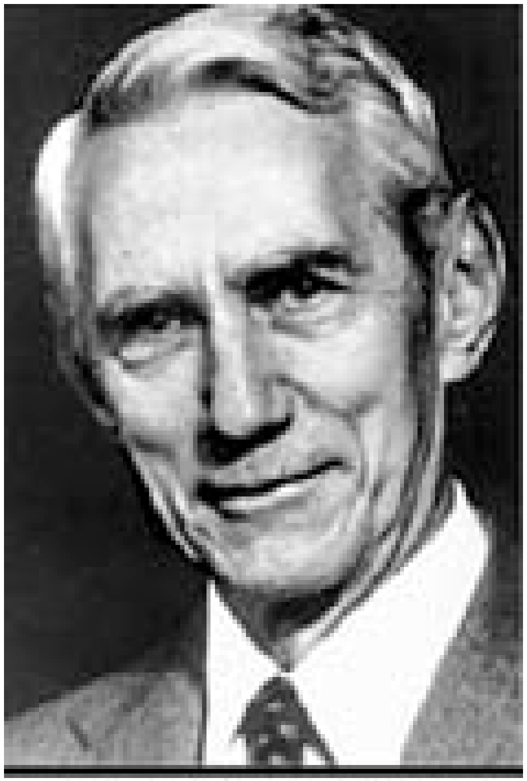
Capacity of Noiseless and Noisy Two-Dimensional Channels

Paul H. Siegel
Electrical and Computer Engineering
Center for Magnetic Recording Research
University of California, San Diego

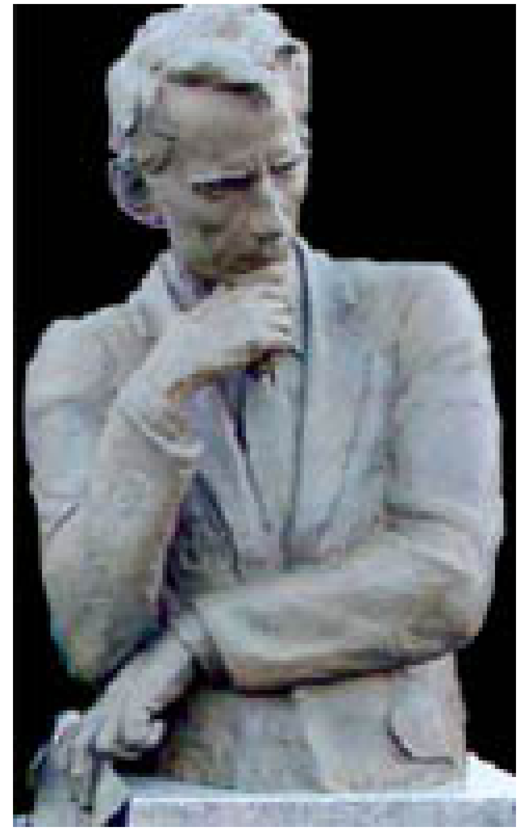
Outline

- Shannon Capacity
- Discrete-Noiseless Channels
  - One-dimensional
  - Two-dimensional
- Finite-State Noisy Channel
  - One-dimensional
  - Two-dimensional
- Summary

Claude E. Shannon



Claude Elwood Shannon
1916 - 2001



The Inscription

CLAUDE ELWOOD SHANNON

1916 – 2001

FATHER OF INFORMATION THEORY

HIS FORMULATION OF THE MATHEMATICAL THEORY OF COMMUNICATION PROVIDED THE FOUNDATION FOR THE DEVELOPMENT OF DATA STORAGE AND TRANSMISSION SYSTEMS THAT LAUNCHED THE INFORMATION AGE.

DEDICATED OCTOBER 16, 2001

EUGENE DAUB, SCULPTOR

The Formula on the "Paper"

Capacity of a discrete channel with noise [Shannon, 1948]

$$C = \mathrm{Max}\,(H(x) - H_y(x))$$

For noiseless channel, $H_y(x)=0$, so:

$$C = \mathrm{Max}\,H(x)$$

Gaylord, MI: C = W log (P+N)/N

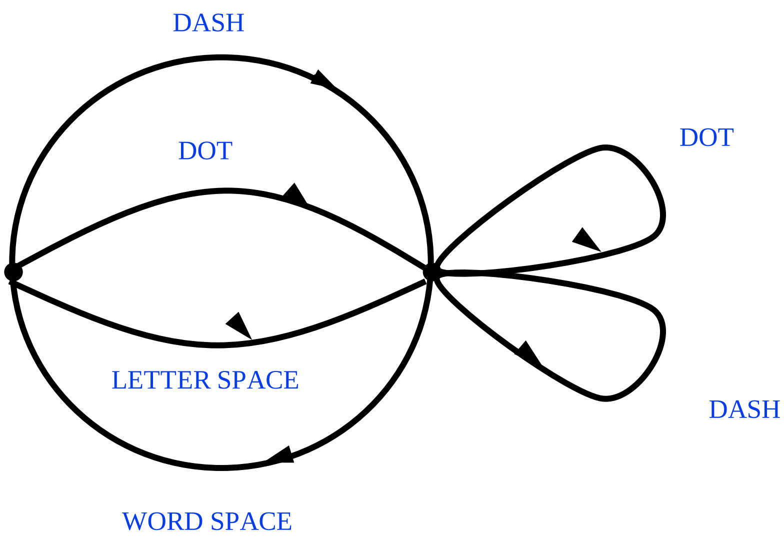
Bell Labs: no formula on paper

("H = – p log p – q log q" on plaque)

Discrete Noiseless Channels (Constrained Systems)

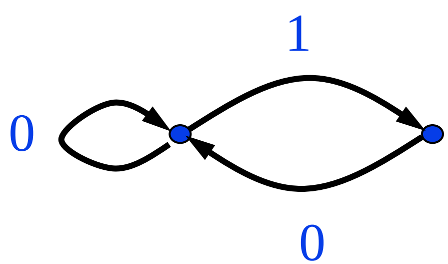
- A constrained system S is the set of sequences generated by walks on a labeled, directed graph G.

Telegraph channel constraints [Shannon, 1948]

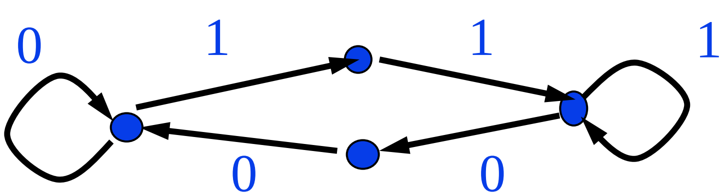


Magnetic Recording Constraints

Runlength constraints
("finite-type": determined by finite list F of forbidden words)



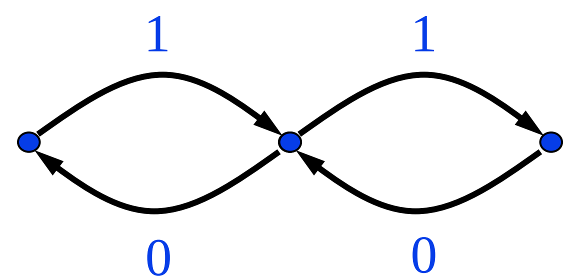
Forbidden word F={11}



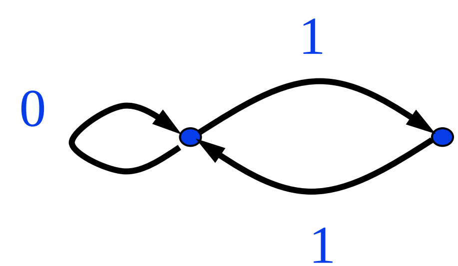
Forbidden words F={101, 010}

Spectral null constraints
("almost-finite-type")

Biphase



Even



(d,k) runlength-limited constraints

- For $0 \le d \le k$, a (d,k) runlength-limited sequence is a binary string such that:

$$d \le \#0\text{'s between consec}$$

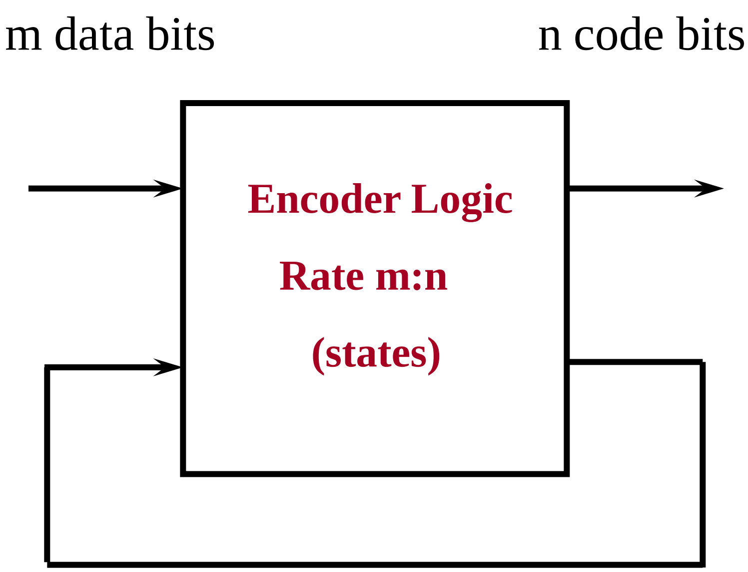
- F={11} forbidden list corresponds to $(d,k) = (1,\infty)$

1 0 0 0 1 0 1 0 0 1 0 1 0 0 0 1 0 1 0 0 0 0 1 0

Practical Constrained Codes

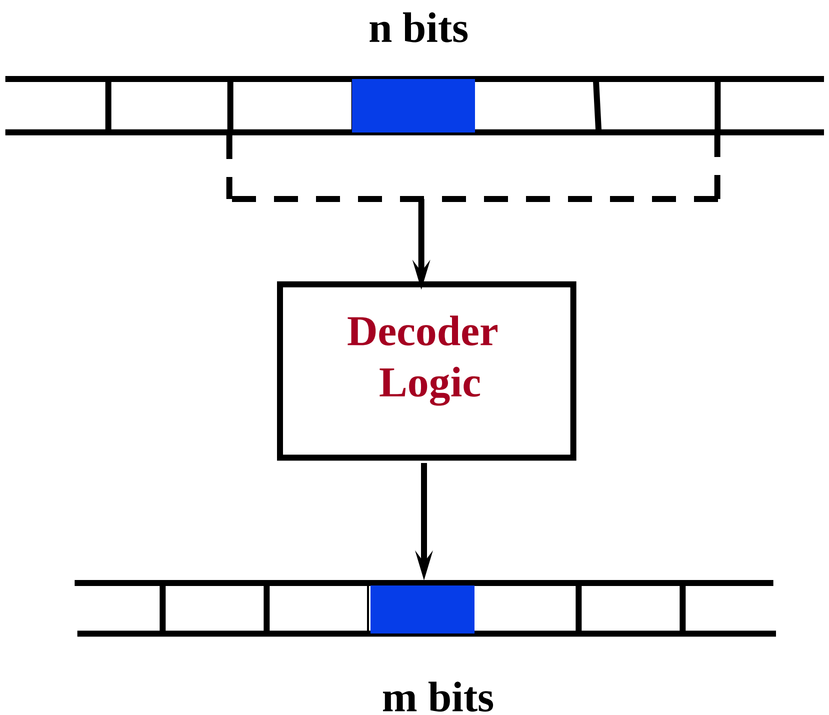
Finite-state encoder

(from binary data into S)



Sliding-block decoder

(inverse mapping from S to data)



We want: high rate R=m/n
low complexity

Codes and Capacity

- How high can the code rate be?
- Shannon defined the capacity of the constrained system ***S*:**

$$C = \lim_{n \to \infty} \frac{1}{n} \log N(S,n)$$

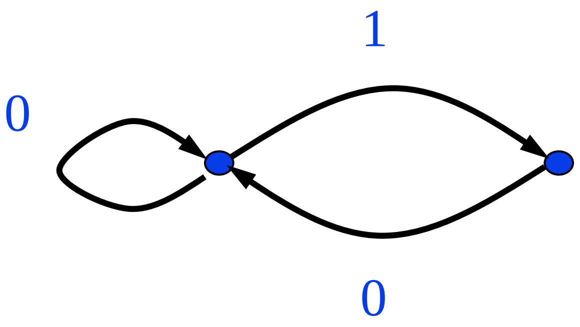
where ***N*(*S*,*n*)** is the number of sequences in ***S*** of length ***n***.

Theorem [Shannon,1948] : If there exists a decodable code at rate ***R*= *m*/*n*** from binary data to ***S***, then $R \leq C$.

Theorem [Shannon,1948] : For any rate ***R*=*m*/*n* < *C*** there exists a block code from binary data to ***S*** with rate ***km*:*kn***, for some integer $k \geq 1$.

Computing Capacity: Adjacency Matrices

- Let A_G be the adjacency matrix of the graph ***G*** representing ***S***.



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

- The entries in $A_G{}^n$ correspond to paths in *G* of length ***n***.

Computing Capacity (cont.)

- Shannon showed that, for suitable representing graphs ***G*** ,

$$C = log \ \rho\left(A_G\right)$$

where $\rho\left(A_G\right) = max\left\{ \left|\lambda\right| : \lambda \ is \ an \ eigenvalue \ of \ A_G \right\}$, i.e., the spectral radius of the matrix A_G.

- Assigning "transition probabilities" to the edges of ***G***, the constrained system ***S*** becomes a Markov source ***x***, with entropy ***H*(*x*)**. Shannon proved that

$$C = max \ H\left(x\right)$$

and expressed the maximizing probabilities in terms of the spectral radius and corresponding eigenvector of A_G .

Maxentropic Measure

- Let λ denote the largest real eigenvalue of A_G, with corresponding eigenvector $\underline{B} = [B_1, \ldots, B_M]$
- Then the maxentropic (capacity-achieving) transition probabilities are given by

$$P_{ij} = \frac{B_j}{B_i} \cdot \frac{A_{ij}}{\lambda}$$

- The stationary state distribution is expressed in terms of corresponding left and right eigenvectors.

Computing Capacity (cont.)

- Example: $(d,k) = (1,\infty)$

$$C = \log \frac{1+\sqrt{5}}{2} \approx 0.6942$$

- More generally, $C_{d,k} = log\ \lambda_{d,k}$, where $\lambda_{d,k}$ is the largest real root of the polynomial

$$f_{d,k}(x) = x^{k+1} - x^{k-d} - \ldots - x - 1, \text{ for } k < \infty$$

and

$$C_{d,\infty} = C_{d-1,\,2d-1}, \text{ for } d \geq 1.$$

Constrained Coding Theorems

- Stronger coding theorems were motivated by the problem of constrained code design for magnetic recording.

  Theorem[Adler-Coppersmith-Hassner, 1983]

  Let $\boldsymbol{S}$ be a finite-type constrained system. If $\boldsymbol{m/n \leq C}$, then there exists a rate $\boldsymbol{m{:}n}$ sliding-block decodable, finite-state encoder.

  (Proof is constructive: state-splitting algorithm.)

  Theorem[Karabed-Marcus, 1988]

  Ditto if $\boldsymbol{S}$ is almost-finite-type.

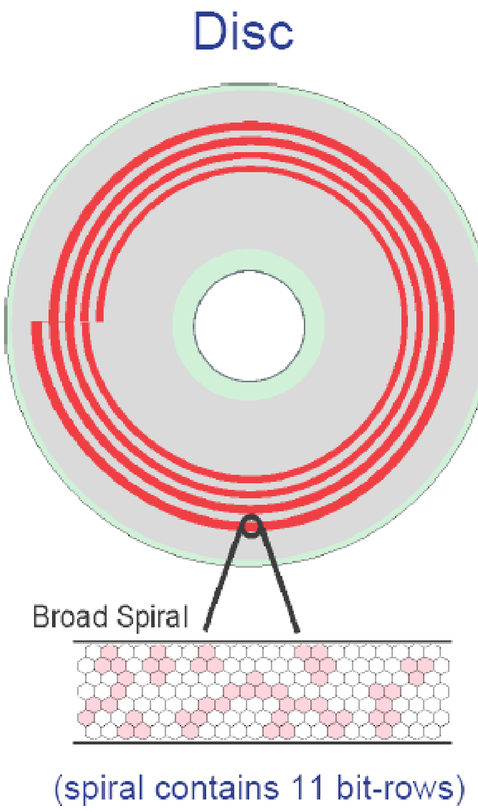
  (Proof not so constructive…)

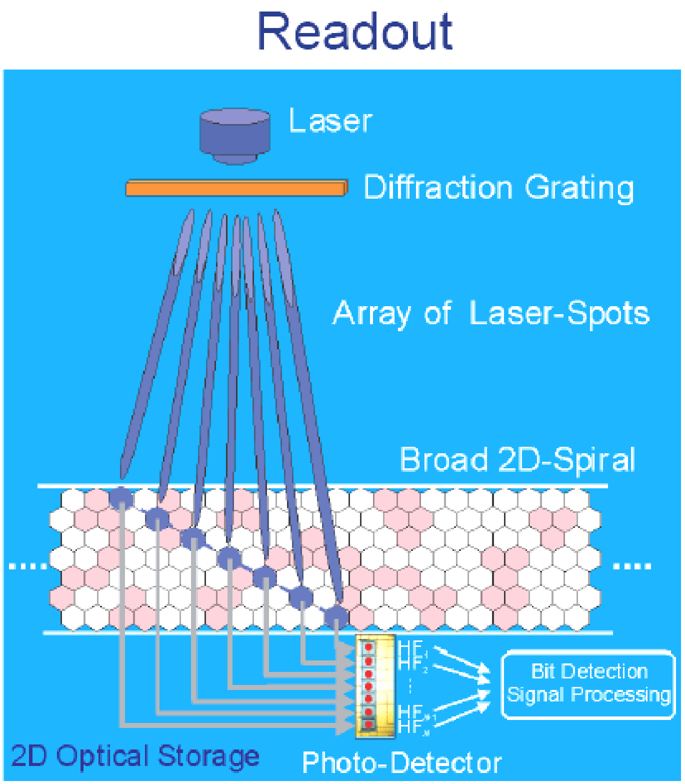
Two-Dimensional Constrained Systems

- Band-recording and page-oriented recording technologies require 2-dimensional constraints, for example:

- Two-Dimensional Optical Storage (TwoDOS) - Philips
- Holographic Storage - InPhaseTechnologies
- Patterned Magnetic Media – Hitachi, Toshiba, …
- Thermo-Mechanical Probe Array – IBM

TwoDOS





Courtesy of Wim Coene, Philips Research

Constraints on the Integer Lattice Z^2

- $S_{sq}^{1,\infty}$: $(d,k)=(1,\infty)$ constraint in x - y directions:

1				**1**		
			1		**1**	
	1			**1**		
1		**1**				**1**
	1		**1**		**1**	
		1		**1**		
1			**1**			

$$F=\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1 & 1\end{bmatrix}\right\}$$

Independent Sets

Hard-Square Model

(d,k) Constraints on the Integer Lattice Z^2

- For 2-dimensional (d,k) constraints $S_{sq}^{d,k}$, the capacity is given by:

$$C^{d,k} = \lim_{m,n\to\infty} \frac{N_{m,n}^{d,k}}{mn}$$

- The only nontrivial (d,k) pairs for which $C^{d,k}$ is known precisely are those with zero capacity, namely [Kato-Zeger, 1999] :

$$C^{d,d+1} = 0 \quad , d>0$$

$$C^{d,k} > 0 \, , k \geq d+2$$

(d,k) Constraints on Z^2 – Capacity Bounds

- Transfer matrix methods provide numerical bounds on $C^{1,\infty}$ [Calkin-Wilf, 1998] , [Nagy-Zeger, 2000]

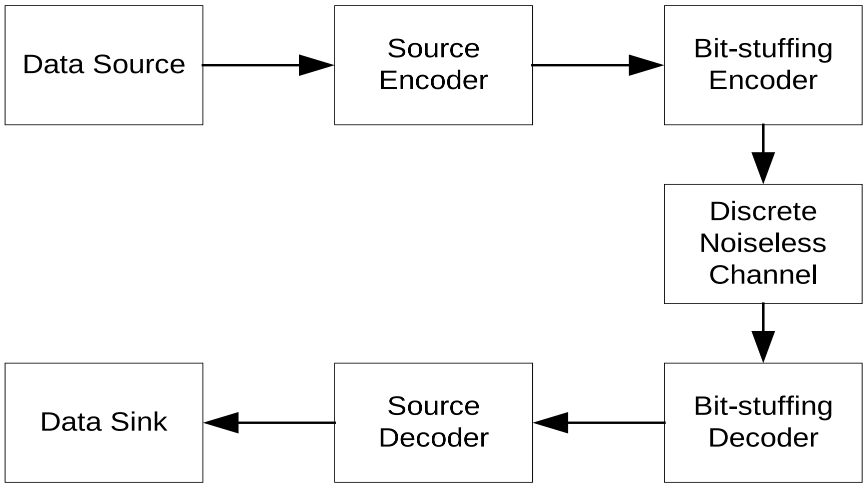
$$0.587891161775 \leq C^{1,\infty} \leq 0.587891161868$$

- Variable-rate "bit-stuffing" encoders for $S_{sq}^{d,\infty}$ yield best known lower bounds on $C^{d,\infty}$ for $d > 1$ [Halevy, et al., 2004]:

$$C^{d,\infty} \geq \lim_{m,n\to\infty} \max_{0<p<1} \frac{h(p)}{1+2dp-p^2\left(1-p^{2d-1}\right)} - o_{(\min\{m,n\})/d}(1)$$

d	Lower bound	d	Lower bound
2	0.4267	4	0.2858
3	0.3402	5	0.2464

2-D Bit-Stuffing (d, ∞) *RLL Encoder*



- Source encoder converts binary data to i.i.d bit stream (biased bits) with $\Pr(1) = p, \Pr(0) = 1 - p$, rate penalty $h(p)$.
- Bit-stuffing encoder inserts redundant bits which can be identified uniquely by decoder.
- Encoder rate $R(p)$ is a lower bound of the capacity. (For d=1, we can determine $R(p)$ precisely.)

2-D Bit-Stuffing (1,∞) *RLL Encoder*

- Biased sequence: 1 1 1 0 0 0 1 0 0 1 0 0 0 0 1 1 0 0 0

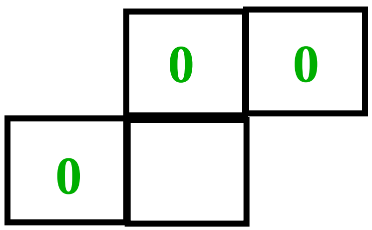
1	0	1	0	0	0	0			
0	1	0	1	0	0				
0	0	0	0	1	0				
0	0	0	1	0					
1	0	0	0						
0	0								
0									

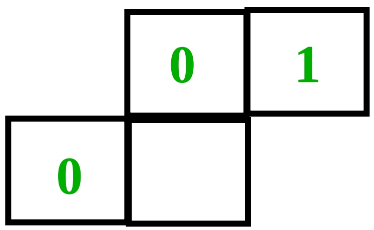
Optimal bias Pr(1) = p = 0.3556

$R(p)$=0.583056 (within 1% of capacity)

Enhanced Bit-Stuffing Encoder

- Use 2 source encoders, with parameters p_0 , p_1 .





Optimal bias

$\Pr(1) = p_0 = 0.328167$

Optimal bias

$\Pr(1) = p_1 = 0.433068$

$R(p_0 , p_1)$=0.587277 (within 0.1% of capacity)

Non-Isolated Bit (n.i.b.) Constraint on Z^2

- The non-isolated bit constraint S_{sq}^{nib} is defined by the forbidden set:

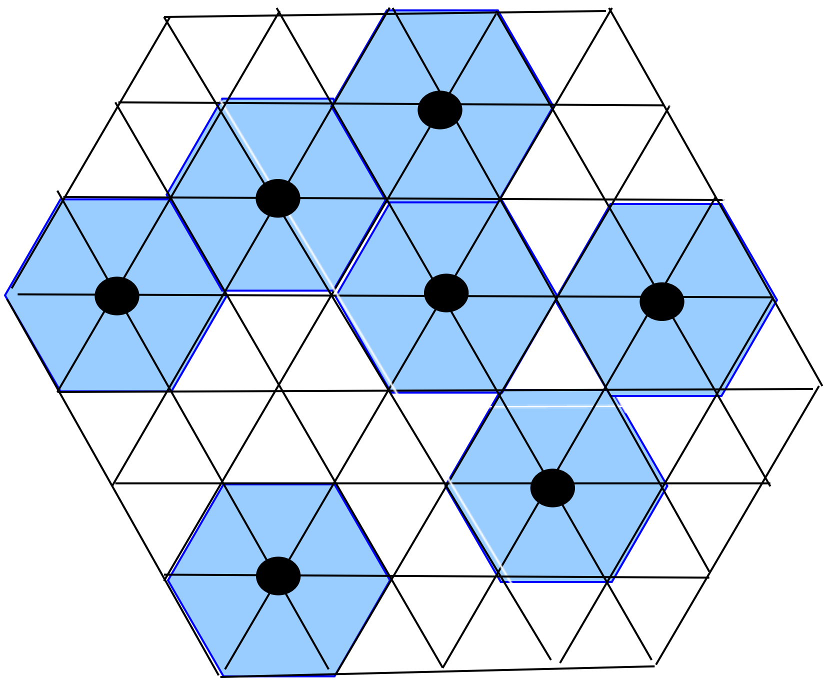
$$F = \left\{ \begin{bmatrix} & 0 & \\ 0 & 1 & 0 \\ & 0 & \end{bmatrix}, \begin{bmatrix} & 1 & \\ 1 & 0 & 1 \\ & 1 & \end{bmatrix} \right\}$$

- Analysis of the coding ratio of a bit-stuffing encoder yields:

$$0.91276 \leq C_{sq}^{nib} \leq 0.93965$$

Constraints on the Hexagonal Lattice A_2

- $S_{hex}^{1,\infty}$: $(d,k)=(1,\infty)$ constraints:



Hard-Hexagon Model

$\Leftrightarrow$

				1		
		1				
1			1		1	
			1			
1						

$$F=\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1 & 1\end{bmatrix},\begin{bmatrix} & 1\\1 & \end{bmatrix}\right\}$$

Hard Hexagon Capacity

- Capacity of hard hexagon model $C_{hex}^{1,\infty}$ is known precisely! [Baxter,1980]*

$C_{hex}^{1,\infty} = \log \kappa_h$, where $\kappa = \kappa_1\kappa_2\kappa_3\kappa_4$ and

$$\kappa_1 = 4^{-1}3^{5/4}11^{-5/12}c^{-2}$$

$$\kappa_2 = \left[1-\sqrt{1-c}+\sqrt{2+c+2\sqrt{1+c+c^2}}\right]^2$$

$$\kappa_3 = \left[-1-\sqrt{1-c}+\sqrt{2+c+2\sqrt{1+c+c^2}}\right]^2$$

$$\kappa_4 = \left[\sqrt{1-a}+\sqrt{2+a+2\sqrt{1+a+a^2}}\right]^{-1/2}$$

$$a = -\frac{124}{363}11^{1/3}$$

$$b = \frac{2501}{11979}33^{1/2}$$

$$c = \left[\frac{1}{4}+\frac{3}{8}a\left[(b+1)^{1/3}-(b-1)^{1/3}\right]\right]^{1/3}$$

So, $$C_{hex}^{1,\infty} \approx 0.480767622$$

Hard Hexagon Capacity

- Alternatively, the hard hexagon entropy constant κ satisfies a degree-24 polynomial with (big!) integer coefficients.

- Baxter does offer this disclaimer regarding his derivation, however:

> *"It is not mathematically rigorous, in that certain analyticity properties of κ are assumed, and the results of Chapter 13 (which depend on assuming that various large-lattice limits can be interchanged) are used. However, I believe that these assumptions, and therefore (14.1.18)-(14.1.24), are in fact correct."

(d,k) Constraints on A_2 – Capacity Bounds

- Zero capacity region partially known [Kukorelly-Zeger, 2001].
- Variable-to-fixed length "bit-stuffing" encoders for $S_{hex}^{d,\infty}$ yield best known lower bounds on $C_{hex}^{d,\infty}$ for $d>1$ [Halevy, et al., 2004]:

$$C_{hex}^{d,\infty} \geq \lim_{m,n\to\infty} \max_{0<p<1} \frac{h(p)}{1+3dp-p^2} - o_{(\min\{m,n\})/d} \quad (1)$$

d	Lower bound	d	Lower bound
2	0.3387	4	0.2196
3	0.2630	5	0.1901

Practical 2-D Constrained Codes

- There is no comprehensive algorithmic theory for constructing encoders and decoders for 2-D constrained systems.
- Very efficient bit-stuffing encoders have been defined and analyzed for several 2-D constraints, but they are not suitable for practical applications [Roth et al., 2001] , [Halevy et al., 2004] , [Nagy-Zeger, 2004].
- Optimal block codes with $m \times n$ rectangular code arrays have been designed for small values of m and n, and some finite-state encoders have been designed, but there is no generally applicable method [Demirkan-Wolf, 2004] .

Concluding Remarks

- The lack of convenient graph-based representations of 2-D constraints prevents the straightforward extension of 1-D techniques for analysis and code design.

- There are strong connections to statistical physics that may open up new approaches to understanding 2-D constrained systems (and, perhaps, vice-versa).

Noisy Finite-State ISI Channels (1-Dim.)

- Binary input process $x[i]$
- Linear intersymbol interference $h[i]$
- Additive, i.i.d. Gaussian noise $n[i] \sim N(0, \sigma^2)$

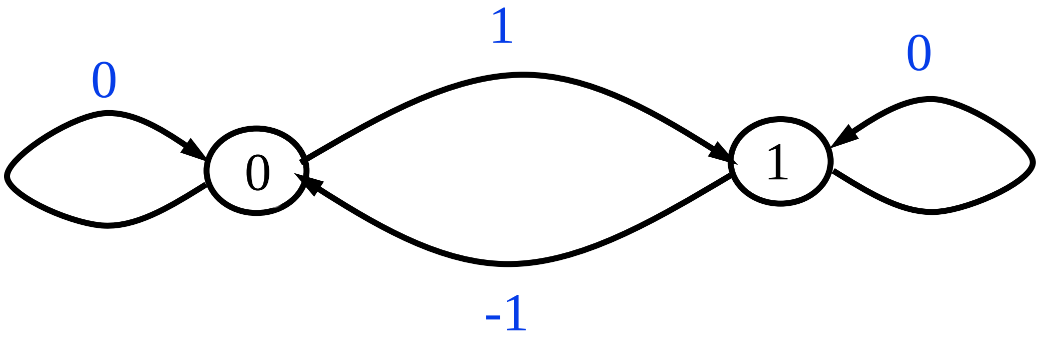
$$y[i] = \sum_{k=0}^{n-1} h[k]x[i-k] + n[i]$$

Example: Partial-Response Channels

- Impulse response:

$$h(D) = \sum_{i=0}^{N} h[i]D^i = (1-D)(1+D)^{N-1}$$

- Example: Dicode channel $\quad h(D) = (1-D)$



Entropy Rates

- Output entropy rate: $H(Y) = \lim_{n\to\infty} \frac{1}{n} H(Y_1^n)$
- Noise entropy rate: $H(N) = \frac{1}{2} log(\pi e N_0)$
- Conditional entropy rate:

$$H(Y \mid X) = \lim_{n\to\infty} \frac{1}{n} H(Y_1^n \mid X_1^n) = H(N)$$

Mutual Information Rates

- Mutual information rate:

$$I(X;Y) = H(Y) - H(Y \mid X) = H(Y) - H(N)$$

- Capacity: $C = \max_{P(X)} I(X;Y)$

- Symmetric information rate (SIR):

  Inputs $X = \{x[i]\}$ are constrained to be independent, identically distributed, and equiprobable binary digits.

Finding the Output Entropy Rate

- For one-dimensional ISI channel model:

$$H(Y) = \lim_{n \to \infty} \frac{1}{n} H(Y_1^n)$$

and

$$H(Y_1^n) = -E\left[\log p\left(Y_1^n = y_1^n\right)\right]$$

where

$$Y_1^n = \left[Y[1], Y[2], \ldots Y[n]\right]$$

Sample Entropy Rate

- If we simulate the channel $\boldsymbol{N}$ times, using inputs with specified (Markovian) statistics and generating output realizations

$$\underline{y}^{(k)} = \left[y[1]^{(k)}, y[2]^{(k)}, \ldots, y[n]^{(k)} \right], k = 1,2,\ldots,N$$

then

$$-\frac{1}{N}\sum_{k=1}^{N} \log p\left(\underline{y}^{(k)}\right)$$

converges to $H\left(Y_1^n\right)$ with probability 1 as $N \rightarrow \infty$.

Computing Sample Entropy Rate

- The forward recursion of the sum-product (BCJR) algorithm can be used to calculate the probability $p(y_1^n)$ of a sample realization of the channel output.

- In fact, we can write

$$-\frac{1}{n} \log p\left(y_1^n\right) = -\frac{1}{n} \sum_{i=1}^{n} \log p\left(y_i \mid y_1^{i-1}\right)$$

where the quantity $p\left(y_i \mid y_1^{i-1}\right)$ is precisely the normalization constant in the (normalized) forward recursion.

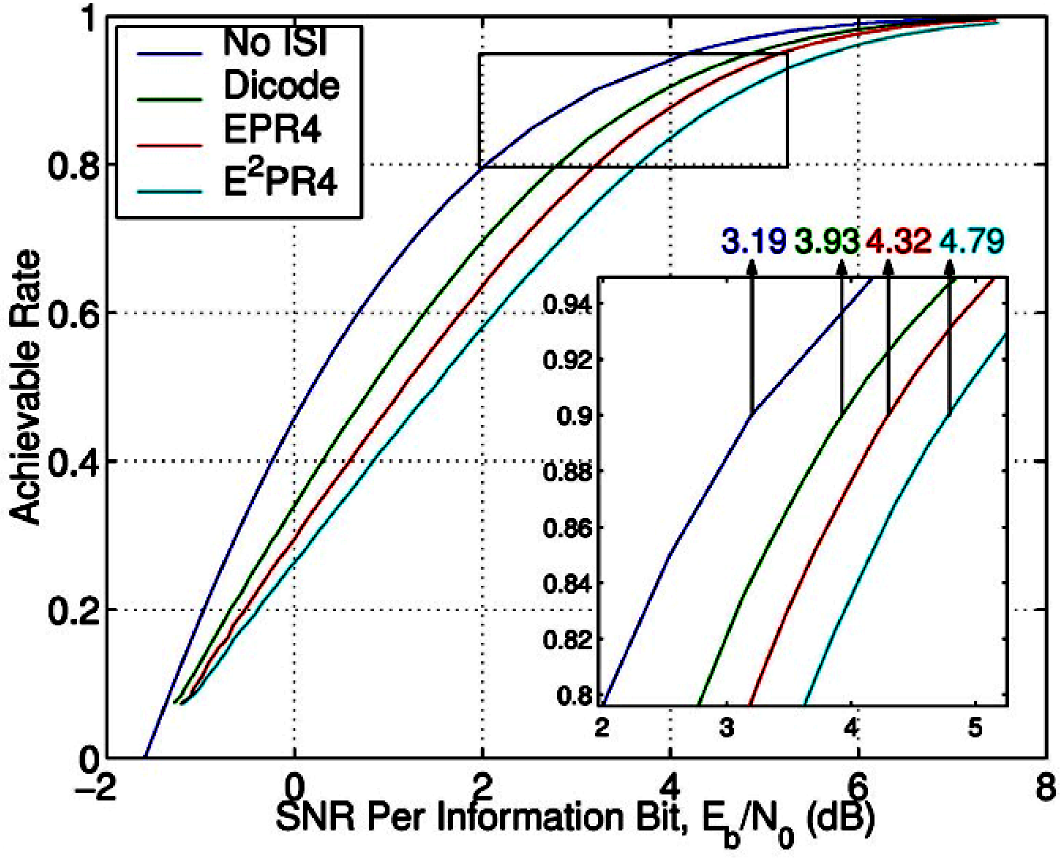
Computing Entropy Rates

- Shannon-McMillan-Breimann theorem implies

$$-\frac{1}{n}\log p\left(y_1^n\right)\underset{a.s.}{\rightarrow} H(Y)$$

as $n \rightarrow \infty$, where y_1^n is a single long sample realization of the channel output process.

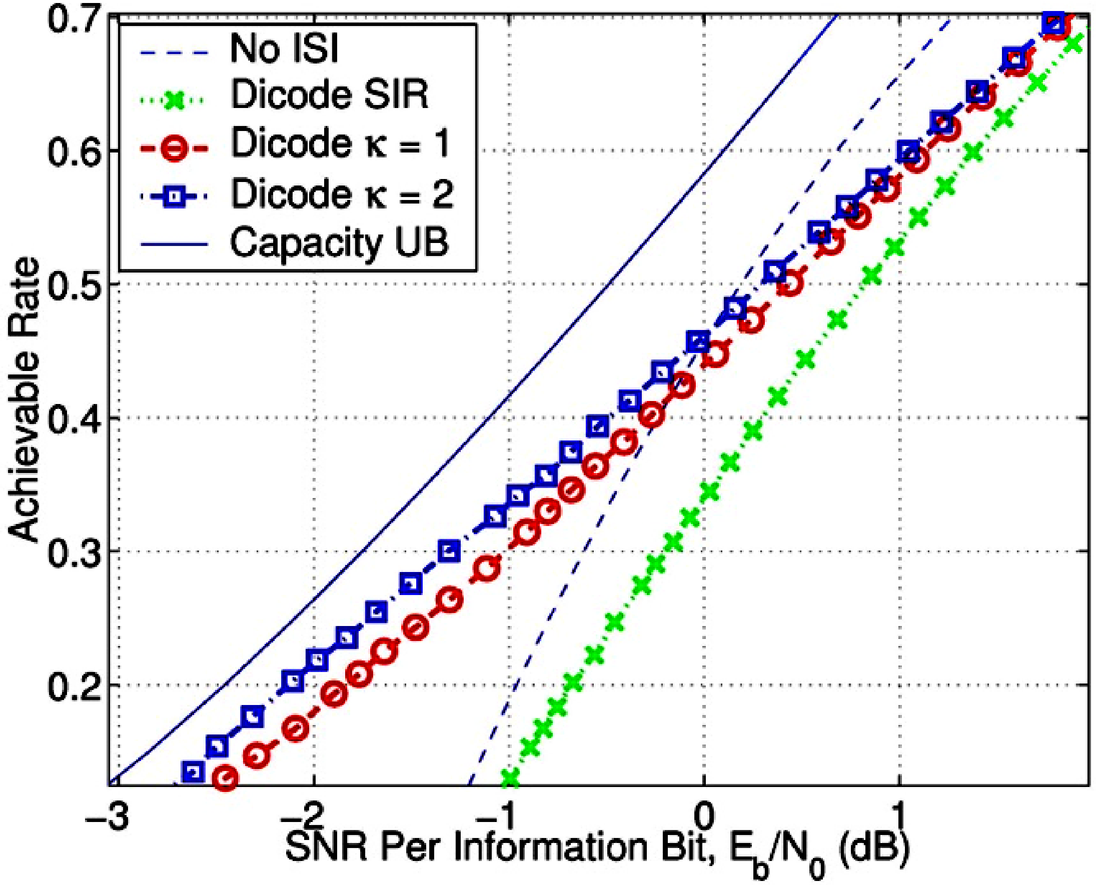
SIR for Partial-Response Channels



Computing the Capacity

- For Markov input process of specified order *r* , this technique can be used to find the mutual information rate. (Apply it to the combined source-channel.)
- For a fixed order *r* , [Kavicic, 2001] proposed a Generalized Blahut-Arimoto algorithm to optimize the parameters of the Markov input source.
- The stationary points of the algorithm have been shown to correspond to critical points of the information rate curve [Vontobel,2002] .

Capacity Bounds for Dicode h(D)=1-D



Markovian Sufficiency

Remark: It can be shown that optimized Markovian processes whose states are determined by their previous r symbols can asymptotically achieve the capacity of finite-state intersymbol interference channels with AWGN as the order r of the input process approaches ∞.

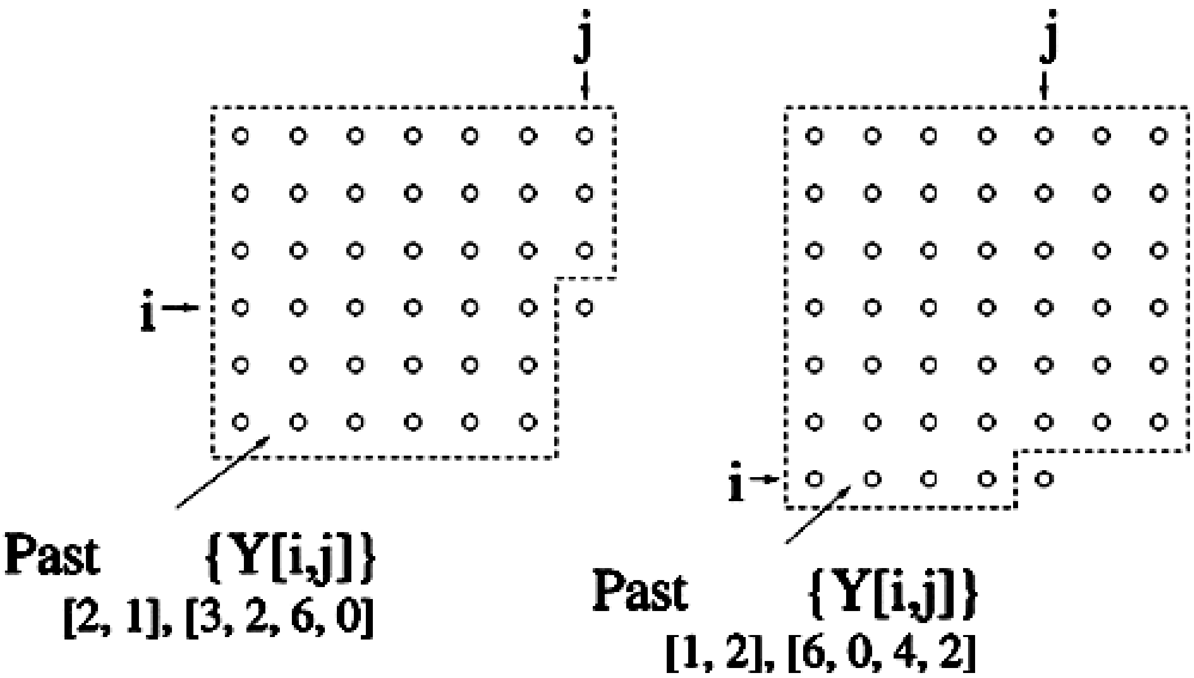
(This generalizes to 2 dimensional channels.)

[Chen-Siegel, 2004]

Capacity and SIR in Two Dimensions

- In two dimensions, we could estimate $H(Y)$ by calculating the sample entropy rate of a very large simulated output array.
- However, there is no counterpart of the BCJR algorithm in two dimensions to simplify the calculation.
- Instead, *conditional entropies can be used* to derive upper and lower bounds on $H(Y)$.

Examples of Past{Y[i,j]}



Conditional Entropies

- For a stationary two-dimensional random field Y on the integer lattice, the entropy rate satisfies:

$$H(Y) = H\left(Y[i,j] \mid Past_{k,\infty}\{Y[i,j]\}\right)$$

(The proof uses the entropy chain rule. See [5-6])

- This extends to random fields on the hexagonal lattice, via the natural mapping to the integer lattice.

Upper Bound on H(Y)

- For a stationary two-dimensional random field *Y*,

$$H(Y) \leq \min_{k} H_{k,l}^{U1}$$

where

$$H_{k,l}^{U1}(Y) = H\left(Y[i,j] \mid Past_{k,l}\{Y[i,j]\}\right)$$

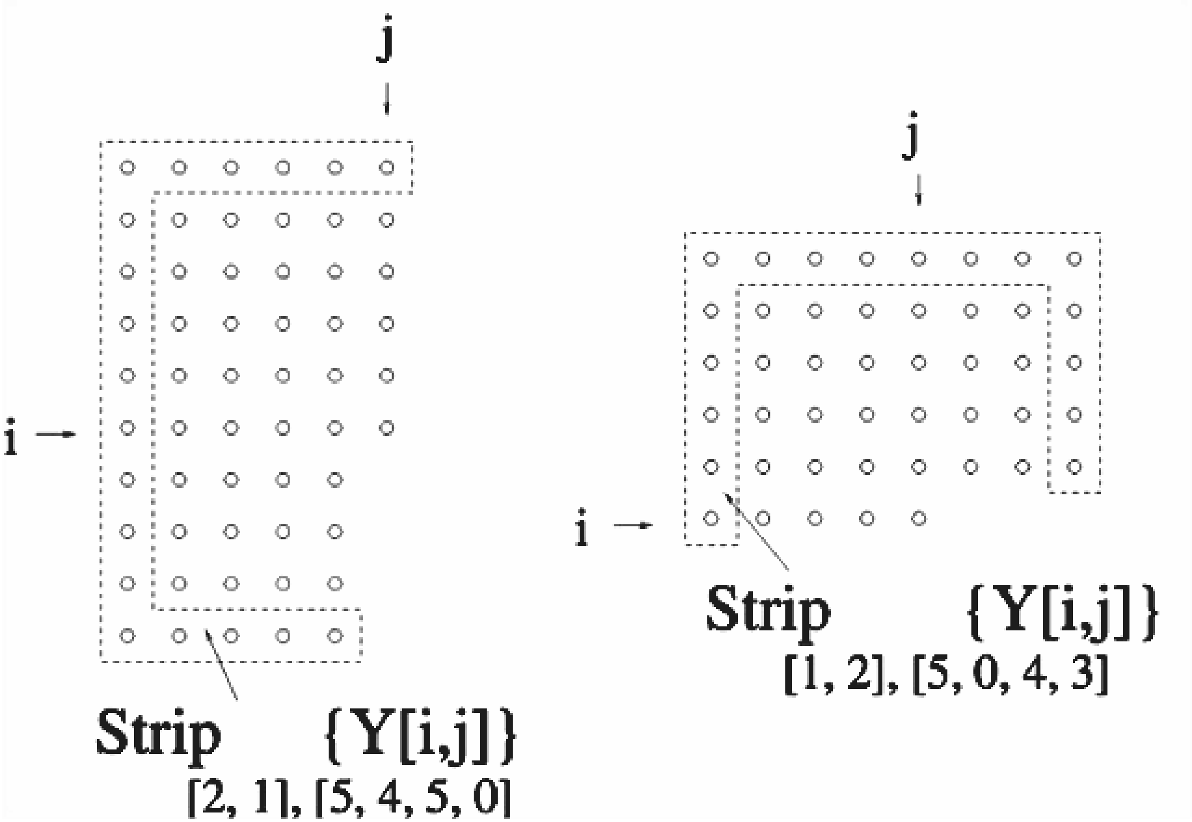
Two-Dimensional Boundary of Past{Y[i,j]}

- Define $Strip_{k,l}\{Y[i,j]\}$ to be the boundary

  of $Past_{k,l}\{Y[i,j]\}$.

- The exact expression for $Strip_{k,l}\{Y[i,j]\}$ is messy, but the geometrical concept is simple.

Two-Dimensional Boundary of Past{Y[i,j]}



Lower Bound on H(Y)

- For a stationary two-dimensional hidden Markov field Y,

where $$H(Y) \geq \max_k H_{k,l}^{L1}$$

$$H_{k,l}^{L1}(Y) = H\left(Y[i,j] \mid Past_{k,l}\{Y[i,j]\}, X\left(St_{k,l}\{Y[i,j]\}\right)\right)$$

and $X\left(St_{k,l}\{Y[i,j]\}\right)$ is the "state information" for

the strip $Strip_{k,l}\{Y[i,j]\}$.

Computing the SIR Bounds

- Estimate the two-dimensional conditional entropies $H(A|B)$ over a small array.

- Calculate $P(A,B), P(B)$ to get $P(A|B)$ for many realizations of output array.

- For column-by-column ordering, treat each row $\underline{Y}_i$ as a variable and calculate the joint probability $P\{\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_m\}$ row-by-row using the BCJR forward recursion.

2x2 Impulse Response

- "Worst-case" scenario - large ISI:

$$h_1[i,j] = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

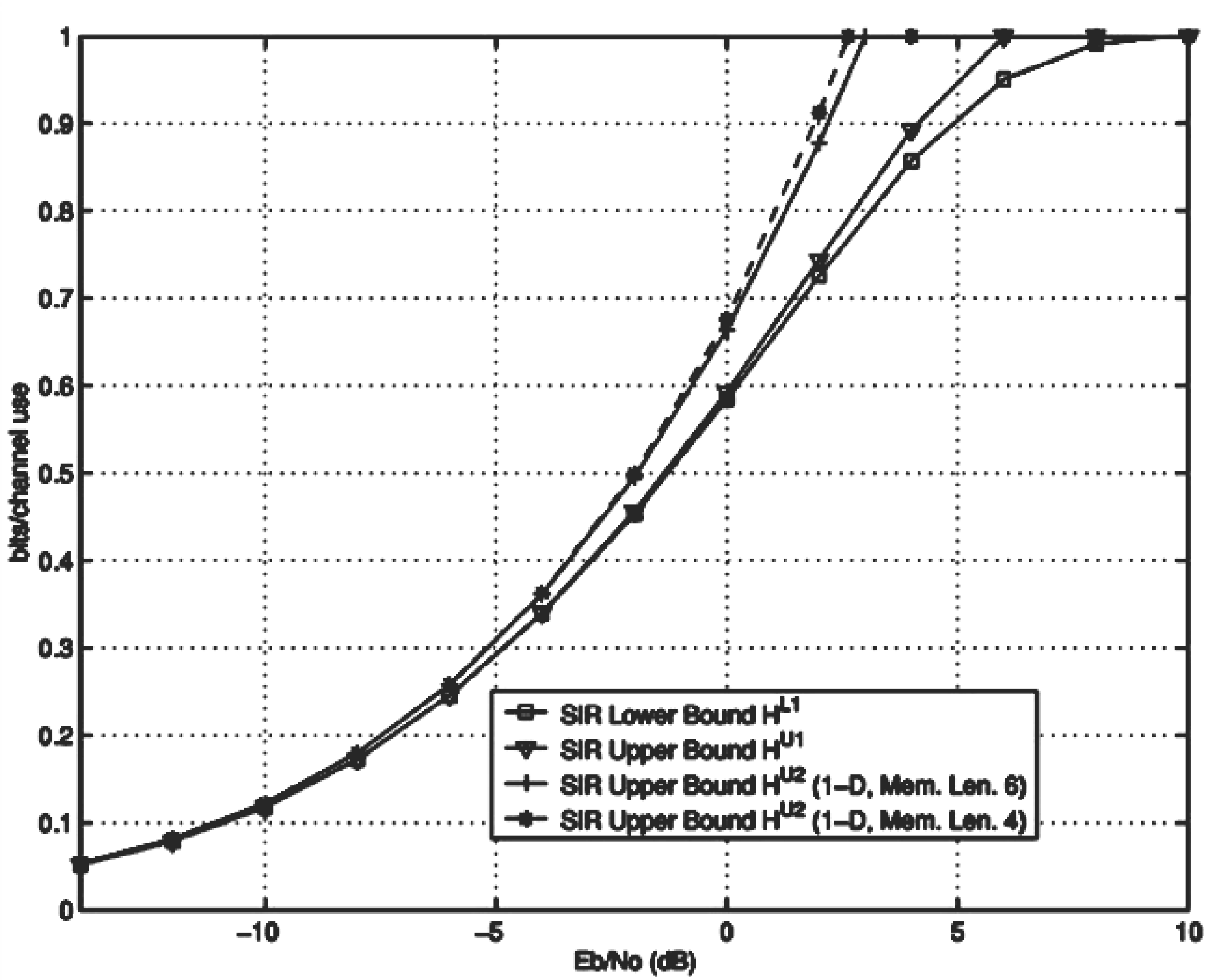
- Conditional entropies computed from 100,000 realizations.

- Upper bound: $$\min\left\{ H^{U1}_{[2,1],[7,7,3,0]} - \frac{1}{2}\log(\pi e N_0), 1 \right\}$$

- Lower bound: $$H^{L1}_{[2,1],[7,7,3,0]} - \frac{1}{2}\log(\pi e N_0)$$

(corresponds to element in middle of last column)

SIR Bounds for 2x2 Channel



Computing the SIR Bounds

- The number of states for each variable increases exponentially with the number of columns in the array.
- This requires that the two-dimensional impulse response have a small support region.
- It is desirable to find other approaches to computing bounds that reduce the complexity, perhaps at the cost of weakening the resulting bounds.

Alternative Upper Bound

- Modified BCJR approach limited to small impulse response support region.
- Introduce "auxiliary ISI channel" and bound

$$H(Y) \leq H_{k,l}^{U2}$$

where

$$H_{k,l}^{U2} = \int \cdots \int_{-\infty}^{\infty} -p\Big(y[i,j], Past_{k,l}\{y[i,j]\}\Big) \log q\Big(y[i,j] \,|\, Past_{k,l}\{y[i,j]\}\Big) d\underline{y}$$

and $q\Big(y[i,j] \,|\, Past_{k,l}\{y[i,j]\}\Big)$ is an arbitrary conditional

probability distribution.

3x3 Impulse Response

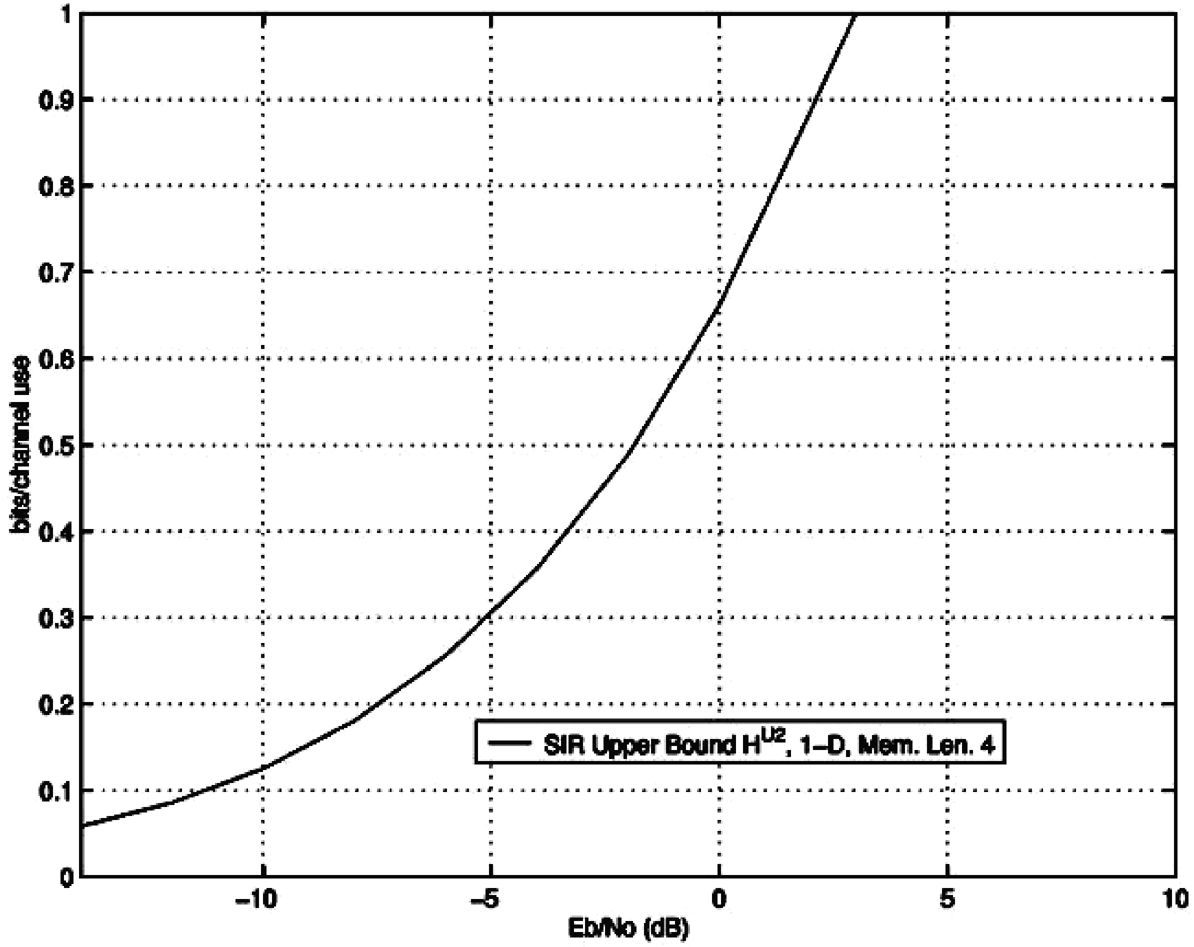
- Two-DOS transfer function

$$h_2[i,j] = \frac{1}{\sqrt{10}} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

- Auxiliary one-dimensional ISI channel with memory length 4.

- Useful upper bound up to $E_b/N_0 = 3$ dB.

SIR Upper Bound for 3x3 Channel



Concluding Remarks

- Recent progress has been made in computing information rates and capacity of 1-dim. noisy finite-state ISI channels.
- As in the noiseless case, the extension of these results to 2-dim. channels is not evident.
- Upper and lower bounds on the SIR of two-dimensional finite-state ISI channels have been developed.
- Monte Carlo methods were used to compute the bounds for channels with small impulse response support region.
- Bounds can be extended to multi-dimensional ISI channels.
- Further work is required to develop computable, tighter bounds for general multi-dimensional ISI channels.

References

1. D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," IEEE International Conference on Communications, Helsinki, Finland, June 2001, vol. 9, pp.2692-2695.
2. H.D. Pfister, J.B. Soriaga, and P.H. Siegel, "On the achievable information rate of finite state ISI channels," Proc. Globecom 2001, San Antonio, TX, November2001, vol. 5, pp. 2992-2996.
3. V. Sharma and S.K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," Proc. IEEE International Symposium on Information Theory, Washington, DC, June 2001, p. 283.
4. A. Kavcic, "On the capacity of Markov sources over noisy channels," Proc. Globecom 2001, San Antonio, TX, November2001, vol. 5, pp. 2997-3001.
5. D. Arnold, H.-A. Loeliger, and P.O. Vontobel, "Computation of information rates from finite-state source/channel models," Proc.40th Annual Allerton Conf. Commun., Control, and Computing, Monticello, IL, October 2002, pp. 457-466.

References

6. Y. Katznelson and B. Weiss, “Commuting measure-preserving transformations,” Israel J. Math., vol. 12, pp. 161-173, 1972.
7. D. Anastassiou and D.J. Sakrison, “Some results regarding the entropy rates of random fields,” IEEE Trans. Inform. Theory, vol. 28, vol. 2, pp. 340-343, March 1982.