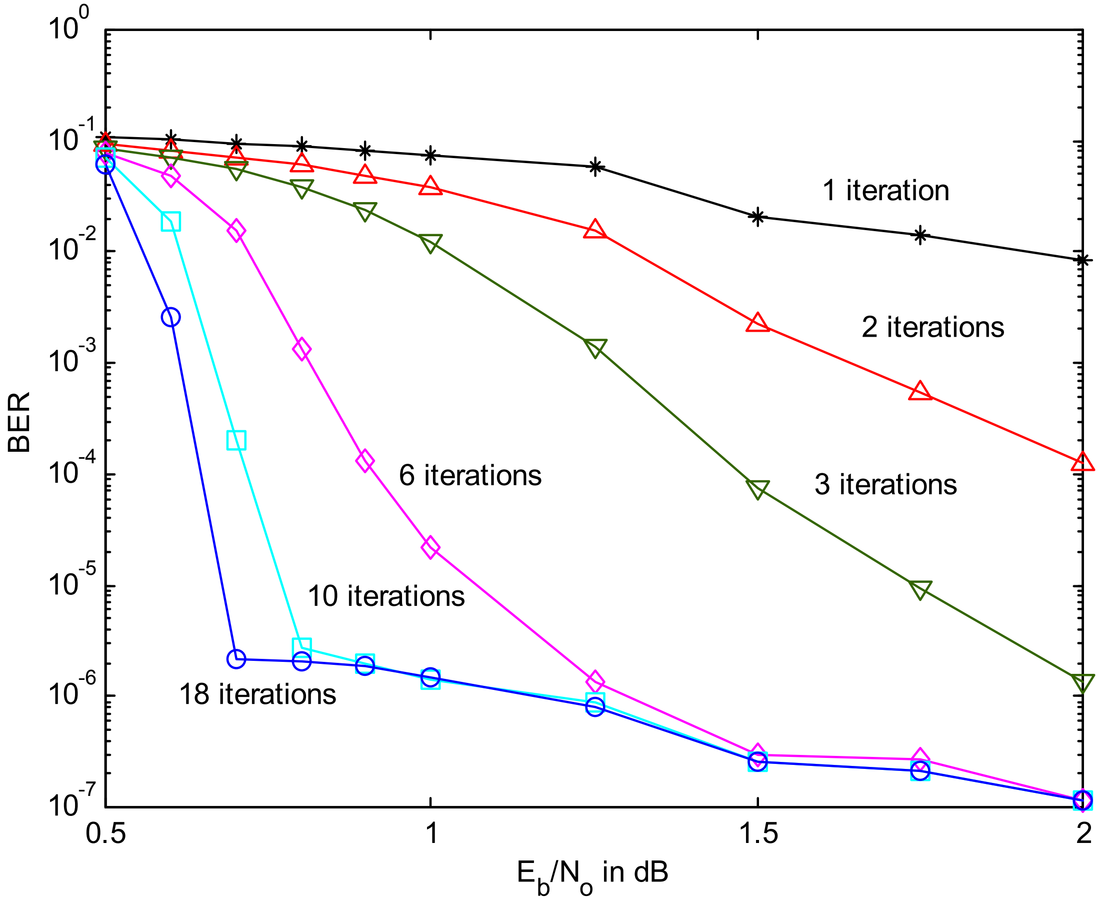
Mutual Information as a Tool for the Design, Analysis, and Testing of Modern Communication Systems

June 8, 2007

Matthew Valenti
Associate Professor
West Virginia University
Morgantown, WV 26506-6109
mvalenti@wvu.edu

Motivation: Turbo Codes



- Berrou et al 1993
  - Rate ½ code.
  - 65,536 bit message.
  - Two K=5 RSC encoders.
  - Random interleaver.
  - Iterative decoder.
  - BER = 10^{-5} at 0.7 dB.
- Comparison with Shannon capacity:
  - Unconstrained: 0 dB.
  - With BPSK: 0.2 dB.

Key Observations and Their Implications

- Key observations:
  - Turbo-like codes closely approach the channel capacity.
  - Such codes are complex and can take a long time to simulate.
- Implications:
  - If we know that we can find a code that approaches capacity, why waste time simulating the actual code?
  - Instead, let's devote our design effort towards determining capacity and optimizing the system with respect to capacity.
  - Once we are done with the capacity analysis, we can design (select?) and simulate the code.

Challenges

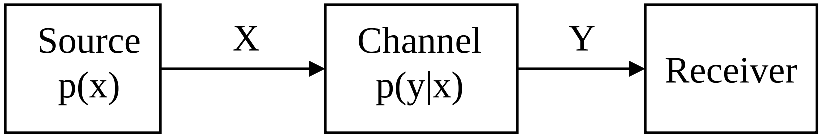
- How to efficiently find capacity under the constraints of:
  - Modulation.
  - Channel.
  - Receiver formulation.
- How to optimize the system with respect to capacity.
  - Selection of free parameters, e.g. code rate, modulation index.
  - Design of the code itself.
- Dealing with nonergodic channels
  - Slow and block fading.
  - hybrid-ARQ systems.
  - Relaying networks and cooperative diversity.
  - Finite-length codewords.

Overview of Talk

- The capacity of AWGN channels
  - Modulation constrained capacity.
  - Monte Carlo methods for determining constrained capacity.
  - CPFSK: A case study on capacity-based optimization.
- Design of binary codes
  - Bit interleaved coded modulation (BICM) and off-the-shelf codes.
  - Custom code design using the EXIT chart.
- Nonergodic channels.
  - Block fading and Information outage probability.
  - Hybrid-ARQ.
  - Relaying and cooperative diversity.
  - Finite length codeword effects.

Noisy Channel Coding Theorem (Shannon 1948)

- Consider a memoryless channel with input X and output Y



  - The channel is completely characterized by p(x,y)
- The ***capacity*** C of the channel is

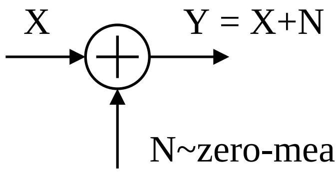
$$C = \max_{p(x)}\{I(X;Y)\} = \max_{p(x)}\left\{\iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy\right\}$$

  - where I(X,Y) is the (average) ***mutual information*** between X and Y.
- 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.

Capacity of the AWGN Channel with Unconstrained Input

- Consider the one-dimensional AWGN channel

The input X is drawn
from ***any*** distribution
with average energy
$E[X^2] = E_s$



N~zero-mean white Gaussian
with energy $E[N^2]= N_0/2$

- The capacity is

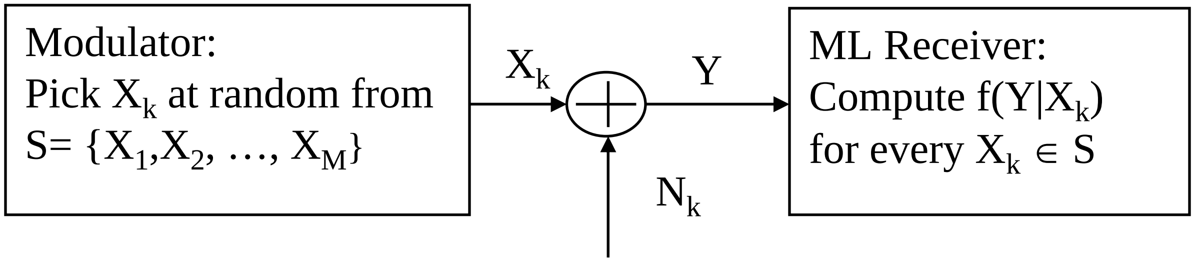
$$C = \max_{p(x)} \left\{ I(X;Y) \right\} = \frac{1}{2} \log_2 \left(\frac{2E_s}{N_o} + 1 \right)$$

bits per channel use

- The X that attains capacity is Gaussian distributed.
  - Strictly speaking, Gaussian X is not practical.

Capacity of the AWGN Channel with a Modulation-Constrained Input

- Suppose X is drawn with equal probability from the finite set S = $\{X_1, X_2, \ldots, X_M\}$



Modulator:
Pick X_k at random from
S= $\{X_1, X_2, \ldots, X_M\}$

ML Receiver:
Compute $f(Y|X_k)$
for every $X_k \in S$

  - where $f(Y|X_k) = \kappa\, p(Y|X_k)$ for any κ common to all X_k
- Since p(x) is now fixed

$$C = \max_{p(x)} \left\{ I(X;Y) \right\} = I(X;Y)$$

  - i.e. calculating capacity boils down to calculating mutual info.

Entropy and Conditional Entropy

- Mutual information can be expressed as:

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

- Where the ***entropy*** of X is

$$H(X) = E[h(X)] = \int p(x)h(x)dx$$

$$\text{where} \quad \underbrace{h(x) = \log\frac{1}{p(x)} = -\log p(x)}_{\text{self-information}}$$

- And the ***conditional entropy*** of X given Y is

$$H(X|Y) = E[h(X|Y)] = \iint p(x,y)h(x|y)dxdy$$

$$\text{where} \quad h(x|y) = -\log p(x|y)$$

Calculating Modulation-Constrained Capacity

- To calculate:

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

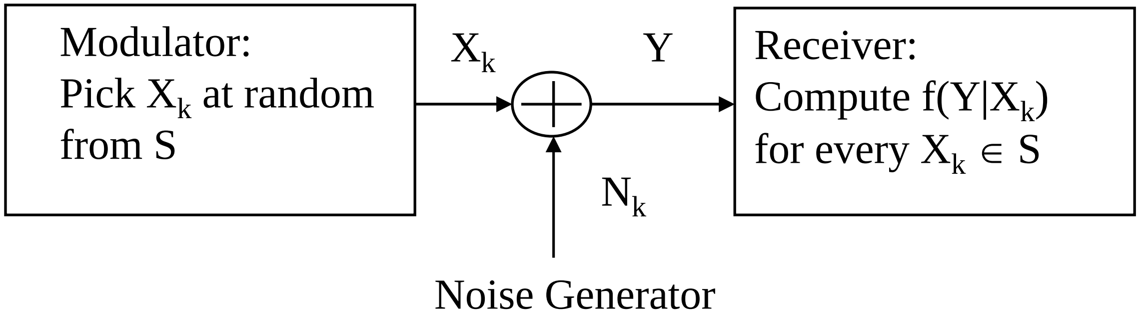
- We first need to compute H(X)

$$\begin{aligned} H(X) &= E[h(X)] \\ &= E\left[\log \frac{1}{p(X)}\right] \\ &= E[\log M] \\ &= \log M \end{aligned}$$

$$p(X) = \frac{1}{M}$$

- Next, we need to compute H(X|Y)=E[h(X|Y)]
  - This is the “hard” part.
  - In some cases, it can be done through numerical integration.
  - Instead, let’s use Monte Carlo simulation to compute it.

Step 1: Obtain p(x|y) from f(y|x)



Modulator:
Pick X_k at random
from S

X_k

Y

N_k

Noise Generator

Receiver:
Compute $f(Y|X_k)$
for every $X_k \in S$

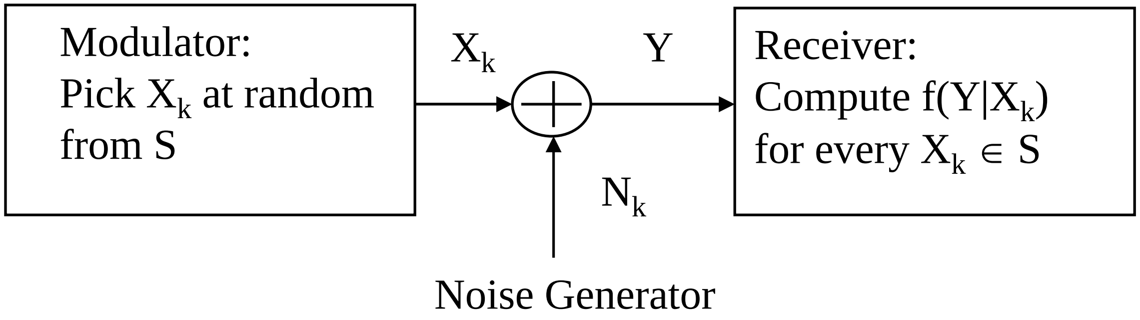
- Since

$$\sum_{x' \in S} p(x'|y) = 1$$

- We can get p(x|y) from

$$p(x|y) = \frac{p(x|y)}{\sum_{x' \in S} p(x'|y)} = \frac{\dfrac{p(y|x)p(x)}{p(y)}}{\sum_{x' \in S} \dfrac{p(y|x')p(x')}{p(y)}} = \frac{f(y|x)}{\sum_{x' \in S} f(y|x')}$$

Step 2: Calculate h(x|y)



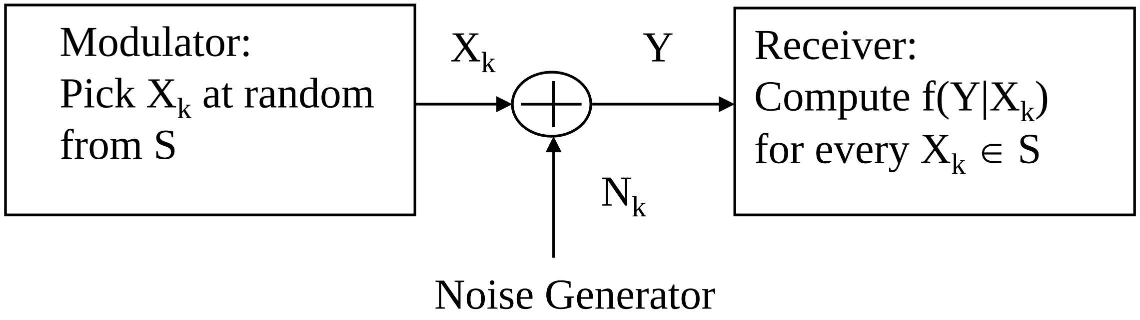
- Given a value of x and y (from the simulation) compute

$$p(x\,|\,y) = \frac{f(y\,|\,x)}{\sum_{x' \in S} f(y\,|\,x')}$$

- Then compute

$$h(x\,|\,y) = -\log p(x\,|\,y) = -\log f(y\,|\,x) + \log \sum_{x' \in S} f(y\,|\,x')$$

Step 3: Calculating H(X|Y)



Modulator:
Pick X_k at random
from S

X_k

Y

N_k

Receiver:
Compute $f(Y|X_k)$
for every $X_k \in S$

Noise Generator

- Since:

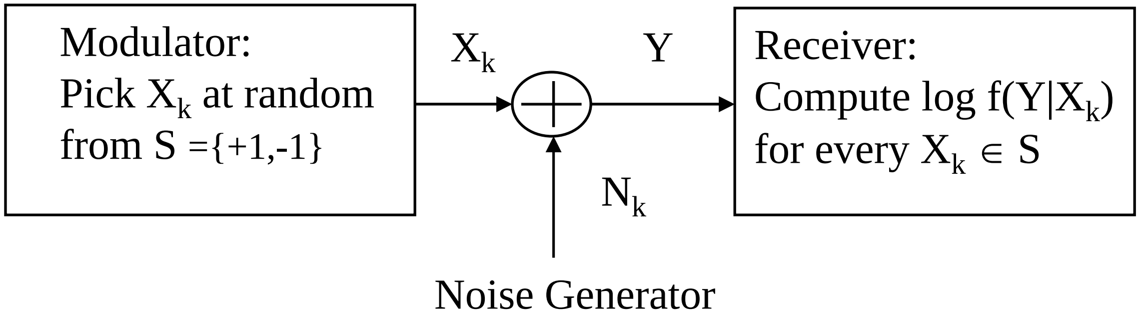
$$H(X\,|\,Y) = E[h(X\,|\,Y)] = \iint p(x,y)h(x\,|\,y)dxdy$$

- Because the simulation is ergodic, H(X|Y) can be found by taking the sample mean:

$$H(X\,|\,Y) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} h(X^{(n)}\,|\,Y^{(n)})$$

- where $(X^{(n)},Y^{(n)})$ is the nth realization of the random pair (X,Y).
  - i.e. the result of the nth simulation trial.

Example: BPSK



Modulator:
Pick X_k at random
from S ={+1,-1}

X_k

Y

N_k

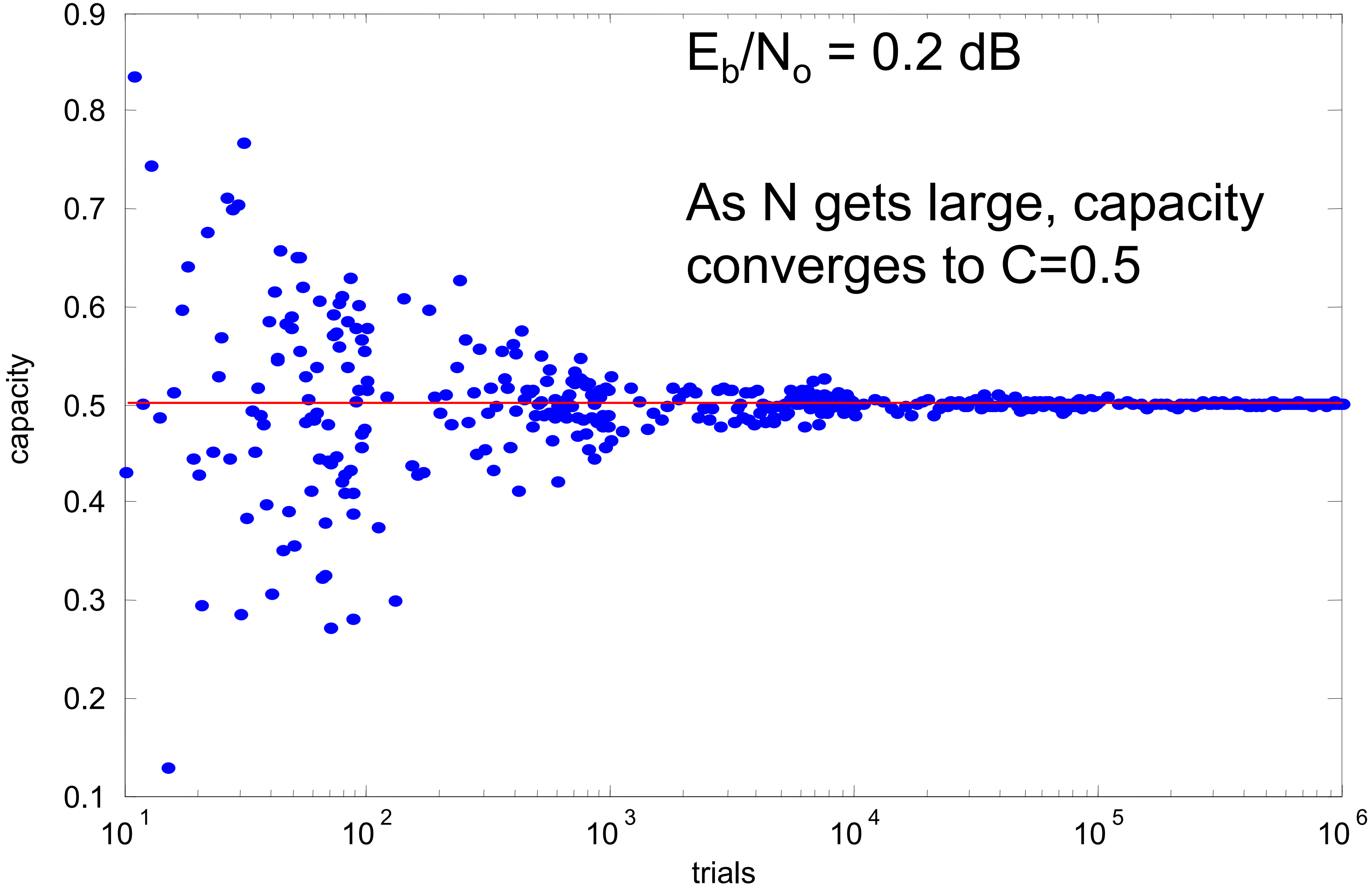
Noise Generator

Receiver:
Compute log $f(Y|X_k)$
for every $X_k \in S$

- Suppose that S ={+1,-1} and N has variance $N_0/2E_s$
- Then:

$$\log f(y\,|\,x) = -\frac{E_s}{N_o}\|y-x\|^2$$

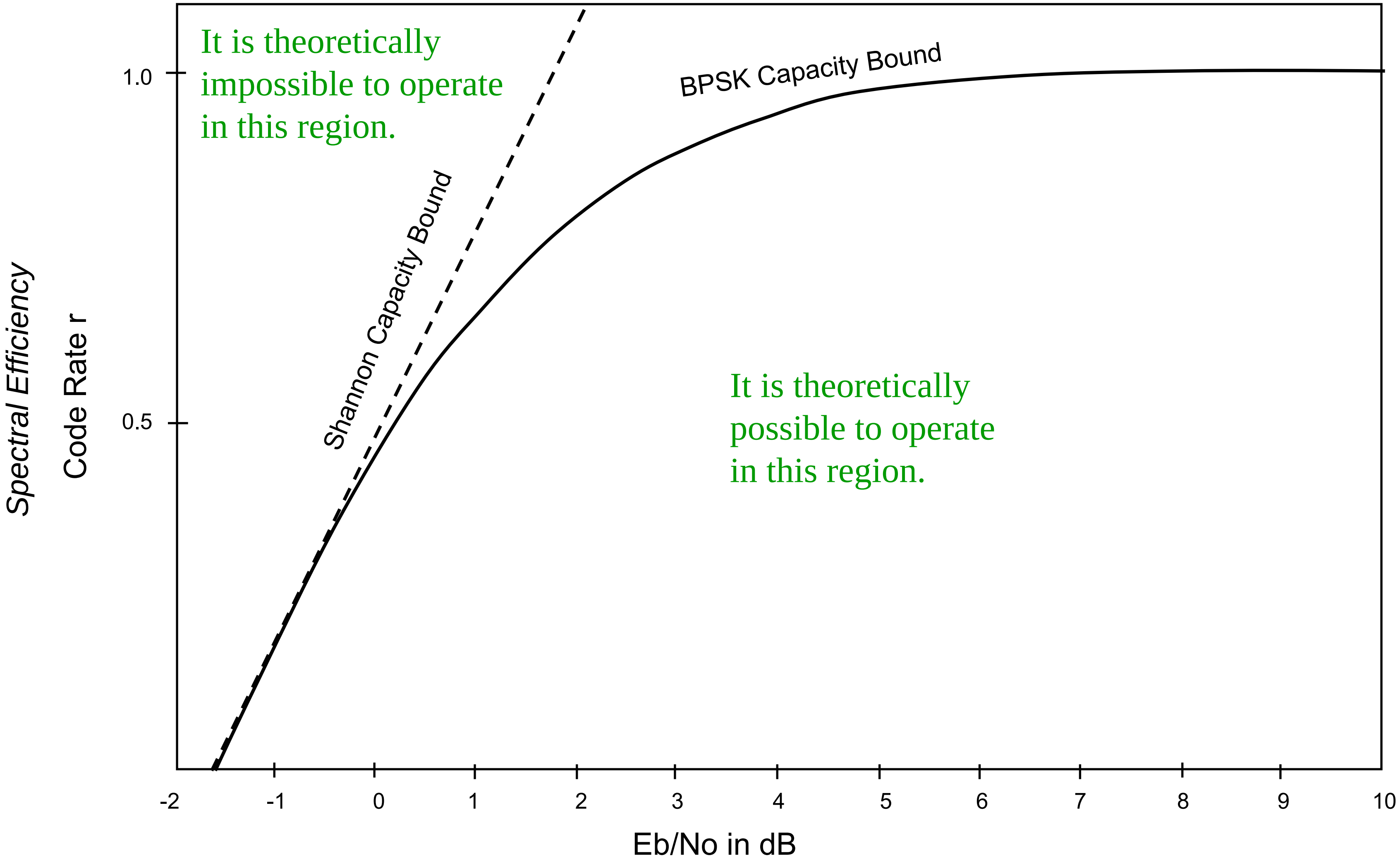
BPSK Capacity as a Function of Number of Simulation Trials



E_b/N_o = 0.2 dB

As N gets large, capacity converges to C=0.5

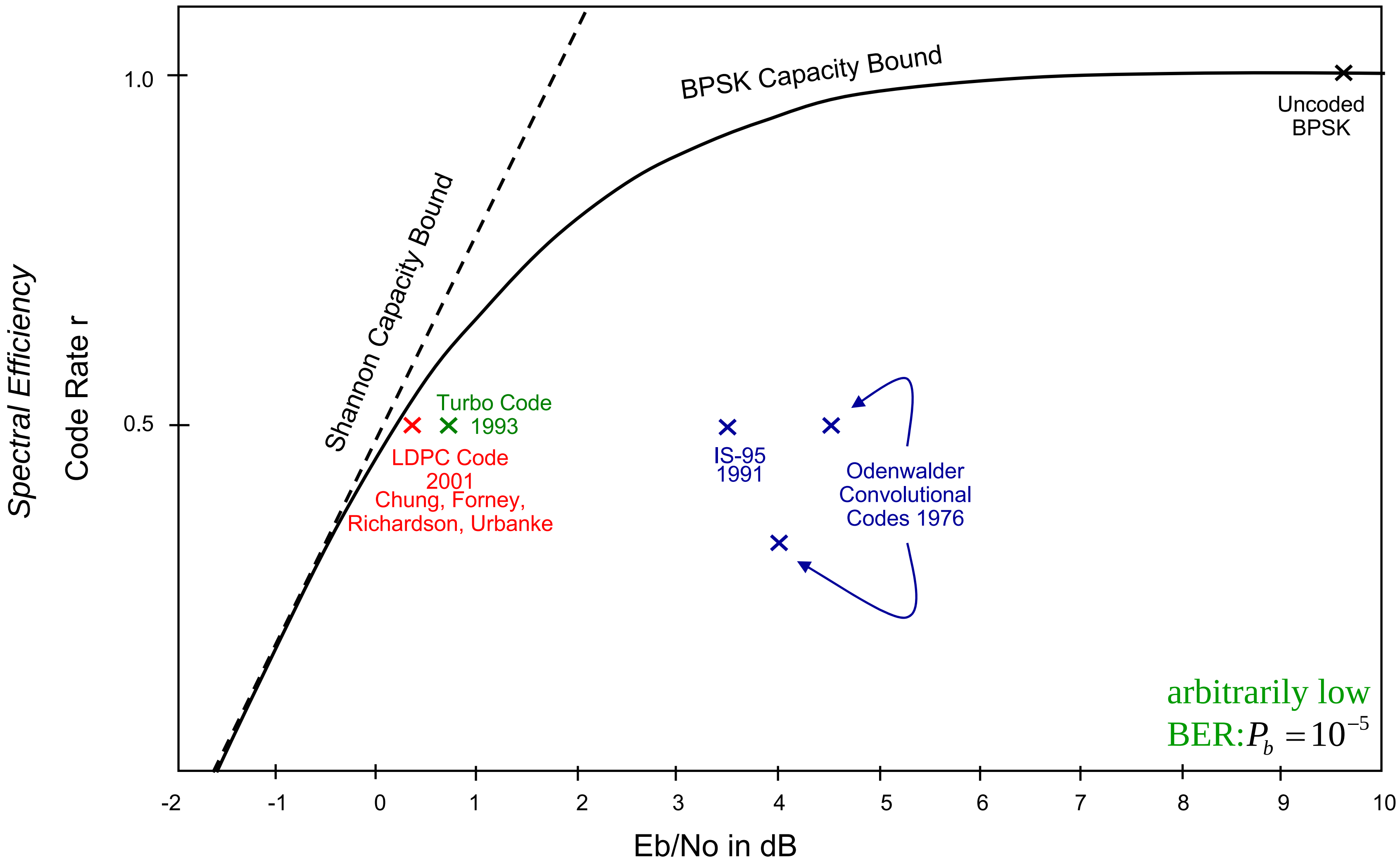
Unconstrained vs.
BPSK Constrained Capacity



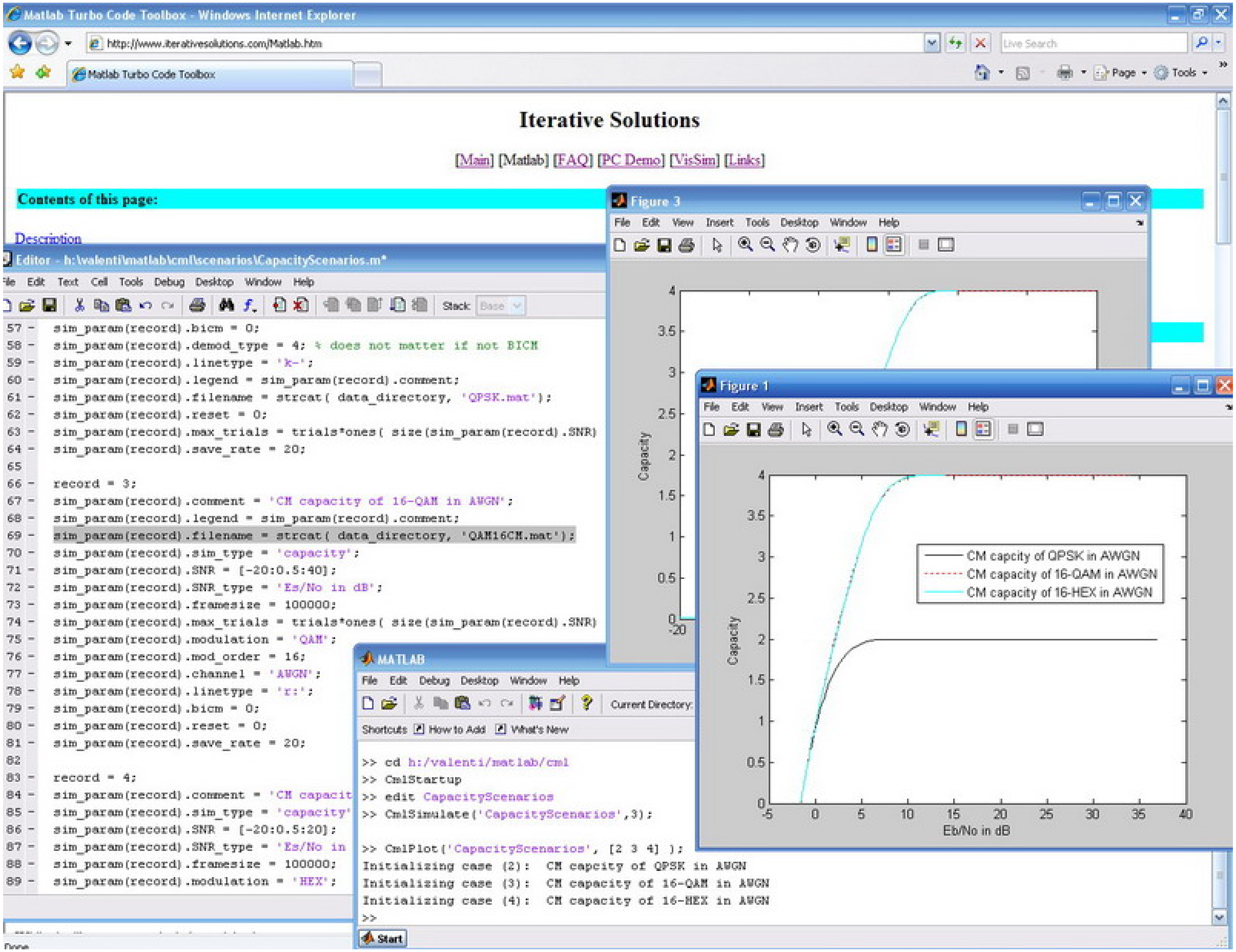
It is theoretically impossible to operate in this region.

It is theoretically possible to operate in this region.

Power Efficiency of Standard Binary Channel Codes

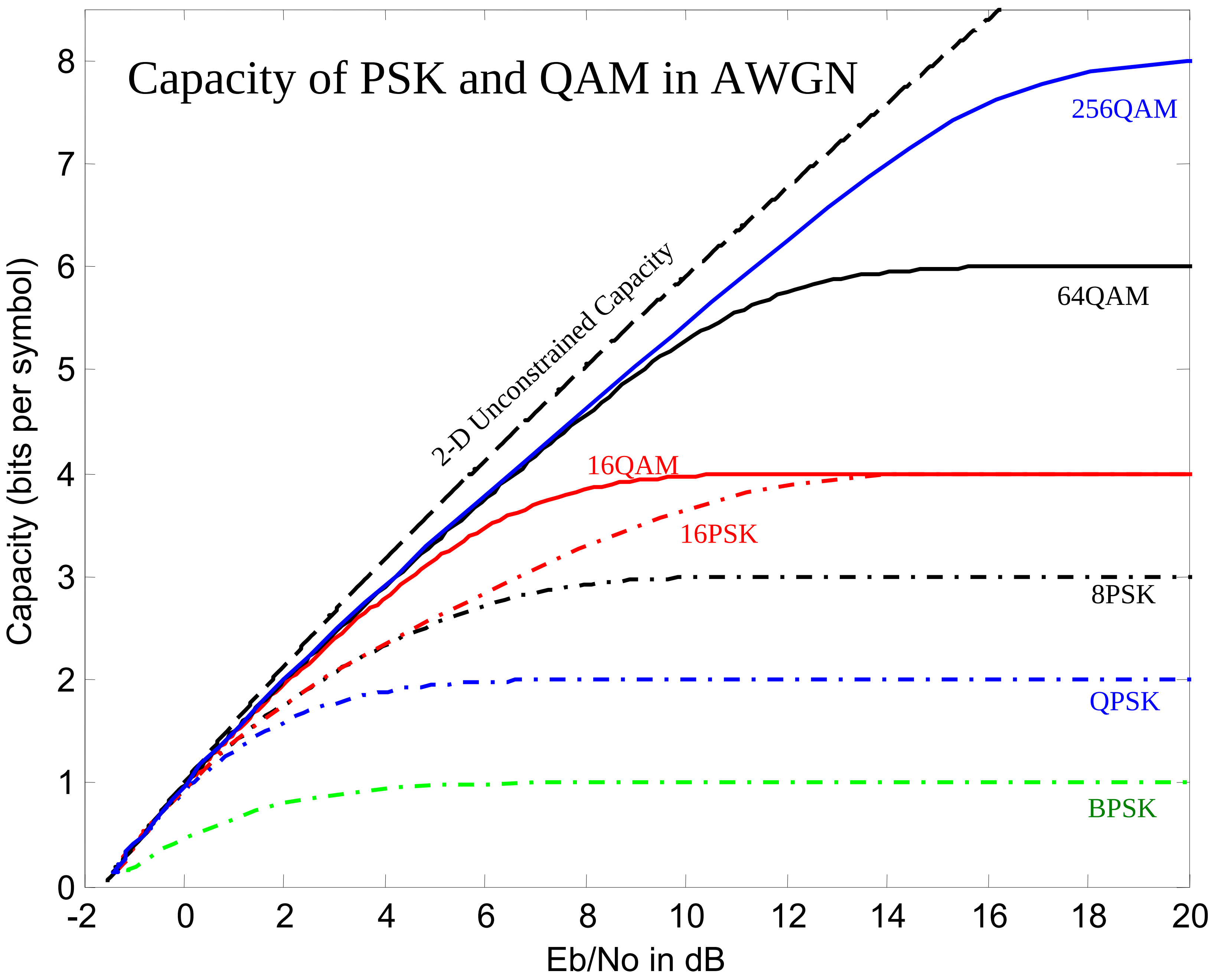


Software to Compute Capacity
www.iterativesolutions.com



```
sim_param(record).bicm = 0;
sim_param(record).demod_type = 4; % does not matter if not BICM
sim_param(record).linetype = 'k-';
sim_param(record).legend = sim_param(record).comment;
sim_param(record).filename = strcat( data_directory, 'QPSK.mat');
sim_param(record).reset = 0;
sim_param(record).max_trials = trials*ones( size(sim_param(record).SNR)
sim_param(record).save_rate = 20;

record = 3;
sim_param(record).comment = 'CM capacity of 16-QAM in AWGN';
sim_param(record).legend = sim_param(record).comment;
sim_param(record).filename = strcat( data_directory, 'QAM16CM.mat');
sim_param(record).sim_type = 'capacity';
sim_param(record).SNR = [-20:0.5:40];
sim_param(record).SNR_type = 'Es/No in dB';
sim_param(record).framesize = 100000;
sim_param(record).max_trials = trials*ones( size(sim_param(record).SNR)
sim_param(record).modulation = 'QAM';
sim_param(record).mod_order = 16;
sim_param(record).channel = 'AWGN';
sim_param(record).linetype = 'r:';
sim_param(record).bicm = 0;
sim_param(record).reset = 0;
sim_param(record).save_rate = 20;

record = 4;
sim_param(record).comment = 'CM capacit
sim_param(record).sim_type = 'capacity'
sim_param(record).SNR = [-20:0.5:20];
sim_param(record).SNR_type = 'Es/No in
sim_param(record).framesize = 100000;
sim_param(record).modulation = 'HEX';
```

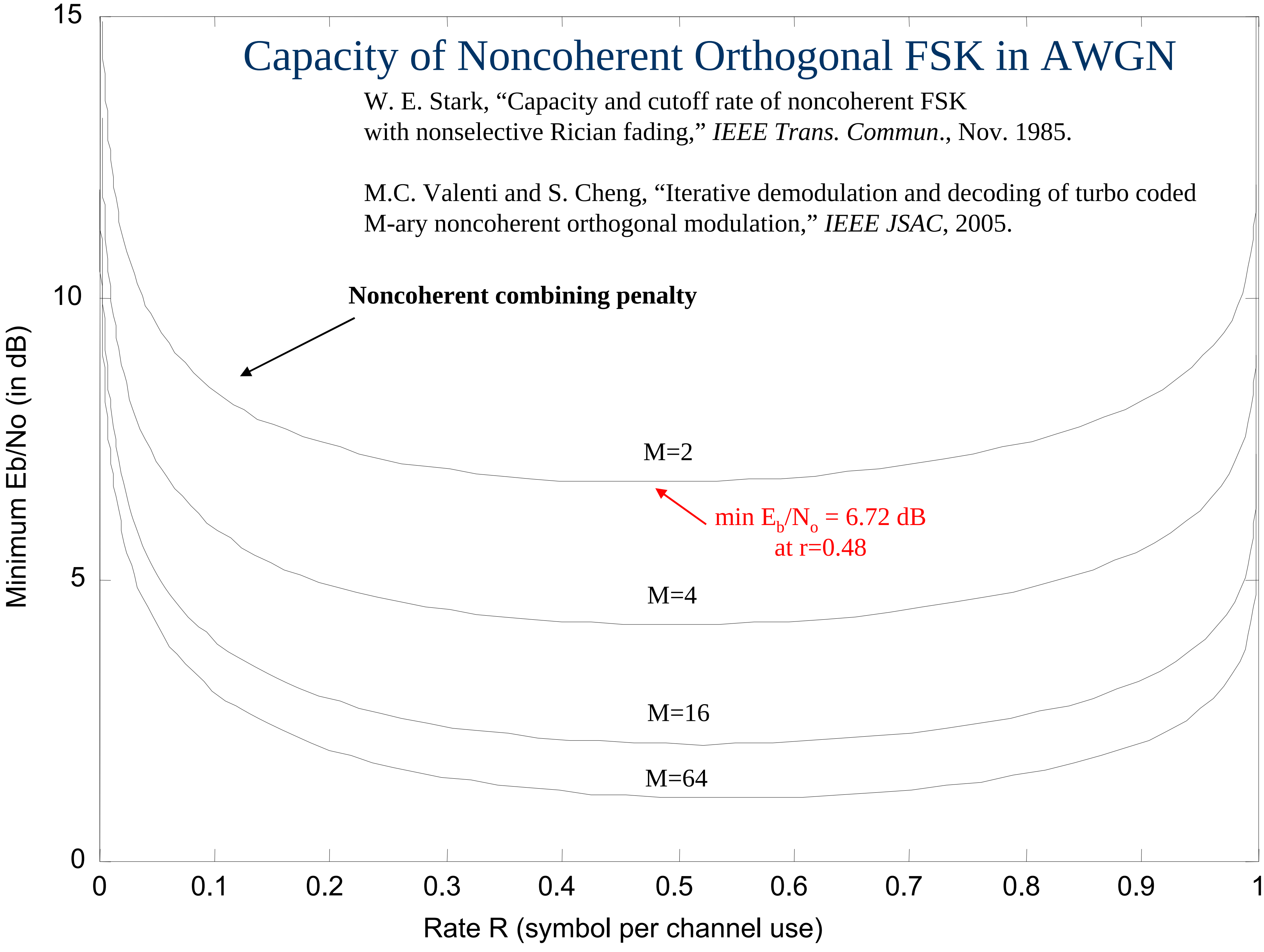
```
>> cd h:/valenti/matlab/cml
>> CmlStartup
>> edit CapacityScenarios
>> CmlSimulate('CapacityScenarios',3);

>> CmlPlot('CapacityScenarios', [2 3 4] );
Initializing case (2):  CM capcity of QPSK in AWGN
Initializing case (3):  CM capacity of 16-QAM in AWGN
Initializing case (4):  CM capacity of 16-HEX in AWGN
>>
```


Capacity of PSK and QAM in AWGN
2-D Unconstrained Capacity
256QAM
64QAM
16QAM
16PSK
8PSK
QPSK
BPSK
Capacity (bits per symbol)
0
1
2
3
4
5
6
7
8
-2
0
2
4
6
8
10
12
14
16
18
20
Eb/No in dB

Capacity of Noncoherent Orthogonal FSK in AWGN

W. E. Stark, “Capacity and cutoff rate of noncoherent FSK with nonselective Rician fading,” *IEEE Trans. Commun.*, Nov. 1985.

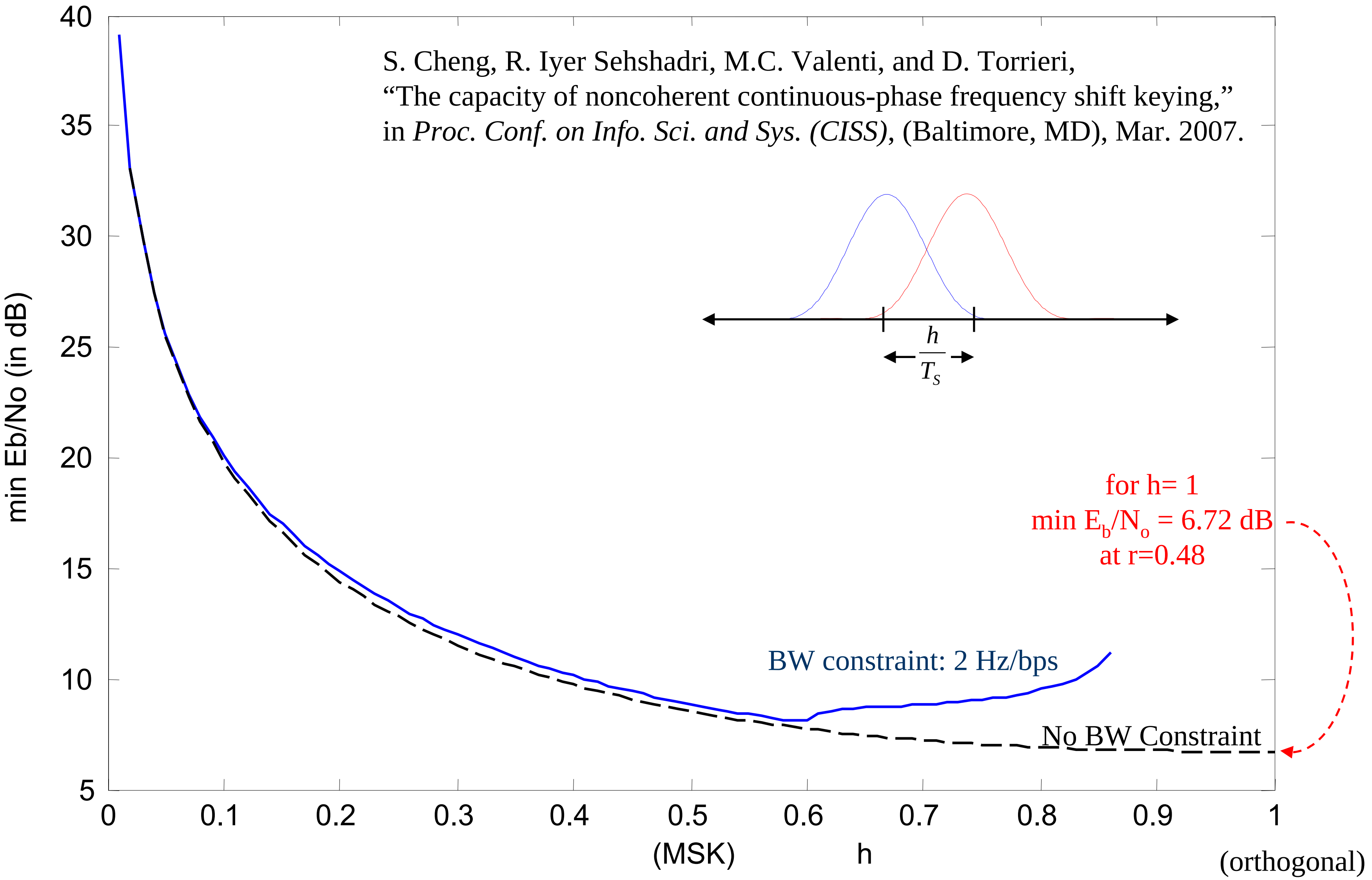
M.C. Valenti and S. Cheng, “Iterative demodulation and decoding of turbo coded M-ary noncoherent orthogonal modulation,” *IEEE JSAC*, 2005.



Noncoherent combining penalty

min E_b/N_o = 6.72 dB at r=0.48

Capacity of Nonorthogonal CPFSK



S. Cheng, R. Iyer Sehshadri, M.C. Valenti, and D. Torrieri,
“The capacity of noncoherent continuous-phase frequency shift keying,”
in *Proc. Conf. on Info. Sci. and Sys. (CISS)*, (Baltimore, MD), Mar. 2007.

for h= 1
min E_b/N_o = 6.72 dB
at r=0.48

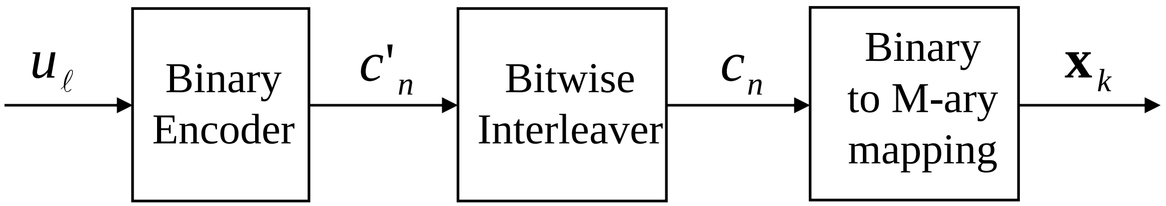
Overview of Talk

- The capacity of AWGN channels
  - Modulation constrained capacity.
  - Monte Carlo methods for determining constrained capacity.
  - CPFSK: A case study on capacity-based optimization.
- Design of binary codes
  - Bit interleaved coded modulation (BICM).
  - Code design using the EXIT chart.
- Nonergodic channels.
  - Block fading: Information outage probability.
  - Hybrid-ARQ.
  - Relaying and cooperative diversity.
  - Finite length codeword effects.

BICM

(Caire 1998)

- Coded modulation (CM) is required to attain the aforementioned capacity.
  - Channel coding and modulation handled jointly.
  - Alphabets of code and modulation are matched.
  - 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).



u_ℓ

Binary Encoder

c'_n

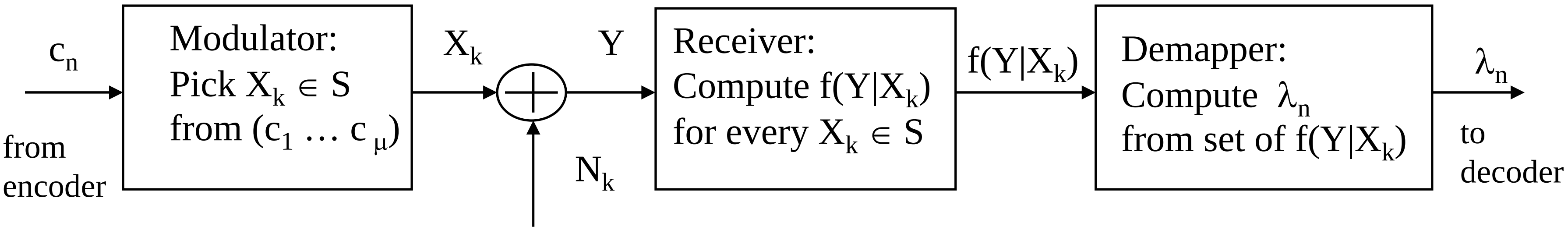
Bitwise Interleaver

c_n

Binary to M-ary mapping

$\mathbf{x}_k$

BICM Receiver



c_n

from encoder

Modulator:
Pick $X_k \in S$
from $(c_1 \ldots c_\mu)$

X_k

Y

N_k

Receiver:
Compute $f(Y|X_k)$
for every $X_k \in S$

$f(Y|X_k)$

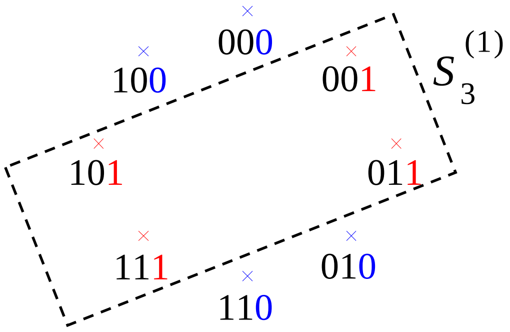
Demapper:
Compute λ_n
from set of $f(Y|X_k)$

λ_n

to decoder

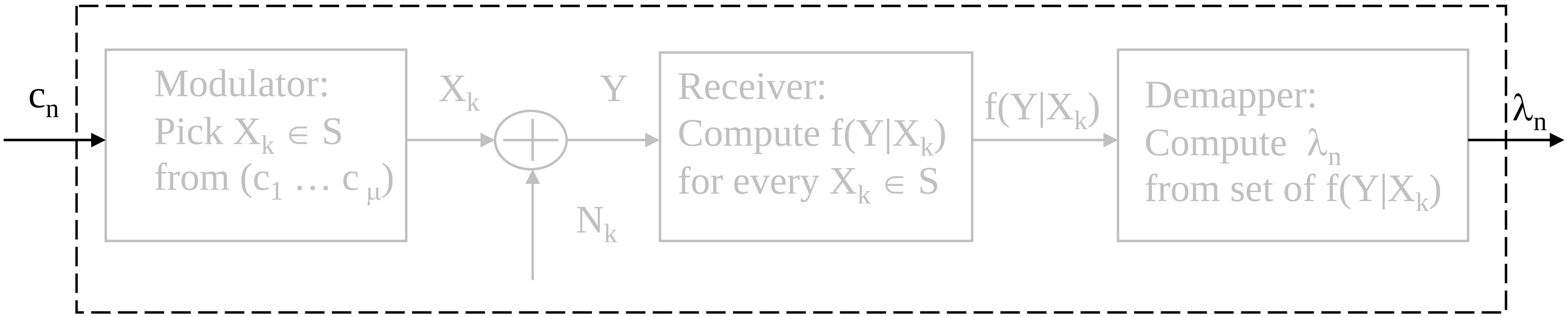
- The symbol likelihoods must be transformed into bit log-likelihood ratios (LLRs):

$$\lambda_n = \log \frac{\sum_{X_k \in S_n^{(1)}} f\left(Y \mid X_k\right)}{\sum_{X_k \in S_n^{(0)}} f\left(Y \mid X_k\right)}$$



  - where $S_n^{(1)}$ represents the set of symbols whose nth bit is a 1.
  - and $S_n^{(0)}$ is the set of symbols whose nth bit is a 0.

BICM Capacity



- Can be viewed as $\mu=\log_2 M$ binary parallel channels, each with capacity

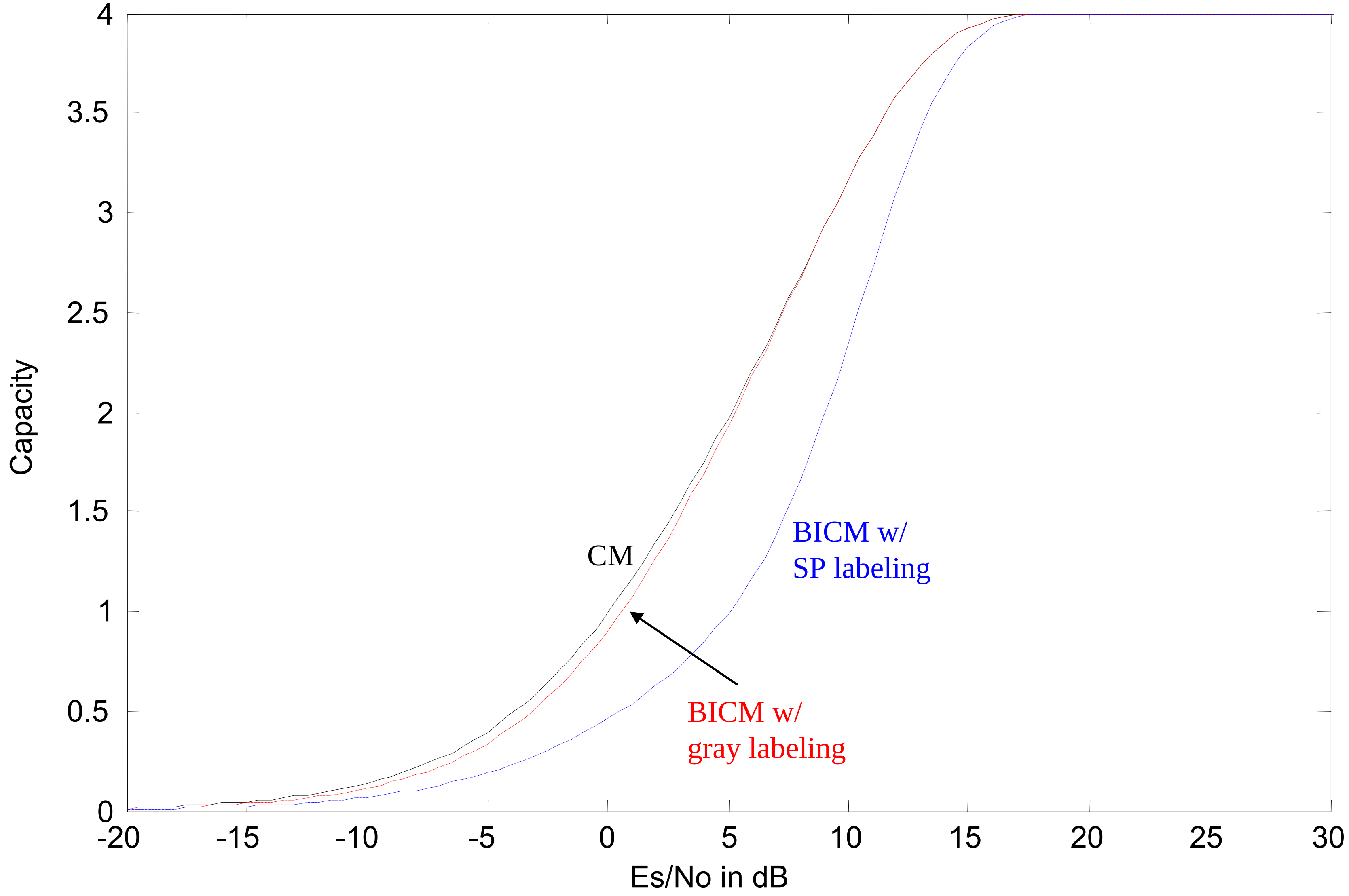
$$C_n = I(c_n, \lambda_n)$$

- Capacity over parallel channels adds:

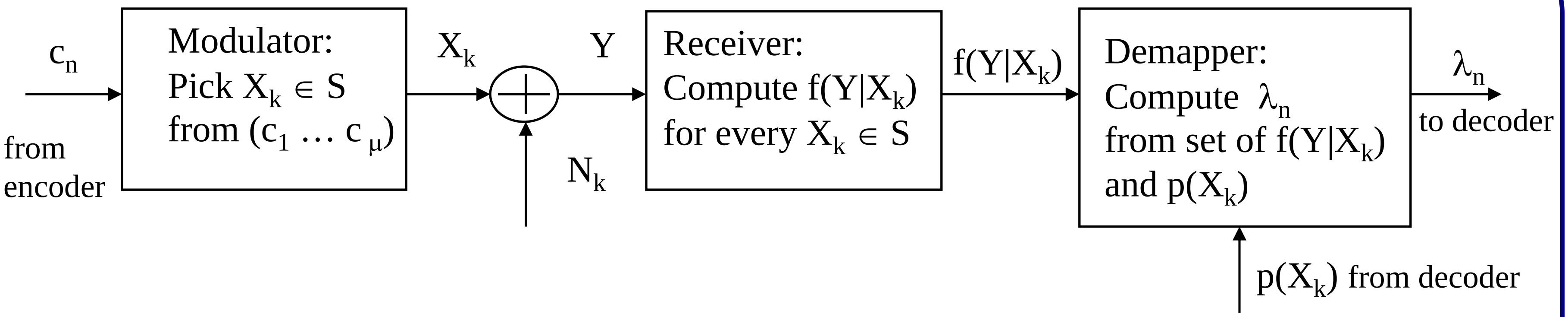
$$C = \sum_{n=1}^{\mu} C_n$$

- As with the CM case, Monte Carlo integration may be used.

CM vs. BICM Capacity for 16QAM



BICM-ID
(Li & Ritcey 1997)



c_n from encoder

Modulator: Pick $X_k \in S$ from $(c_1 \ldots c_\mu)$

X_k

Y

N_k

Receiver: Compute $f(Y|X_k)$ for every $X_k \in S$

$f(Y|X_k)$

Demapper: Compute λ_n from set of $f(Y|X_k)$ and $p(X_k)$

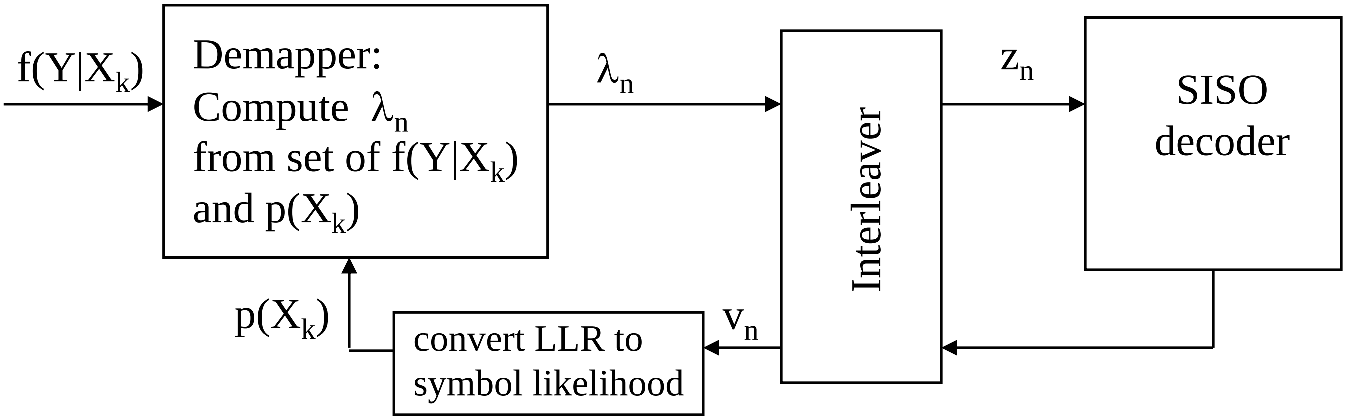
λ_n to decoder

$p(X_k)$ from decoder

- A SISO decoder can provide side information to the demapper in the form of a priori symbol likelihoods.
  - BICM with Iterative Detection The demapper’s output then becomes

$$\lambda_n = \log \frac{\sum_{X_k \in S_n^{(1)}} f\left(Y \mid X_k\right) p(X_k)}{\sum_{X_k \in S_n^{(0)}} f\left(Y \mid X_k\right) p(X_k)}$$

Information Transfer Function (ten Brink 1998)



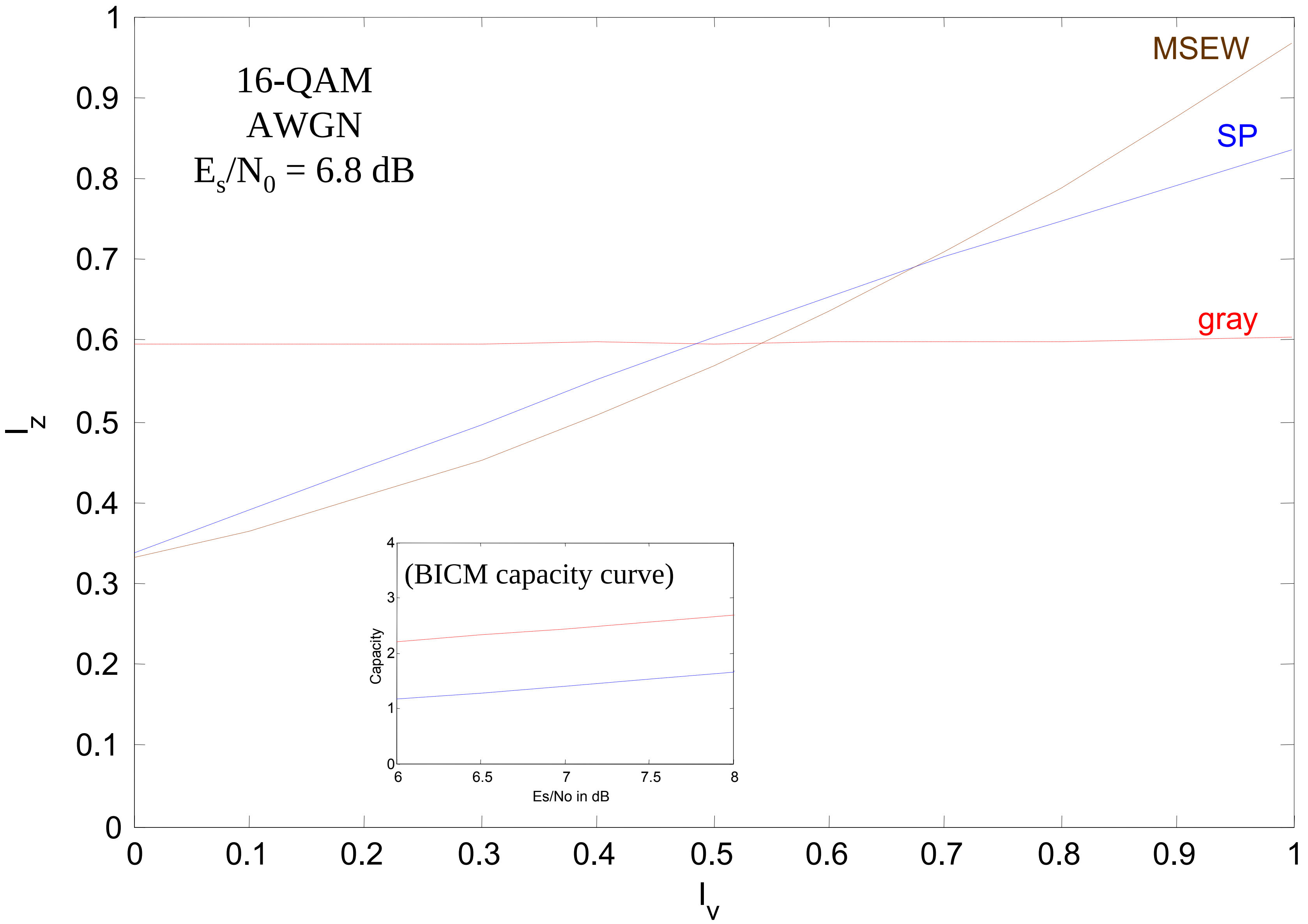
- Assume that v_n is Gaussian and that:

$$I(c_n, v_n) = I_v$$

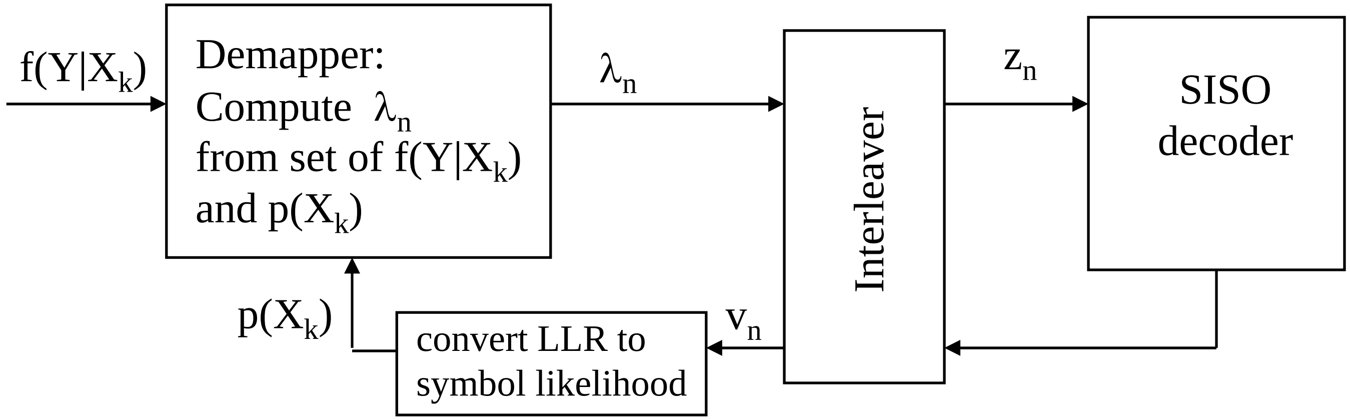
- For a particular channel SNR Es/No, randomly generate a priori LLR's with mutual information I_v.
- Measure the resulting capacity:

$$C = \sum_{n=1}^{\mu} I(c_n, \lambda_n) = \mu I_z$$

Information Transfer Function



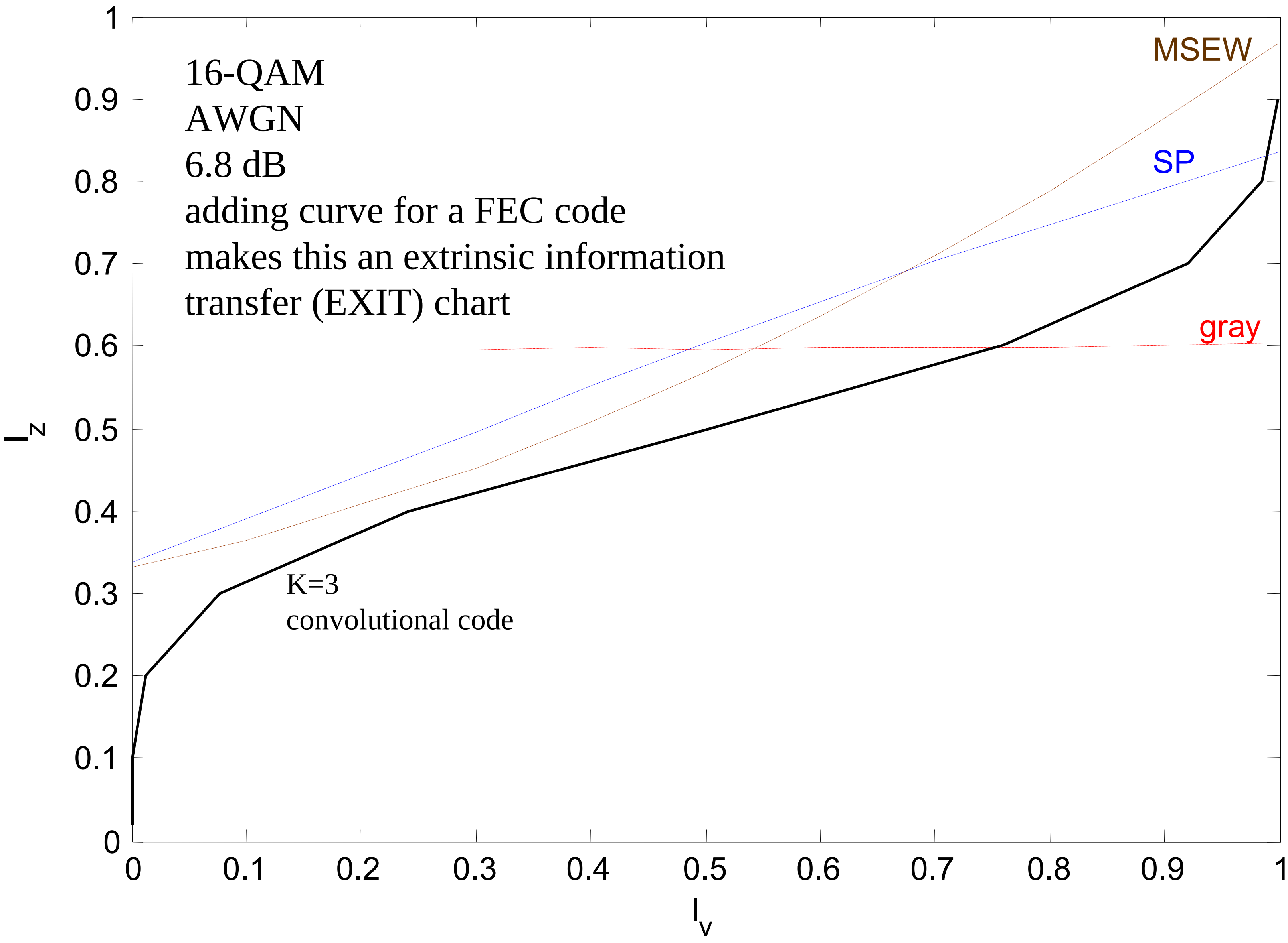
Information Transfer Function for the Decoder



- Similarly, generate a simulated Gaussian decoder input z_n with mutual information I_z.
- Measure the resulting mutual information I_v at the decoder output.

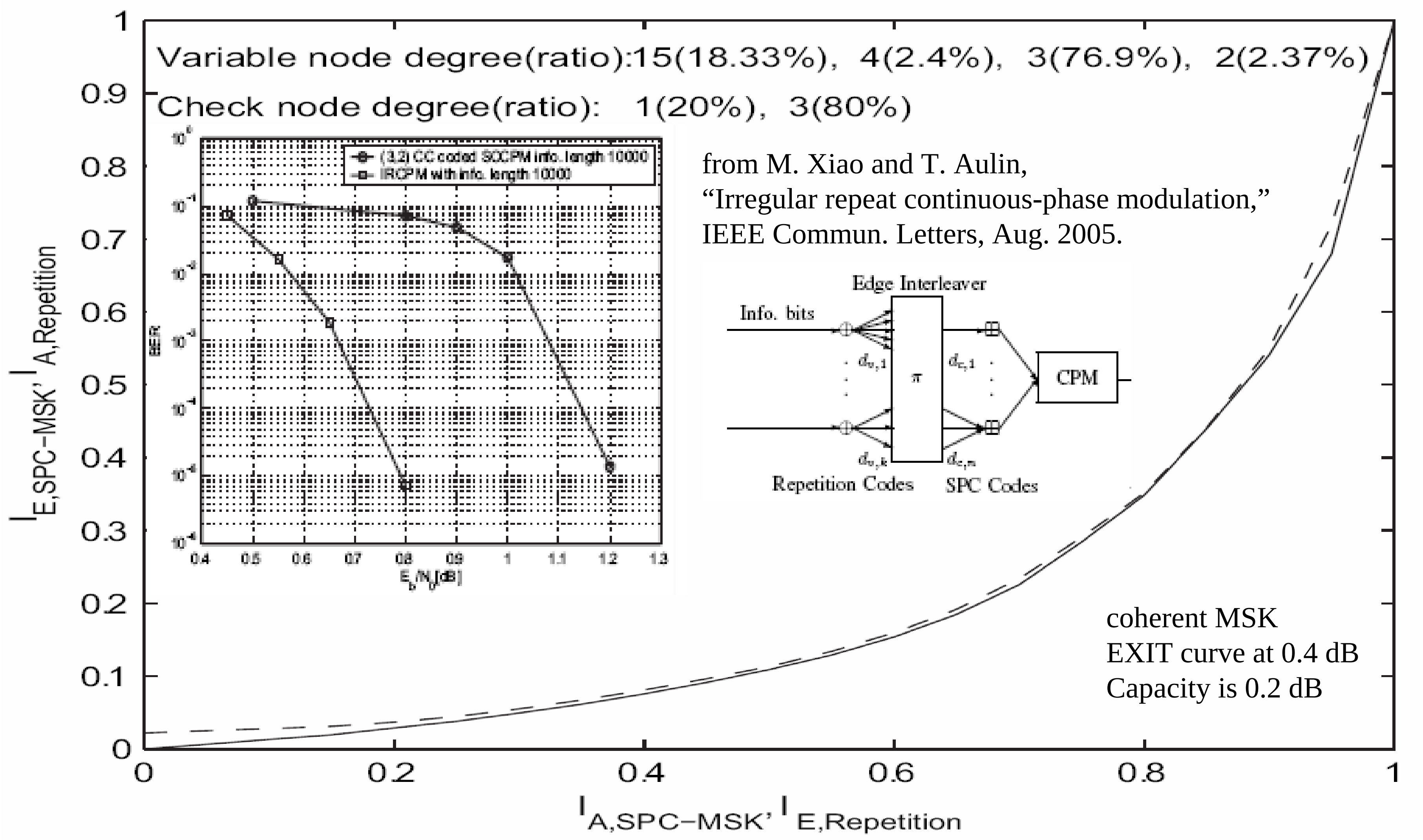
$$I_v = I(c_n, v_n)$$

EXIT Chart



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

Code Design by Matching EXIT Curves



from M. Xiao and T. Aulin,
“Irregular repeat continuous-phase modulation,”
IEEE Commun. Letters, Aug. 2005.

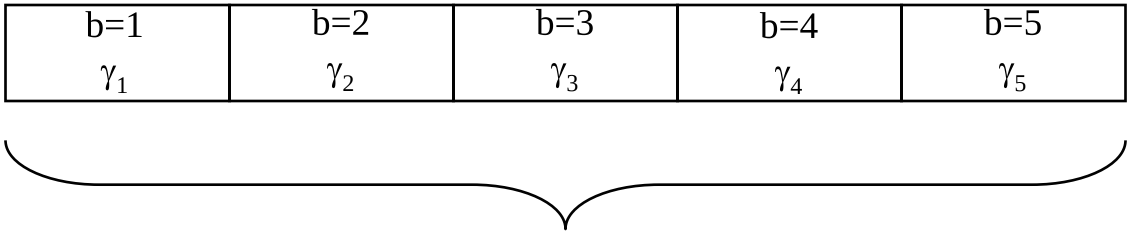
coherent MSK
EXIT curve at 0.4 dB
Capacity is 0.2 dB

Overview of Talk

- The capacity of AWGN channels
  - Modulation constrained capacity.
  - Monte Carlo methods for determining constrained capacity.
  - CPFSK: A case study on capacity-based optimization.
- Design of binary codes
  - Bit interleaved coded modulation (BICM).
  - Code design using the EXIT chart.
- Nonergodic channels.
  - Block fading: Information outage probability.
  - Hybrid-ARQ.
  - Relaying and cooperative diversity.
  - Finite length codeword effects.

Ergodicity
vs. Block Fading

- Up until now, we have assumed that the channel is ***ergodic***.
  - The observation window is large enough that the time-average converges to the statistical average.
- Often, the system might be ***nonergodic.***
- Example: ***Block fading***



b=1	b=2	b=3	b=4	b=5
γ_1	γ_2	γ_3	γ_4	γ_5

The codeword is broken into B equal length blocks
The SNR changes randomly from block-to-block
The channel is conditionally Gaussian
The instantaneous Es/No for block b is γ_b

Accumulating Mutual Information

- The SNR γ_b of block b is a random.
- Therefore, the mutual information I_b for the block is also random.
  - With a complex Gaussian input, $I_b = \log(1+\gamma_b)$
  - Otherwise the modulation constrained capacity can be used for I_b

b=1 $I_1 = \log(1+\gamma_1)$	b=2 I_2	b=3 I_3	b=4 I_4	b=5 I_5

The mutual information of each block is $I_b = \log(1+\gamma_b)$
Blocks are conditionally Gaussian
The entire codeword's mutual info is the sum of the blocks'

$$I_1^B = \sum_{b=1}^{B} I_b \qquad \text{(Code combining)}$$

Information Outage

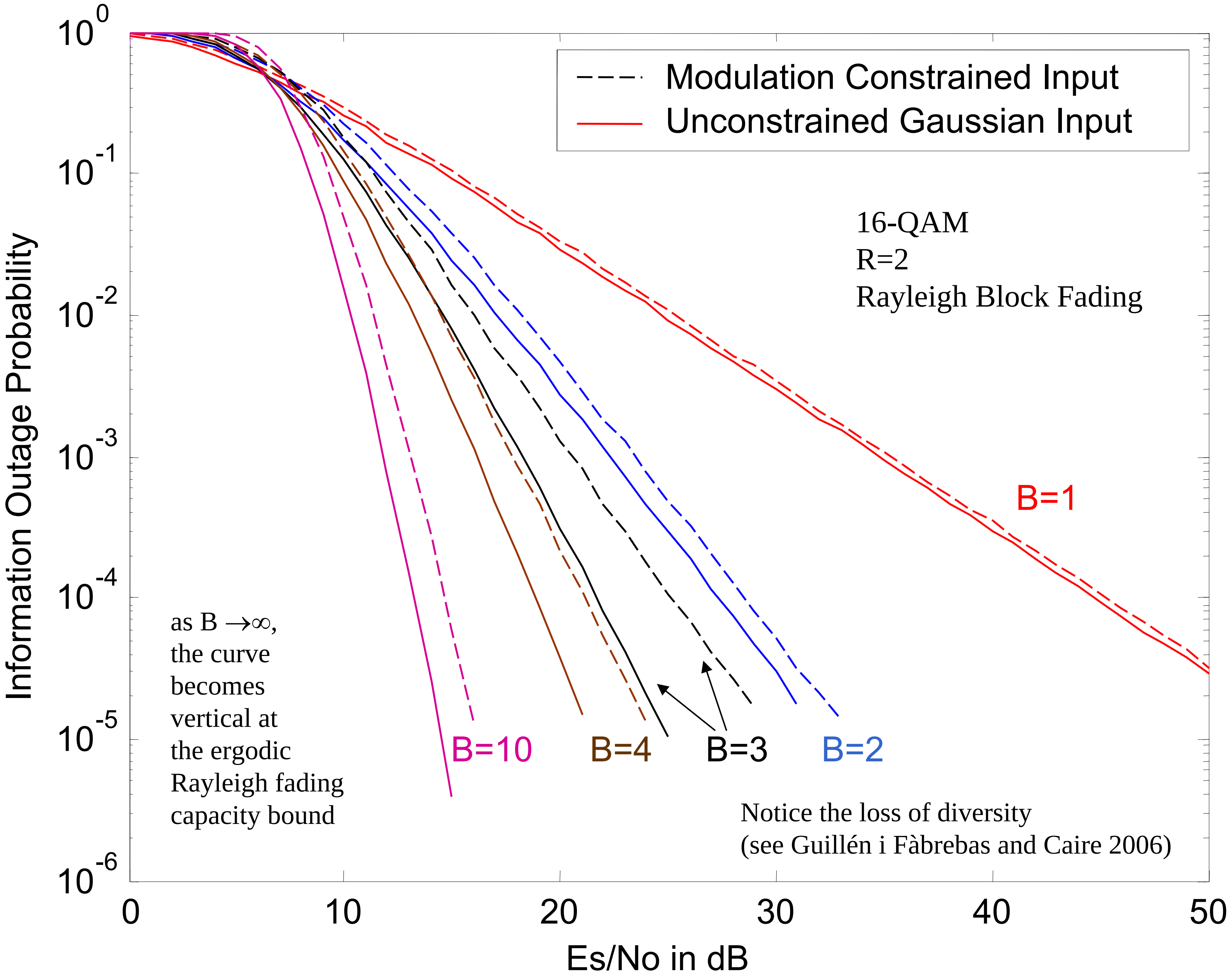
- An ***information outage*** occurs after B blocks if

$$I_1^B < R$$

  - where $R \le \log_2 M$ is the rate of the coded modulation
- An outage implies that no code can be reliable for the particular channel instantiation
- The information outage probability is

$$P_0 = P\left[I_1^B < R\right]$$

  - This is a practical bound on FER for the actual system.



16-QAM
R=2
Rayleigh Block Fading

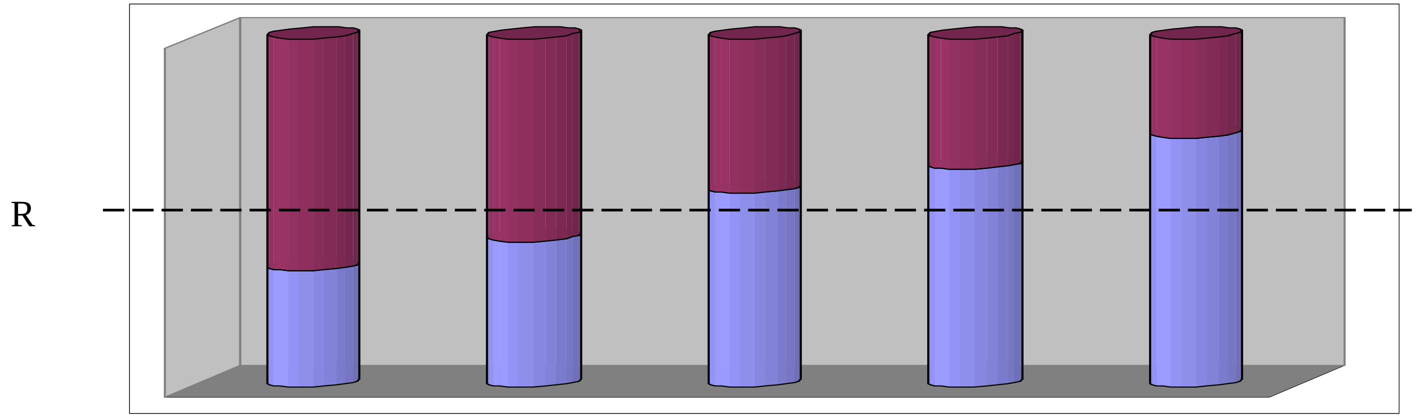
as B →∞,
the curve
becomes
vertical at
the ergodic
Rayleigh fading
capacity bound

Notice the loss of diversity
(see Guillén i Fàbrebas and Caire 2006)

Hybrid-ARQ
(Caire and Tunnineti 2001)

- Once $I_1^B > R$ the codeword can be decoded with high reliability.
- Therefore, why continue to transmit any more blocks?
- With hybrid-ARQ, the idea is to request retransmissions until $I_1^B > R$
  - With hybrid-ARQ, outages can be avoided.
  - The issue then becomes one of latency and throughput.

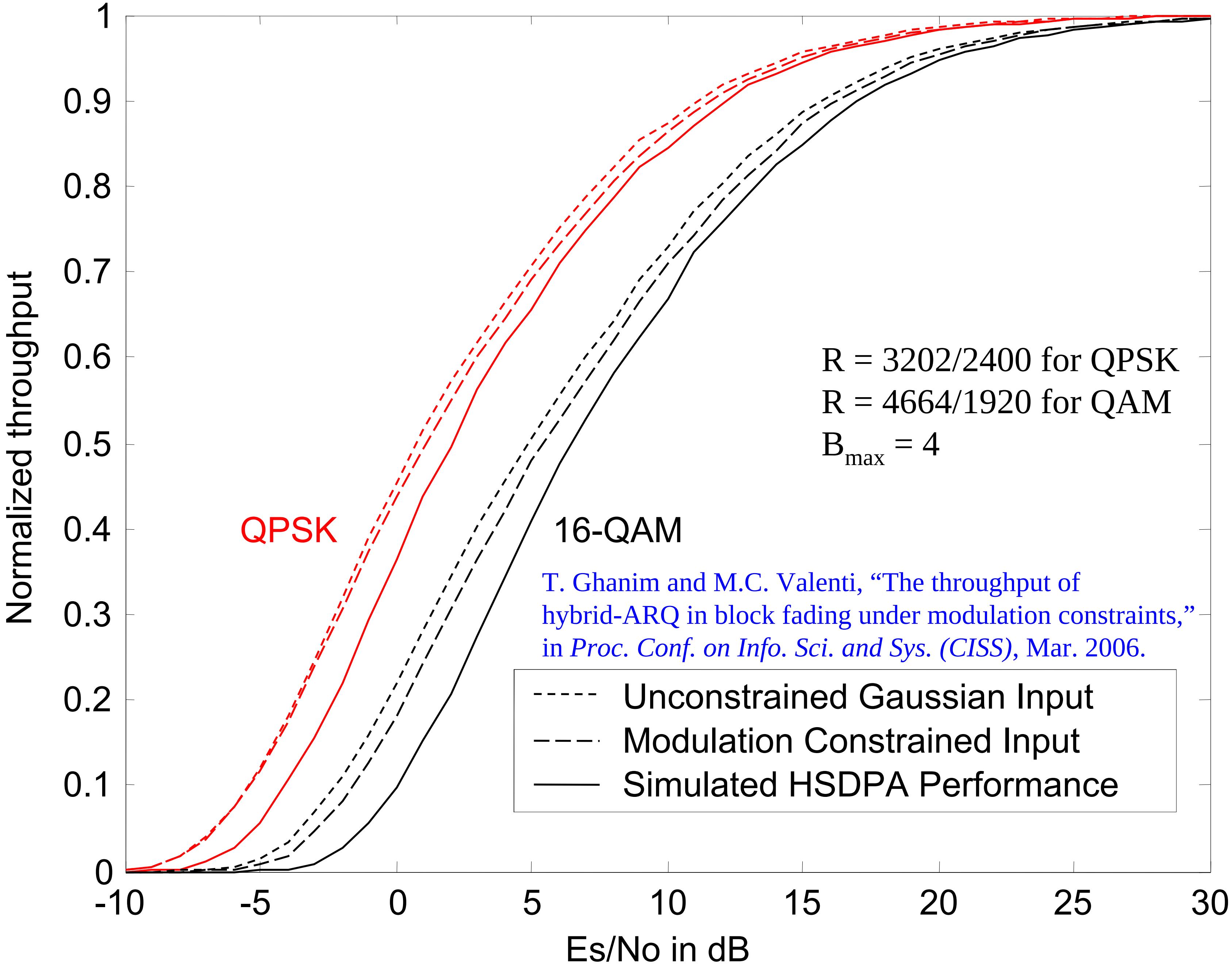
b=1 $I_1 = \log(1+\gamma_1)$	b=2 I_2	b=3 I_3	b=4 I_4	b=5 I_5



NACK NACK ACK {Wasted transmissions}

Latency and Throughput of Hybrid-ARQ

- With hybrid-ARQ B is now a random variable.
  - The average ***latency*** is proportional to E[B].
  - The average ***throughput*** is inversely proportional to E[B].
- Often, there is a practical upper limit on B
  - Rateless coding (e.g. Raptor codes) can allow $B_{max} \rightarrow \infty$
- An example
  - HSDPA: High-speed downlink packet access
  - 16-QAM and QPSK modulation
  - UMTS turbo code
  - HSET-1/2/3 from TS 25.101
  - $B_{max} = 4$

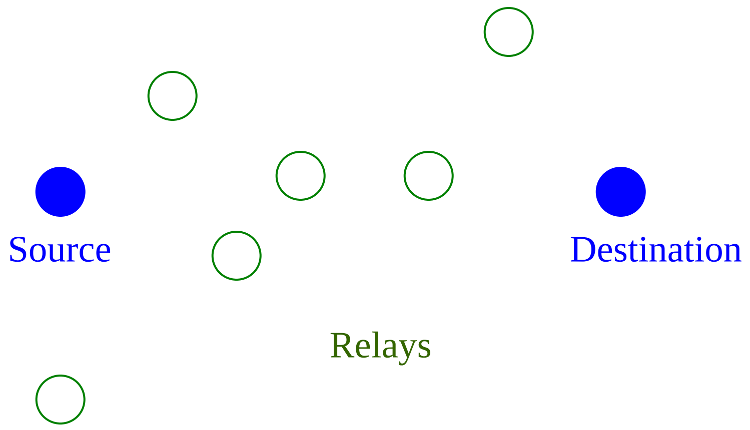


R = 3202/2400 for QPSK
R = 4664/1920 for QAM
$B_{max} = 4$

T. Ghanim and M.C. Valenti, “The throughput of hybrid-ARQ in block fading under modulation constraints,” in *Proc. Conf. on Info. Sci. and Sys. (CISS)*, Mar. 2006.

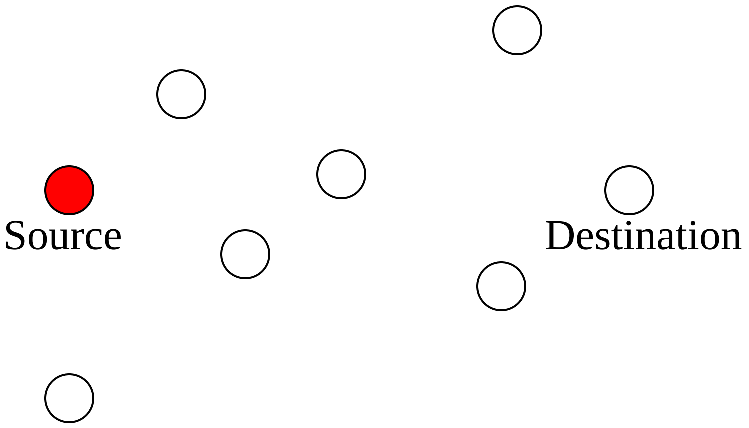
Hybrid-ARQ and Relaying

- Now consider the following ad hoc network:



- We can generalize the concept of hybrid-ARQ
  - The retransmission could be from ***any*** relay that has accumulated enough mutual information.
  - "HARBINGER" protocol
    - Hybrid ARq-Based INtercluster GEographic Relaying
    - ***B. Zhao and M. C. Valenti***. "Practical relay networks: A generalization of hybrid-ARQ," *IEEE JSAC,* Jan. 2005.

HARBINGER: Overview



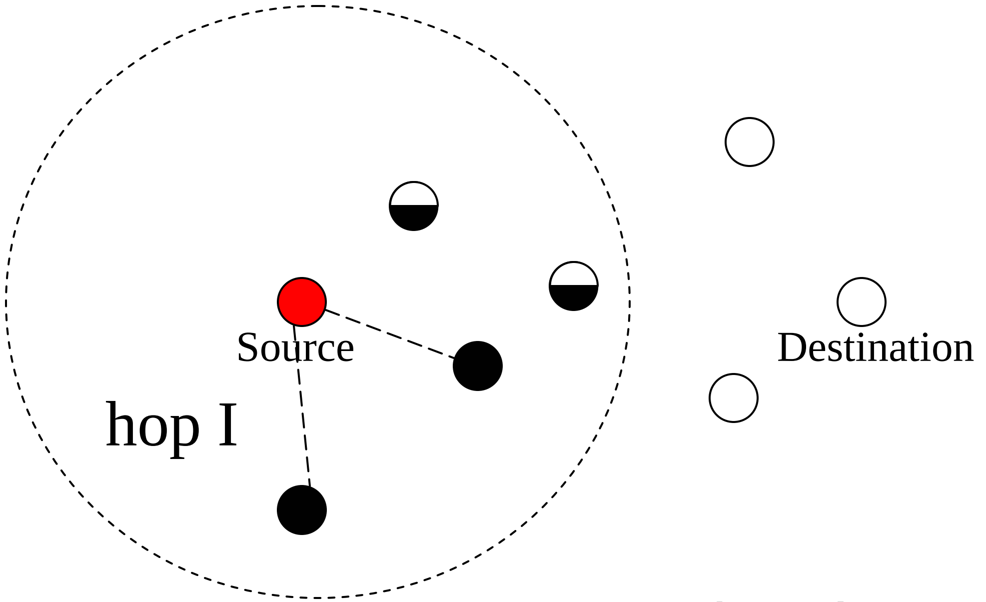
Amount of fill is proportional to the accumulated entropy.

Once node is filled, it is admitted to the decoding set D.

Any node in D can transmit.

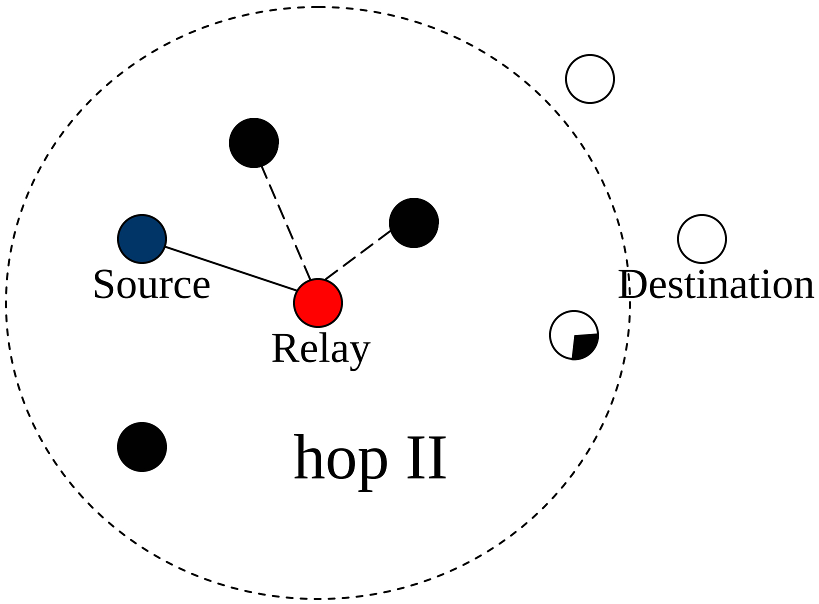
Nodes keep transmitting until Destination is in D.

HARBINGER: Initial Transmission

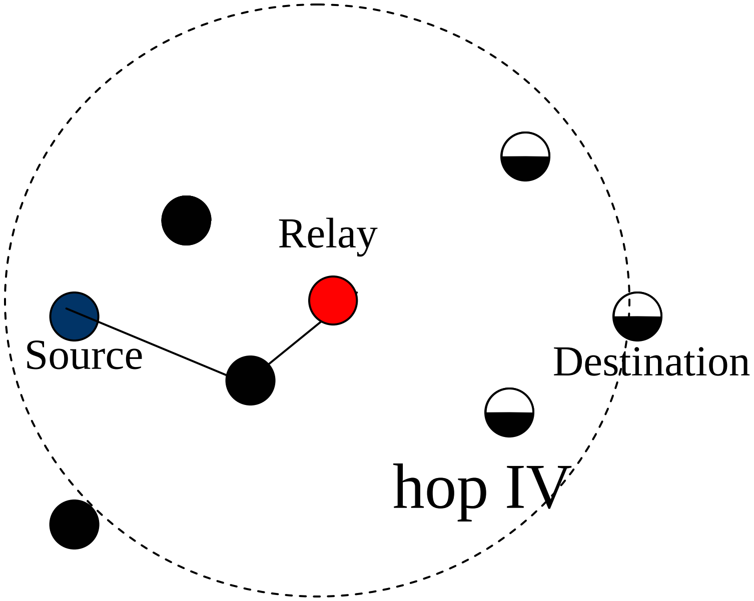


Now D contains three nodes.
Which one should transmit?
Pick the one closest to the destination.

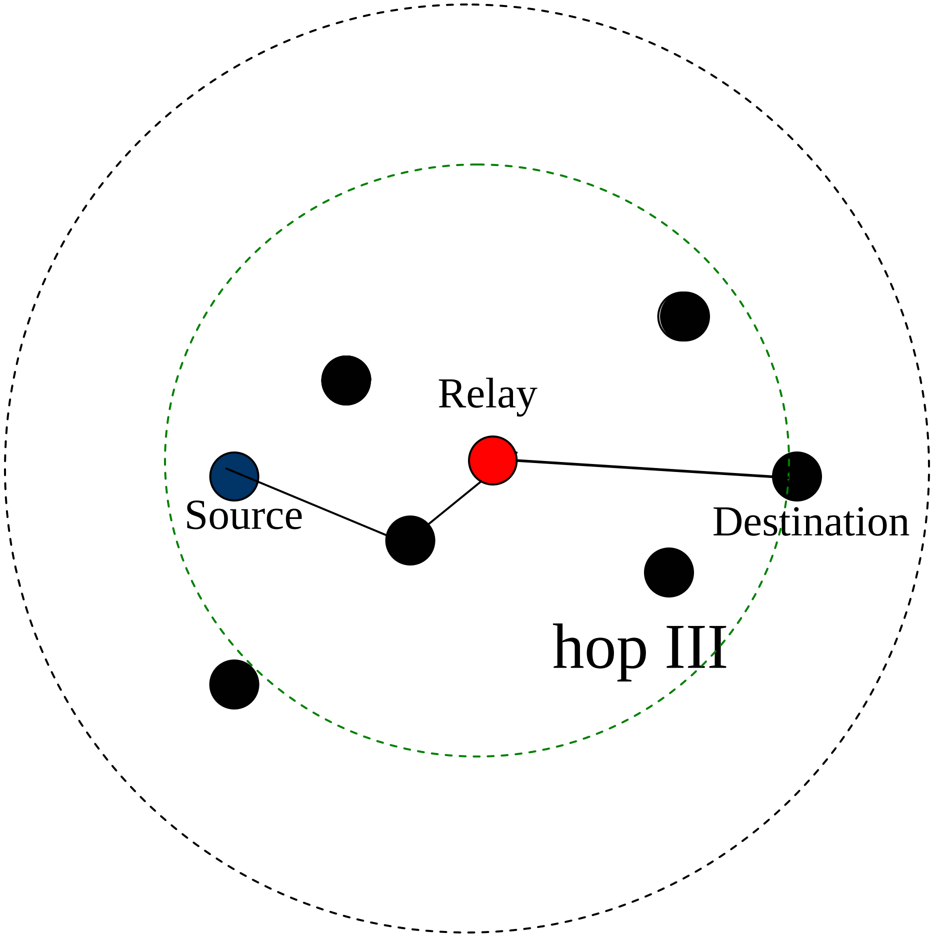
HARBINGER: 2nd Transmission



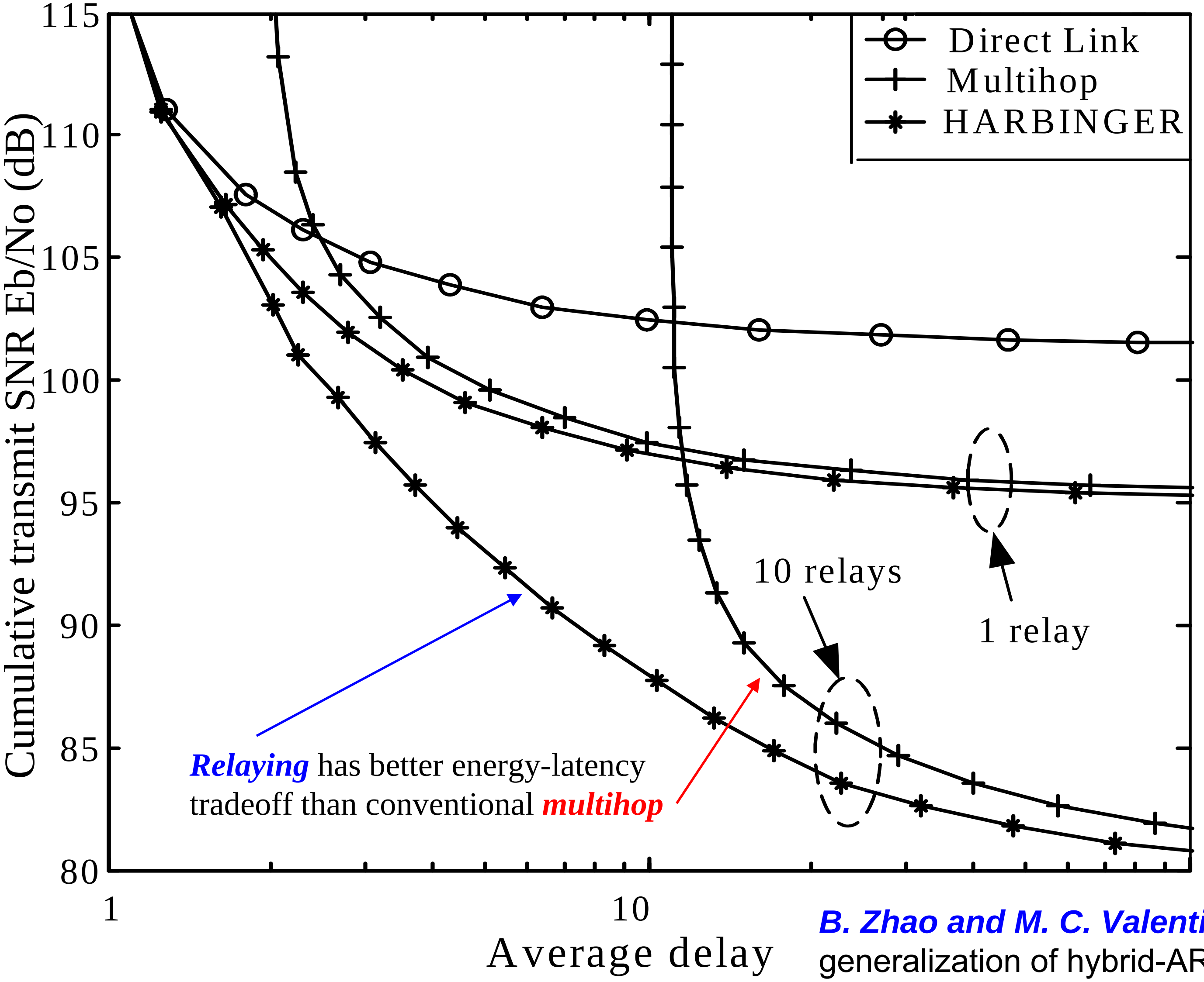
HARBINGER: 3^{rd} Transmission



HARBINGER: 4th Transmission



HARBINGER: Results



Relaying has better energy-latency tradeoff than conventional ***multihop***

Topology:

Relays on straight line
S-D separated by 10 m

Coding parameters:

Per-block rate R=1
No limit on M
Code Combining

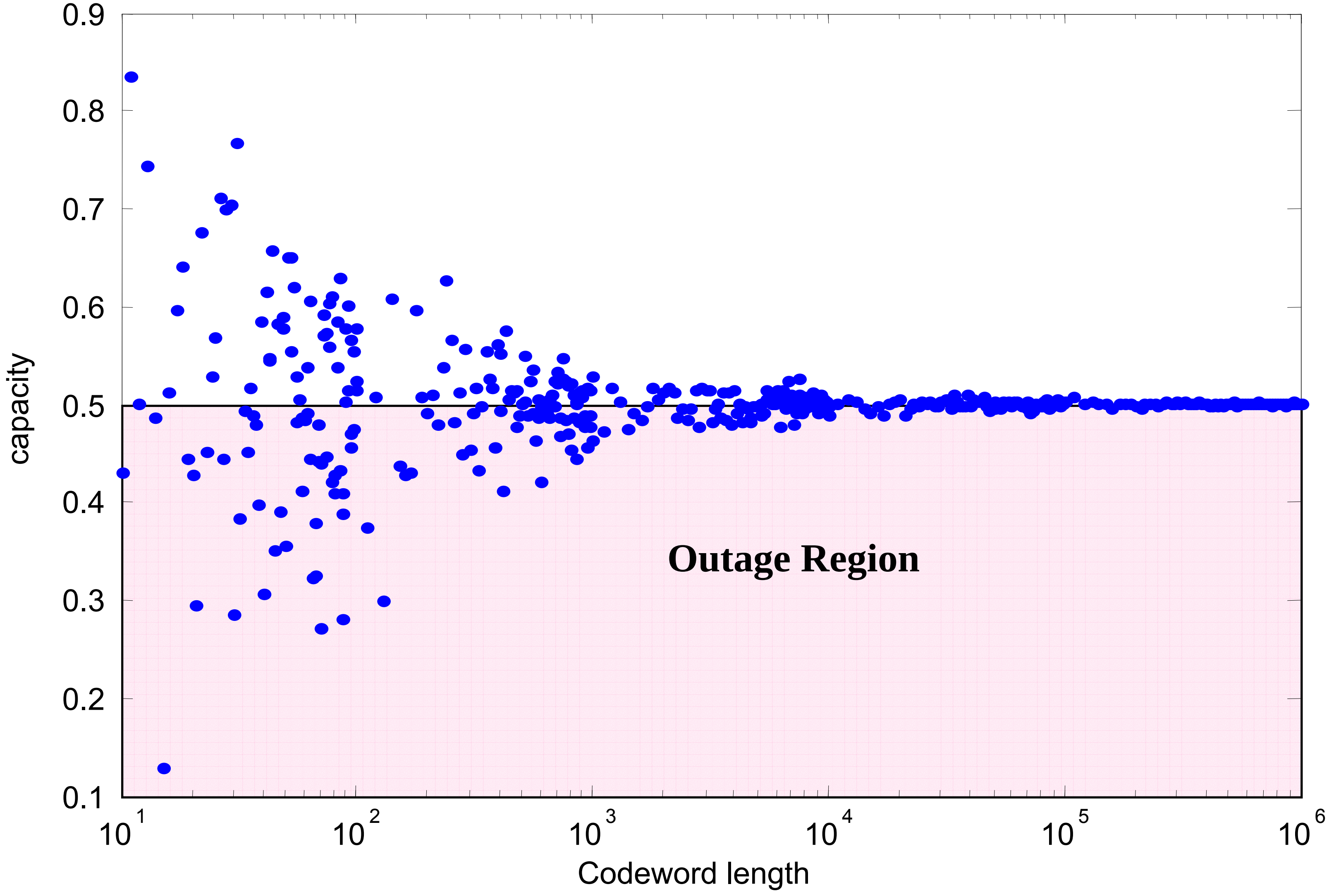
Channel parameters:

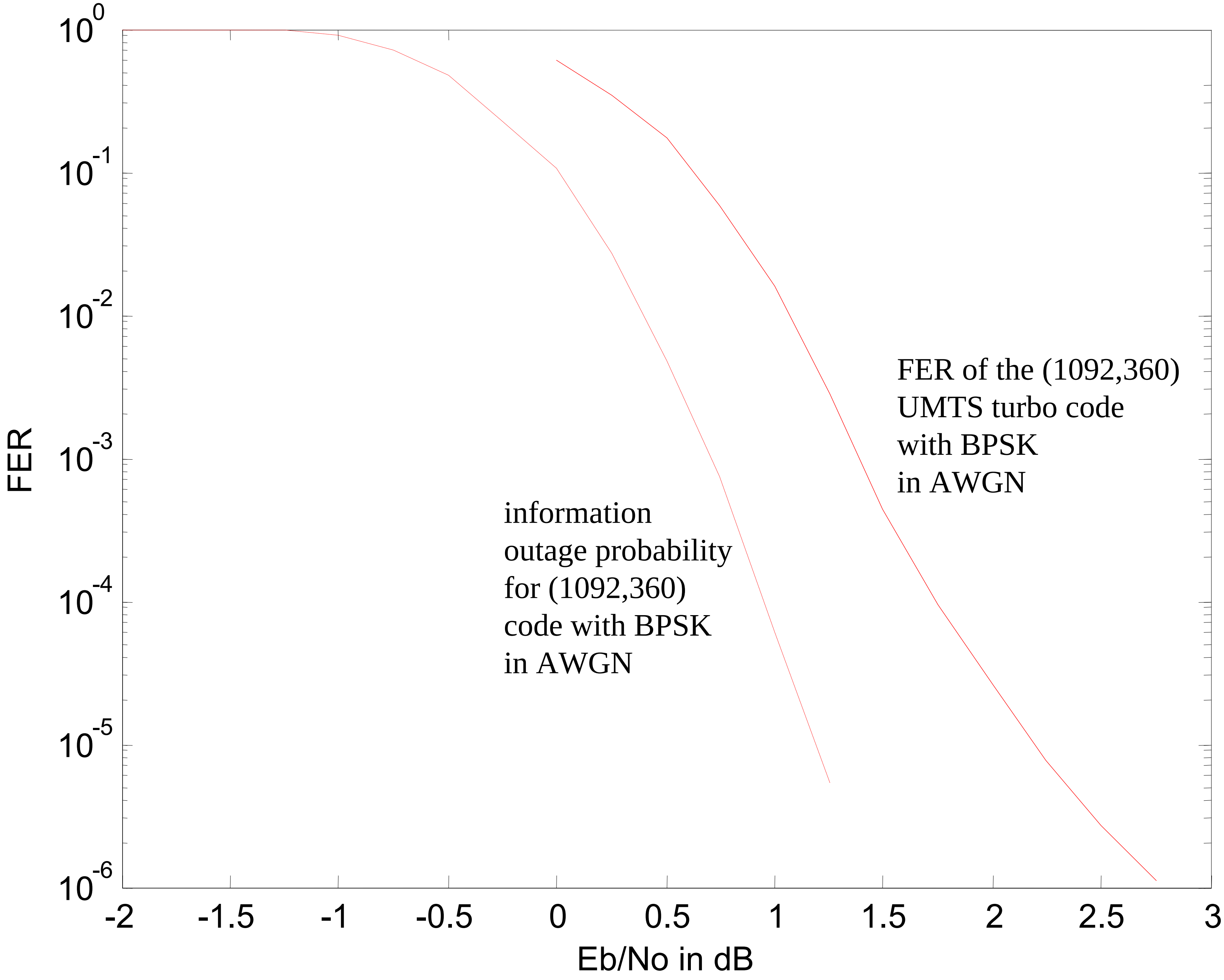
n = 3 path loss exponent
2.4 GHz
d_0 = 1 m reference dist

Unconstrained modulation

B. Zhao and M. C. Valenti. “Practical relay networks: A generalization of hybrid-ARQ,” *IEEE JSAC,* Jan. 2005.

Finite Length Codeword Effects





10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
FER
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Eb/No in dB
FER of the (1092,360)
UMTS turbo code
with BPSK
in AWGN
information
outage probability
for (1092,360)
code with BPSK
in AWGN

Conclusions

- When designing a system, first determine its capacity.
  - Only requires a slight modification of the modulation simulation.
  - Does not require the code to be simulated.
  - Allows for optimization with respect to free parameters.
- After optimizing with respect to capacity, design the code.
  - BICM with a good off-the-shelf code.
  - Optimize code with respect to the EXIT curve of the modulation.
- Information outage analysis can be used to characterize:
  - Performance in slow fading channels.
  - Delay and throughput of hybrid-ARQ retransmission protocols.
  - Performance of multihop routing and relaying protocols.
  - Finite codeword lengths.

Thank You

- For more information and publications
  - http://www.csee.wvu.edu/~mvalenti
- Free software
  - http://www.iterativesolutions.com
  - Runs in matlab but implemented mostly in C
  - Modulation constrained capacity
  - Information outage probability
  - Throughput of hybrid-ARQ
  - Standardized codes: UMTS, cdma2000, and DVB-S2