X)$



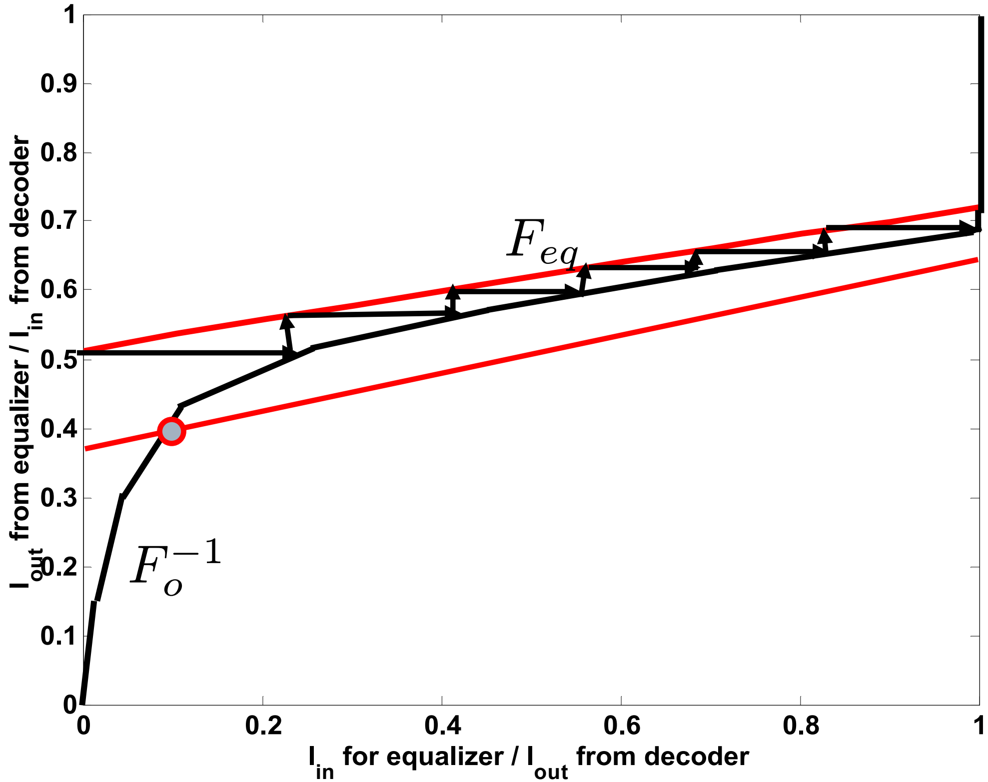
EXIT Curve for Outer Decoder

- Similar curve for the outer decoder also

$$L_{ap,o}(x_k) \sim \mathcal{N}(mx_k, 2m) \qquad L_o(x_k)$$

F_o - Plot of I_{out} vs I_{in} - transfer function of decoder

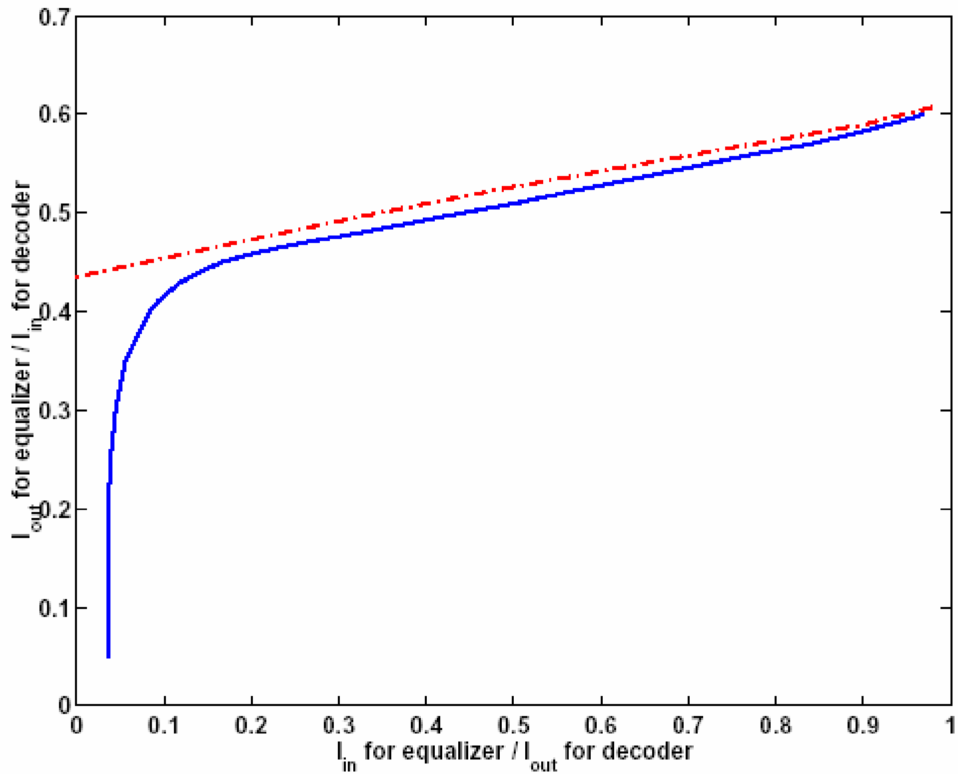
EXIT Chart



Convergence: $F_o^{-1} < F_{eq}$ for convergence

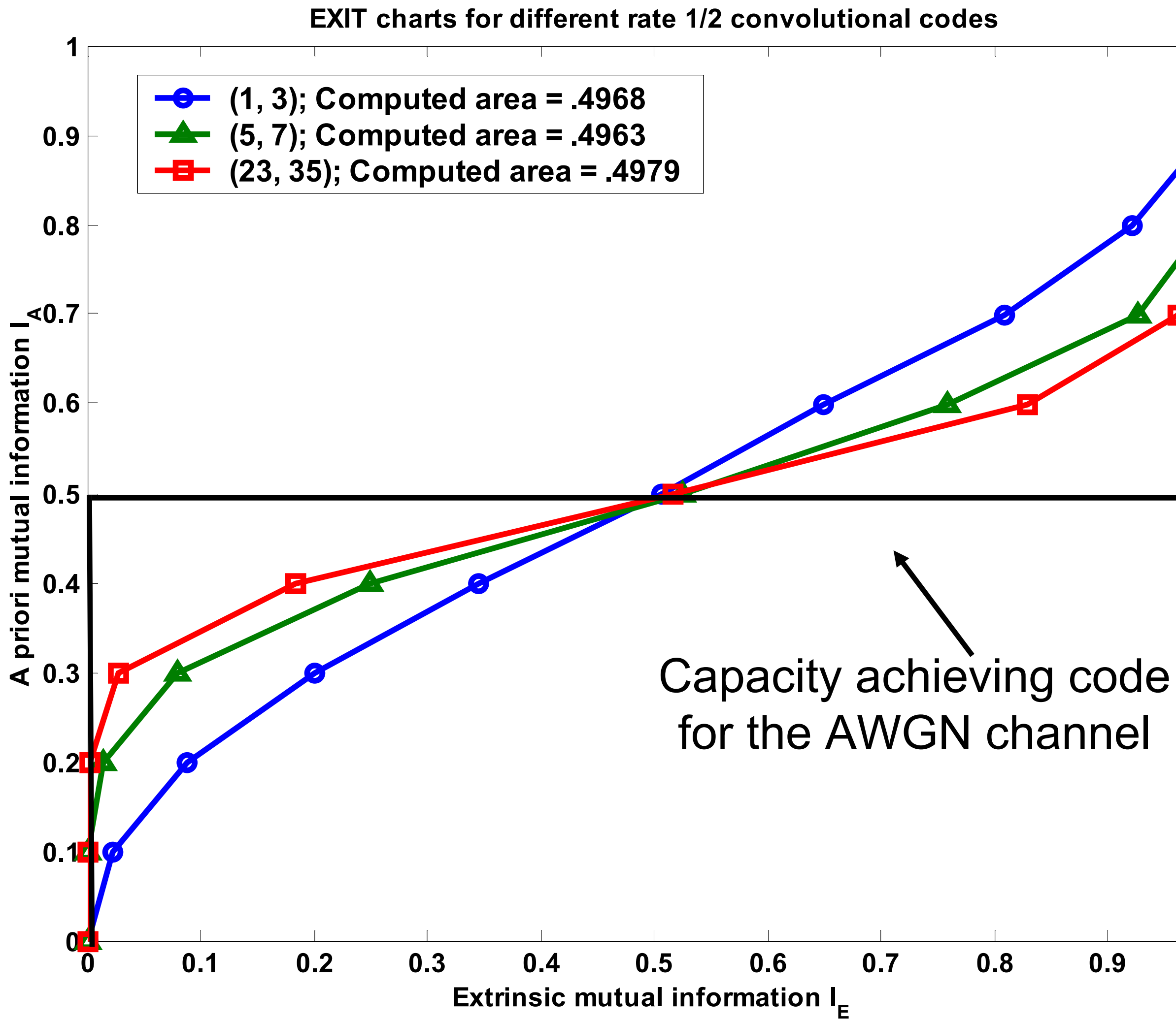
Area Property: [Ashikmin 02] If extrinsic info is from an erasure channel, area under $F_o^{-1}(x) = R$, Rate of the code

Optimal Code Design with Turbo Equalization

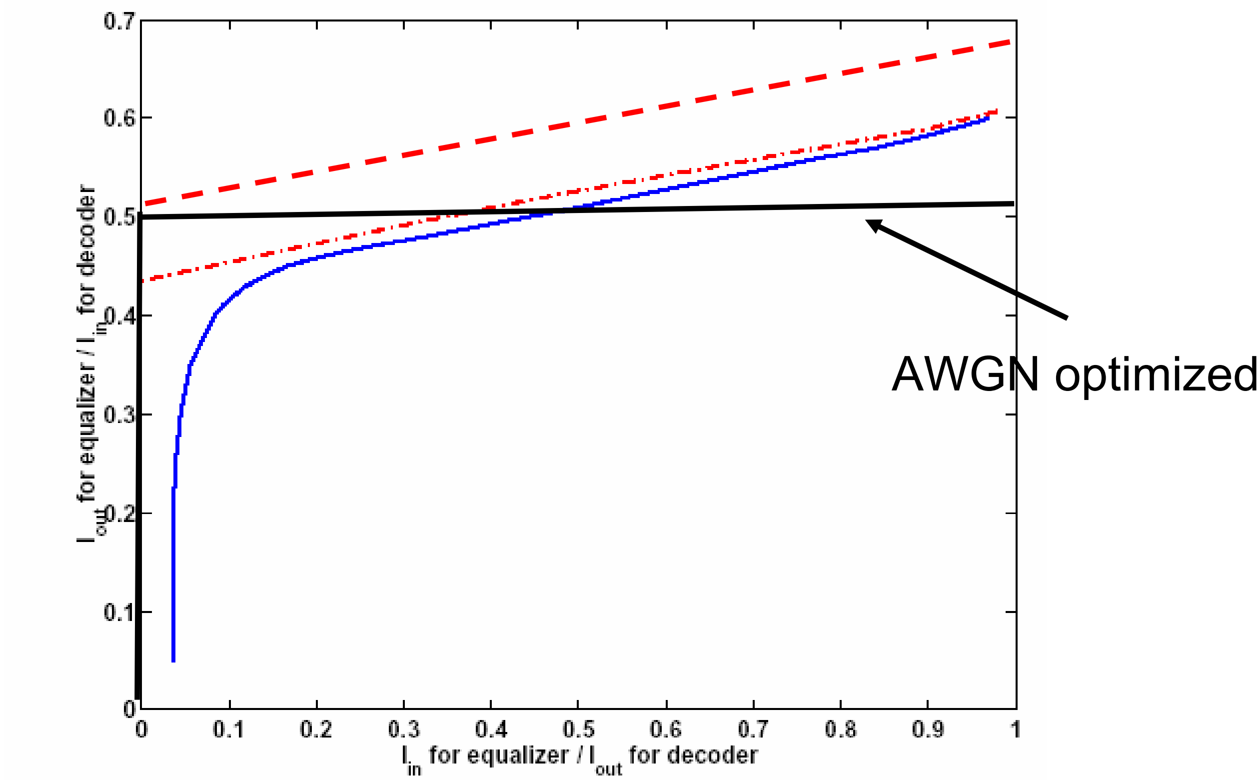


- Optimal Code: $F_{eq} = F_o^{-1}$ - Match the EXIT Curves
- Area under F_{eq} achievable information rate
- Extension to Gaussian a prioris [Our result, Allerton 2004]

Area Property

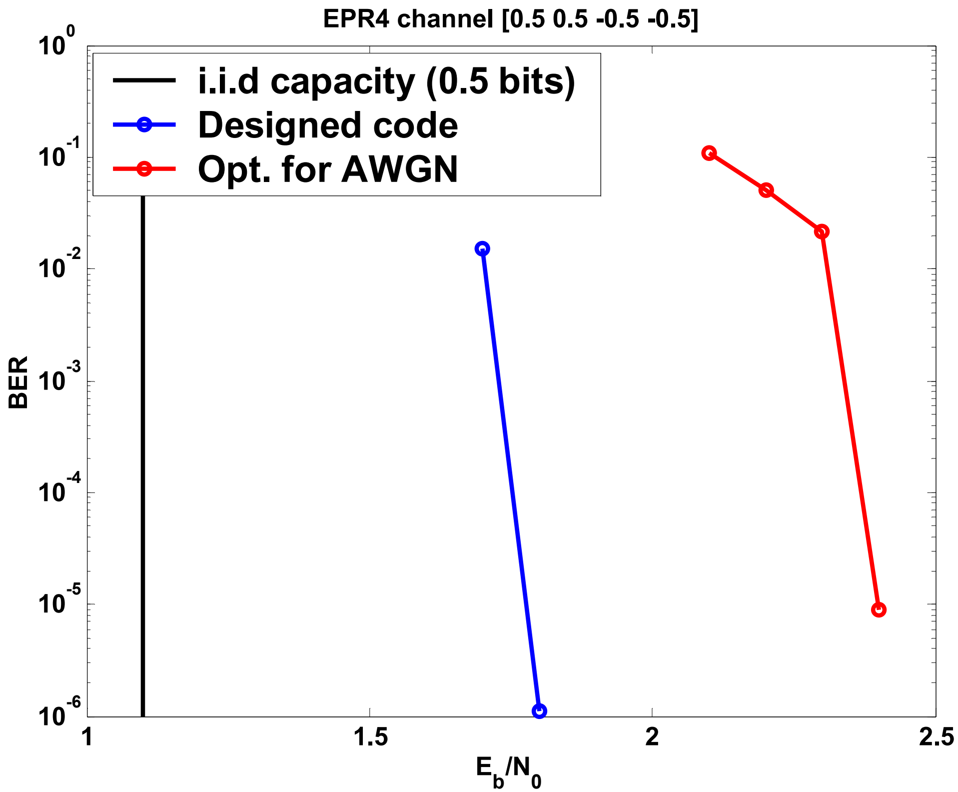


How bad are codes optimized for other channels?



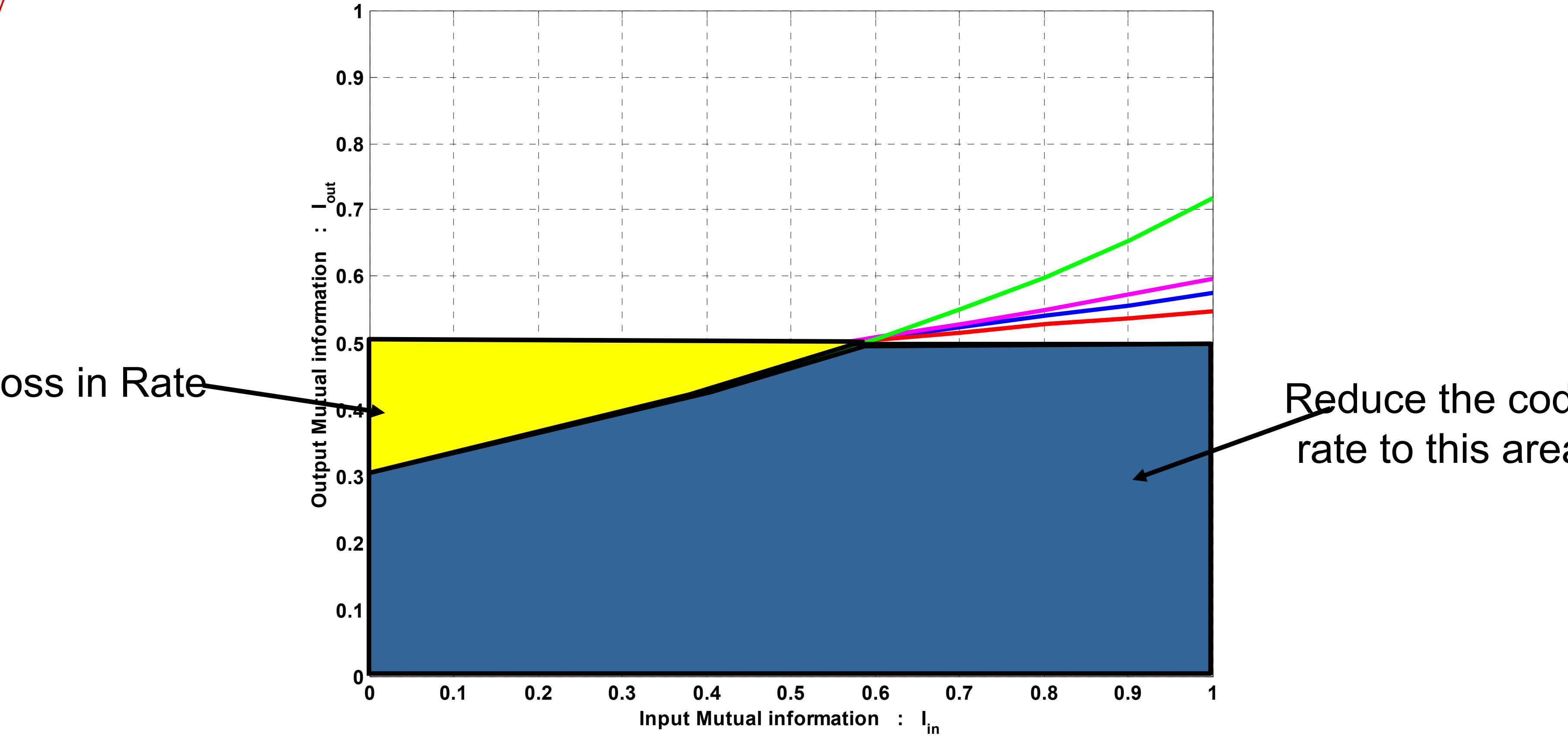
- AWGN optimized code is worse by 0.85 dB
- Or, there is a loss in rate

Simulation Results



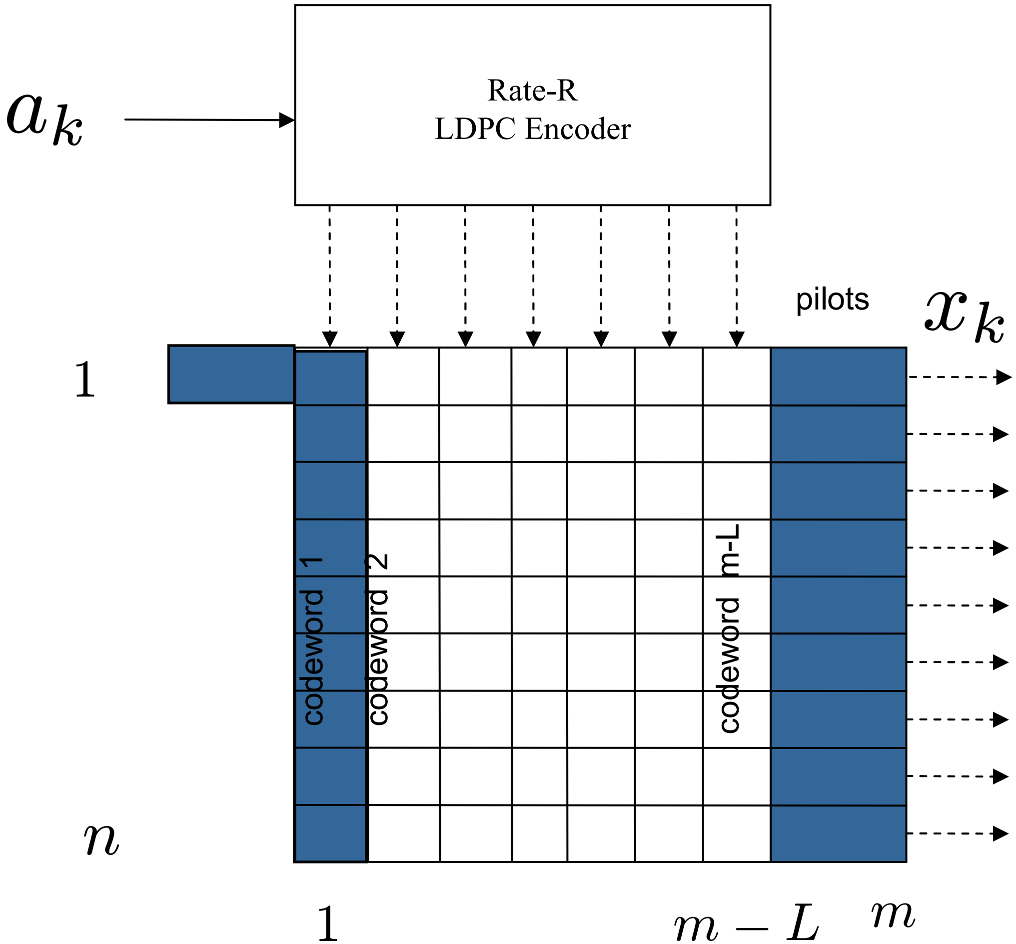
- Optimized code is better by 0.65 dB
- Fairly consistent with EXIT charts

EXIT Charts for Different Channels



- Area under the hull is the information rate achievable
- No guarantee that matched codes exist

Proposed Encoder/Decoder Structure

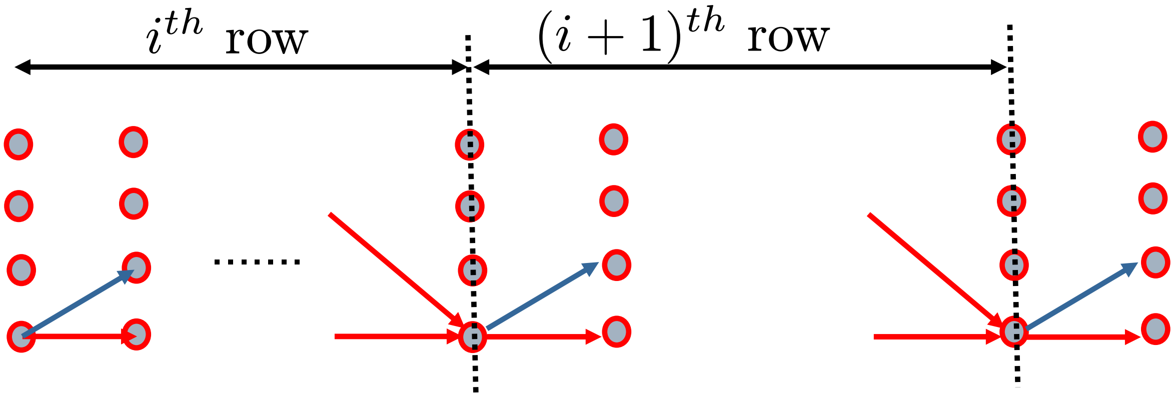


Prior Work

- Cioffi, Dudevoir, Eyuboglu and Forney
  - MMSE-DFE is canonical
  - Transmitter precoding
- Varanasi and Guess
  - Optimality of DFE-MUD
  - Extension to ISI does not produce universal structure
  - Deals only with Gaussian inputs
- Our results are for the finite alphabet case with Markov memory

The BCJR-DFE Receiver

- Produces $\lambda(x_k) = P(x_k = 1 | x_{k-1}, \ldots, x_{k-L}, \boldsymbol{Y})$ in an efficient way



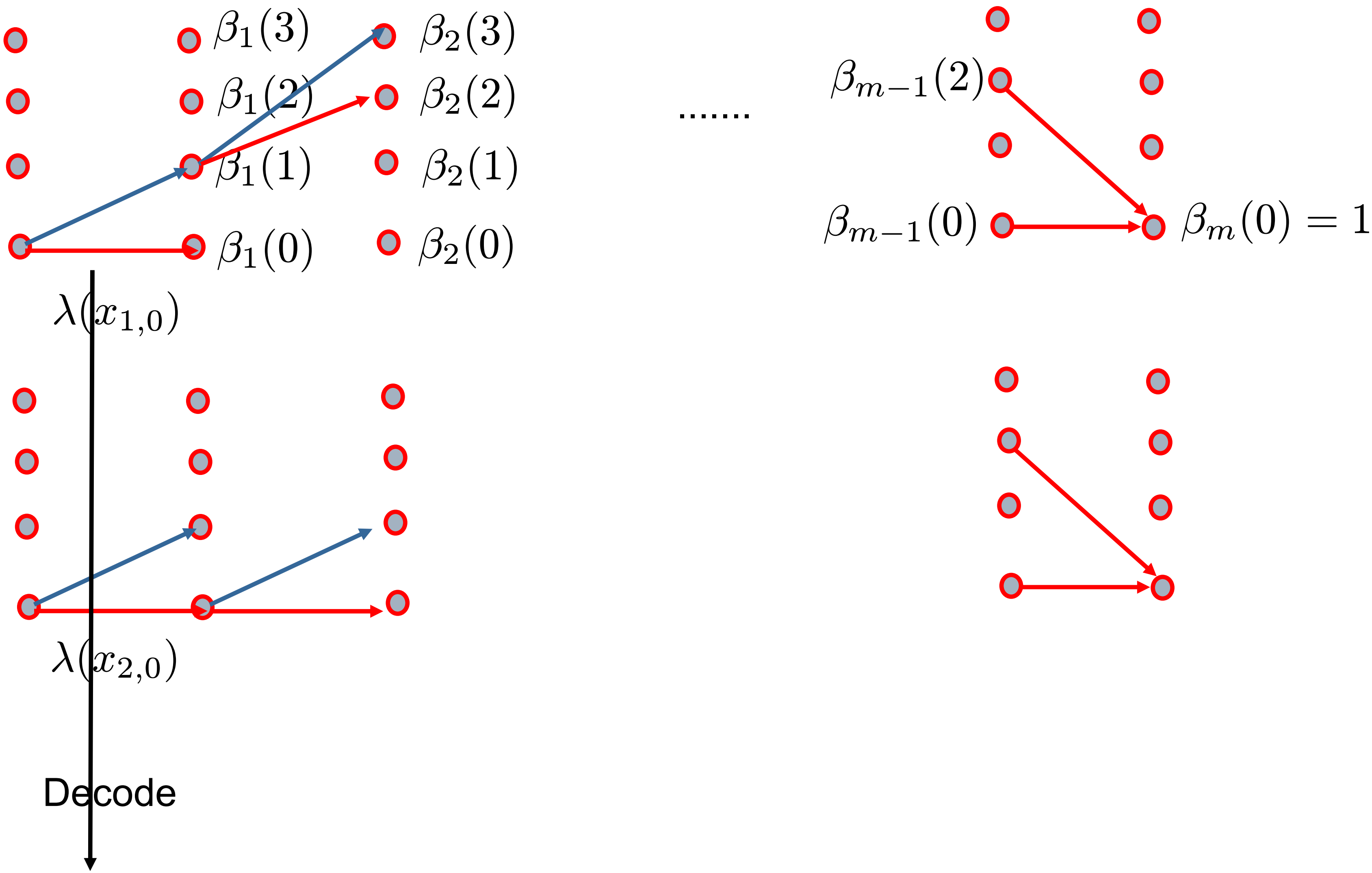
$$P(x_k = 1 | x_{k-1}, \ldots, x_{k-L}, \boldsymbol{Y}) = P(x_k | S_k, \boldsymbol{Y}) = P(x_k = 1 | S_k, y_{i,k}^{i,m}$$

$$= \gamma_1(s_k, s_{k+1}, y_{i,k}) \cdot \beta_{k+1}(s)$$

$\gamma_i(s_k, s_{k+1}, y_{i,k})$ - Branch transition probability

$$\beta_{k+1}(s) = P(y_{i,k+1}^{i,m} | S_{k+1} = s)$$

BCJR-DFE Algorithm



Optimality of the Encoder/Decoder Structure

- For m,n $\to \infty$, required code rate R = $C_{i.i.d}$

$$\mathrm{C}_{i.i.d} = I(\boldsymbol{X};\boldsymbol{Y}) = I(x_1, x_2, \ldots, x_{mn}; \boldsymbol{Y})$$

$$C_{i.i.d} = \sum_{k=1}^{mn} I(x_k; \boldsymbol{Y} | x_{k-1}, x_{k-2}, \ldots, x_1)$$

$$p(x_k|\boldsymbol{Y}, x_{k-1}, \ldots, x_1) = p(x_k|\boldsymbol{Y}, x_{k-1}, \ldots, x_{k-L})$$

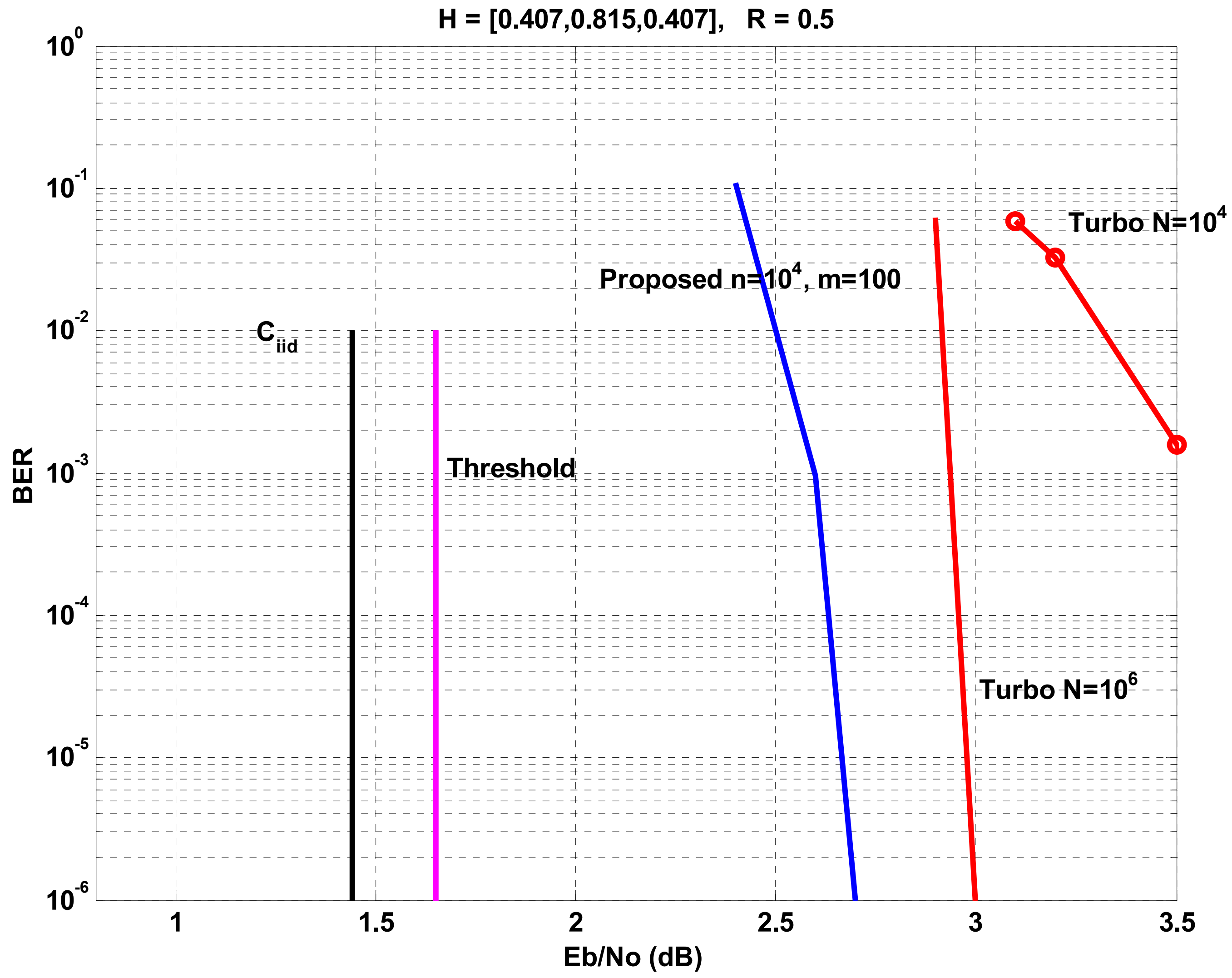
$$I(x_k; \boldsymbol{Y}|x_{k-1}, \ldots, x_1) = I(x_k; \boldsymbol{Y}|x_{k-1}, \ldots, x_{k-L}).$$

$$C_{i.i.d} = m \times n \times I(x_k; \boldsymbol{Y}|x_{k-1}, \ldots, x_{k-L})$$

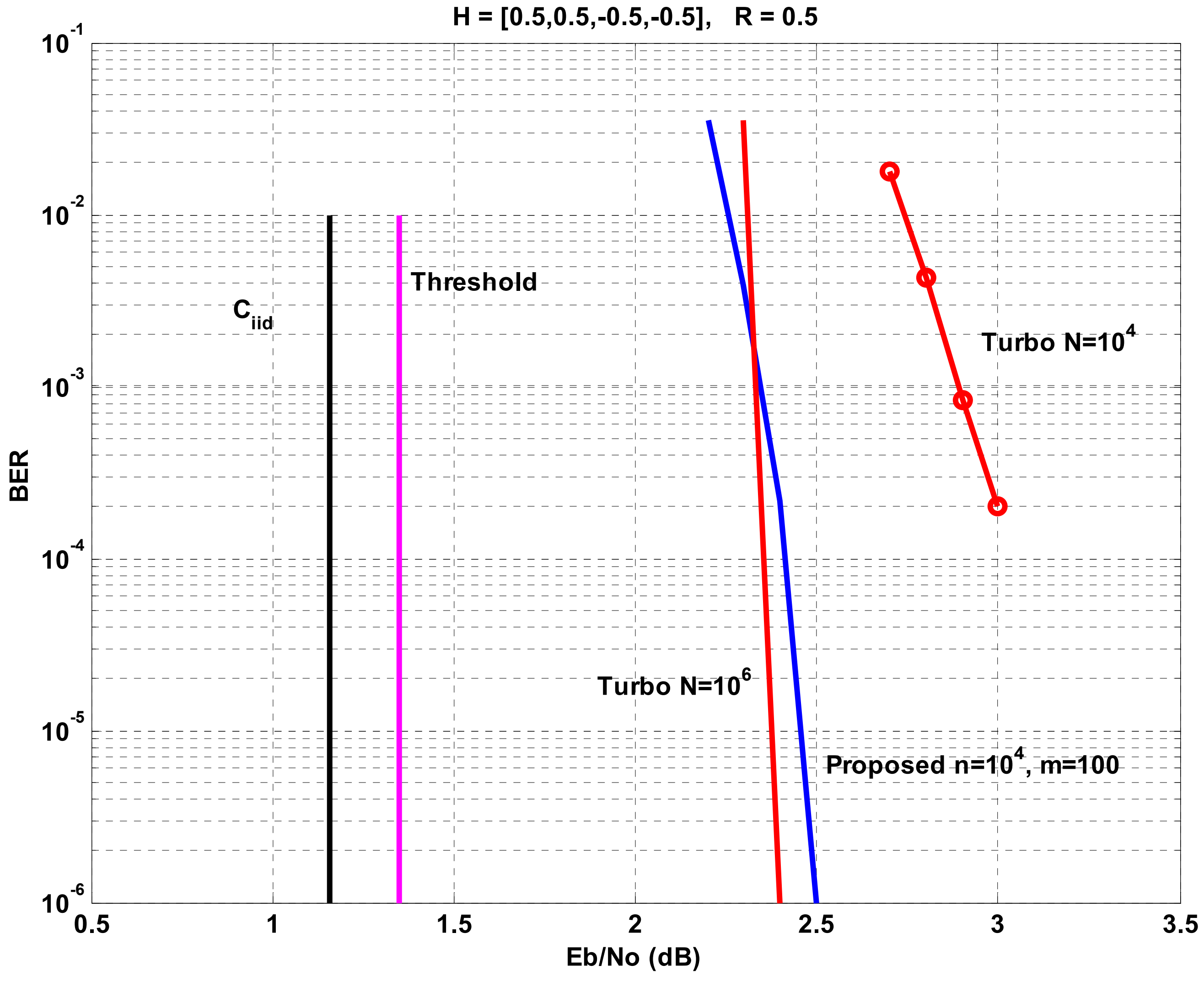
equivalent channel $x_k \to \lambda(x_k) = p(x_k = 1|\boldsymbol{Y}, x_{k-1}, \ldots, x_{k-L})$

Information rate for this channel is exactly $C_{i.i.d}$

Results for a 3-tap channel [0.407 0.815 0.407]



Another channel [0.5 0.5 -0.5 -0.5]



Universality

- The equivalent channel is memoryless

$$x_k \rightarrow \lambda(x_k)$$

$$r_{i,k} = x_{i,k} + w_{i,k}, \quad w_{i,k} \text{ and } w_{j,k} \text{ are independent}$$

- Pdf of $w_{i,k}$ depends on the channel realization
- Universal codes for the memoryless channel will work well
- The distribution is even well approximated as being Gaussian

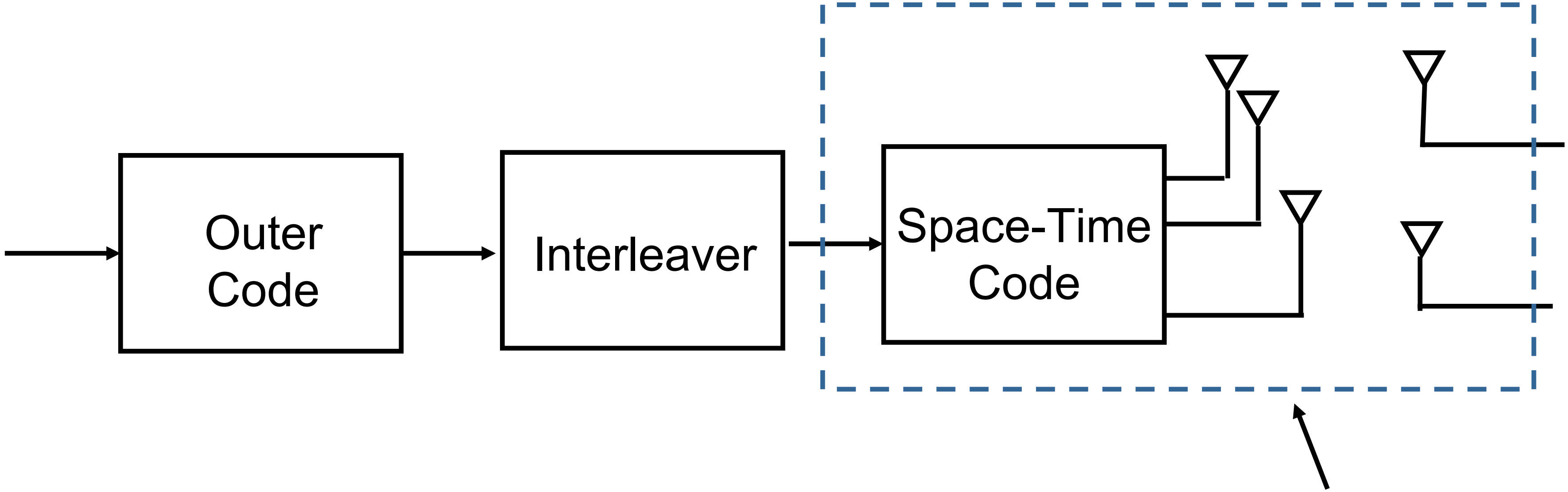
Results to show Universality

Gap from C_{iid} dB

Channel	BCJR-DFE	Turbo Equalization
AWGN	0.2	0.2
1-D	0.2	0.6
$1+D-D^2-D^3$	0.2	1.24
$0.407+0.815D+0.407D^2$	0.21	1.6
100 realizations of 3 tap Rayleigh fading taps	0.21	

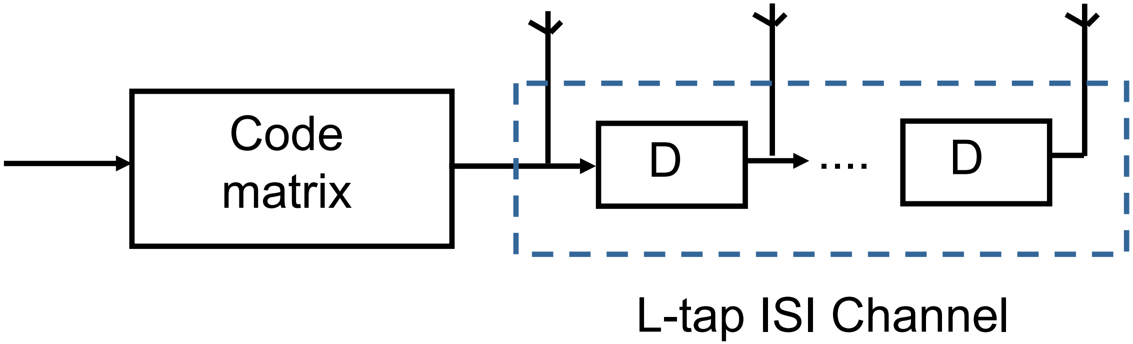
Extension to Other Channels

Concatenated Space-Time Codes



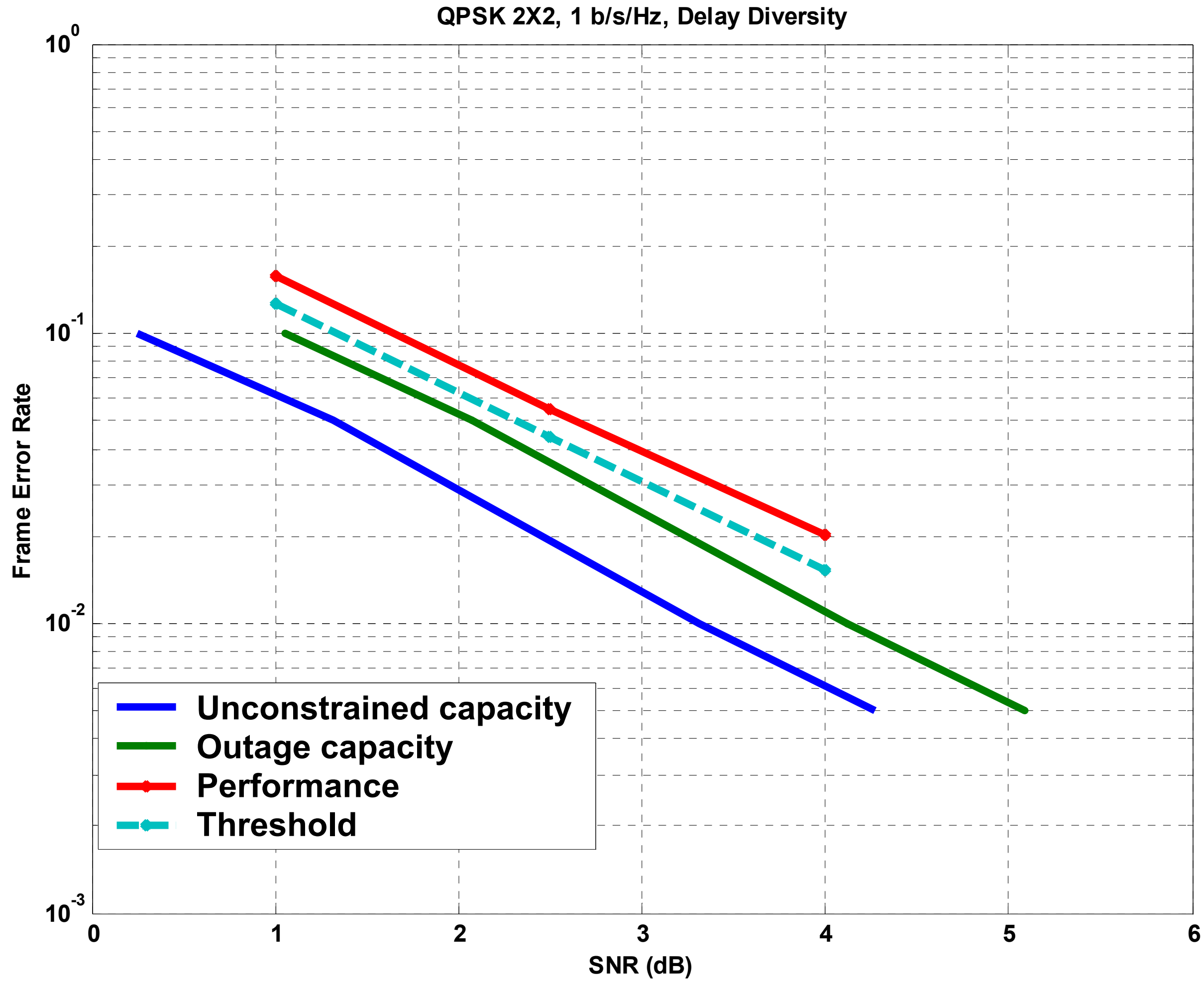
- Consider quasi-static fading
- Find a good space-time code and fix it
- How do we design good outer codes?
- How to perform joint decoding and space-time demodulation ?

MIMO Channels – Delay Diversity

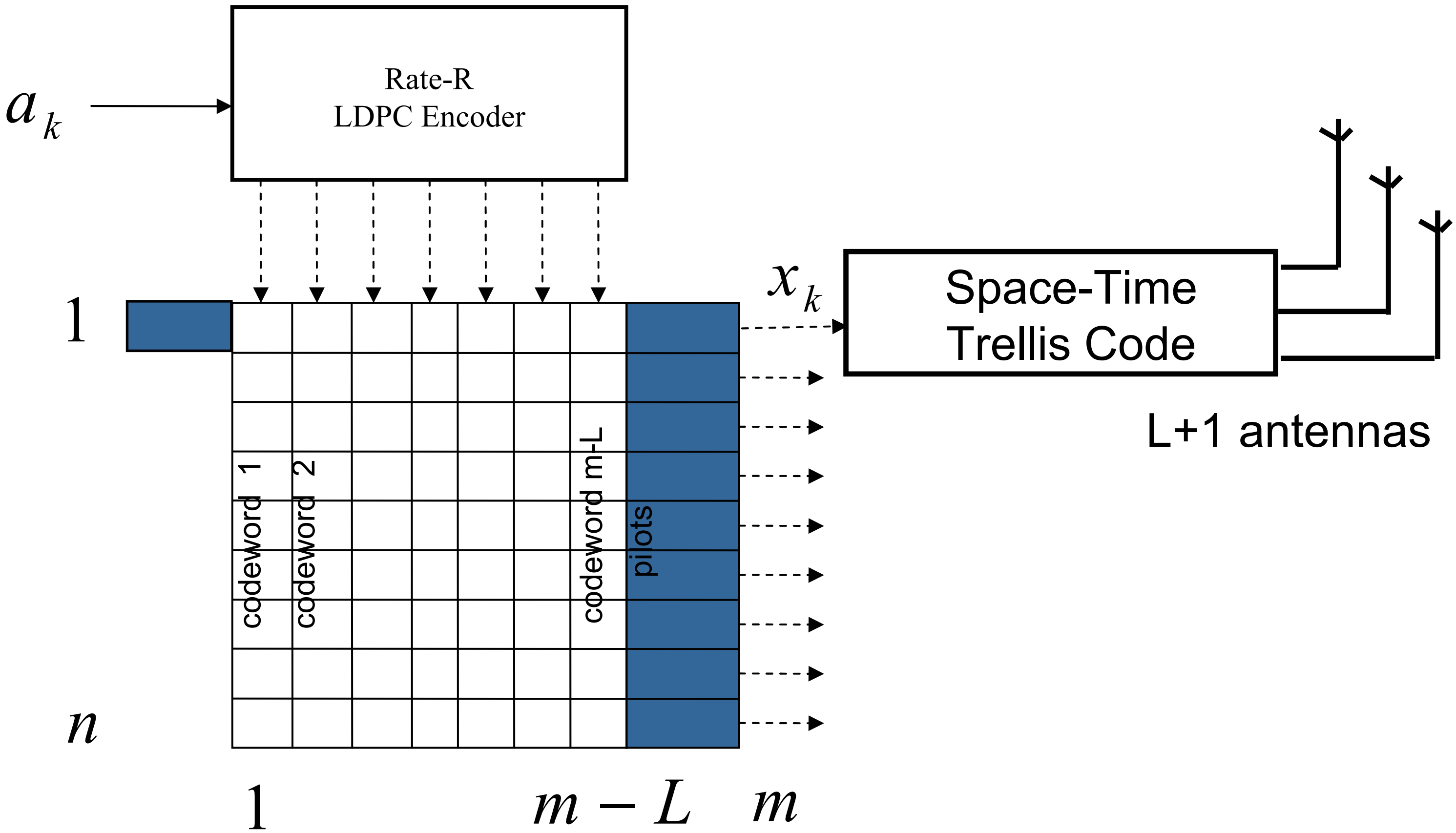


- Extension of MMSE-BLAST to finite constellations
  - MMSE part is replaced by BCJR part
  - BLAST detection is replaced by an optimum decoder
  - Silent periods are replaced by known bits to pin the state
- Decoded decisions prevent error propagation in D-BLAST
- DFE is universal whereas Turbo-BLAST is not

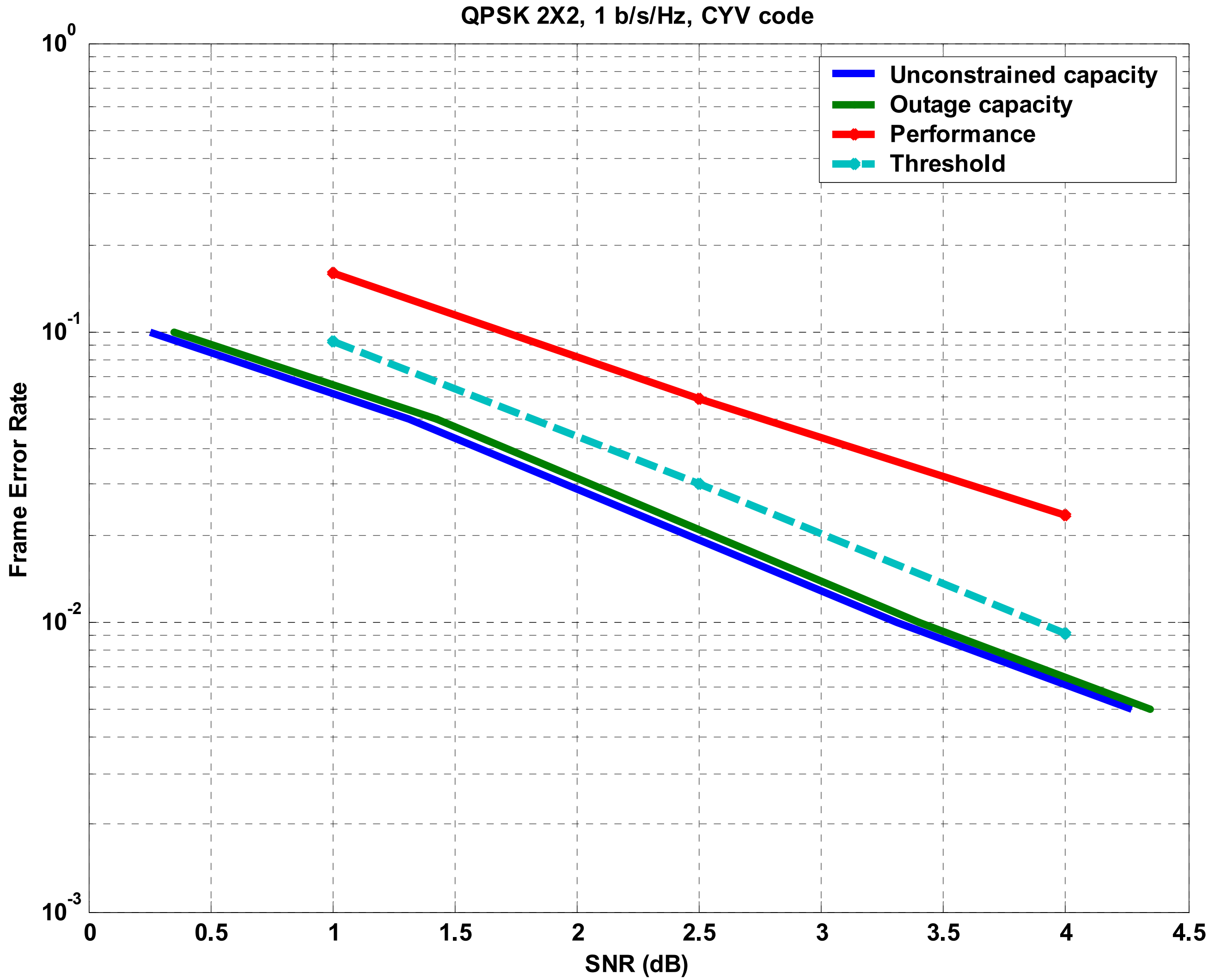
Delay Diversity Results



Any Space-Time Trellis Code Can be Used



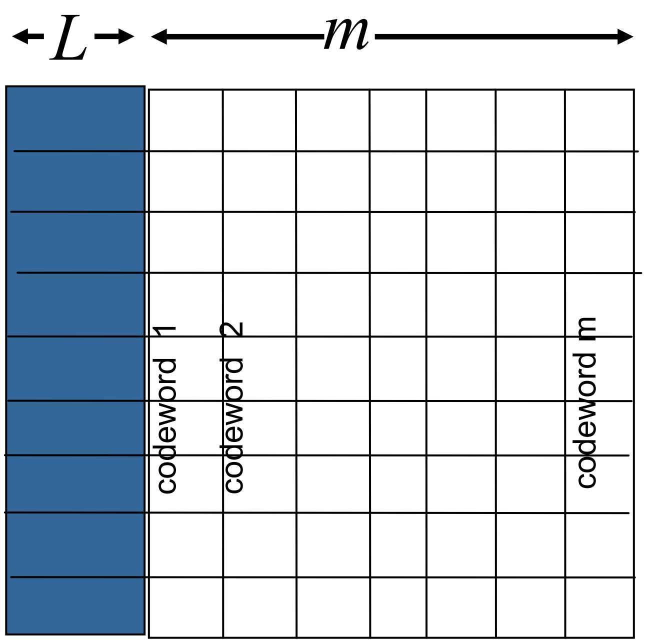
Another QPSK Space-Time Code



Non Coherent Detection

$$y_k = \alpha_k \; x_k + n_k \qquad x_k \in \text{PSK set}$$

Flat Correlated Rayleigh Fading



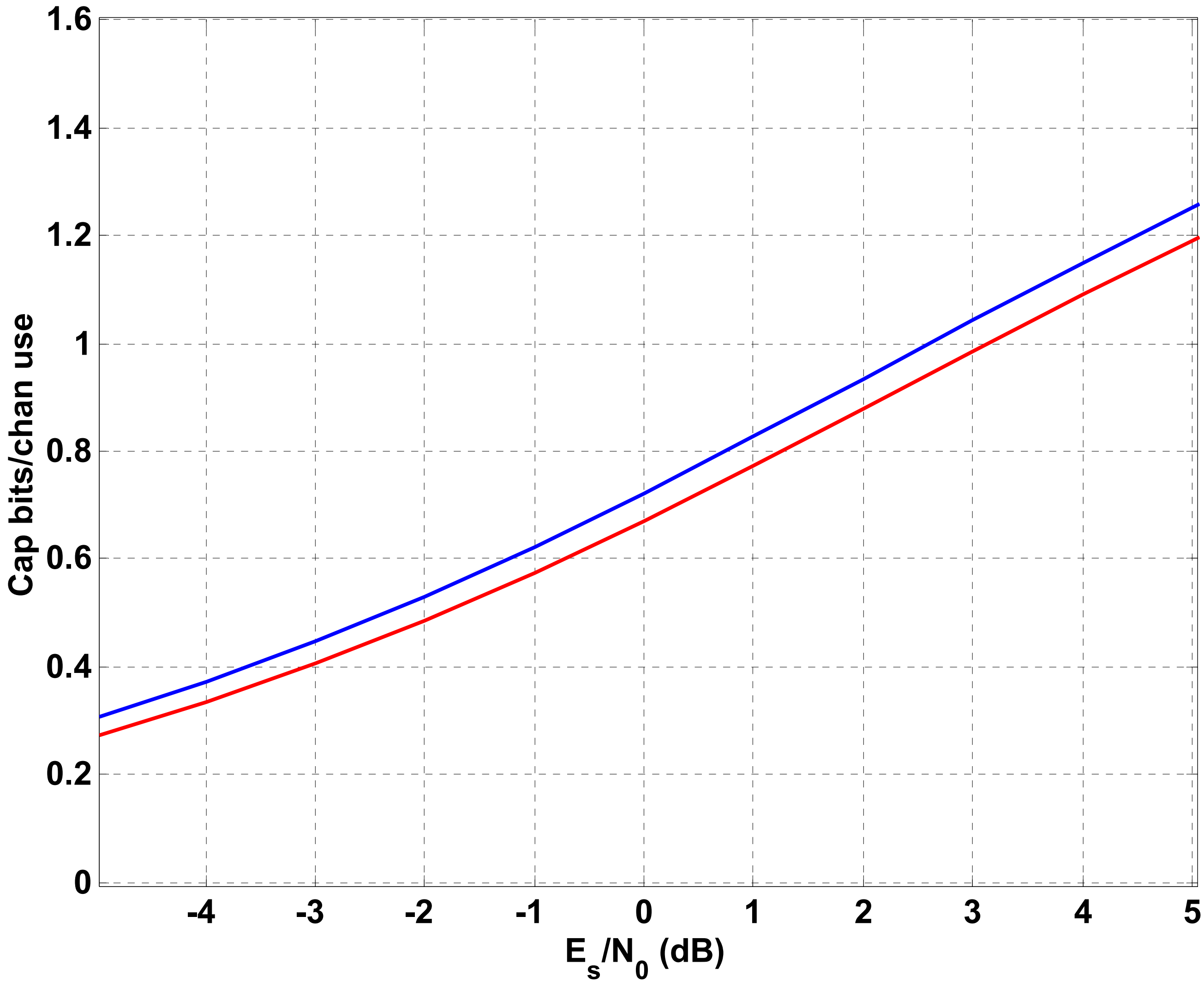
Predict α_k based on past values of $\boldsymbol{y}$ and $\boldsymbol{x}$

Vertically channel is memoryless

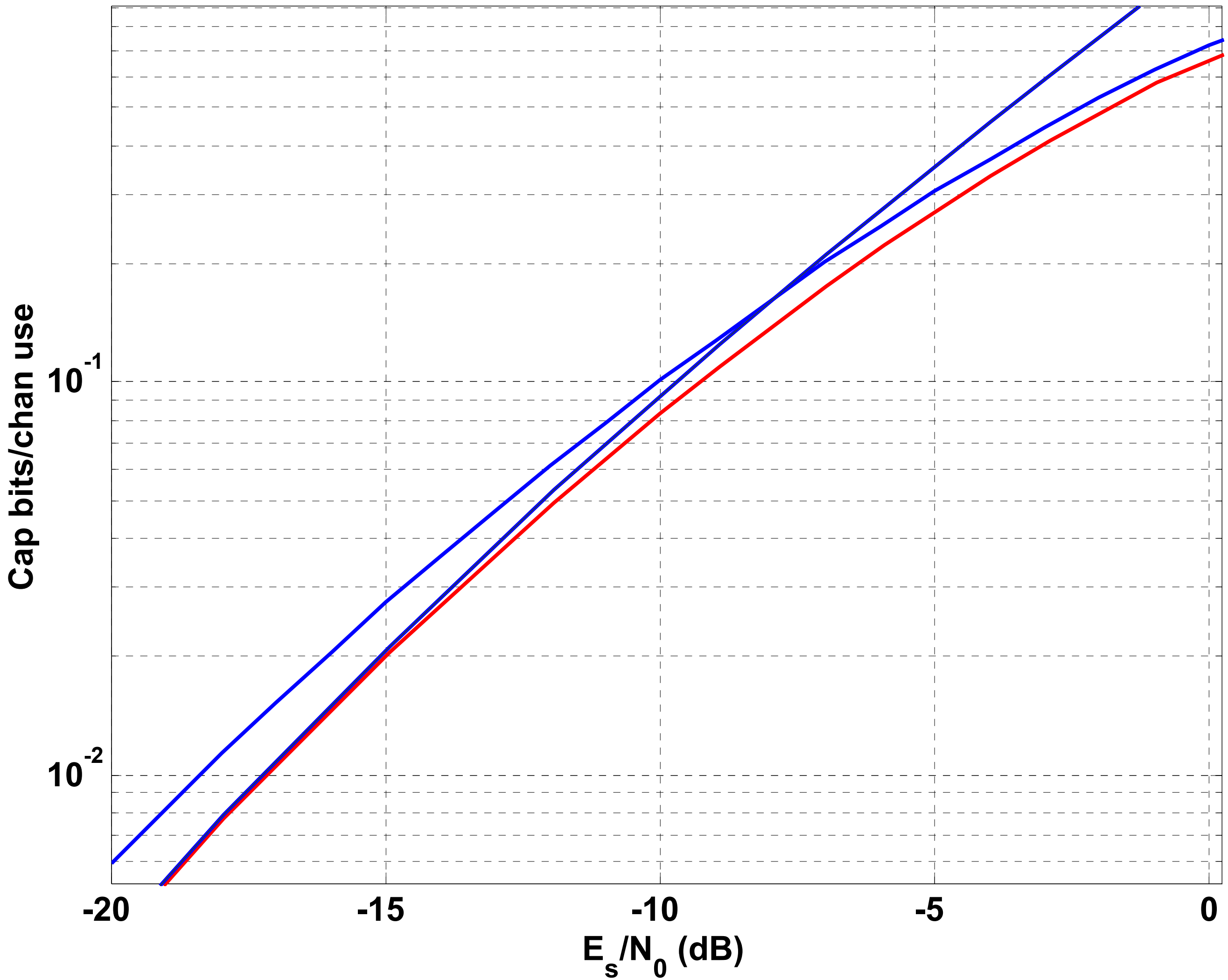
Lower bound on C that is easy to compute

$$C \leq I(x_k; y_k | \boldsymbol{Y}_{-\infty}^{k-1}, \boldsymbol{X}_{-\infty}^{k-1})$$

Results



Two different upper bounds and lower bound



Gaussian Signaling at high SNR

X_k is Gaussian and $\hat{\alpha}_k$ is the estimate of α_k based on the past

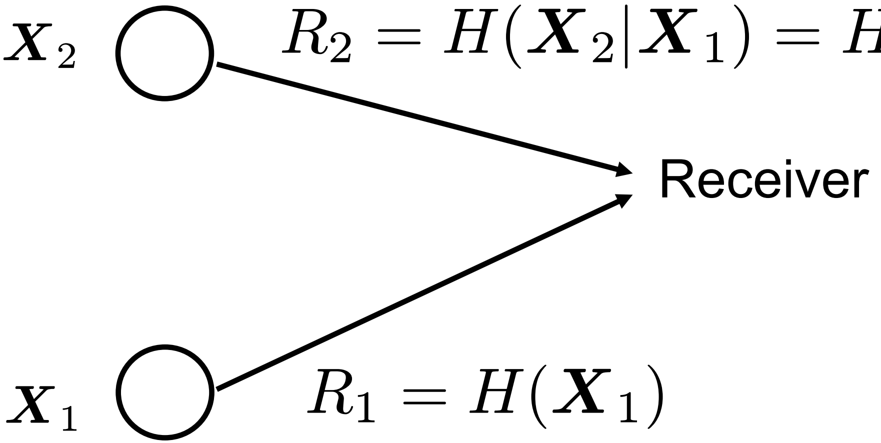
$$I(X_k; Y_k, \hat{\alpha}_k) \leq I(X; Y) + I(\alpha_k; \hat{\alpha}_k) \text{ [Lapidoth and Shamai]}$$

$$I(\alpha_k; \hat{\alpha}_k) = \log \frac{E[|\alpha_k|^2]}{E[|\alpha_k - E[\alpha|\hat{\alpha}_k]|^2]}$$

$I(\alpha_k; \hat{\alpha}_k)$ increases nearly logarithmically for practical SNR

Compression of Correlated Sources

We want to compress X_2 to R_2 bits

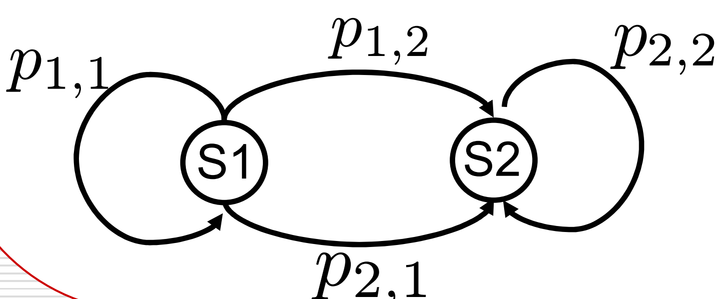


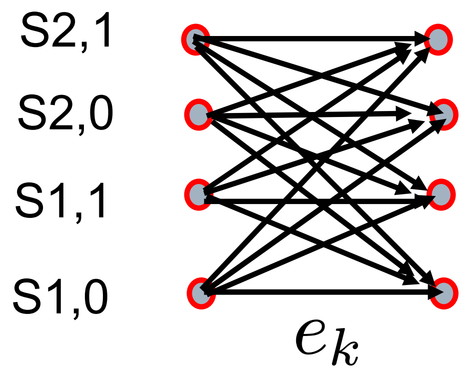
$$R_2 = H(\boldsymbol{X}_2|\boldsymbol{X}_1) = H(\boldsymbol{e})$$

$$R_1 = H(\boldsymbol{X}_1)$$

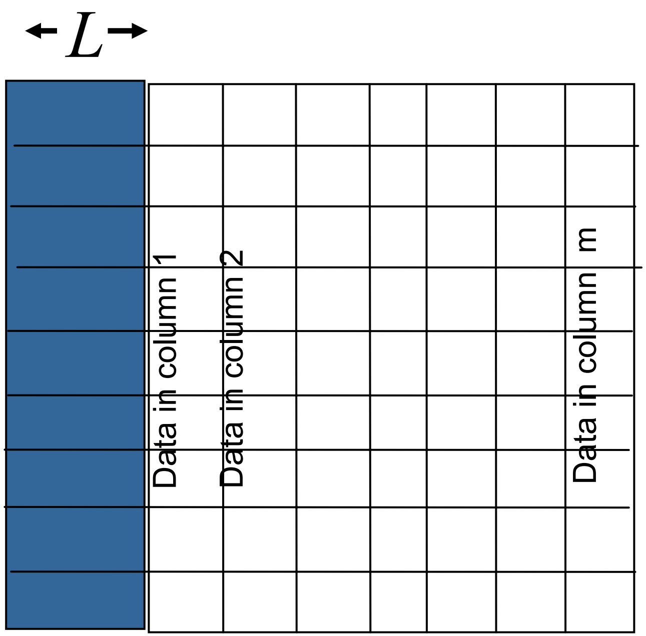
$$x_{2k} = x_{1k} \oplus e_k$$

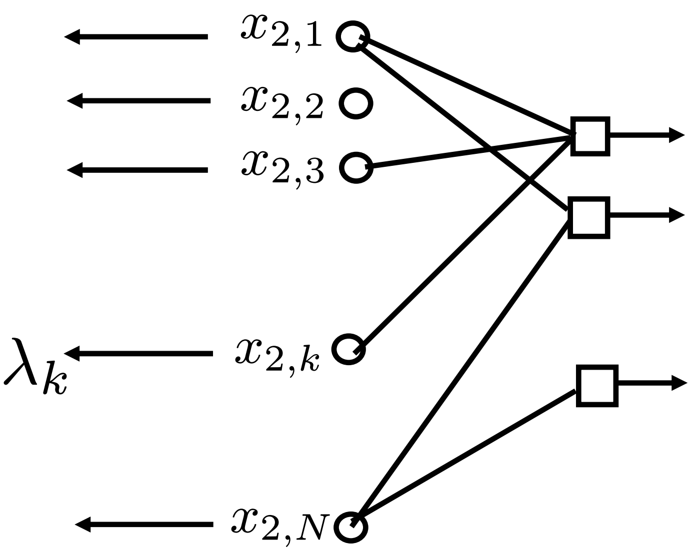
e_ks are the output of a HMM





Compression using LDPC codes





k^{th} row: $e_{k,1}, e_{k,2}, \ldots, e_{k,L} \rightarrow \lambda_{k,L+1} = P(e_{k,L+1} | e_{k,1}, \ldots, e_{k,L})$

$$= P(x_{2k,L+1} | e_{k,1}, \ldots, e_{k,L})$$

Compression achieved $R = H(e_{L+1} | e_1, e_2, \ldots, e_L)$

Possible to achieve $R = \frac{1}{N} \sum_{i=1}^{N} H(e_i | e_1, \ldots, e_{i-1}) \rightarrow H(\boldsymbol{e})$

Numerical Results

- Existing Approach is Iterative [Garcia Frias et al]
  - Iterate between the HMM and the LDPC decoder
  - Makes the code design problem quite difficult

- Example,

Model	$H(X_2 \| X_1)$	Δ_{DF}	Δ_{Turbo}
[0.01,0.065,0.95,0.925]	0.52	0.03	0.08
[0.97,0.967,0.93,0.973]	0.45	0.03	0.11
[0.99,0.989,0.945,0.9895]	0.28	0.03	0.13

- DF structure is more universal

Conclusion

- Capacity achieving coding schemes for channels with memory
- Iterative processing vs Decision Feedback Processing
  - Both can be used to achieve near capacity performance
  - DF is universal and simplifies the code design process
  - Rateless codes for channels with memory
  - Latency is a problem (it may not be insurmountable)
- It is not true that signal processing becomes harder as the codes get stronger
- Put the complexity in the decoder and use DF

Contact Information

- I am in Room 018 until Mid May
- krn@ee.tamu.edu
- http://ee.tamu.edu/~krn
- Also visit http://ee.tamu.edu and follow communications group