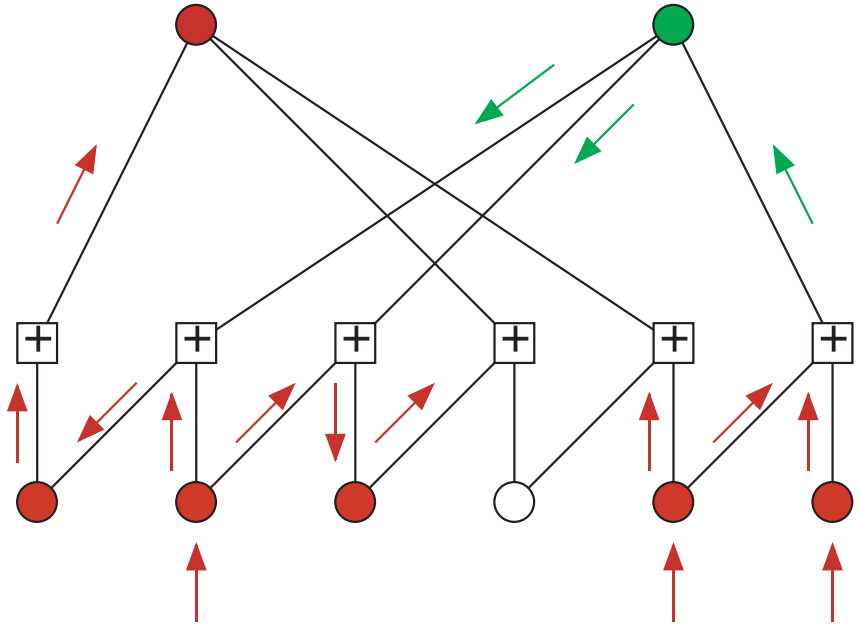
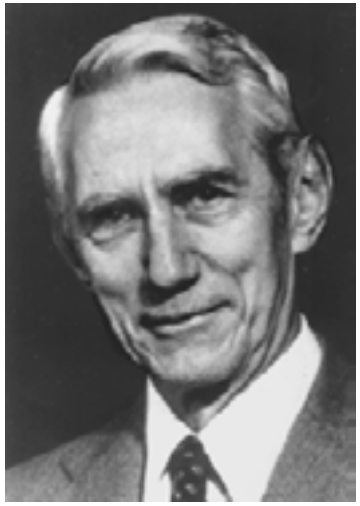
Approaching the Shannon Limit: A Progress Report

Robert J. McEliece
California Institute of Technology



AMS Annual Meeting
January 7, 2005
Atlanta, Georgia

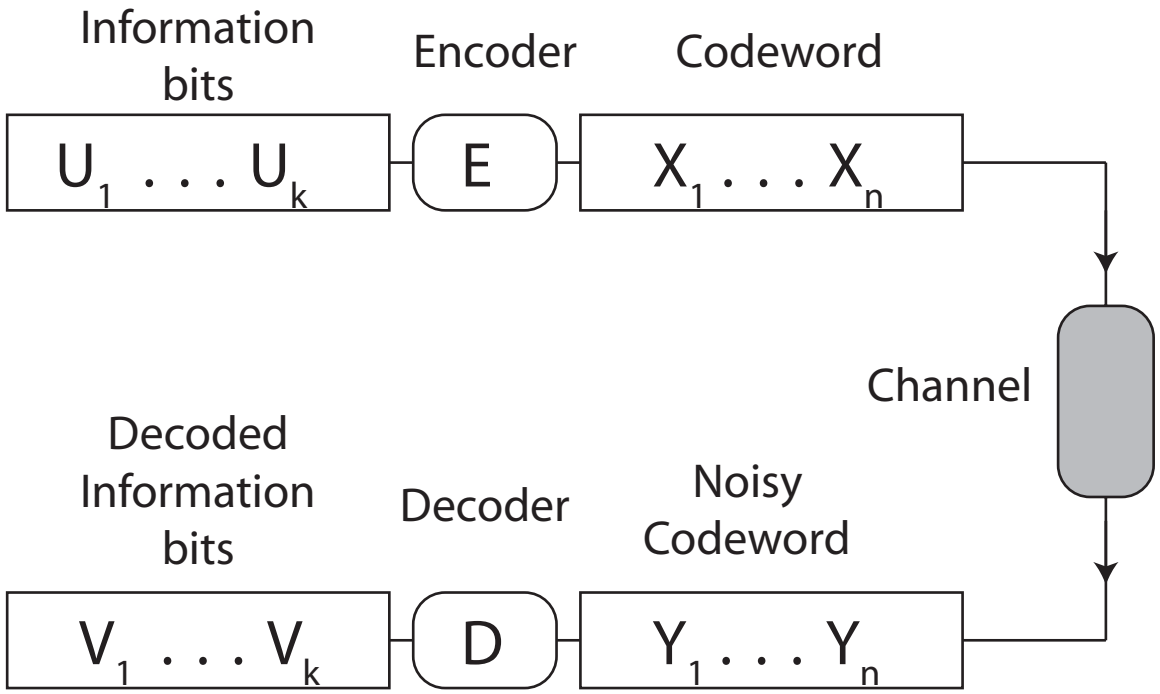
(A Tale of Two Claudes)



Shannon
1948

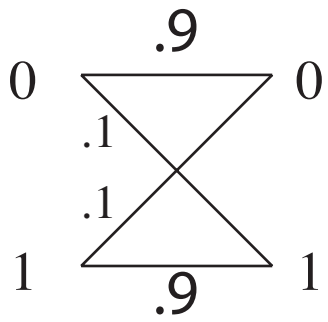
Berrou
1993

A General Communication system.

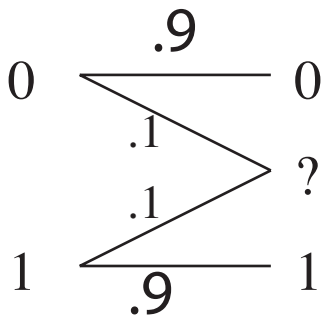


Here the rate is $R = \frac{k}{n}$ bits per channel input, and the decoded error probability is $P_b = \frac{1}{k}\sum_{i=1}^{k} \Pr\{V_i \neq U_i\}$.

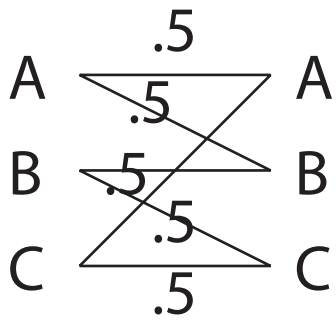
Four Discrete Memoryless Channels.



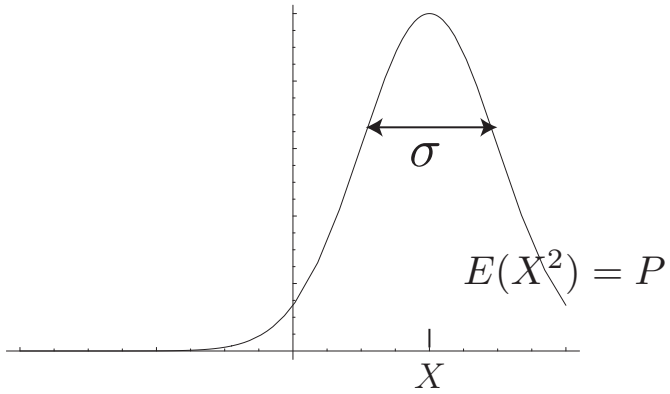
Binary
Symmetric
Channel



Binary
Erasure
Channel

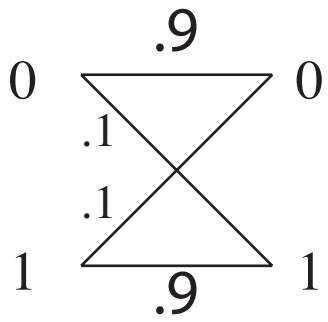


Innominate
Ternary
Channel

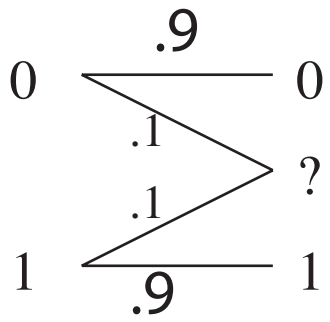


Additive White
Gaussian Noise Channel

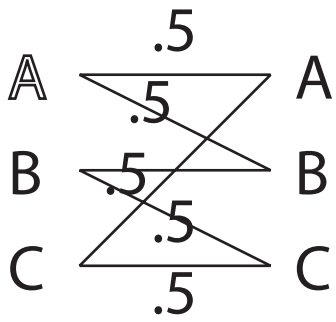
Every Channel has a Capacity C



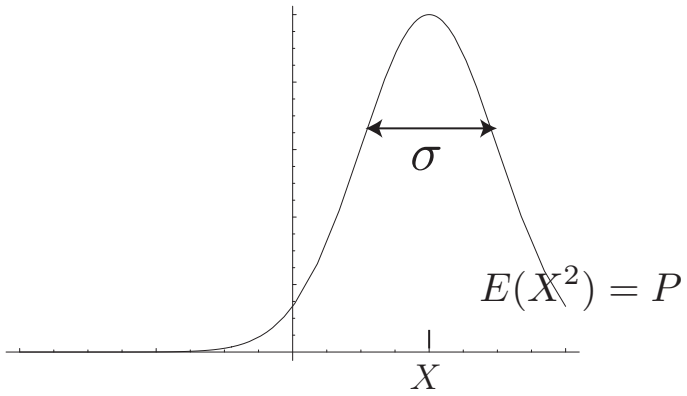
C = 0.531



C = 0.9

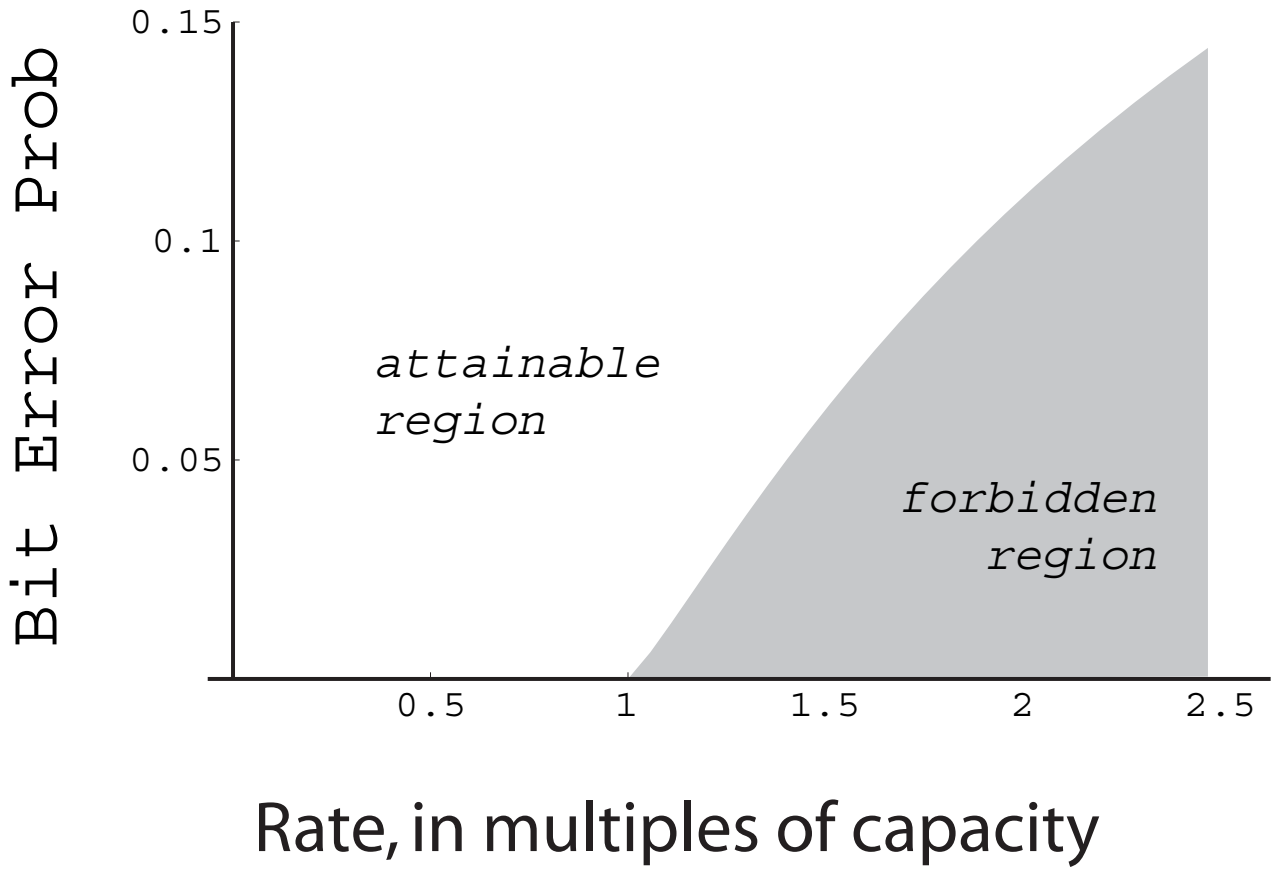


C = log 3/2

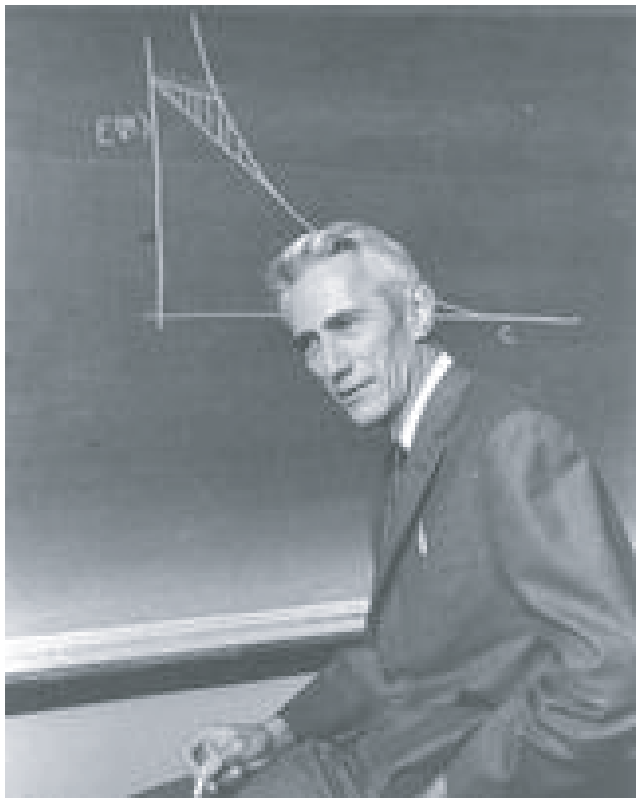


$$C = \frac{1}{2}\log(1 + \frac{P}{\sigma^2})$$

Shannon's Theorem 11 and a Bit More



The Shannon Challenge



How close can you get to C in practice?

Classic Practitioners

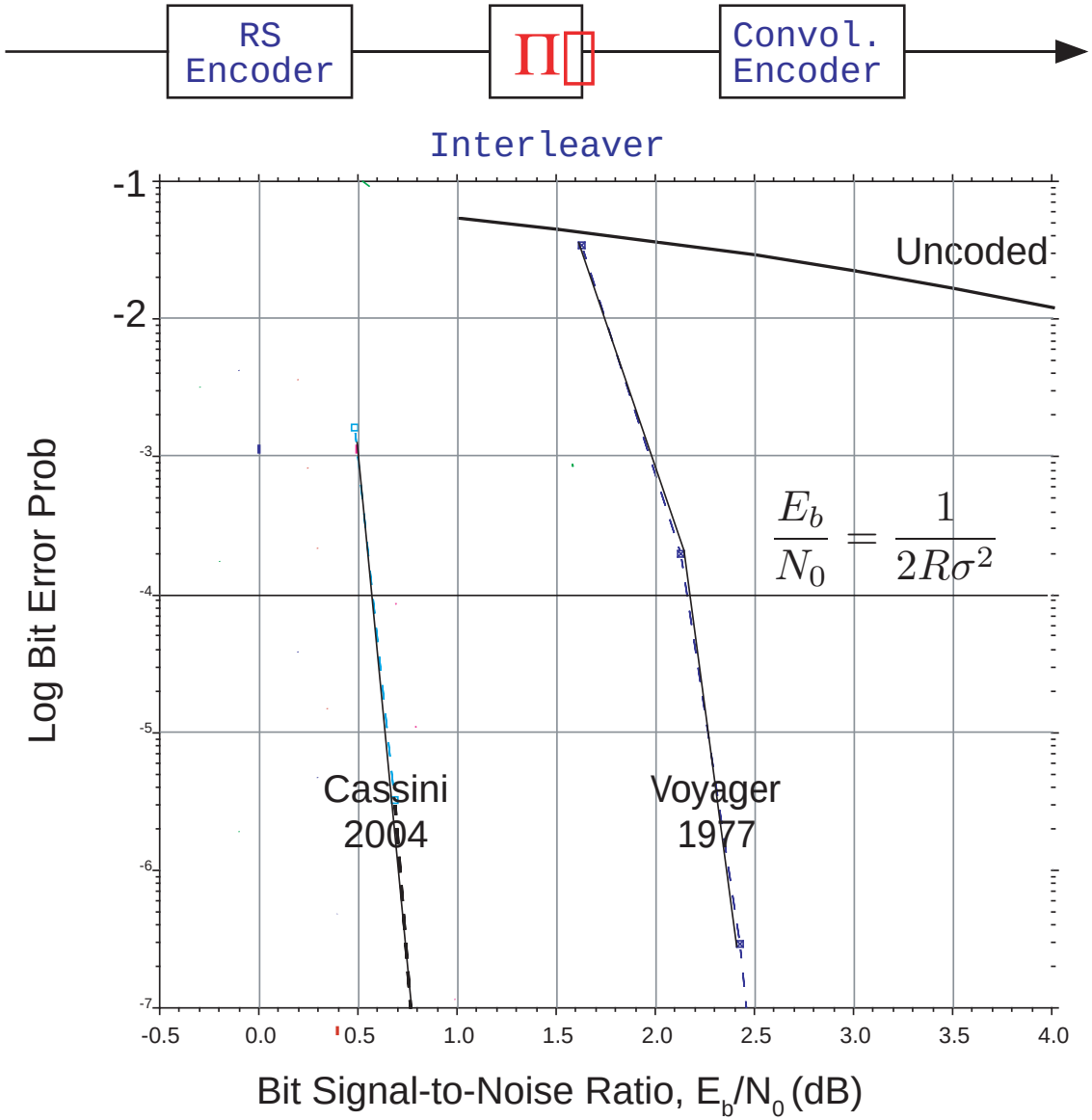
Richard Hamming
Hamming Code

Andrew Viterbi
Convolutional codes

Irving Reed
Reed-Solomon codes

Gus Solomon
Reed-Solomon codes

Pre-1993 State of the Art on the AWGN Channel



Jupiter from Cassini

May 1993: And Then Came. . .

NEAR SHANNON LIMIT ERROR - CORRECTING CODING AND DECODING : TURBO-CODES (1)

Claude Berrou, Alain Glavieux and Punya Thitimajshima

Claude Berrou, Integrated Circuits for Telecommunication Laboratory

Alain Glavieux and Punya Thitimajshima, Digital Communication Laboratory

Ecole Nationale Supérieure des Télécommunications de Bretagne, France

(1) Patents N° 9105279 (France), N° 92460011.7 (Europe), N° 07/870,483 (USA)

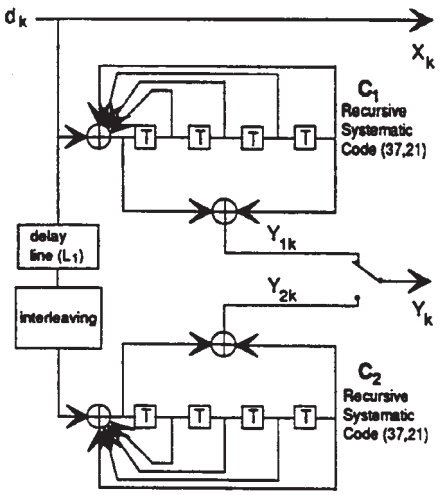


Fig. 2 Recursive Systematic codes with parallel concatenation.

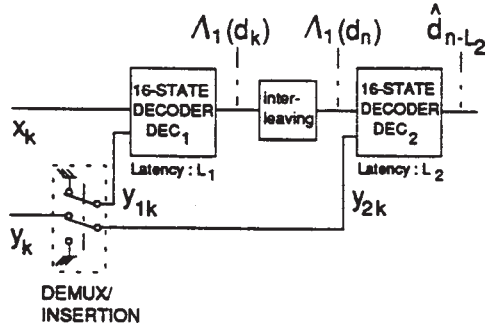
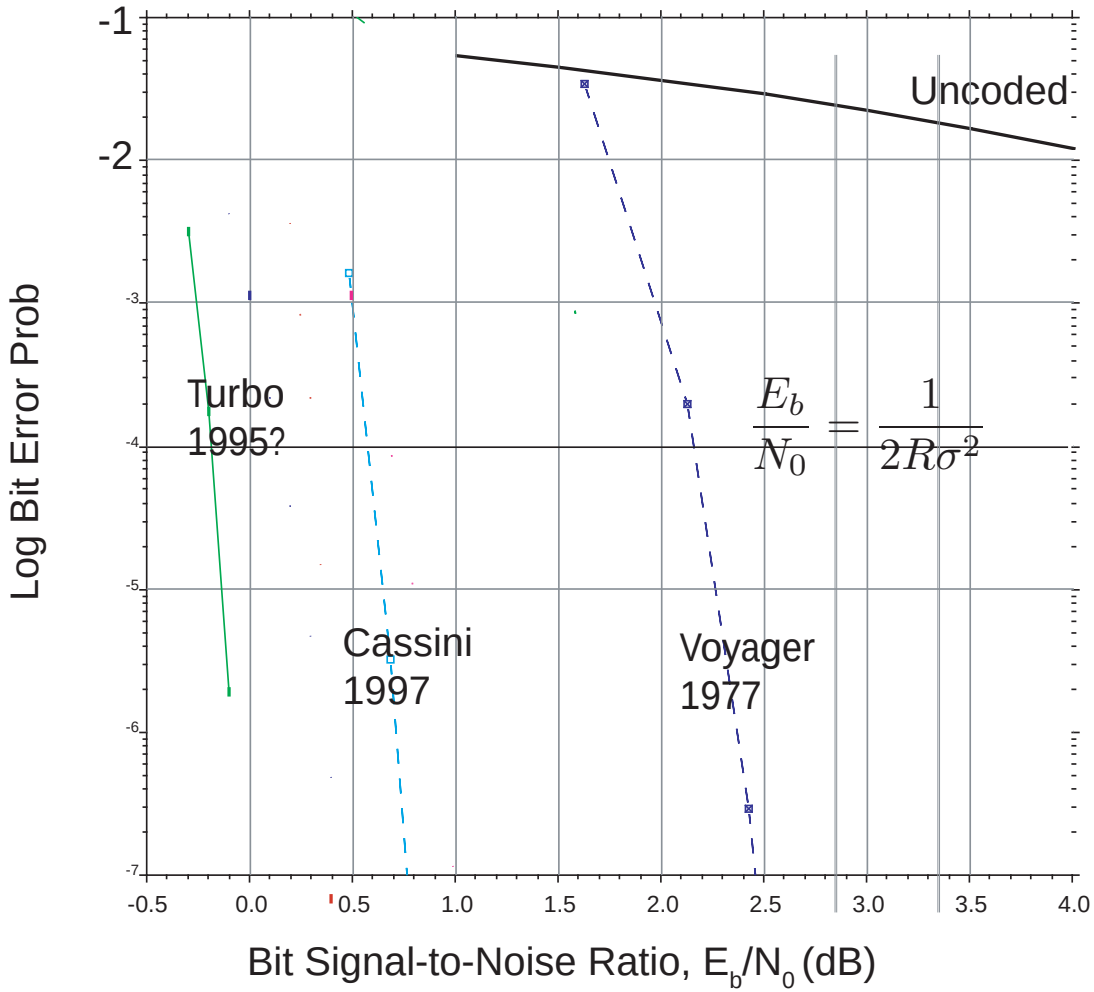


Fig. 3a Principle of the decoder according to a serial concatenation scheme.

The Turbo-Era State of the Art on the AWGN Channel



Overview

With hindsight it is clear that pre-1993 coding theory and practice was hopelessly mired in a maximum-likelihood (exact inference) paradigm. The justly celebrated **turbo decoding algoritm** is a low-complexity **iterative approximation** to maximum a posteriori probability decoding, whose performance, while demonstrably suboptimal, has nevertheless proved to be nearly optimal in an impressive array of experiments around the world.

The Original Turbo-Code.

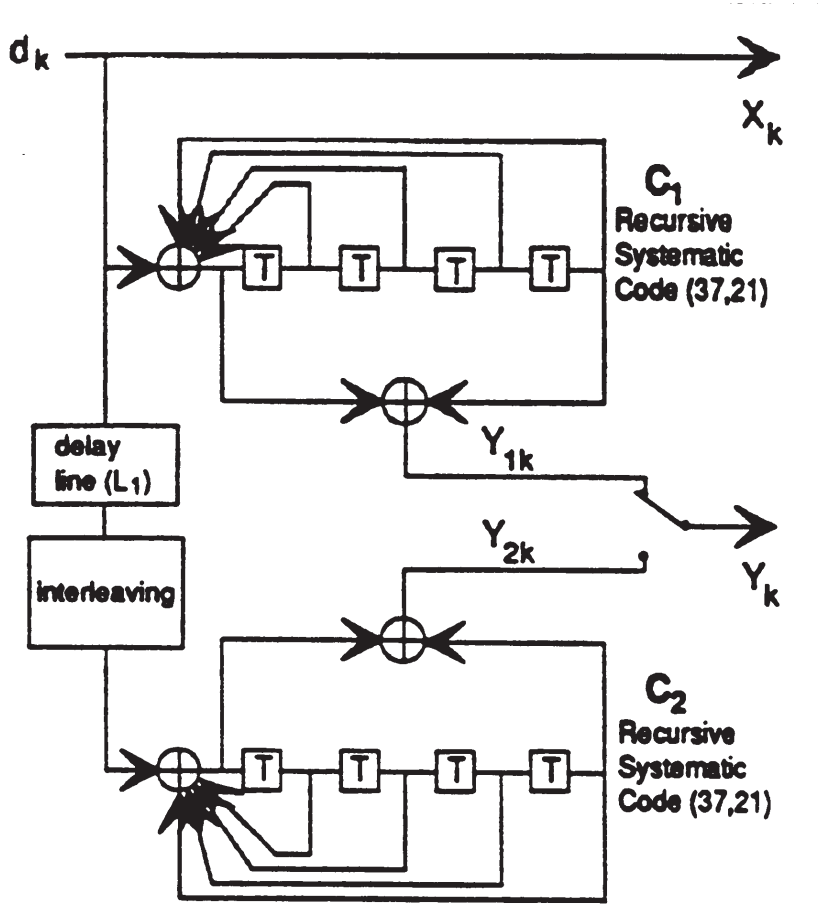
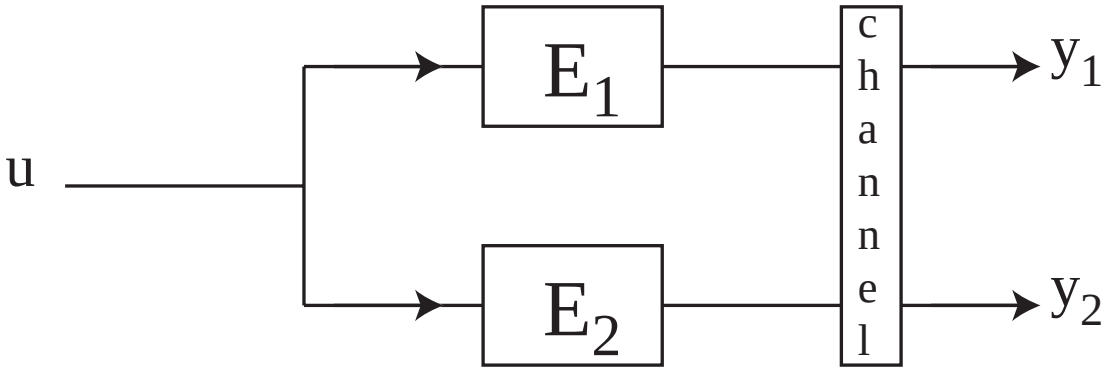


Fig. 2 Recursive Systematic codes with parallel concatenation.

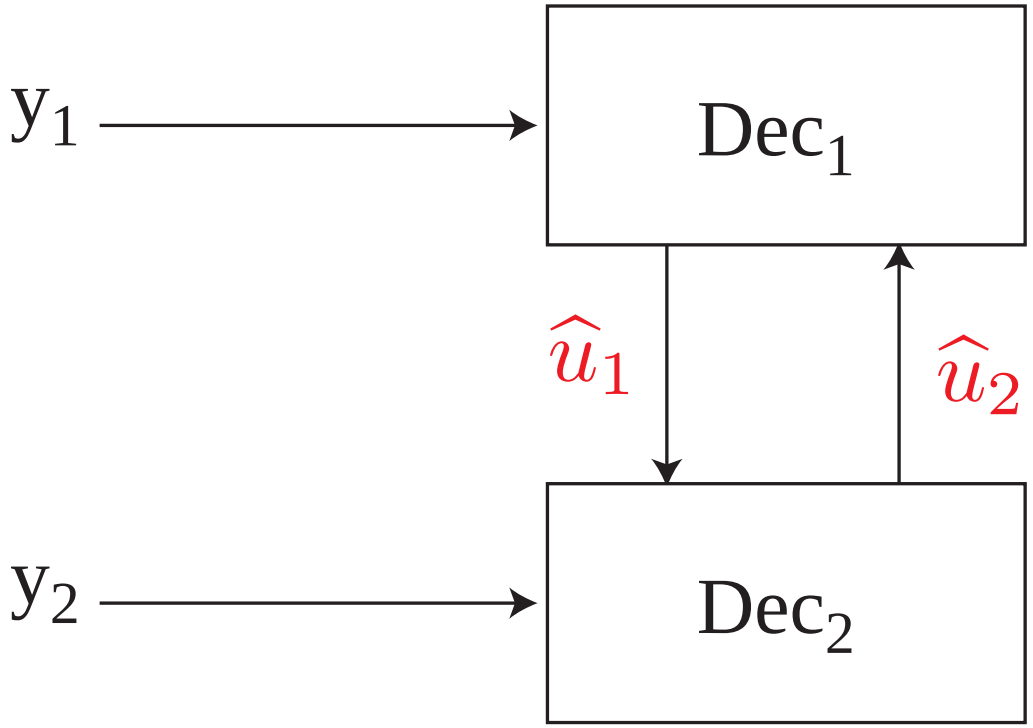
The Turbo Decoding Problem (Simplified)



Infer $\boldsymbol{u}$ from $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$, i.e., calculate

$$\Pr\{U_i = 0 | Y_1 = \boldsymbol{y}_1, Y_2 = \boldsymbol{y}_2\}, \quad (\text{for } i = 1, \ldots, k).$$

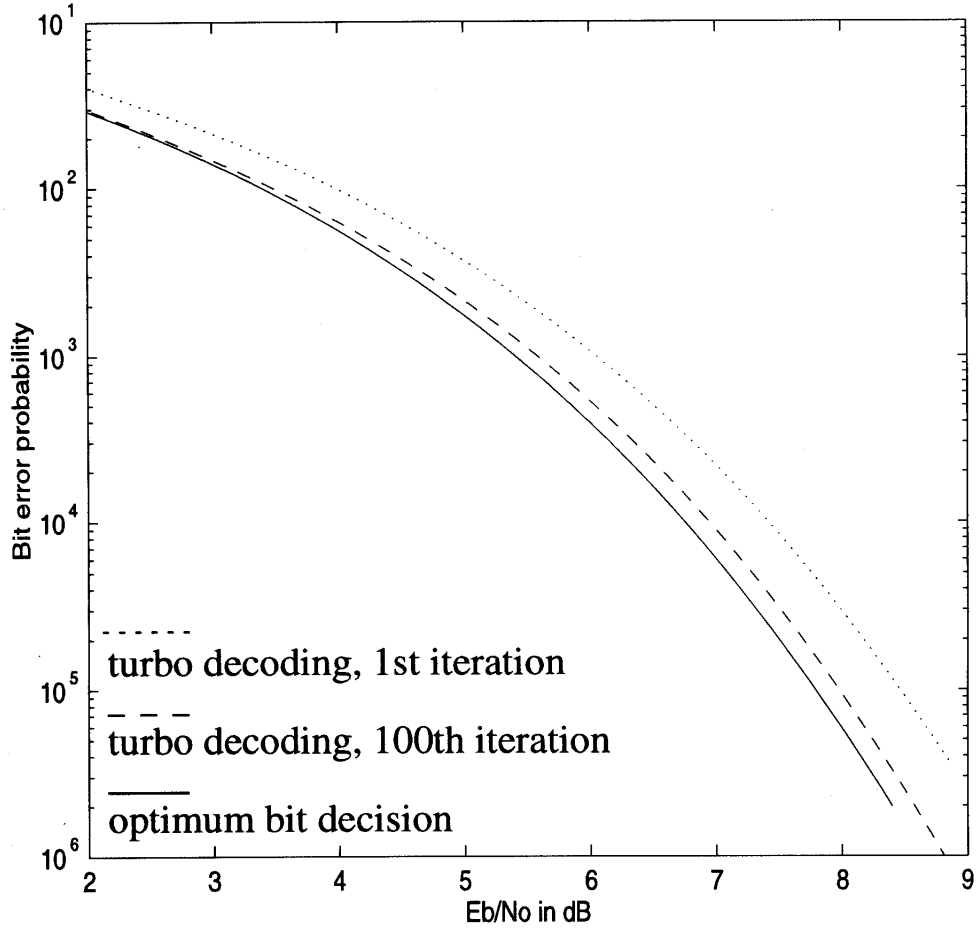
The Turbo Decoder Structure



• Dec_1 and Dec_2 communicate their results to each other, updating their estimates of $\boldsymbol{u}$ as they go, until a consensus is reached

Typical Performance of Turbo Decoding

vs. maximum a posteriori decoding




INTERMISSION

Turbo Codes are Being Replaced by LDPC Codes

LOW-DENSITY PARITY-CHECK CODES

PUBLISHED 1963 BY THE M.I.T. PRESS, CAMBRIDGE, MASSACHUSETTS

ROBERT G. GALLAGER

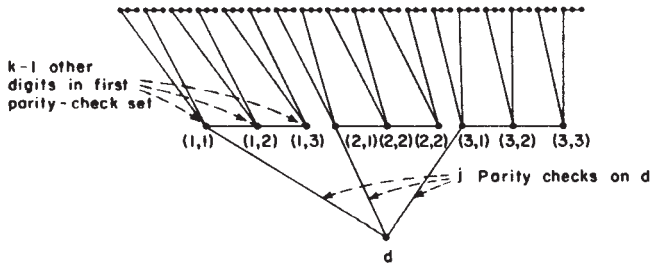


Figure 4.1. Parity-check set tree.

Assume now that both digit d and several of the digits in the first tier are transmission errors. Then on the first decoding attempt, the error-free digits in the second tier and their parity-check constraints will allow correction of the errors in the first tier. This in turn will allow correction of digit d on the second decoding attempt. Thus digits and parity-check equations can aid in decoding a digit seemingly unconnected with them. The probabilistic decoding scheme to be described next utilizes these extra digits and extra parity-check equations more systematically.

4.2 Probabilistic Decoding

Assume that the code words from an (n, j, k) code are used with equal probability on an arbitrary binary-input channel. For any digit d, using the notation of Figure 4.1, an iteration process will be derived that on the m^{th} iteration computes the probability that the transmitted digit in position d is a 1 conditional on the received symbols out to and including the m^{th} tier. For the first iteration, we can consider digit d and the digits in the first tier to form a subcode in which all sets of these digits that satisfy the j parity-check equations in the tree have equal probability of transmission.*

Turbo Codes are Being Replaced by LDPC Codes

Gallager

MacKay

This is certainly a startling development, since LDPC codes were invented by Robert Gallager in 1962! However, LDPC codes were largely forgotten until their rediscovery by David Mackay in 1998, who not only rediscovered them but used powerful modern computers (which were not available to Gallager) to simulate their performance and thereby demonstrate their astonishing power.

Another Landmark Paper (1997)

(LDPC Codes for the Binary Erasure Channel)

Practical Loss-Resilient Codes

Michael G. Luby* Michael Mitzenmacher† M. Amin Shokrollahi‡

Daniel A. Spielman§ Volker Stemann¶

Abstract

We present randomized constructions of linear-time encodable and decodable codes that can transmit over lossy channels at rates extremely close to capacity. The encoding and decoding algorithms for these codes have fast and simple software implementations. Partial implementations of our algorithms are faster by orders of magnitude than the best software implementations of any previous algorithm for this problem. We expect these codes will be extremely useful for applications such as real-time audio and video transmission over the Internet, where lossy channels are common and fast decoding is a requirement.

Despite the simplicity of the algorithms, their design and analysis are mathematically intricate. The design requires the ficients determined by the graph structure. Based on these polynomials, we design a graph structure that guarantees successful decoding with high probability.

1 Introduction

Studies show that the Internet exhibits packet loss, and the measurements in [10] show that the situation has become worse over the past few years. A standard solution to this problem is to request retransmission of data that is not received. When some of this retransmission is lost, another request is made, and so on. In some applications, this introduces technical difficulties. For real-time transmission this solution can lead to unacceptable delays caused by several rounds of communication between sender and receiver. For

The Parity-Check Matrix

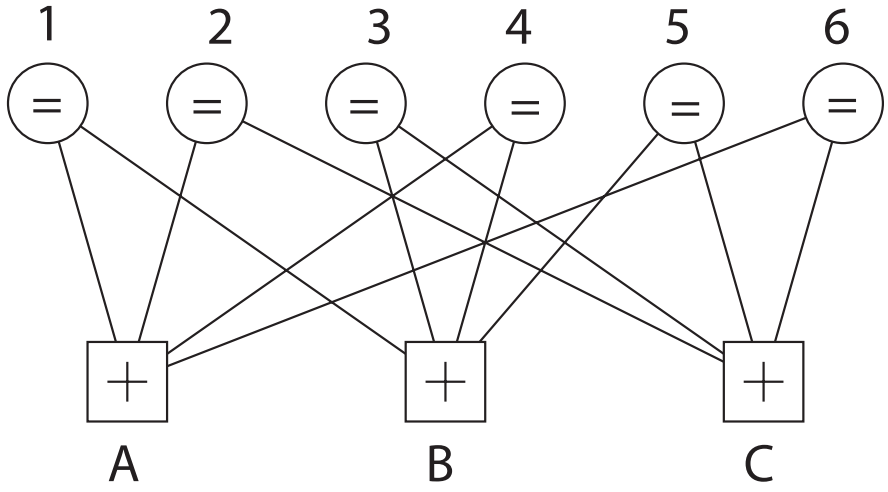
$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

The valid codewords are required to satisify the parity-checks: $H\boldsymbol{x}^T = 0$.

The Tanner Graph

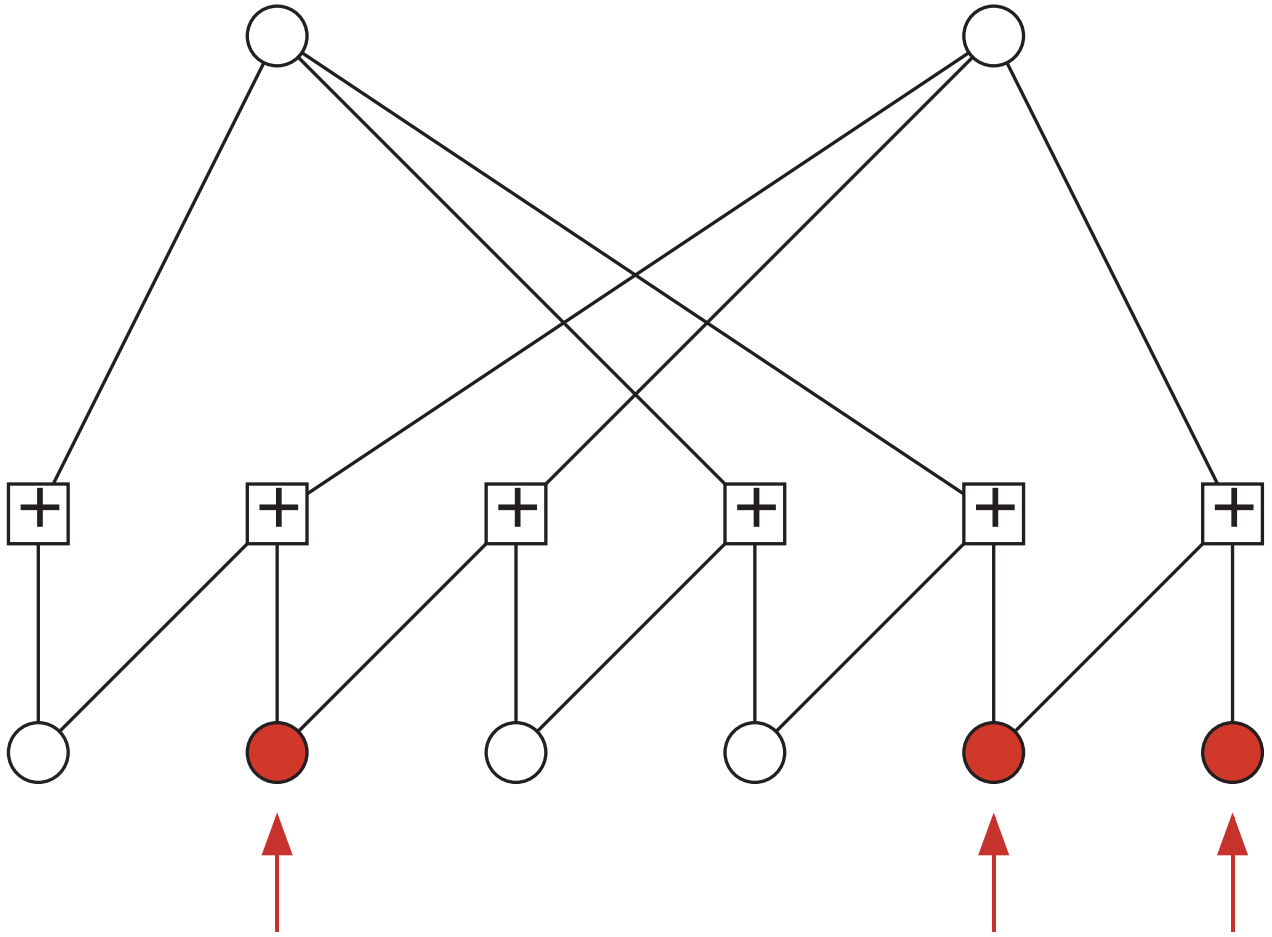
$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array}.$$

Tanner



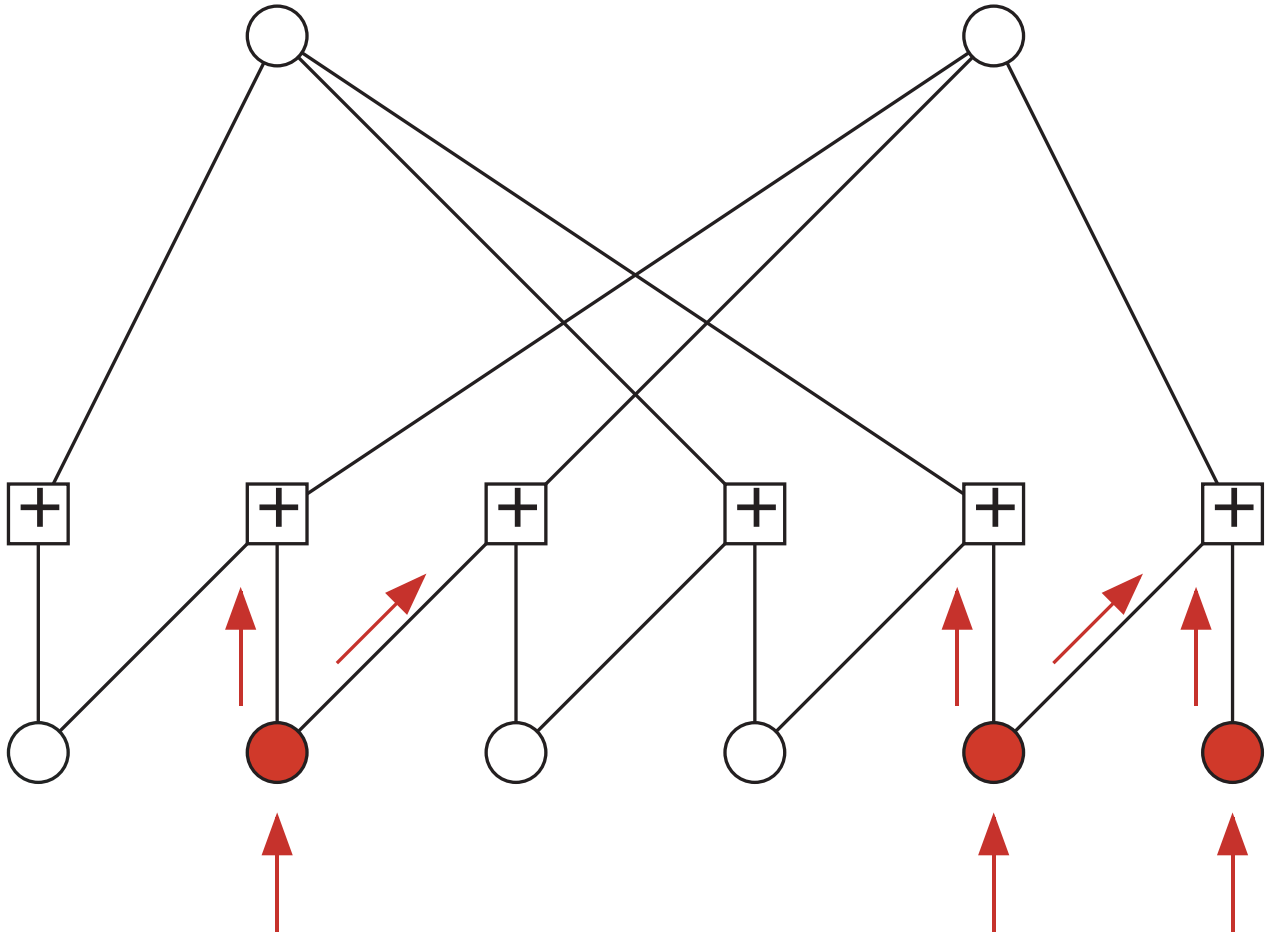
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



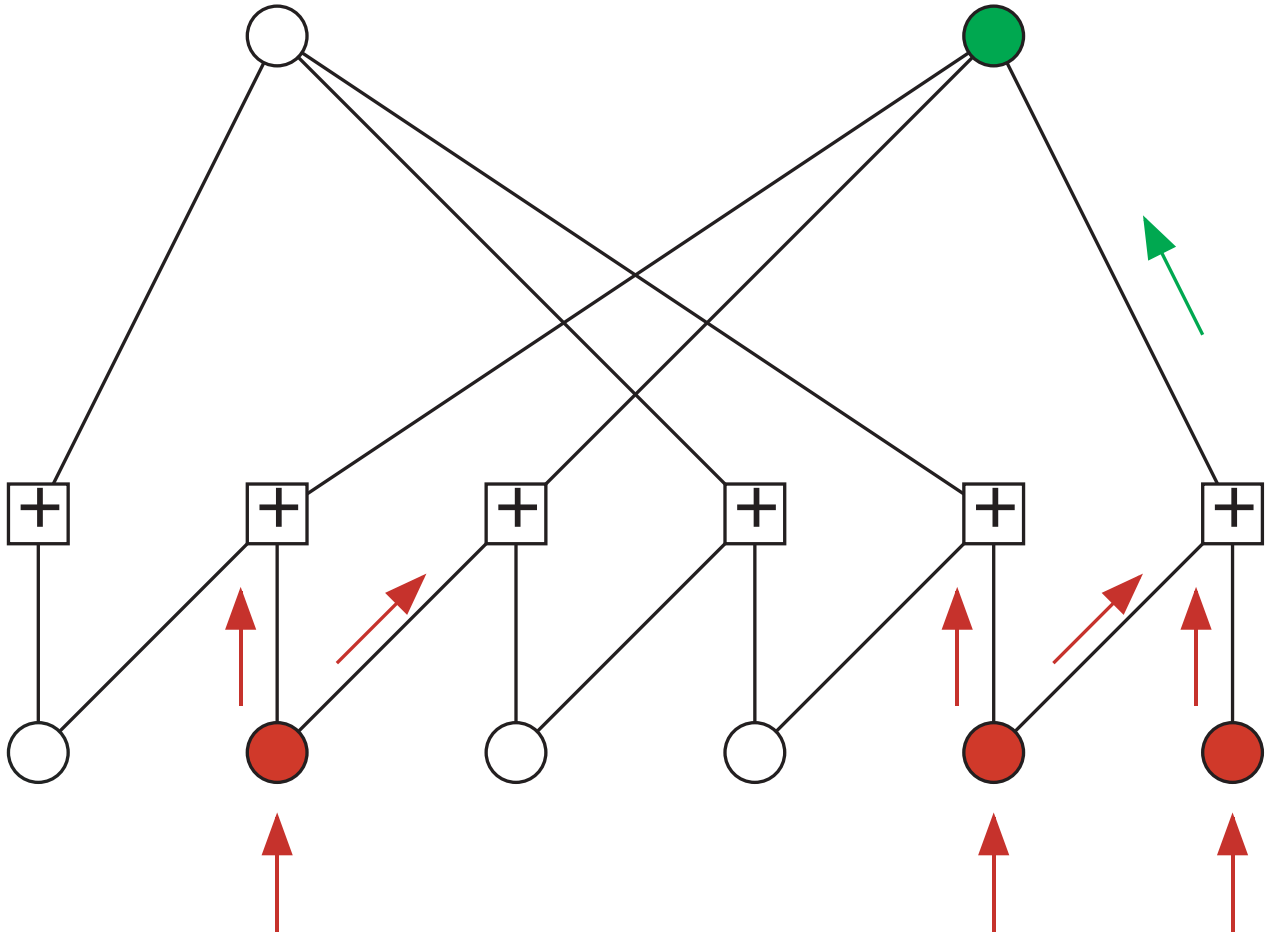
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



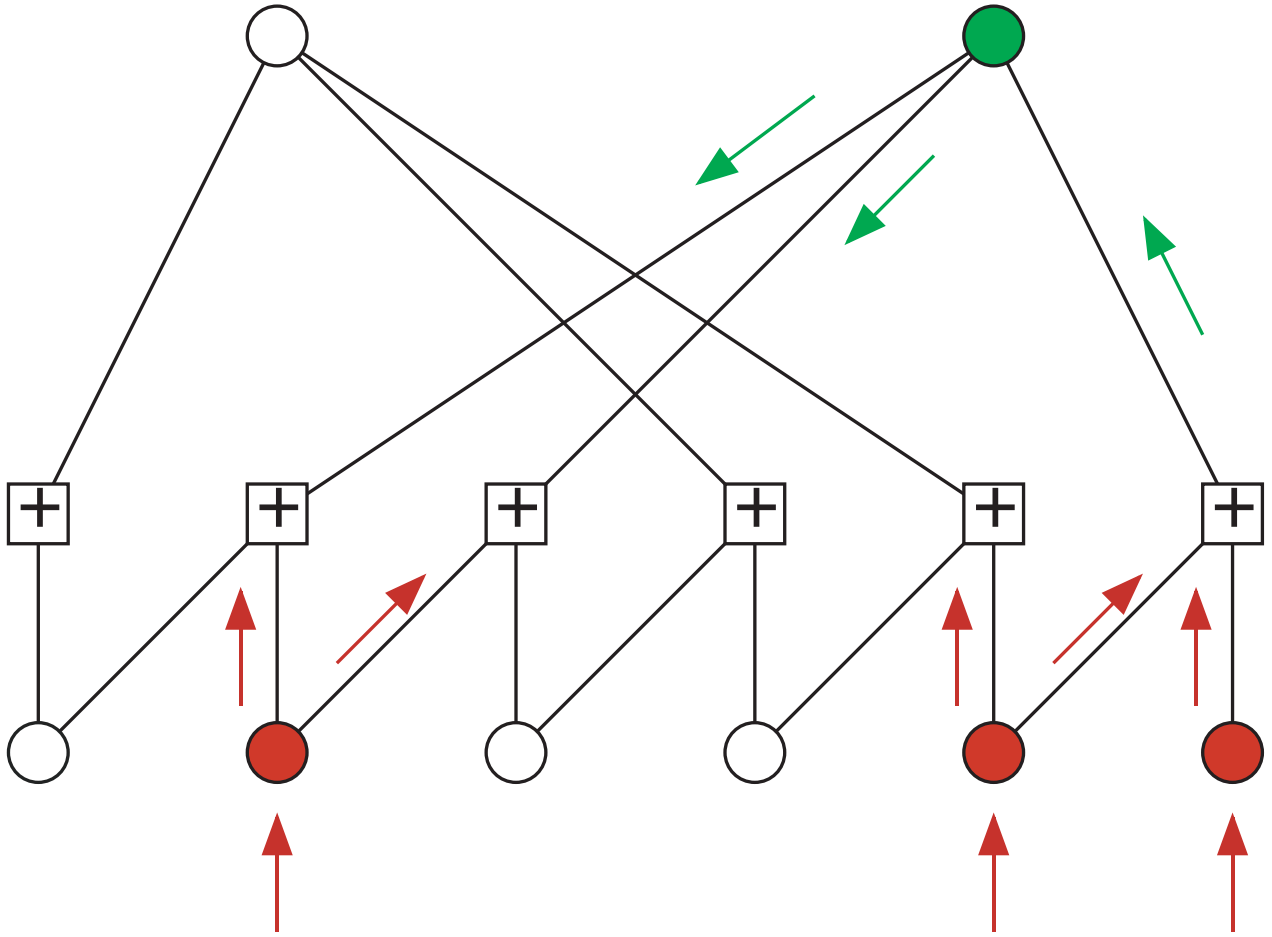
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



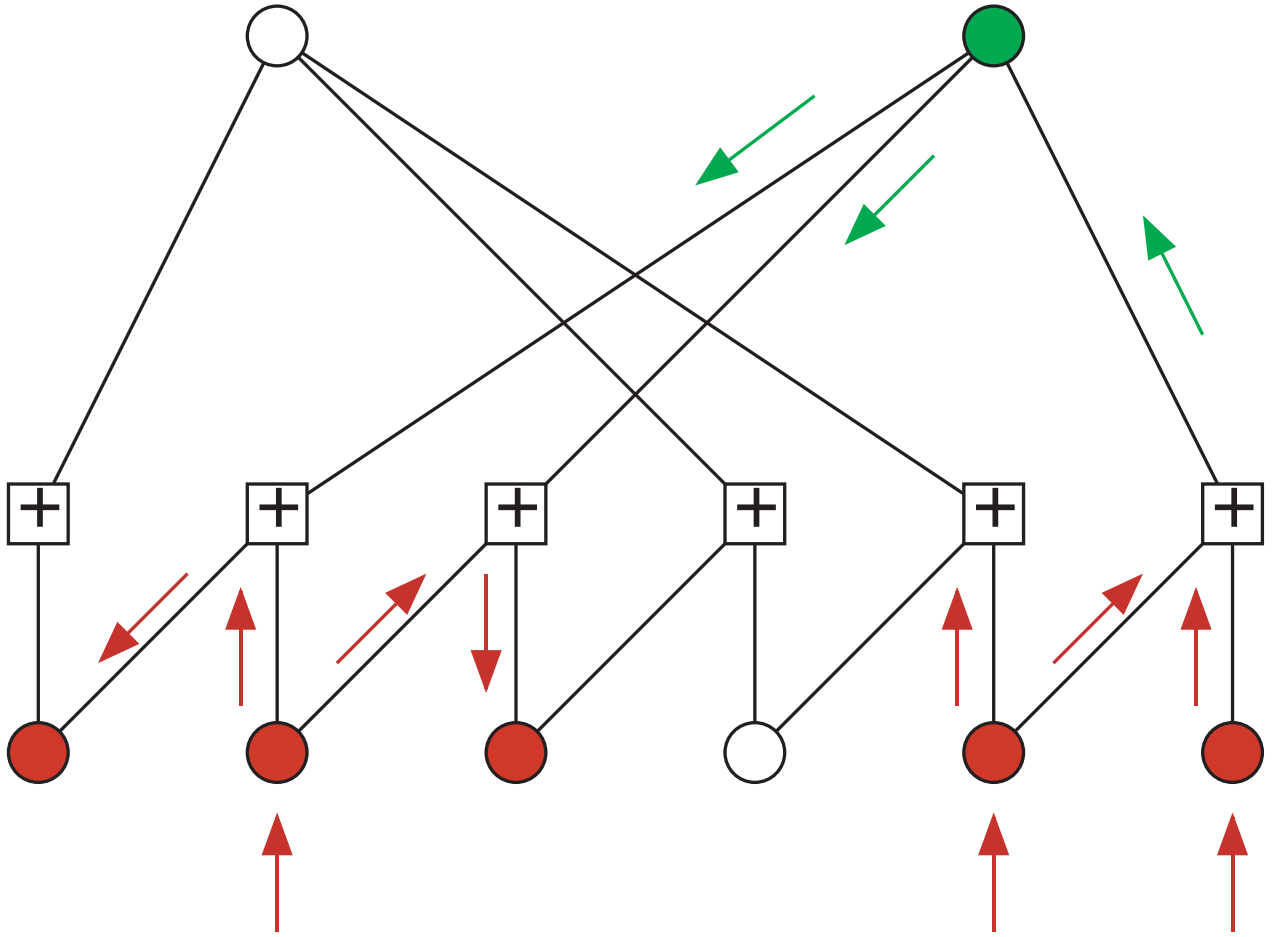
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



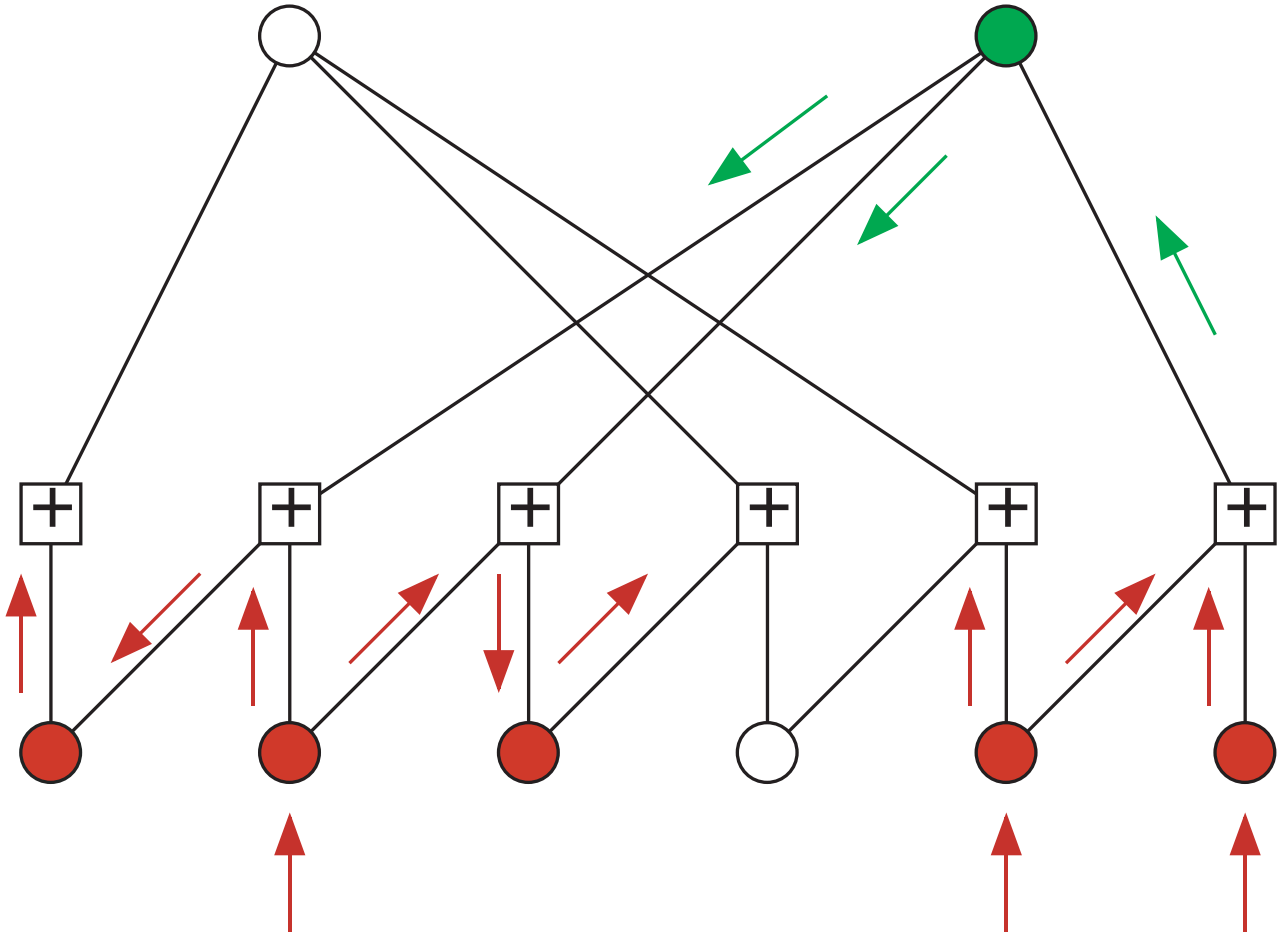
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



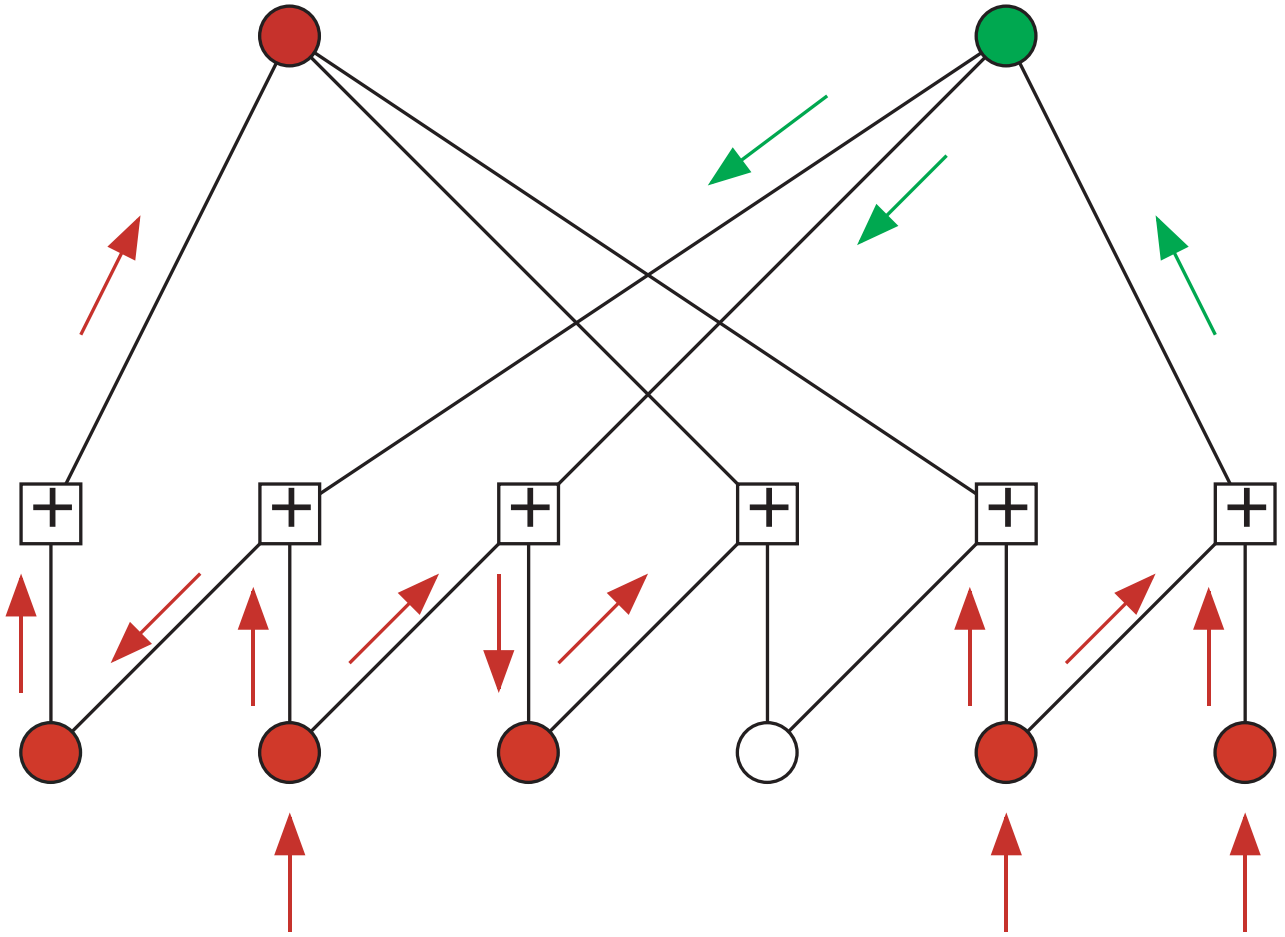
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



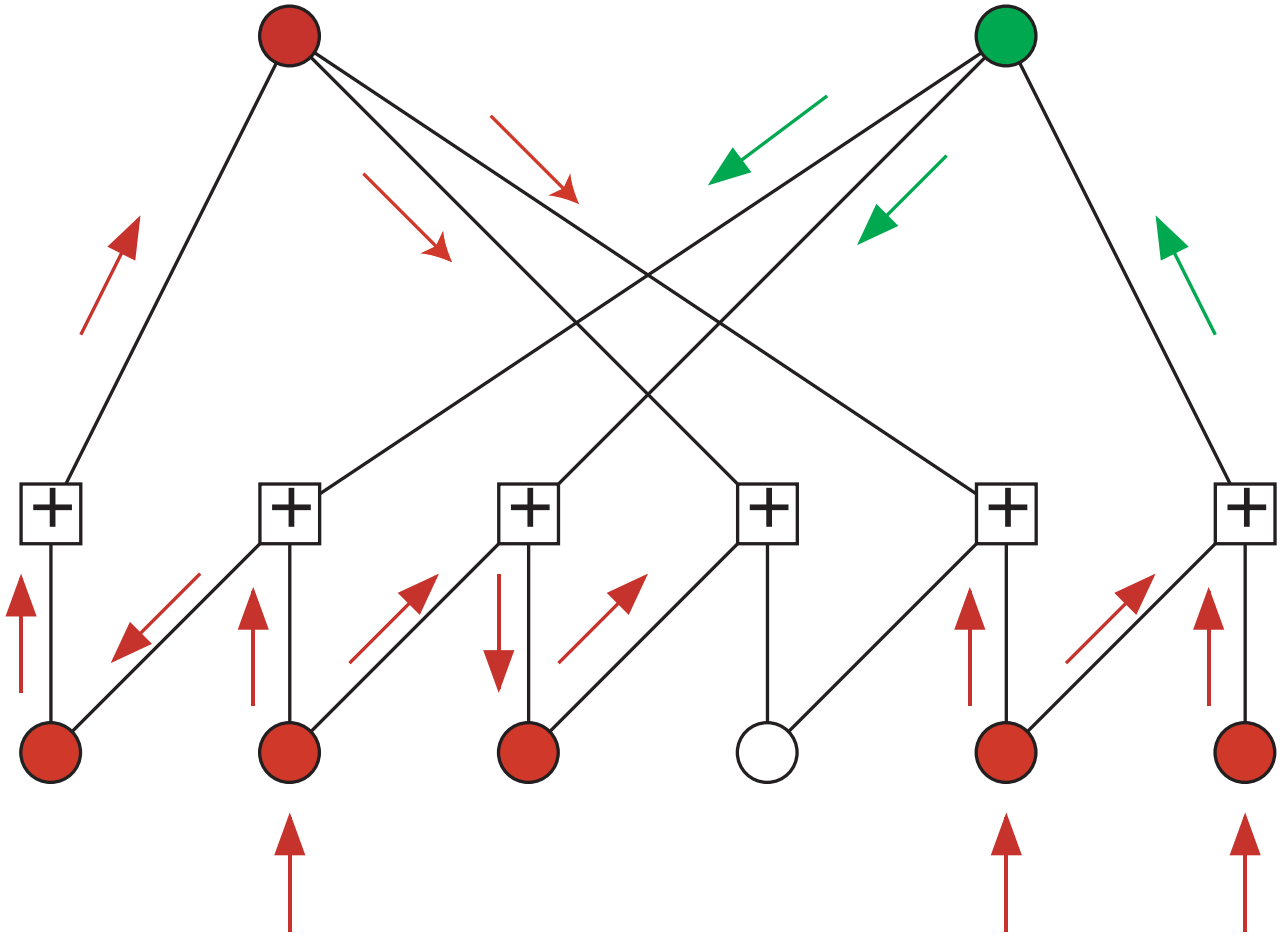
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



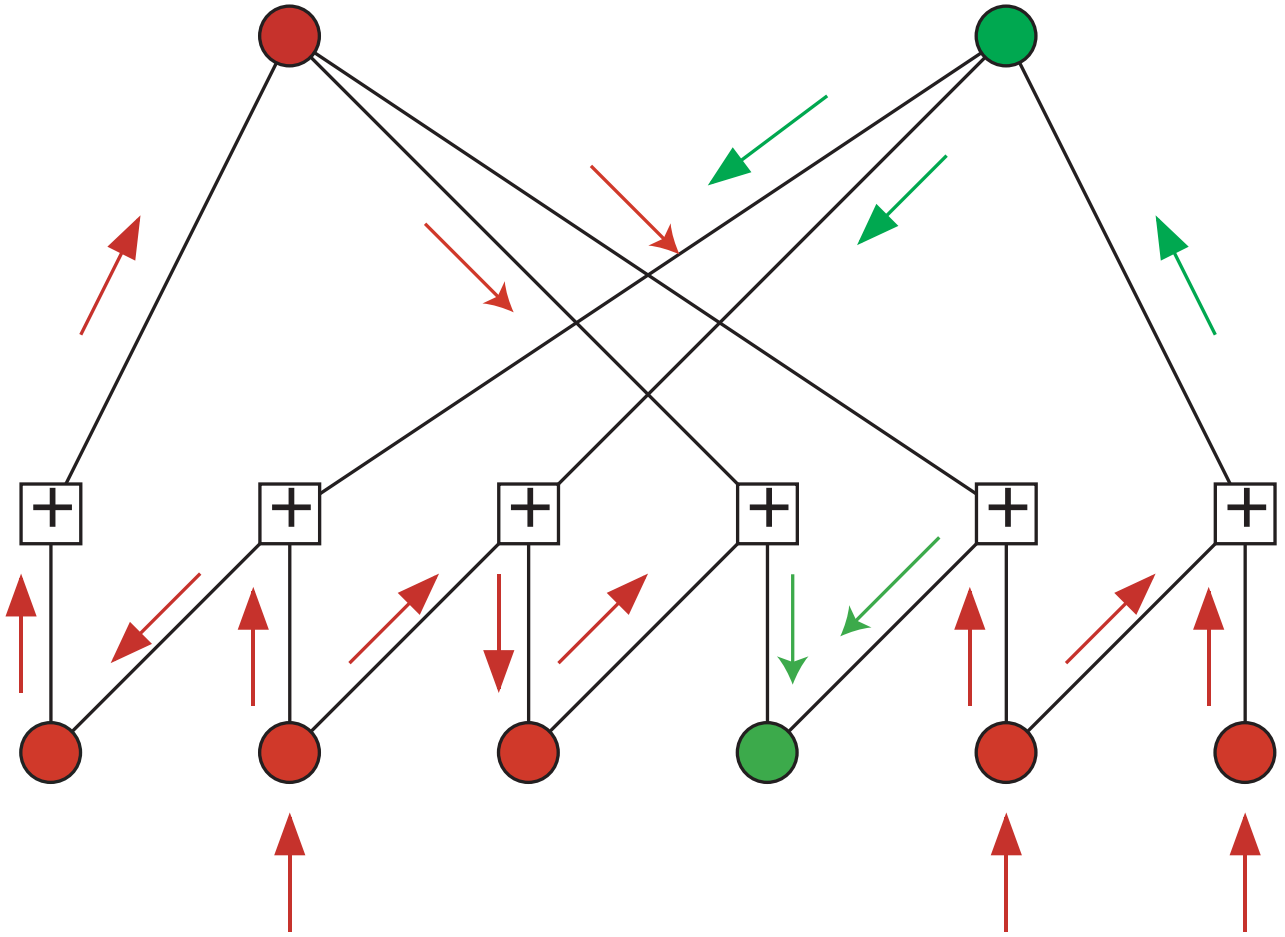
Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



Decoding an LDPC Code on the BEC Using Message Passing

$\longrightarrow = 1$, $\longrightarrow = 0$



What is the Complexity of Iterative Message-Passing Decoding?

- Complexity *per iteration*:

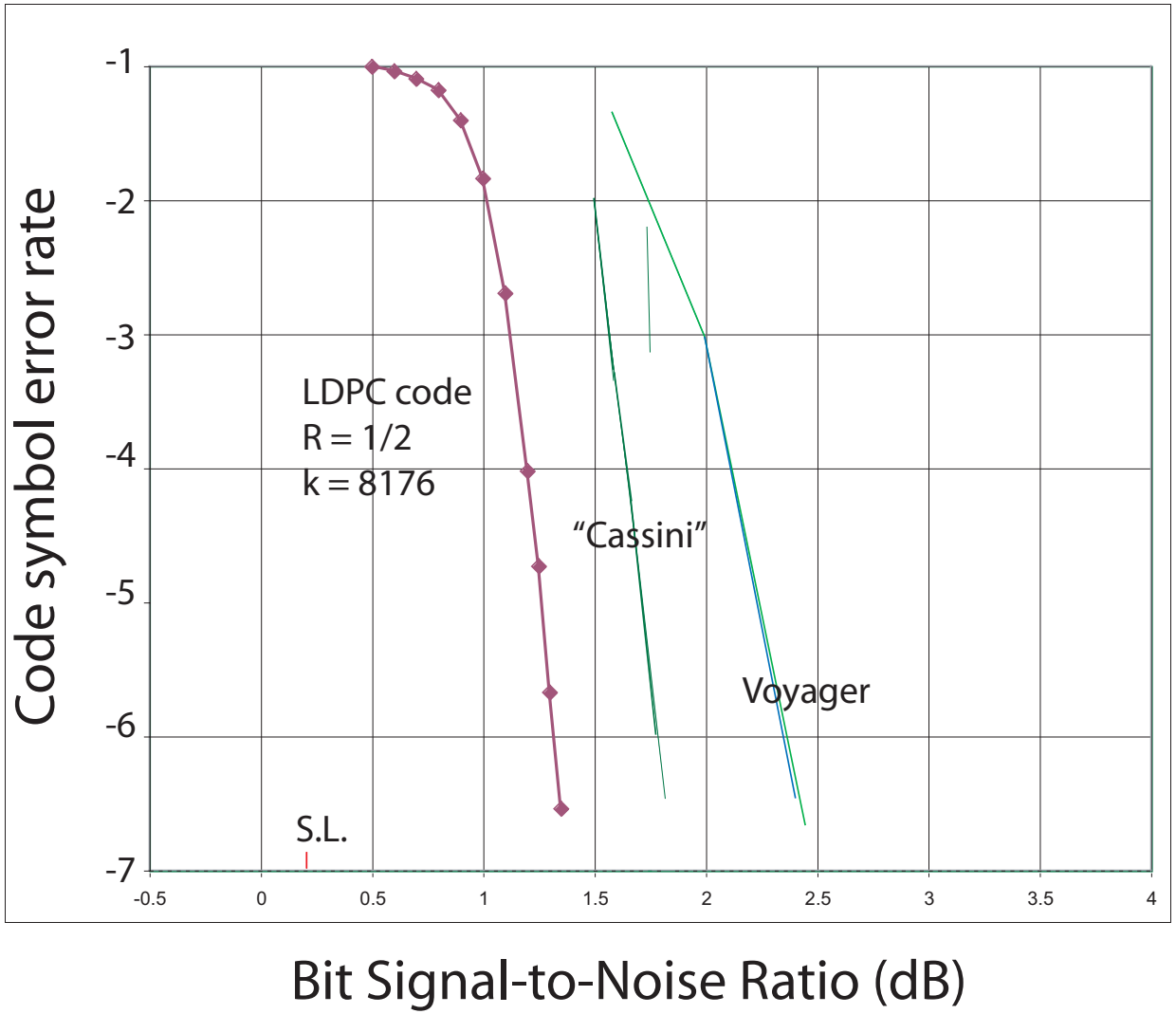
$$\chi_{IT} = 2\frac{E}{k},$$

where E is the number of edges in the Tanner graph, and k is the number of information bits (χ_{IT} is an ensemble invariant).

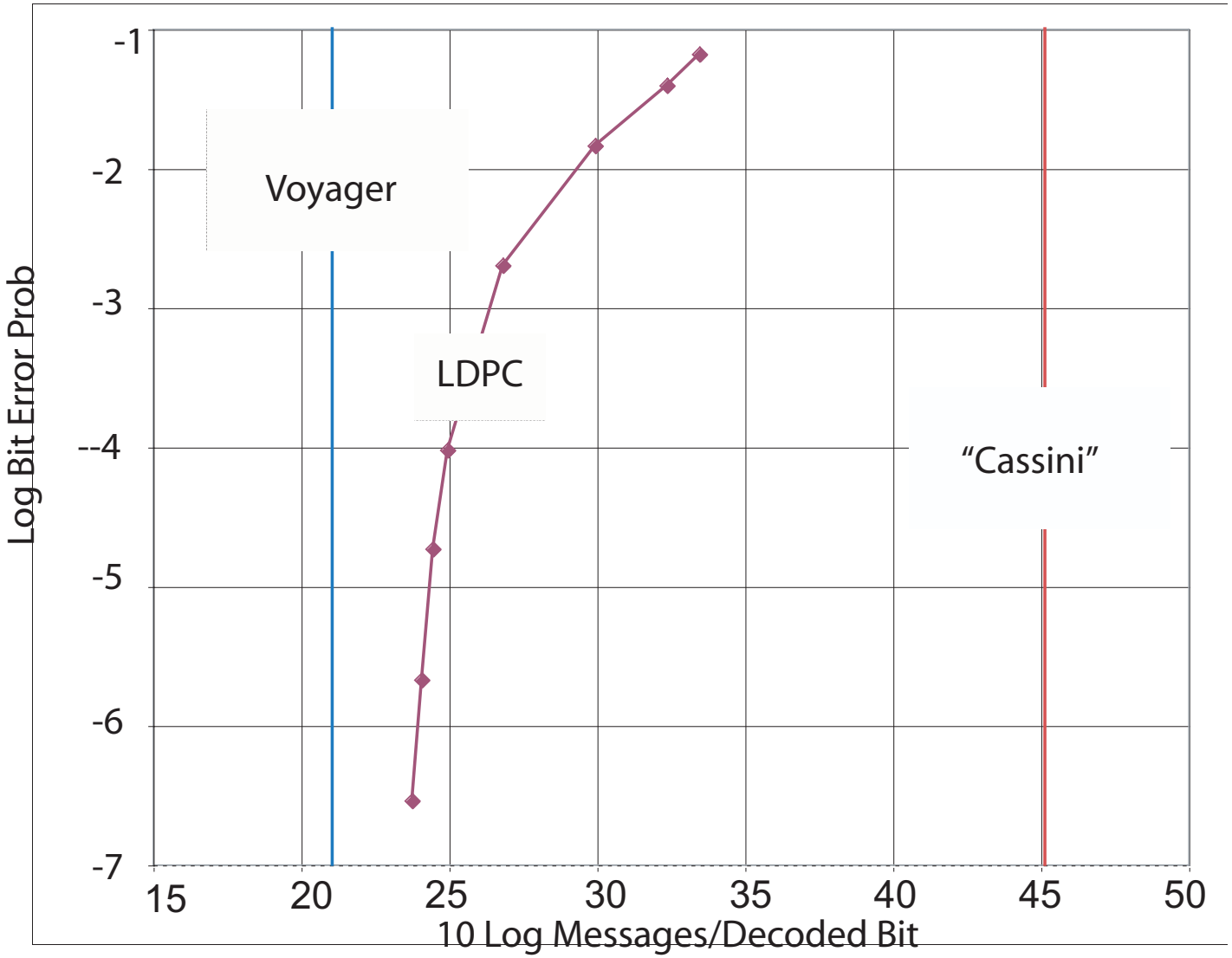
- $N(\epsilon, \pi)$ = Number of iterations needed to achieve error probability π.

$$\boxed{\chi_D(\epsilon, \pi) = \chi_{IT} \cdot N(\epsilon, \pi).}$$

An Example


Code symbol error rate
-1
-2
-3
-4
-5
-6
-7
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
LDPC code
R = 1/2
k = 8176
"Cassini"
Voyager
S.L.
Bit Signal-to-Noise Ratio (dB)

An Example


-1
-2
-3
--4
-5
-6
-7
Log Bit Error Prob
Voyager
LDPC
"Cassini"
15
20
25
30
35
40
45
50
10 Log Messages/Decoded Bit

Theory is Available!

Richardson

Urbanke

"The capacity of low-density parity-check codes under message-passing decoding"

"Density Evolution" is the Tool

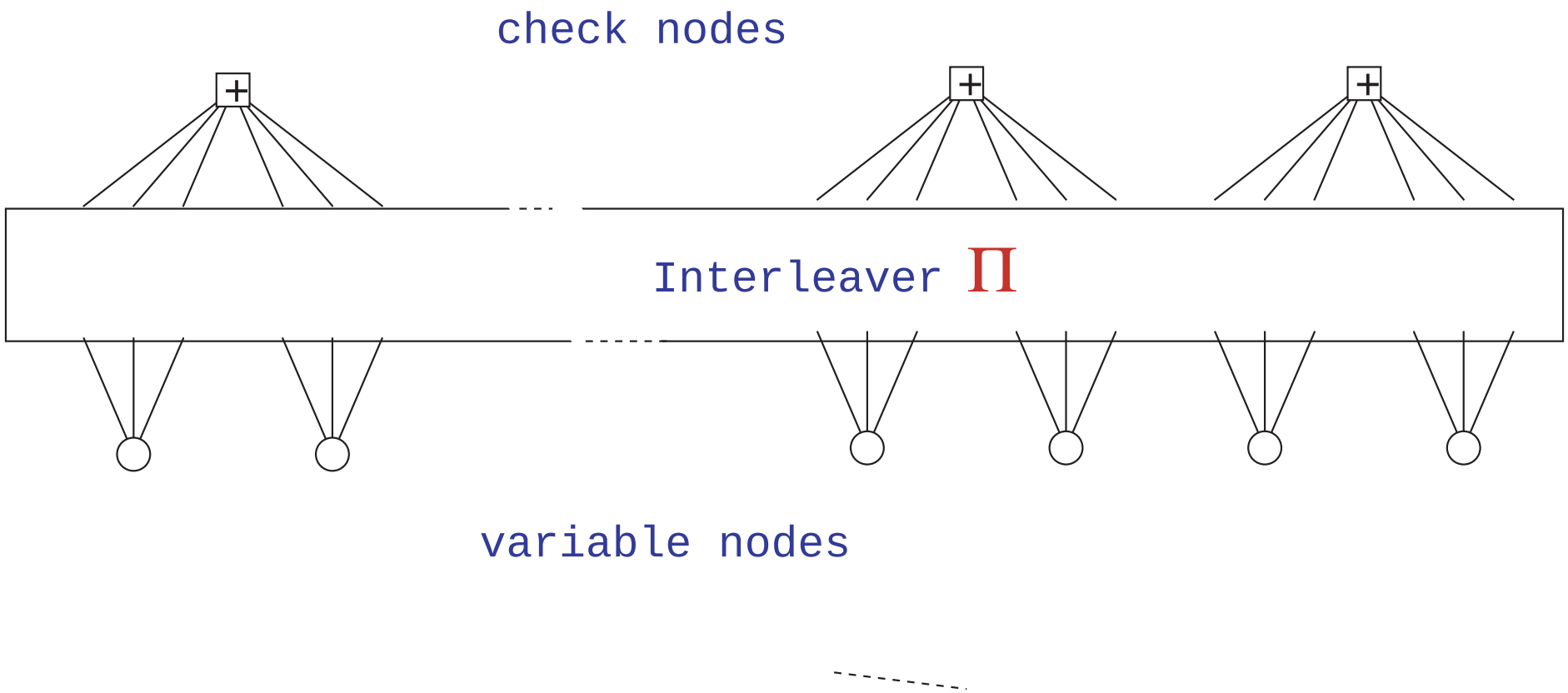
In density evolution, the idea is to treat the messages sent as random variables, and to track the probability density function of the messages. For example, on the binary erasure channel, one can simply track the probability that a given message is an "erasure."

The Fine Print

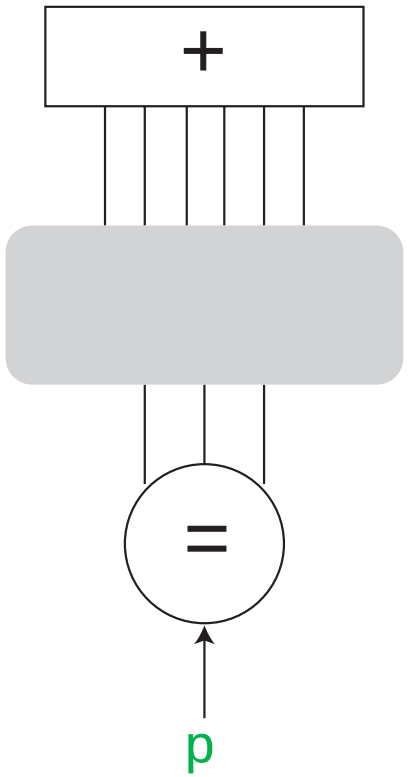
• The ensemble of codes must satisfiy the **RU condition:** For any fixed L, the probability that the depth-L neighborhood of a randomly selected edge contains a cycle goes to zero as $k \to \infty$.

• Therefore L-fold *density evolution* gives the limiting value ($k \to \infty$) of the ensemble bit error probability after L iterations. This limiting value will depend on the "noise parameter" of the channel. The largest noise parameter for which the limiting bit error probability is zero is called the *ensemble noise threshold*.

Tanner Graph for a (regular) $(3, 6)$ LDPC Code Ensemble

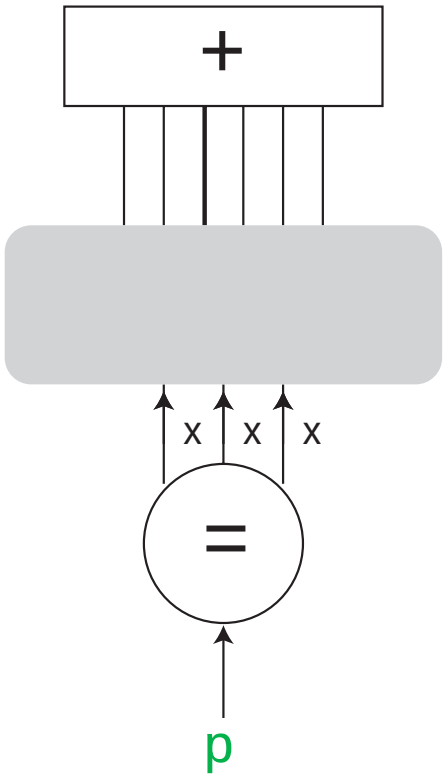


Density Evolution



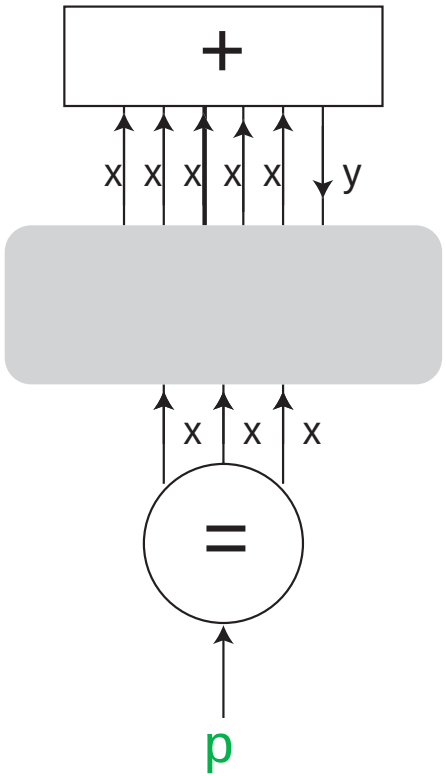
The $(3, 6)$ Ensemble, showing just one degree-3 variable node and one degree-6 check node. Evidence from the channel arrives from below. This evidence, which "seeds" the decoder, is absent (erased) with probability p.

Density Evolution, First Iteration



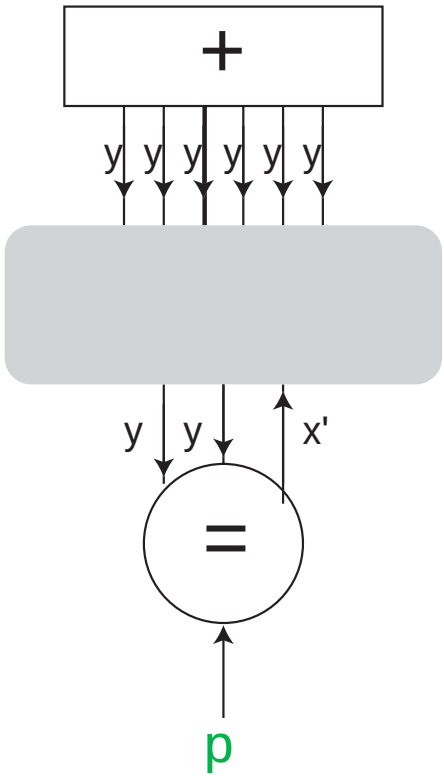
x = probability that the indicated message is "erasure." (On the first iteration, $x = p$.)

Density Evolution, First Iteration



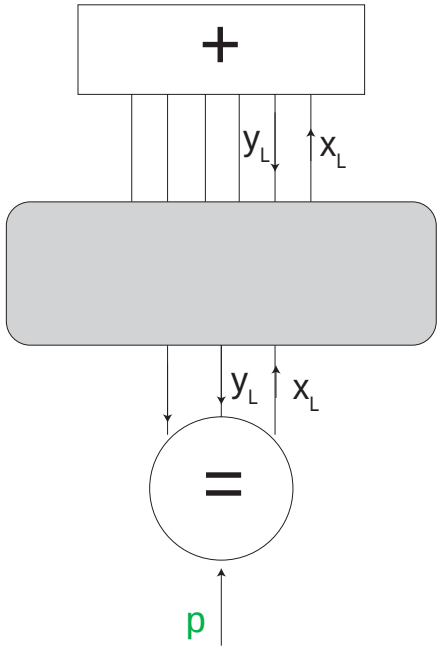
y is an erasure iff at least one of the x's is an erasure. Thus $y = 1 - (1 - x)^5$.

Density Evolution, First Iteration



x' is an erasure iff both y's and the channel input are: $x' = py^2 = p(1-(1-x)^5)^2$.

Density Evolution, Lth Iteration



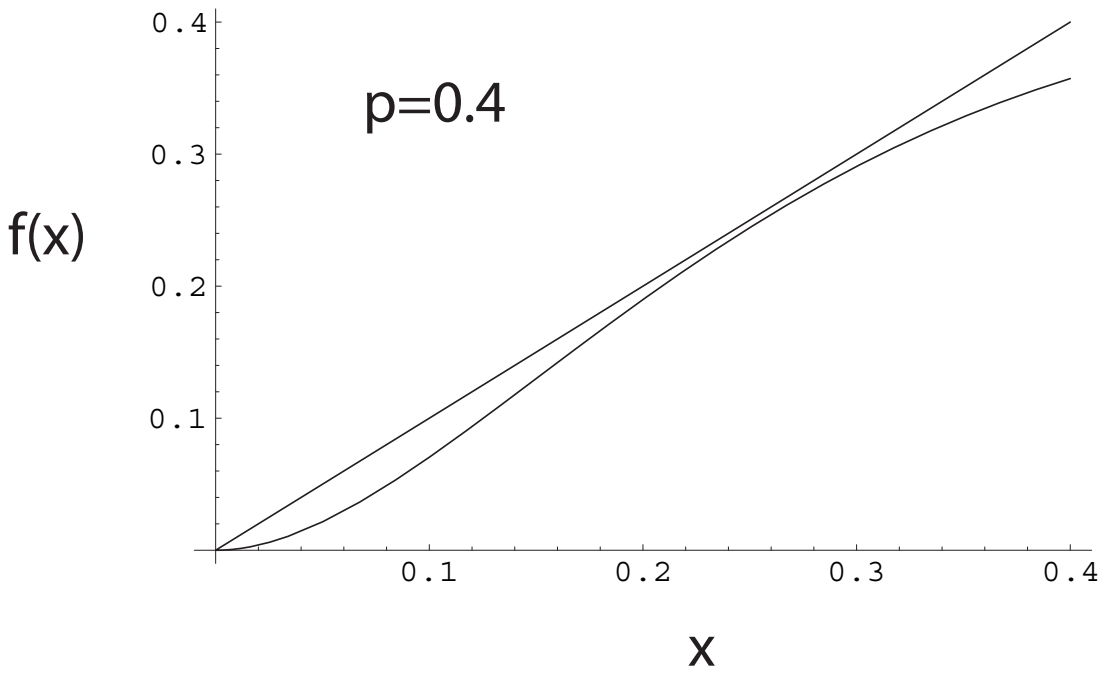
If the probability of an erased message is x_L on the Lth iiteration, then

$$x_{L+1} = f(x_L),$$

where

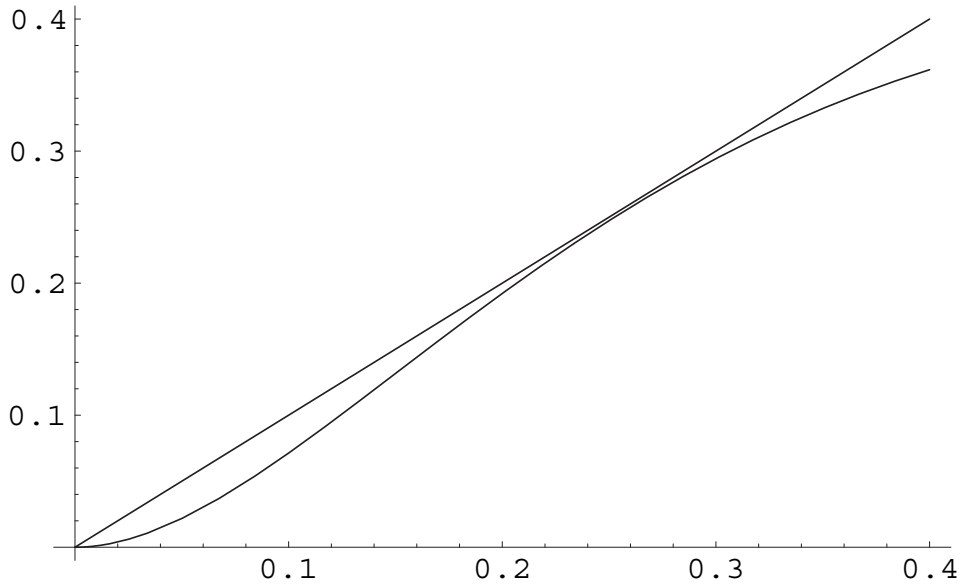
$$f(x) = p(1 - (1 - x)^5)^2.$$

Density Evolution, the Payoff



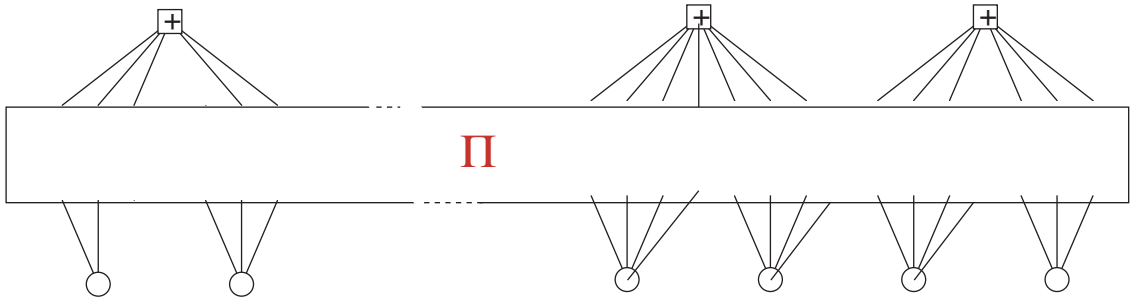
With $p = 0.4$, the only solution to the equation $f(x) = x$ is $x = 0$.

Density Evolution, the Payoff



With $p = 0.42944$, the curves just touch. Therefore the noise threshold for the $(3, 6)$ ensemble is 0.42944. On the other hand with $p = 0.42944$, channel capacity is 0.57056. In summary, $R = 0.5$, $C = .57056$, or 87.6% of capacity.

To do Better, We Need Irregular Ensembles



Richardson

Urbanke

Shokrollahi

How We May Appear to Future Generations

Claude Shannon — *Born on the planet Earth (Sol III) in the year 1916 A.D. Generally regarded as the father of the Information Age, he formulated the notion of channel capacity in 1948 A.D. Within several decades, mathematicians and engineers had devised practical ways to communicate reliably at data rates within 1% of the Shannon limit ...*

Encyclopedia Galactica, 166th ed.