On the Achievable Rates of Finite State ISI Channels

Henry D. Pfister, Joseph B. Soriaga, and Paul H. Siegel
Signal Transmission and Recording (STAR) Group
University of California, San Diego

{hpfister,jsoriaga,psiegel}@ucsd.edu

GLOBECOM 2001
November 27th, 2001

University of California
San Diego

Overview

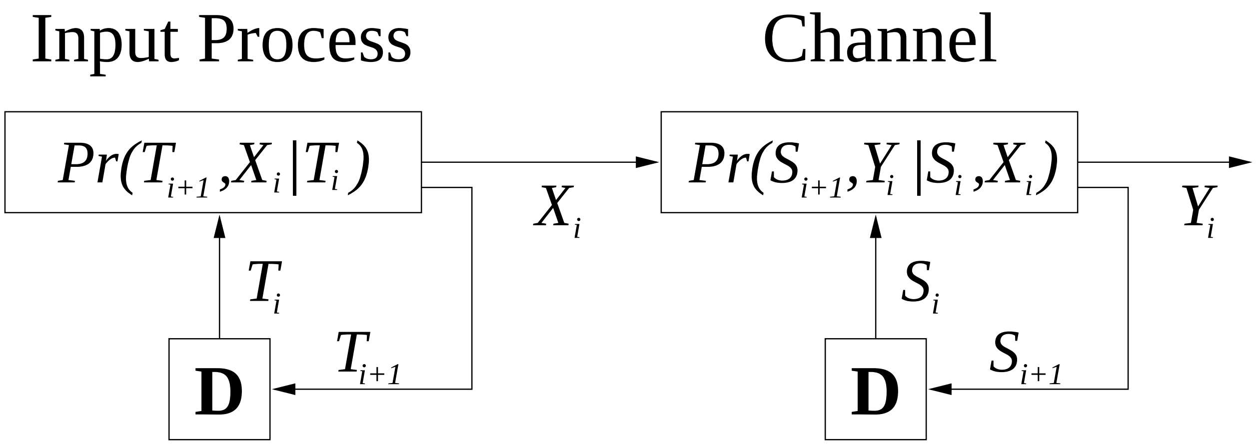
- Motivation
  - Turbo/LDPC codes nearly achieve capacity in AWGN
  - The Capacity of most ISI channels is unknown
    - Lower bounded by Hirt using Multi-D Integration (PhD Thesis 1988)
- Estimating achievable information rates
  - A Simple Monte Carlo Algorithm for Finite State Channels
    - Also reported by Arnold & Loeliger (ICC 2001) and Sharma & Singh (ISIT 2001)
  - A Constructive Lower Bound based on multistage encoding/decoding
    - Achievable rate approaches the Symmetric Information Rate (SIR)
- Maximizing over Markov input distributions gives tighter lower bounds on capacity

Achievable Information Rates

- Mutual Information Rate, $I(\mathcal{X};\mathcal{Y})$, between channel input $\mathcal{X}$ and output $\mathcal{Y}$
  - At rates less than $I(\mathcal{X};\mathcal{Y})$, there exist coding systems which have vanishing word error probability as $n \to \infty$
  - Channel Capacity is the supremum of $I(\mathcal{X};\mathcal{Y})$ over all input distributions
  - Symmetric Information Rate (SIR) is $I(\mathcal{X};\mathcal{Y})$ when all inputs are equiprobable

- Finite State (FS) Channels
  - FS Fading channels (Ex. Gilbert-Elliot channel)
    - State transitions are independent of the channel inputs
  - FS Intersymbol Interference (FSISI) channels
    - State transitions are a deterministic function of the inputs
    - Discrete-input linear ISI channels with AWGN (Ex. idealized PR)

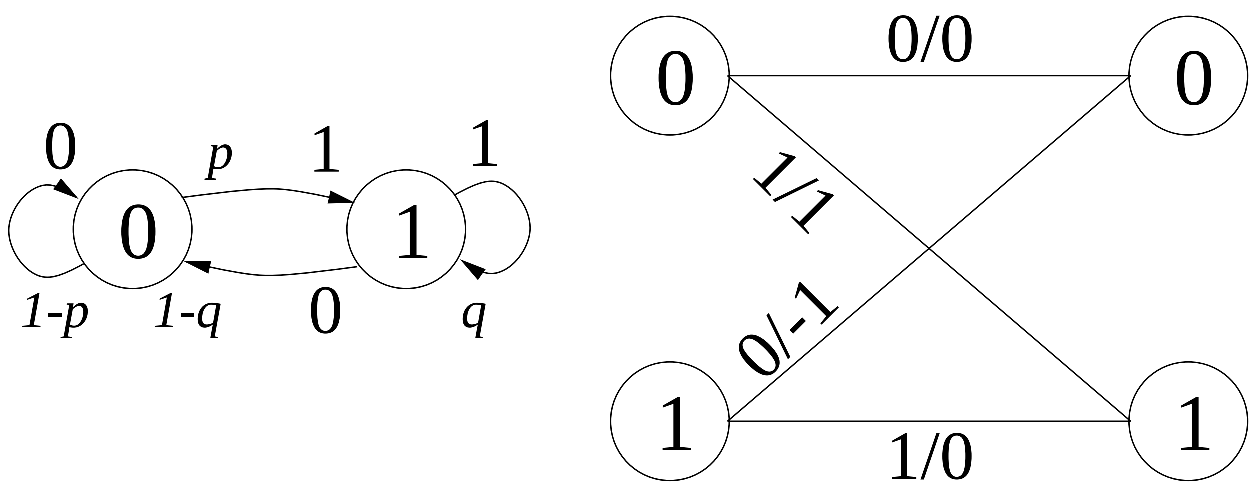
Finite State Channels with Markov Inputs



- Here is the mathematical system model: (note: $\mathbf{X}_1^n = (X_1, \ldots, X_n)$)
  - Input process generates $\mathbf{X}_1^n$ from state sequence $\mathbf{T}_1^{n+1}$ $\quad (T_i = X_{i-\kappa}^{i-1})$
  - FS Channel has input $\mathbf{X}_1^n$, state sequence $\mathbf{S}_1^{n+1}$, and output $\mathbf{Y}_1^n$
    - It is causal and Markov: $Pr(Y_i, S_{i+1}|\mathbf{Y}_1^{i-1}, \mathbf{S}_1^i, \mathbf{X}_1^n) = Pr(Y_i, S_{i+1}|S_i, X_i)$
  - We can join the two FS machines, $Q_i = (T_i, S_i)$, into a single FS machine

Example State Diagrams

- State diagram of input and FSISI channel
  - Memory 1 channel ⇒ channel state is equal to the last input
  - Markov memory 1 input ⇒ channel state is equal to the last output (p, q completely define the distribution of X)



(a) Input process (b) Standard Dicode Trellis

Mutual Information Rates

- The mutual information rate for this system can be written as

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X},\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X})$$

  - Markov inputs $\Rightarrow$ $H(\mathcal{X})$ easily computed in closed form
  - FSISI channel in AWGN $\Rightarrow$ $H(\mathcal{Y}|\mathcal{X}) = \frac{1}{2}\log(2\pi e\sigma^2)$
    - $H(\mathcal{Y}|\mathcal{X})$ is in closed form $\Rightarrow$ All we need for $I(\mathcal{X};\mathcal{Y})$ is $H(\mathcal{Y})$

- The Shannon-McMillan-Breiman Theorem says (for almost all $\mathbf{Y}_1^n$)

$$\lim_{n\to\infty} -\frac{1}{n}\log Pr(Y_1,\ldots,Y_n) = H(\mathcal{Y})$$

- So we define the sample entropy rate, for a realization $y_1,\ldots,y_n$ of $\mathbf{Y}_1^n$, as

$$\hat{H}_n(\mathcal{Y}) = -\frac{1}{n}\log Pr(y_1,\ldots,y_n) = -\frac{1}{n}\sum_{i=1}^{n}\log Pr(y_i|y_1,\ldots,y_{i-1})$$

A Simple Monte Carlo Method

- Recall that the forward APP recursion, with $\alpha_i(q) = Pr(Q_i = q|\mathbf{y}_1^{i-1})$, is

$$\alpha_{i+1}(q) = \frac{1}{A_{i+1}} \sum_{q' \in \mathbb{Q}} \alpha_i(q') Pr(Y_i = y_i, Q_{i+1} = q | Q_i = q')$$

  - A_{i+1} is the normalization coefficient which forces $\sum_{q \in \mathbb{Q}} \alpha_{i+1}(q) = 1$

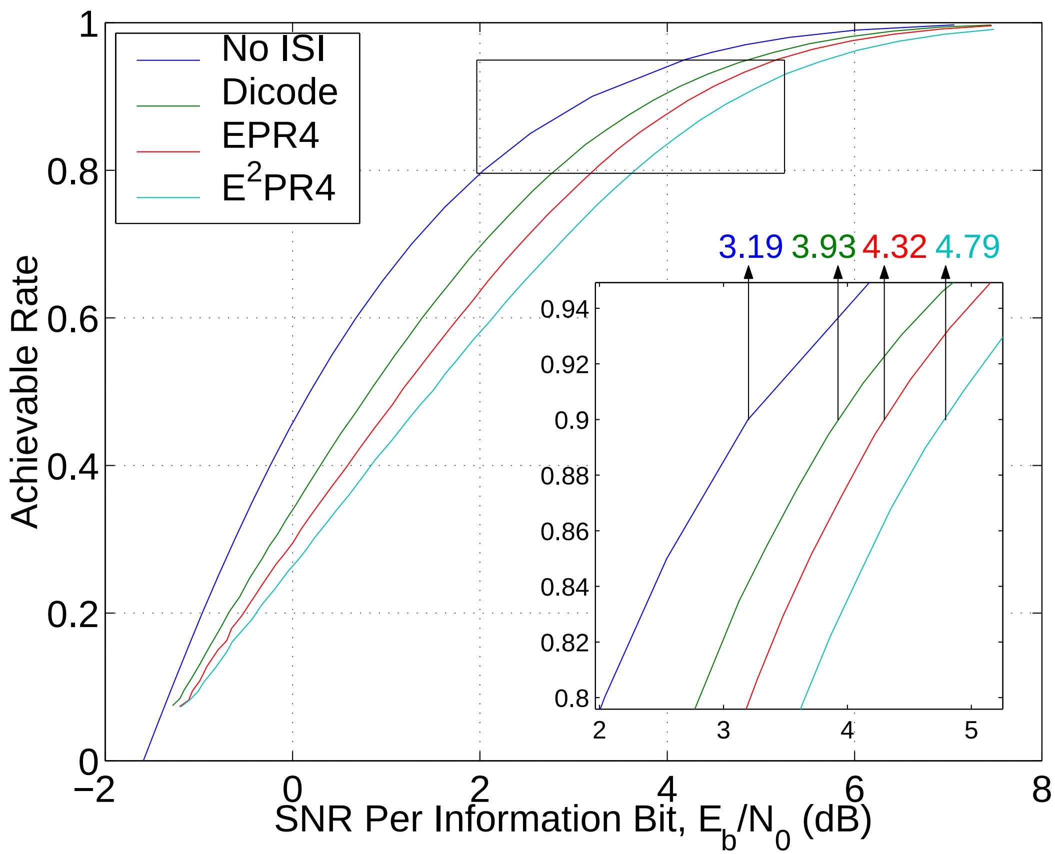
- But $A_{i+1} = Pr(y_i | y_1, \ldots, y_{i-1})$ is exactly what we need!

$$\hat{H}_n(\mathcal{Y}) = -\frac{1}{n} \sum_{i=1}^{n} \log Pr(y_i | y_1, \ldots, y_{i-1}) = -\frac{1}{n} \sum_{i=2}^{n+1} \log A_i$$

  - Initialization: $\hat{H}_0(\mathcal{Y}) = 0$, $\alpha_1(\cdot)$ given, and q_1 chosen randomly $\sim \alpha_1(\cdot)$
  - In many cases, a CLT holds $\Rightarrow \hat{H}_n(\mathcal{Y})$ is asymptotically Gaussian

Results

- The Symmetric Information Rate (SIR) for various PR channels



Optimizing the Input Distribution

- Let M_κ be the set of Markov input distributions with memory κ
  - We define the sequence $\{C_\kappa\}$ with

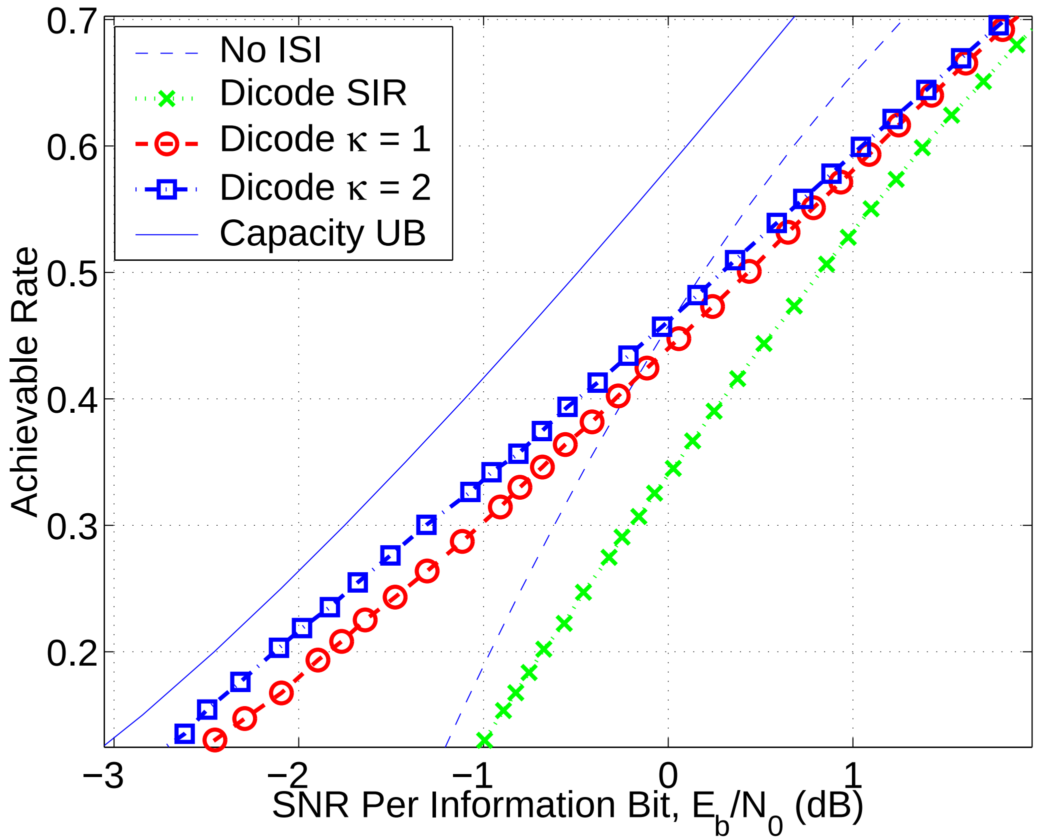
$$C_\kappa = \max_{Pr(\mathbf{X}) \in M_\kappa} \lim_{n\to\infty} \frac{1}{n} I(\mathbf{X}_1^n; \mathbf{Y}_1^n)$$

  - Gives a sequence of non-decreasing lower bounds on channel capacity

- Numerical optimization of $I(\mathcal{X};\mathcal{Y})$
  - Monte Carlo method used to estimate $I(\mathcal{X};\mathcal{Y})$ for any $Pr(\mathbf{X}) \in M_\kappa$
  - Gradient ascent finds the maximum (because $I(\mathcal{X};\mathcal{Y})$ concave in $Pr(\mathbf{X})$)
  - Elegant and efficient Arimoto-Blahut type algorithm by Kavcic (Globecom 2001)

Results

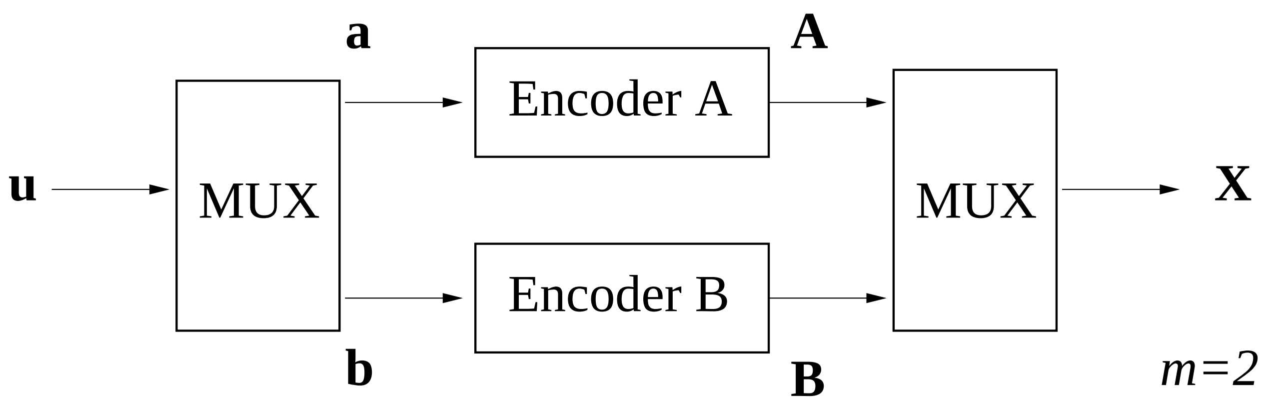
❍ Bounds on the capacity of the dicode channel



A Constructive Lower Bound

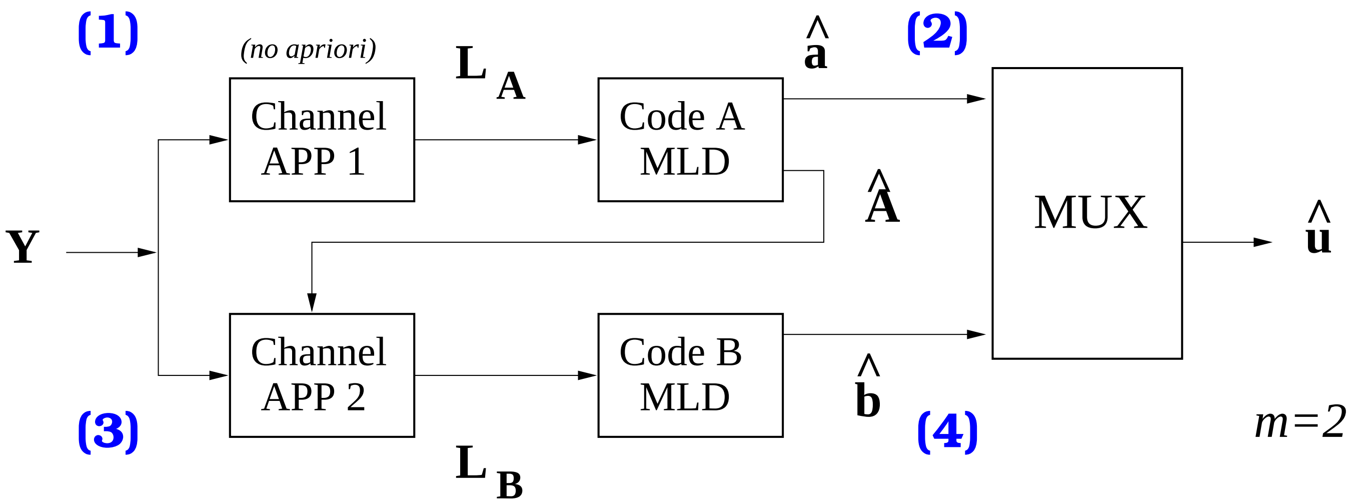
- Approaching the Symmetric Information Rate
  - Encoder: Interleaves multiple independent codes of varying rates
  - Decoder: Multistage decoding using decisions from previous stages
- Practical Advantages
  - Only uses a channel *a posteriori* probability detector (APP) and binary codes (no joint decoding components).
  - Uses channel APP detector a fixed number of times, and quantifies the gain from additional uses.

Multistage Encoding



- Block interleave m independent codes
  (in above example, $\mathbf{X} = (A_1, B_1, \ldots, A_n, B_n)$).
- Design codes such that $R_1 \leq R_2 \leq \cdots \leq R_m$
  ($R_A \leq R_B$).
- Overall code rate is $R_{av} = \frac{1}{m} \sum_{i=1}^{m} R_i$.

Multistage Decoding



1. $\mathbf{Y}$ is decoded with Channel APP 1
2. Channel input estimates $\mathbf{L}_A$ are fed into Code A MLD to recover $\hat{\mathbf{a}}$
3. $\mathbf{Y}$ is decoded again with additional *a priori* info $\hat{\mathbf{A}}$.
4. Code B MLD recovers remaining information bits $\hat{\mathbf{b}}$ from $\mathbf{L}_B$

Generalization to m interleaved codes is straightforward

Calculating Achievable Rates of Codes A and B

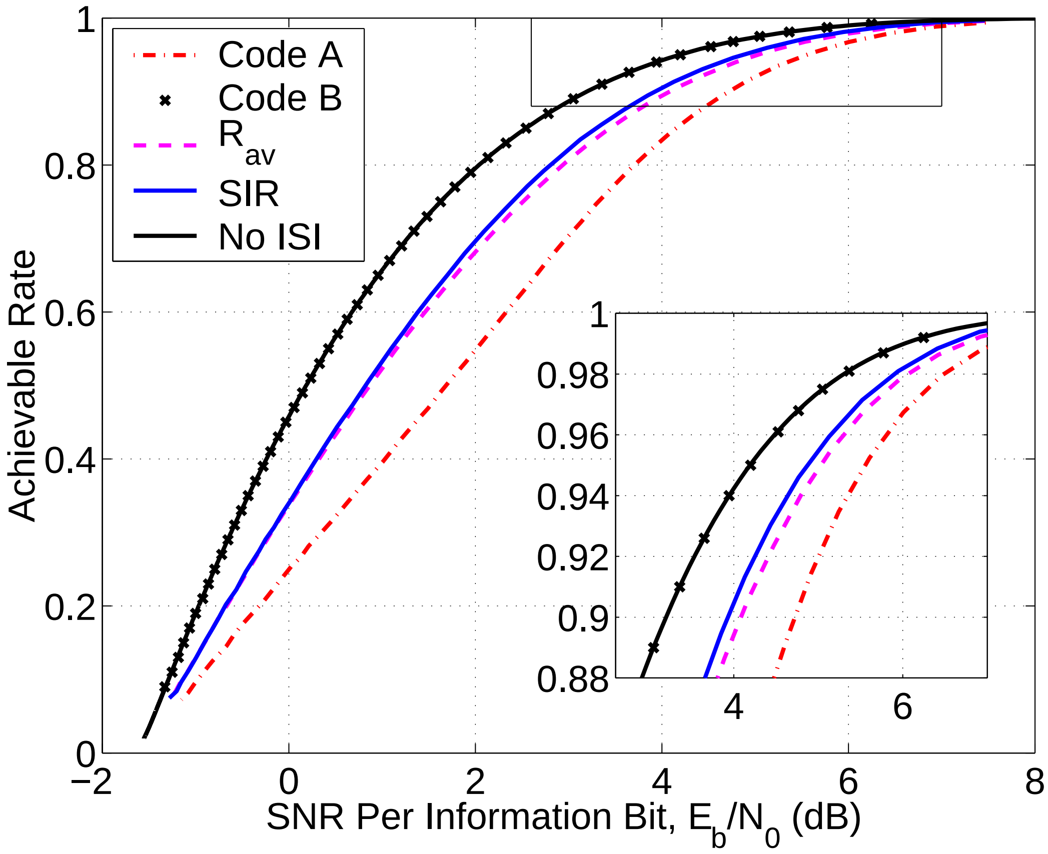
- *Find largest rate R_A^*, such that if $R_A < R_A^*$, we have reliable communication*
  - Consider the channel/detector subchannel: $A_k \longrightarrow L_{A,k}$
  - Lower bound $R_A^* \geq I(A_k; L_{A,k})$ by ignoring correlation in $L_{A,k}$
    - Assume that A_k i.i.d. $B(1/2)$, so computing $I(A_k; L_{A,k})$ requires only

$$f(l|\alpha) \stackrel{def}{=} f(L_{A,k} = l \mid A_k = \alpha)$$

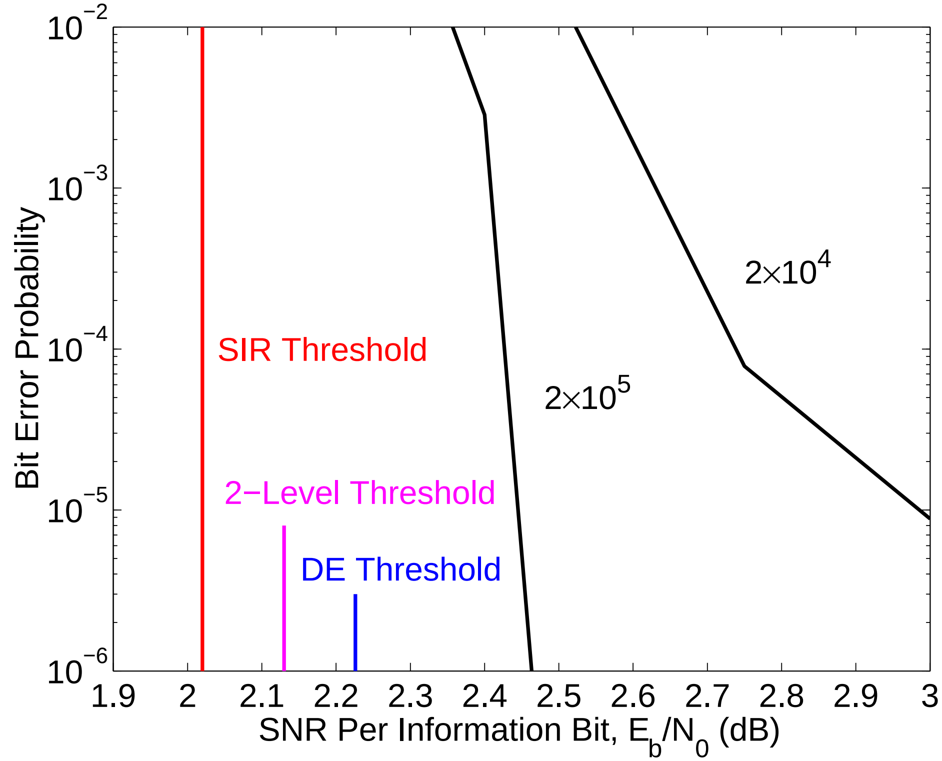
    - Difficult to solve $\Rightarrow$ approximate $f(l|\alpha)$ with a histogram
- Similarly, we have $R_B^* \geq I(B_k; L_{B,k})$ with code bits B_k drawn i.i.d. $B(1/2)$
  - Approximate $f(l|\beta)$ with histogram of odd time outputs:
    - Use channel APP with perfect *apriori* info at even times $\Rightarrow R_B^* \geq R_A^*$
- We believe $I(A_k; L_{A,k})$ is the max rate for single channel APP and a linear code

Simulation Results on Dicode Channel for $m = 2$

- Monte Carlo evaluation of rate distribution and achievable R_{av}.



Optimized LDPC Codes for the Dicode Channel with $m = 2$



Codes optimized for R=0.7 (0.63,0.77,$d_l = 50$)

Summary

- The Capacity of most ISI channels is (was?) unknown
  - The development of very powerful codes makes it interesting
- Introduced a simple method to estimate $I(\mathcal{X};\mathcal{Y})$ for FSISI channels
  - Requires only the ability to run an APP/BCJR decoder on long sequences
  - Efficient enough to allow optimization over Markov input distributions
- Introduced a constructive lower bound on $I(\mathcal{X};\mathcal{Y})$
  - Based on multistage coding and histograms of APP/BCJR output statistics
  - Actual coding on the Dicode channel is $0.2 - 0.5\, dB$ from SIR
  - EPR4 Channel: $m = 3$ threshold $\sim 0.2\, dB$ from the SIR