A Brief Introduction to Spatially-Coupled Codes and Threshold Saturation

Henry D. Pfister

Based on joint work with Yung-Yih Jian, Santhosh Kumar,
Krishna R. Narayanan, Phong S. Nguyen, and Arvind Yedla

2014 North American School of Information Theory
Toronto, Canada
June 20th, 2014

Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

Chalkboard Proof

General Factor Graphs

Wyner-Ziv and Gelfand-Pinsker

Conclusions

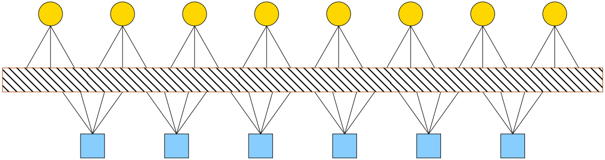
Outline

- Review of LDPC Codes and Density Evolution
- Spatially-Coupled Graphical Models
- Universality for Multiuser Scenarios
- Abstract Formulation of Threshold Saturation
- Chalkboard Proof
- General Factor Graphs
- Wyner-Ziv and Gelfand-Pinsker
- Conclusions

Point-to-Point Communication

- Coding for Discrete-Time Memoryless Channels
  - Transition probability: $P_{Y|X}(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$
  - Transmit a length-n codeword $\underline{x} \in \mathcal{C} \subset \mathcal{X}^n$
- Shannon Capacity
  - Code rate: $R \triangleq \frac{1}{n} \log_2 |\mathcal{C}|$ (bits per channel use)
  - As $n \to \infty$, reliable transmission if $R < C \triangleq \max_{p(x)} I(X;Y)$
- Example: the binary erasure channel BEC(ε)
  - Bits sent perfectly (with prob. $1 - \varepsilon$) or erased (with prob. ε)
  - Capacity: $C = 1 - \varepsilon =$ fraction unerased bits
  - Roughly one info bit transmitted for each unerased reception

Low-Density Parity-Check (LDPC) Codes



code bits

permutation

parity
checks

- Linear codes with a sparse parity-check matrix H
  - Regular (l, r): H has l ones per column and r ones per row
  - Irregular: number of ones given by degree distribution (λ, ρ)
  - Introduced by Gallager in 1960; largely forgotten until 1995
- Tanner Graph
  - An edge connects check node i to bit node j if $H_{ij} = 1$
  - Naturally leads to message-passing iterative (MPI) decoding

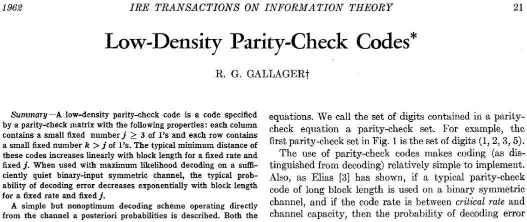
Decoding LDPC Codes

- Belief-Propagation (BP) Decoder
  - Low-complexity message-passing decoder by Gallager
  - Probability estimates are passed along edges in the Tanner graph
  - Updates based on assuming incoming estimates are independent
- Density Evolution (DE)
  - Tracks distribution of messages during iterative decoding
  - BP noise threshold can be computed via DE
  - Long codes decode almost surely if DE predicts success
- Maximum A Posteriori (MAP) Decoder
  - Optimum decoder that chooses the most likely codeword
  - Infeasible in practice due to enormous number of codewords
  - MAP noise threshold can be bounded using EXIT curves

A Little History

Robert Gallager

introduced LDPC codes in 1962 paper



1962 *IRE TRANSACTIONS ON INFORMATION THEORY* 21

Low-Density Parity-Check Codes*

R. G. GALLAGER†

Summary—A low-density parity-check code is a code specified by a parity-check matrix with the following properties: each column contains a small fixed number $j \geq 3$ of 1's and each row contains a small fixed number $k > j$ of 1's. The typical minimum distance of these codes increases linearly with block length for a fixed rate and fixed j. When used with maximum likelihood decoding on a sufficiently quiet binary-input symmetric channel, the typical probability of decoding error decreases exponentially with block length for a fixed rate and fixed j.

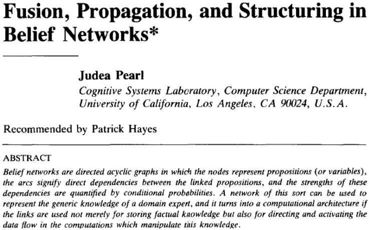
A simple but nonoptimum decoding scheme operating directly from the channel a posteriori probabilities is described. Both the

equations. We call the set of digits contained in a parity-check equation a parity-check set. For example, the first parity-check set in Fig. 1 is the set of digits (1, 2, 3, 5).

The use of parity-check codes makes coding (as distinguished from decoding) relatively simple to implement. Also, as Elias [3] has shown, if a typical parity-check code of long block length is used on a binary symmetric channel, and if the code rate is between *critical rate* and channel capacity, then the probability of decoding error

Judea Pearl

defined general belief-propagation in 1986 paper



Fusion, Propagation, and Structuring in Belief Networks*

Judea Pearl
Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles, CA 90024, U.S.A.

Recommended by Patrick Hayes

ABSTRACT

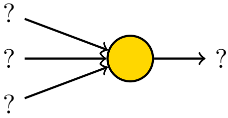
Belief networks are directed acyclic graphs in which the nodes represent propositions (or variables), the arcs signify direct dependencies between the linked propositions, and the strengths of these dependencies are quantified by conditional probabilities. A network of this sort can be used to represent the generic knowledge of a domain expert, and it turns into a computational architecture if the links are used not merely for storing factual knowledge but also for directing and activating the data flow in the computations which manipulate this knowledge.

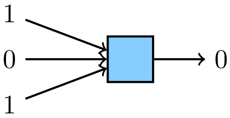
Message-Passing Decoding for the BEC (1)

- Bit and check nodes define the set of valid codewords
    - Circles represent a single bit value shared by checks
    - Squares assert attached bits sum to 0 mod 2
- Iterative decoding on the binary erasure channel (BEC)
    - Estimates of bit values are passed along edges in phases
    - 1st phase: bits pass messages to adjacent checks
    - 2nd phase: checks pass messages to adjacent bits
    - Each output message depends on other input messages
    - Messages are always either the correct value or an erasure

Message-Passing Decoding for the BEC (2)

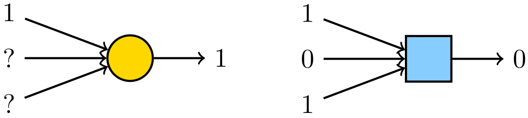
- Message passing rules for the BEC
  - Bits pass an erasure only if all other inputs are erased
  - Checks pass the correct value only if all other inputs are correct





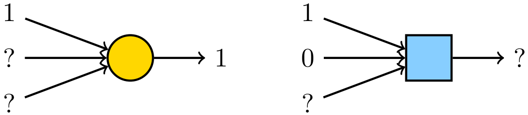
Message-Passing Decoding for the BEC (2)

- Message passing rules for the BEC
  - Bits pass an erasure only if all other inputs are erased
  - Checks pass the correct value only if all other inputs are correct



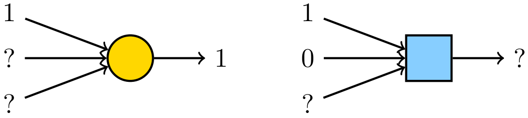
Message-Passing Decoding for the BEC (2)

- Message passing rules for the BEC
  - Bits pass an erasure only if all other inputs are erased
  - Checks pass the correct value only if all other inputs are correct

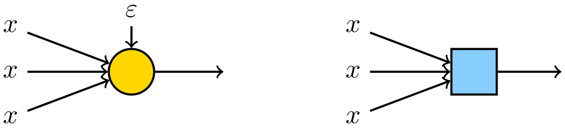


Message-Passing Decoding for the BEC (2)

- Message passing rules for the BEC
  - Bits pass an erasure only if all other inputs are erased
  - Checks pass the correct value only if all other inputs are correct

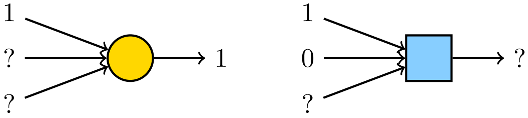


- If input messages are independently correct/erased with prob. x

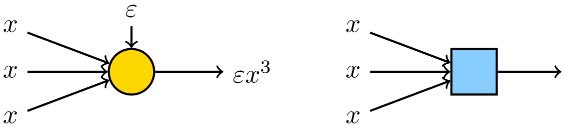


Message-Passing Decoding for the BEC (2)

- Message passing rules for the BEC
  - Bits pass an erasure only if all other inputs are erased
  - Checks pass the correct value only if all other inputs are correct

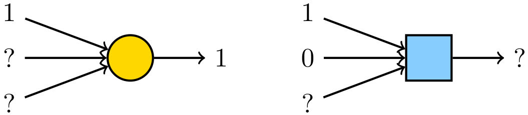


- If input messages are independently correct/erased with prob. x

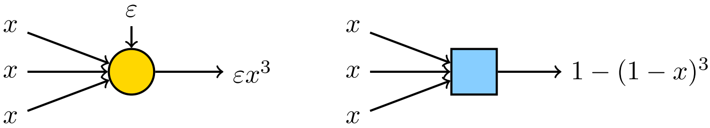


Message-Passing Decoding for the BEC (2)

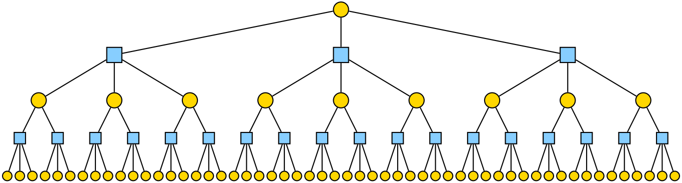
- Message passing rules for the BEC
  - Bits pass an erasure only if all other inputs are erased
  - Checks pass the correct value only if all other inputs are correct



- If input messages are independently correct/erased with prob. x



Computation Graph and Density Evolution



$\tilde{x}_3 = \varepsilon y_2^3$

$y_2 = 1-(1-x_2)^3$

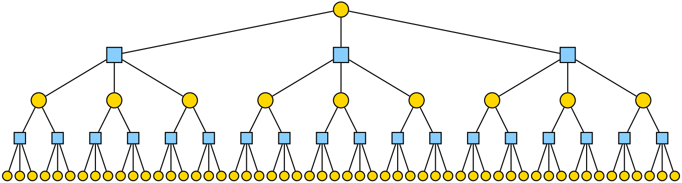
$x_2 = \varepsilon y_1^2$

$y_1 = 1-(1-x_1)^3$

$x_1 = \varepsilon$

- Computation graph for a (3,4)-regular LDPC code
  - Illustrates decoding from the perspective of a single bit-node
  - For long random LDPC codes, the graph is typically a tree
  - Allows density evolution to track message erasure probability
  - On each level, messages are independently erased with fixed prob.

Computation Graph and Density Evolution



$\tilde{x}_3 = \varepsilon y_2^3$

$y_2 = 1-(1-x_2)^3$

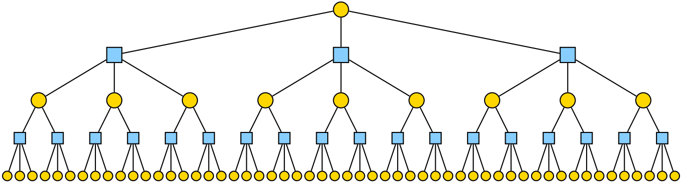
$x_2 = \varepsilon y_1^2$

$y_1 = 1-(1-x_1)^3$

$x_1 = 0.600$

- Computation graph for a (3,4)-regular LDPC code
  - Illustrates decoding from the perspective of a single bit-node
  - For long random LDPC codes, the graph is typically a tree
  - Allows density evolution to track message erasure probability
  - On each level, messages are independently erased with fixed prob.

Computation Graph and Density Evolution



$\tilde{x}_3 = \varepsilon y_2^3$

$y_2 = 1-(1-x_2)^3$

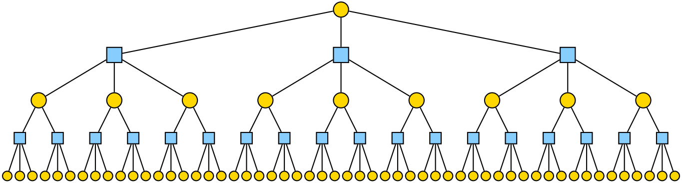
$x_2 = \varepsilon y_1^2$

$y_1 = 0.936$

$x_1 = 0.600$

- Computation graph for a (3,4)-regular LDPC code
    - Illustrates decoding from the perspective of a single bit-node
    - For long random LDPC codes, the graph is typically a tree
    - Allows density evolution to track message erasure probability
    - On each level, messages are independently erased with fixed prob.

Computation Graph and Density Evolution



$\tilde{x}_3 = \varepsilon y_2^3$

$y_2 = 1-(1-x_2)^3$

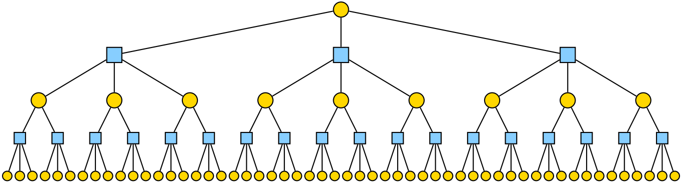
$x_2 = 0.526$

$y_1 = 0.936$

$x_1 = 0.600$

- Computation graph for a (3,4)-regular LDPC code
  - Illustrates decoding from the perspective of a single bit-node
  - For long random LDPC codes, the graph is typically a tree
  - Allows density evolution to track message erasure probability
  - On each level, messages are independently erased with fixed prob.

Computation Graph and Density Evolution



$\tilde{x}_3 = \varepsilon y_2^3$

$y_2 = 0.894$

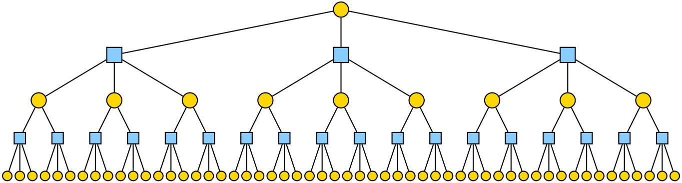
$x_2 = 0.526$

$y_1 = 0.936$

$x_1 = 0.600$

- Computation graph for a (3,4)-regular LDPC code
  - Illustrates decoding from the perspective of a single bit-node
  - For long random LDPC codes, the graph is typically a tree
  - Allows density evolution to track message erasure probability
  - On each level, messages are independently erased with fixed prob.

Computation Graph and Density Evolution



$\tilde{x}_3 = 0.429$

$y_2 = 0.894$

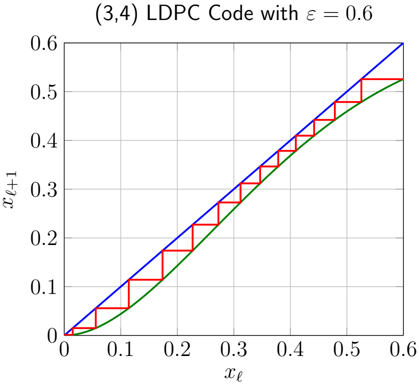
$x_2 = 0.526$

$y_1 = 0.936$

$x_1 = 0.600$

- Computation graph for a (3,4)-regular LDPC code
  - Illustrates decoding from the perspective of a single bit-node
  - For long random LDPC codes, the graph is typically a tree
  - Allows density evolution to track message erasure probability
  - On each level, messages are independently erased with fixed prob.

Density Evolution (DE) for LDPC Codes



Density evolution for a $(3,4)$-regular LDPC code:

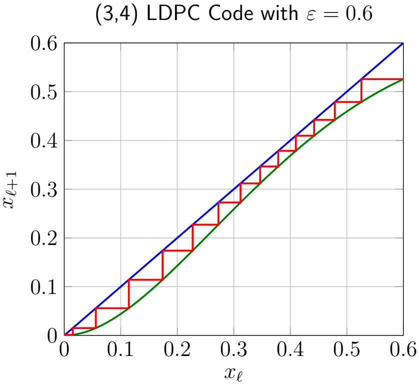
$$x_{\ell+1} = \varepsilon \left(1 - (1 - x_\ell)^3\right)^2$$

Decoding Thresholds:

$$\varepsilon^{\mathrm{BP}} \approx 0.647$$
$$\varepsilon^{\mathrm{MAP}} \approx 0.746$$
$$\varepsilon^{\mathrm{Sh}} = 0.750$$

- DE tracks bit-to-check msg erasure rate x_ℓ after ℓ iterations

Density Evolution (DE) for LDPC Codes



Density evolution for a $(3,4)$-regular LDPC code:

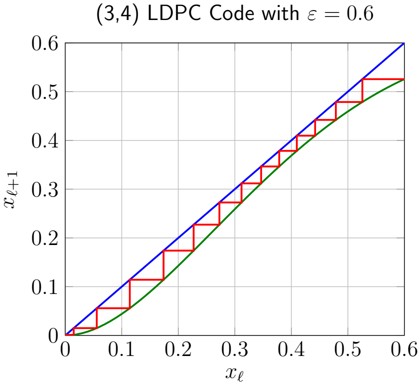
$$x_{\ell+1} = \varepsilon \left(1 - (1 - x_\ell)^3\right)^2$$

Decoding Thresholds:

$$\varepsilon^{\mathrm{BP}} \approx 0.647$$
$$\varepsilon^{\mathrm{MAP}} \approx 0.746$$
$$\varepsilon^{\mathrm{Sh}} = 0.750$$

- DE tracks bit-to-check msg erasure rate x_ℓ after ℓ iterations
- x_ℓ decreases to a limit $x_\infty(\varepsilon)$ that depends on ε

Density Evolution (DE) for LDPC Codes



Density evolution for a $(3,4)$-regular LDPC code:

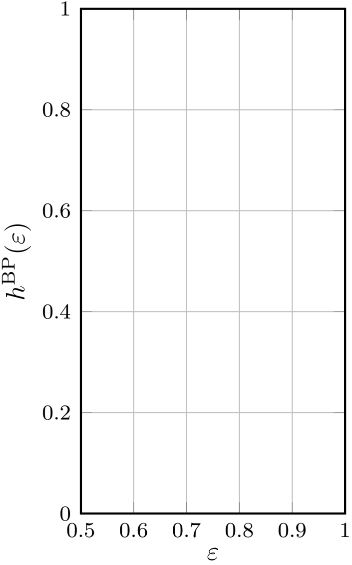
$$x_{\ell+1} = \varepsilon\left(1-(1-x_\ell)^3\right)^2$$

Decoding Thresholds:

$$\varepsilon^{\mathrm{BP}} \approx 0.647$$
$$\varepsilon^{\mathrm{MAP}} \approx 0.746$$
$$\varepsilon^{\mathrm{Sh}} = 0.750$$

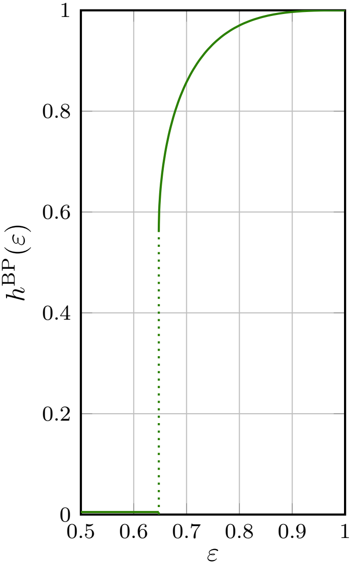
- DE tracks bit-to-check msg erasure rate x_ℓ after ℓ iterations
- x_ℓ decreases to a limit $x_\infty(\varepsilon)$ that depends on ε
- As $n \to \infty$, decoding succeeds if ε less than the BP noise threshold
    - $\varepsilon^{\mathrm{BP}} = \sup\{\varepsilon \in [0,1] \,|\, x_\infty(\varepsilon) = 0\}$ (easily computed numerically)

EXtrinsic Information Transfer (EXIT) Curves



- (3,4)-regular LDPC code
  - Codeword $(X_1, \ldots, X_n)$
  - Received $(Y_1, \ldots, Y_n)$

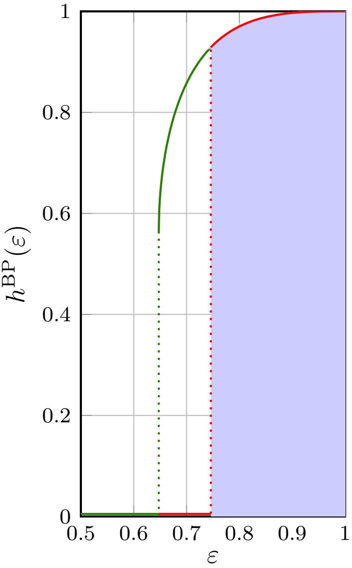
EXtrinsic Information Transfer (EXIT) Curves



- (3,4)-regular LDPC code
  - Codeword $(X_1, \ldots, X_n)$
  - Received $(Y_1, \ldots, Y_n)$

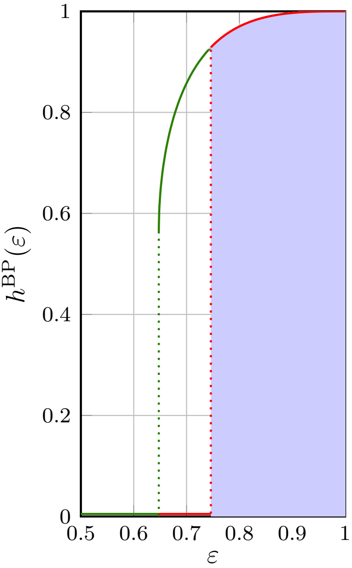
- BP EXIT curve via DE
  - This code: $h^{\mathrm{BP}}(\varepsilon) = (x_\infty(\varepsilon))^3$
  - 0 below BP threshold 0.647

EXtrinsic Information Transfer (EXIT) Curves



- (3,4)-regular LDPC code
  - Codeword $(X_1, \ldots, X_n)$
  - Received $(Y_1, \ldots, Y_n)$
- BP EXIT curve via DE
  - This code: $h^{\mathrm{BP}}(\varepsilon) = (x_\infty(\varepsilon))^3$
  - 0 below BP threshold 0.647
- MAP EXIT curve is extrinsic entropy $H(X_i|Y_{\sim i})$ vs. channel ε
  - 0 below MAP threshold 0.746
  - Area under curve equals rate
  - Upper bounded by BP EXIT

EXtrinsic Information Transfer (EXIT) Curves



- (3,4)-regular LDPC code
  - Codeword $(X_1, \ldots, X_n)$
  - Received $(Y_1, \ldots, Y_n)$
- BP EXIT curve via DE
  - This code: $h^{\mathrm{BP}}(\varepsilon) = (x_\infty(\varepsilon))^3$
  - 0 below BP threshold 0.647
- MAP EXIT curve is extrinsic entropy $H(X_i|Y_{\sim i})$ vs. channel ε
  - 0 below MAP threshold 0.746
  - Area under curve equals rate
  - Upper bounded by BP EXIT
- MAP threshold upper bound $\overline{\varepsilon}^{\mathrm{MAP}}$
  - ε: area under BP EXIT equals rate

Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

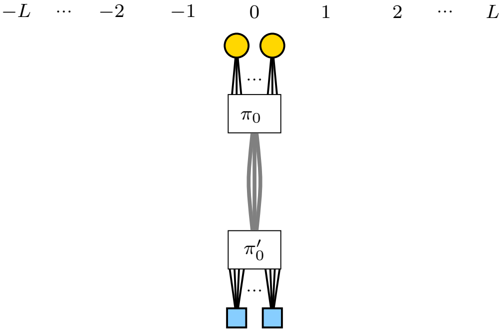
Chalkboard Proof

General Factor Graphs

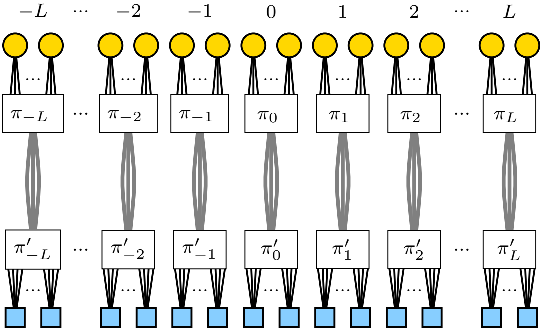
Wyner-Ziv and Gelfand-Pinsker

Conclusions

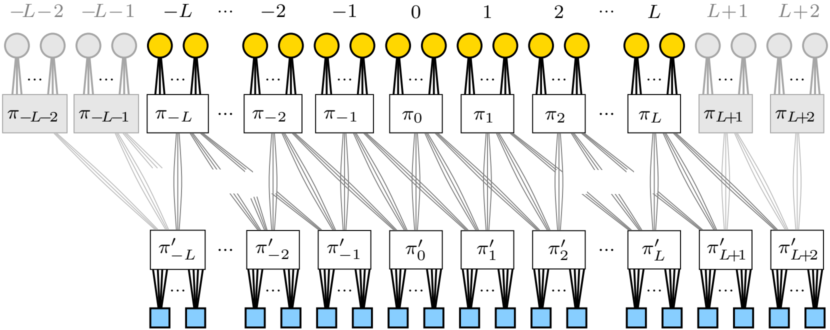
Spatially-Coupled LDPC Codes: (l, r, L, w) Ensemble



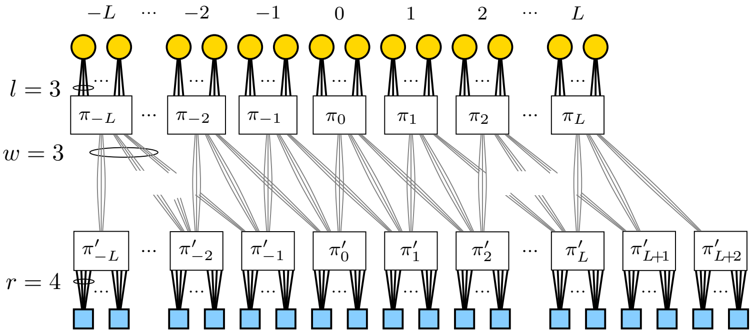
Spatially-Coupled LDPC Codes: (l, r, L, w) Ensemble



Spatially-Coupled LDPC Codes: (l, r, L, w) Ensemble



Spatially-Coupled LDPC Codes: (l, r, L, w) Ensemble



- Historical Notes
  - LDPC convolutional codes introduced by FZ in 1999
  - Shown to have near optimal noise thresholds by LSZC in 2005
  - (l, r, L, w) ensemble proven to achieve capacity by KRU in 2011

The LDPCC Gang

5274 IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 10, OCTOBER 2010

Iterative Decoding Threshold Analysis for LDPC Convolutional Codes

Michael Lentmaier, *Member, IEEE*, Arvind Sridharan, *Member, IEEE*, Daniel J. Costello, Jr., *Life Fellow, IEEE*,
and Kamil Sh. Zigangirov, *Fellow, IEEE*

The Spatial Coupling KRU

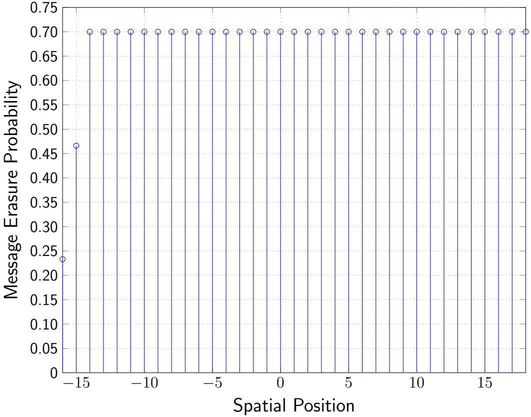
IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 57, NO. 2, FEBRUARY 2011 803

Threshold Saturation via Spatial Coupling: Why Convolutional LDPC Ensembles Perform So Well over the BEC

Shrinivas Kudekar, *Member, IEEE*, Thomas J. Richardson, *Fellow, IEEE*, and Rüdiger L. Urbanke

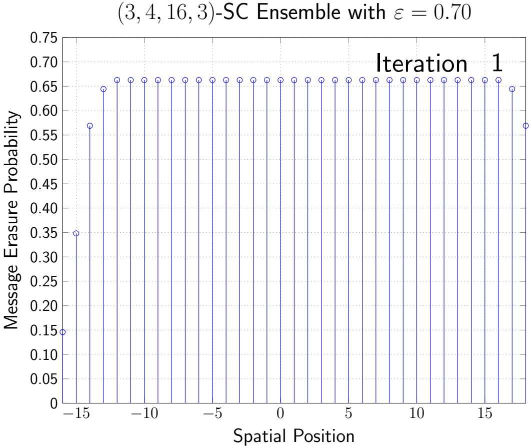
Density Evolution for the (l, r, L, w)-SC LDPC Ensemble

$(3, 4, 16, 3)$-SC Ensemble with $\varepsilon = 0.70$



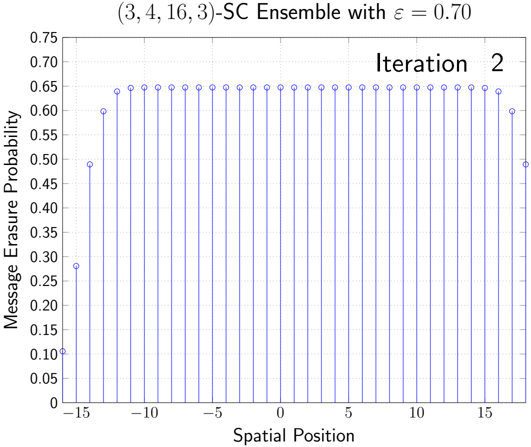
$$x_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w}\sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1}\right) \right)^{l-1} \mathbf{1}_{[-L,L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



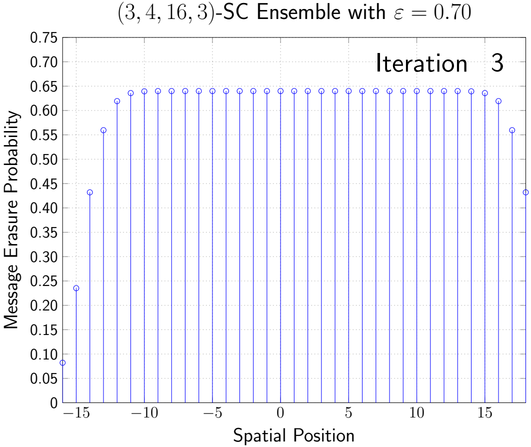
$$x_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w} \sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1} \right) \right)^{l-1} \mathbf{1}_{[-L, L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



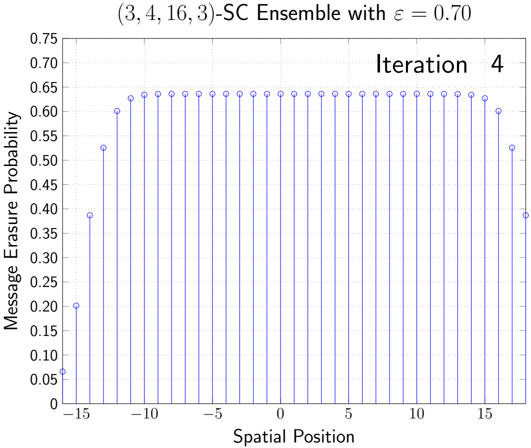
$$x_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w}\sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1}\right)\right)^{l-1} \mathbf{1}_{[-L,L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



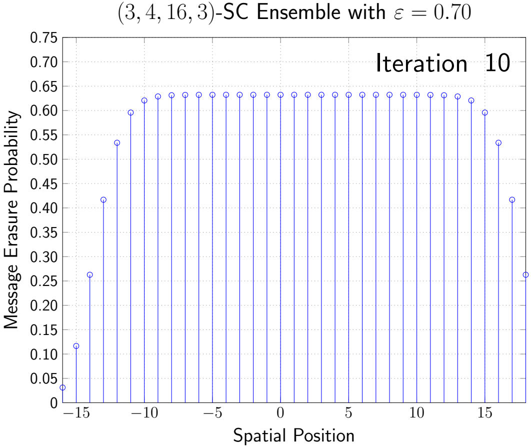
$$x_i^{(\ell+1)} = \frac{1}{w} \sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w} \sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1} \right) \right)^{l-1} \mathbf{1}_{[-L, L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



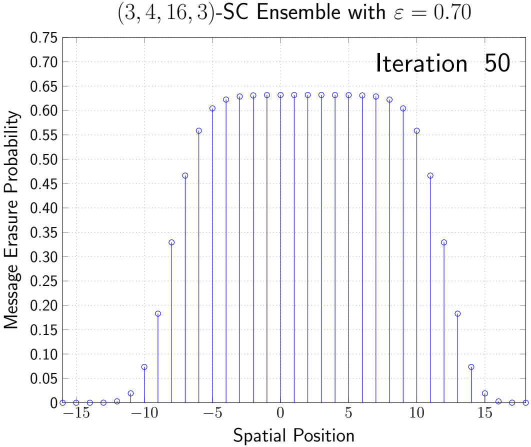
$$x_i^{(\ell+1)} = \frac{1}{w} \sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w} \sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1}\right) \right)^{l-1} \mathbf{1}_{[-L, L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



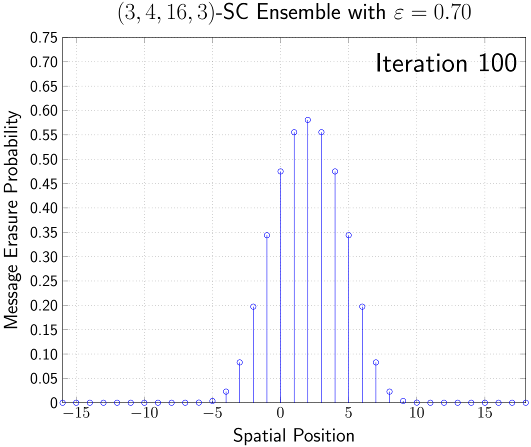
$$x_i^{(\ell+1)} = \frac{1}{w} \sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w} \sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1} \right) \right)^{l-1} \mathbf{1}_{[-L, L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



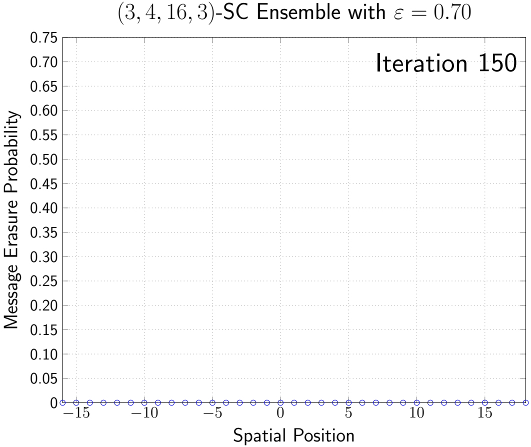
$$x_i^{(\ell+1)} = \frac{1}{w} \sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w} \sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1} \right) \right)^{l-1} \mathbf{1}_{[-L,L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



$$x_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w}\sum_{j=0}^{w-1} \left(1-(1-x_{i+j-k}^{(\ell)})^{r-1}\right)\right)^{l-1} \mathbf{1}_{[-L,L+w-1]}(i-k)$$

Density Evolution for the (l, r, L, w)-SC LDPC Ensemble



$$x_i^{(\ell+1)} = \frac{1}{w} \sum_{k=0}^{w-1} \varepsilon \left(\frac{1}{w} \sum_{j=0}^{w-1} \left(1 - (1 - x_{i+j-k}^{(\ell)})^{r-1}\right) \right)^{l-1} \mathbf{1}_{[-L, L+w-1]}(i-k)$$

Properties of Threshold Saturation

l	r	$\varepsilon^{\mathrm{BP}}$	$\varepsilon^{\mathrm{MAP}}$
3	6	0.4294	0.4882
4	8	0.3834	0.4977
5	10	0.3416	0.4995
6	12	0.3075	0.4999
7	14	0.2798	0.5000

- Spatial coupling achieves the MAP threshold as $w \to \infty$
    - BP threshold typically decreases after $l = 3$
    - MAP threshold is increasing in l, r for fixed rate
- Benefits and Drawbacks
    - For fixed L, minimum distance grows linearly with block length
    - Rate loss of $O(w/L)$ is a big obstacle in practice

Threshold Saturation via Spatial Coupling

- **General Phenomenon** (observed by Kudekar, Richardson, Urbanke)
  - BP threshold of the spatially-coupled system converges to the MAP threshold of the uncoupled system
  - Can be proven rigorously in many cases!

Threshold Saturation via Spatial Coupling

- **General Phenomenon** (observed by Kudekar, Richardson, Urbanke)
  - BP threshold of the spatially-coupled system converges to the MAP threshold of the uncoupled system
  - Can be proven rigorously in many cases!
- Connection to statistical physics
  - Factor graph defines system of coupled particles
  - Valid sequences are ordered crystalline structures

Threshold Saturation via Spatial Coupling

- **General Phenomenon** (observed by Kudekar, Richardson, Urbanke)
  - BP threshold of the spatially-coupled system converges to the MAP threshold of the uncoupled system
  - Can be proven rigorously in many cases!

- Connection to statistical physics
  - Factor graph defines system of coupled particles
  - Valid sequences are ordered crystalline structures

- Between BP and MAP threshold, system acts as supercooled liquid
  - Correct answer (crystalline state) has minimum energy
  - Crystallization (i.e., decoding) does not occur without a seed
  - Ex.: ice melts at $0\,^{\circ}$C but freezing w/o a seed requires $-48.3\,^{\circ}$C

http://www.youtube.com/watch?v=Xe8vJrIvDQM

Why is Spatial Coupling Interesting?

- Breakthroughs: first practical constructions of
    - universal codes for binary-input memoryless channels [KRU12]
    - information-theoretically optimal compressive sensing [DJM11]
    - universal codes for Slepian-Wolf and MAC problems [YJNP11]
    - codes $\rightarrow$ capacity with iterative hard-decision decoding [JNP12]
    - codes $\rightarrow$ rate-distortion limit with iterative decoding [AMUV12]

Why is Spatial Coupling Interesting?

- Breakthroughs: first practical constructions of
  - universal codes for binary-input memoryless channels [KRU12]
  - information-theoretically optimal compressive sensing [DJM11]
  - universal codes for Slepian-Wolf and MAC problems [YJNP11]
  - codes $\rightarrow$ capacity with iterative hard-decision decoding [JNP12]
  - codes $\rightarrow$ rate-distortion limit with iterative decoding [AMUV12]
- It allows rigorous proof in many cases
  - Original proofs [KRU11/12] quite specific to LDPC codes
  - Our proof is for increasing scalar/vector recursions [YJNP12/13]

Why is Spatial Coupling Interesting?

- Breakthroughs: first practical constructions of
  - universal codes for binary-input memoryless channels [KRU12]
  - information-theoretically optimal compressive sensing [DJM11]
  - universal codes for Slepian-Wolf and MAC problems [YJNP11]
  - codes → capacity with iterative hard-decision decoding [JNP12]
  - codes → rate-distortion limit with iterative decoding [AMUV12]

- It allows rigorous proof in many cases
  - Original proofs [KRU11/12] quite specific to LDPC codes
  - Our proof is for increasing scalar/vector recursions [YJNP12/13]

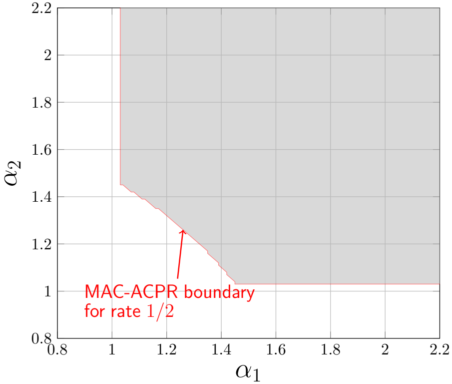
- Spatial coupling as a proof technique [GMU13]
  - For a large random factor graph, construct a coupled version
  - Use DE to analyze BP decoding of coupled system
  - Compare uncoupled MAP with coupled BP via interpolation

Outline

- Review of LDPC Codes and Density Evolution
- Spatially-Coupled Graphical Models
- Universality for Multiuser Scenarios
- Abstract Formulation of Threshold Saturation
- Chalkboard Proof
- General Factor Graphs
- Wyner-Ziv and Gelfand-Pinsker
- Conclusions

Universality over Unknown Parameters

- The Achievable Channel Parameter Region (ACPR)
  - For a sequence of coding schemes involving one or more parameters, the parameter region where decoding succeeds in the limit
  - In contrast, a capacity region is a rate region for fixed channels



Universality over Unknown Parameters

- The Achievable Channel Parameter Region (ACPR)
  - For a sequence of coding schemes involving one or more parameters, the parameter region where decoding succeeds in the limit
  - In contrast, a capacity region is a rate region for fixed channels
- Properties
  - For fixed encoders, the ACPR depends on the decoders
  - For example, one has BP-ACPR $\subseteq$ MAP-ACPR
  - Often, $\exists$ unique maximal ACPR given by information theory

Universality over Unknown Parameters

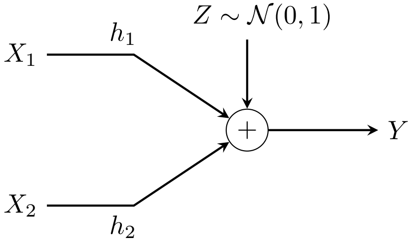
- The Achievable Channel Parameter Region (ACPR)
    - For a sequence of coding schemes involving one or more parameters, the parameter region where decoding succeeds in the limit
    - In contrast, a capacity region is a rate region for fixed channels

- Properties
    - For fixed encoders, the ACPR depends on the decoders
    - For example, one has BP-ACPR $\subseteq$ MAP-ACPR
    - Often, $\exists$ unique maximal ACPR given by information theory

- Universality
    - A sequence of encoding/decoding schemes is called universal if:

      its ACPR equals the optimal ACPR
    - Channel parameters are assumed unknown at the transmitter
    - At the receiver, the channel parameters are easily estimated

2-User Binary-Input Gaussian Multiple Access Channel



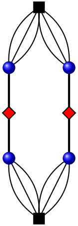
- Fixed noise variance
- Real channel gains h_1 and h_2 not known at transmitter
- Each code has rate R

- MAC-ACPR denotes the information-theoretic optimal region

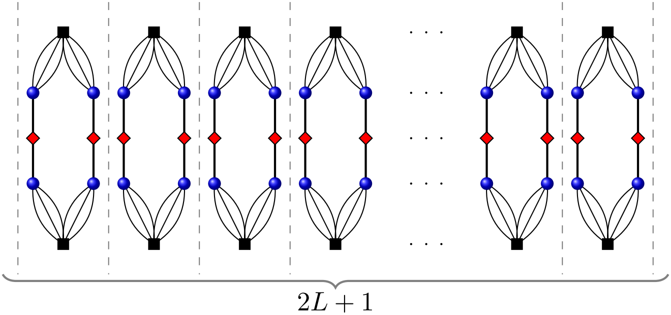
A Little History: SC for Multiple-Access (MAC) Channels

- KK consider a binary-adder erasure channel (ISIT 2011)
    - SC exhibits threshold saturation for the joint decoder
- YNPN consider the Gaussian MAC (ISIT/Allerton 2011)
    - SC exhibits threshold saturation for the joint decoder
    - For channel gains h_1, h_2 unknown at transmitter, SC provides universality
- Others consider CDMA systems without coding
    - TTK show SC improves BP demod of standard CDMA
    - ST prove saturation for a SC protograph-style CDMA

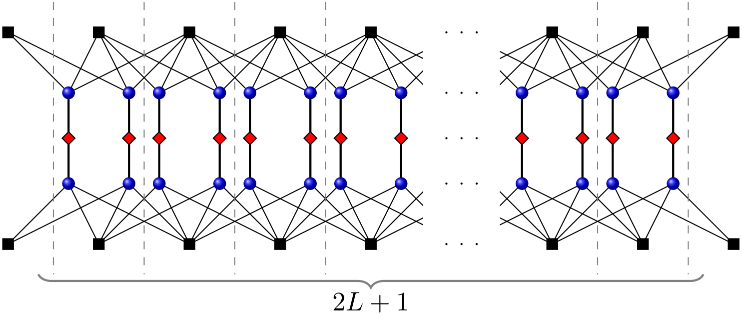
Spatially-Coupled Factor Graph for Joint Decoder



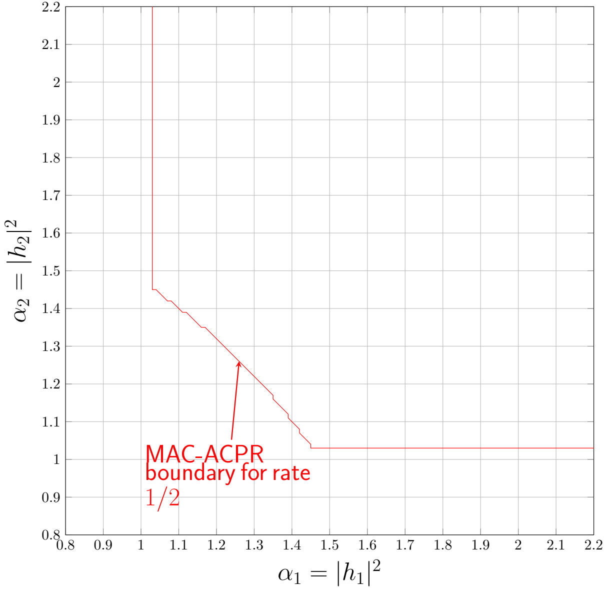
Spatially-Coupled Factor Graph for Joint Decoder



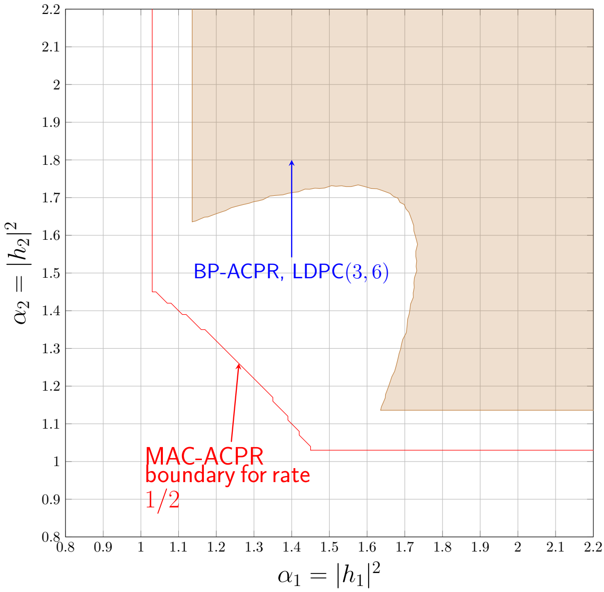
Spatially-Coupled Factor Graph for Joint Decoder



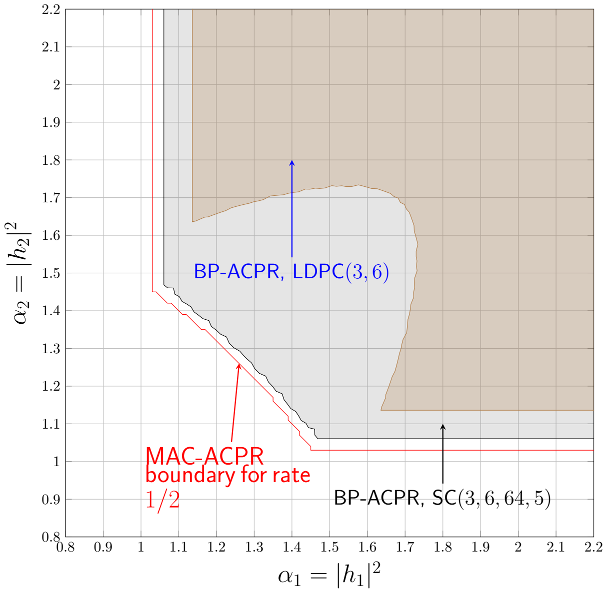
DE Performance of the Joint Decoder



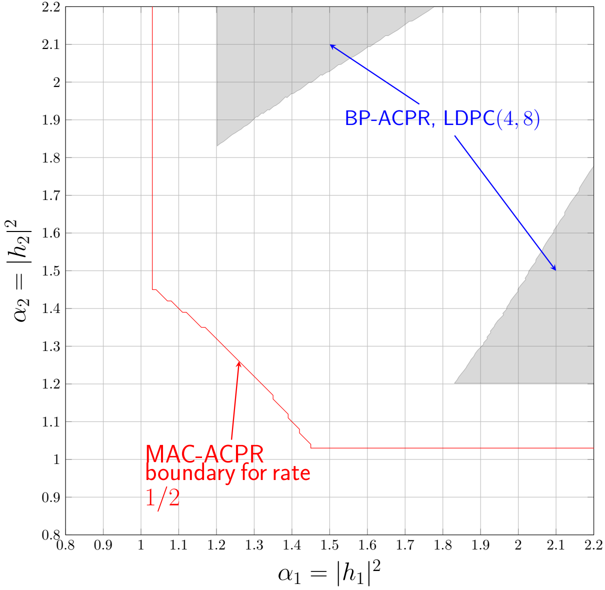
DE Performance of the Joint Decoder



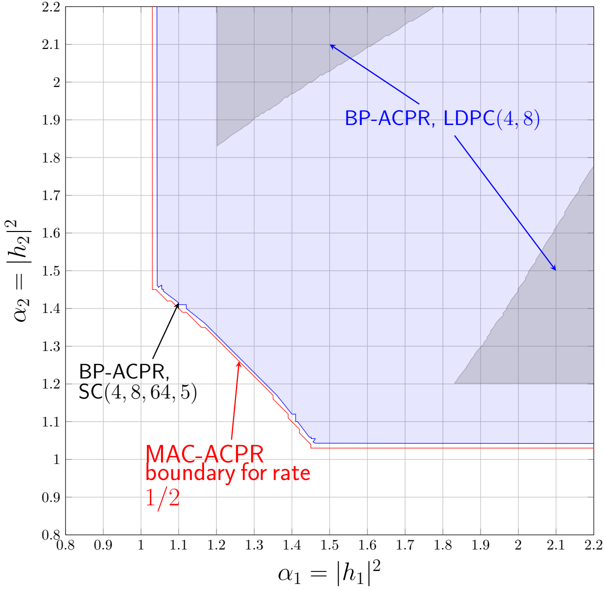
DE Performance of the Joint Decoder



DE Performance of the Joint Decoder



DE Performance of the Joint Decoder



Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

Chalkboard Proof

General Factor Graphs

Wyner-Ziv and Gelfand-Pinsker

Conclusions

An Abstract Approach to Spatial Coupling

Let $f\colon \mathcal{X} \to \mathcal{X}$ and $g\colon \mathcal{X} \to \mathcal{X}$ be strictly increasing C^2 functions on $\mathcal{X} = [0,1]$ with $f(0) = g(0) = 0$. Then, the scalar recursion (from $x^{(0)} = 1$)

$$\begin{aligned} y^{(\ell+1)} &= g\left(x^{(\ell)}\right) \\ x^{(\ell+1)} &= f\left(y^{(\ell+1)}\right) \end{aligned}$$

An Abstract Approach to Spatial Coupling

Let $f\colon \mathcal{X} \to \mathcal{X}$ and $g\colon \mathcal{X} \to \mathcal{X}$ be strictly increasing C^2 functions on $\mathcal{X} = [0,1]$ with $f(0) = g(0) = 0$. Then, the scalar recursion (from $x^{(0)} = 1$)

$$\begin{aligned} y^{(\ell+1)} &= g\left(x^{(\ell)}\right) = 1-(1-x)^3 \\ x^{(\ell+1)} &= f\left(y^{(\ell+1)}\right) = \varepsilon x^2 \end{aligned}$$

Ex. (3,4) LDPC

An Abstract Approach to Spatial Coupling

Let $f\colon \mathcal{X}\to\mathcal{X}$ and $g\colon \mathcal{X}\to\mathcal{X}$ be strictly increasing C^2 functions on $\mathcal{X}=[0,1]$ with $f(0)=g(0)=0$. Then, the scalar recursion (from $x^{(0)}=1$)

$$\begin{aligned} y^{(\ell+1)} &= g\left(x^{(\ell)}\right) = 1-(1-x)^3 \\ x^{(\ell+1)} &= f\left(y^{(\ell+1)}\right) = \varepsilon x^2 \end{aligned}$$

Ex. (3,4) LDPC

characterizes fixed point of the coupled recursion ($x_i^{(0)}=1, i\in[N+w-1]$)

$$\begin{aligned} y_i^{(\ell+1)} &= g\left(x_i^{(\ell)}\right) \\ x_i^{(\ell+1)} &= \sum_{j=1}^{N+w-1} A_{j,i}\, f\left(\sum_{k=1}^{N} A_{j,k}\, y_k^{(\ell+1)}\right) \\ A &= \frac{1}{w}\begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \ddots & 1 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 1 & 1 & \cdots & 1 \end{bmatrix} \end{aligned}$$

An Abstract Approach to Spatial Coupling

Let $f\colon \mathcal{X}\to\mathcal{X}$ and $g\colon \mathcal{X}\to\mathcal{X}$ be strictly increasing C^2 functions on $\mathcal{X}=[0,1]$ with $f(0)=g(0)=0$. Then, the scalar recursion (from $x^{(0)}=1$)

$$\begin{aligned} y^{(\ell+1)} &= g\left(x^{(\ell)}\right) = 1-(1-x)^3 \\ x^{(\ell+1)} &= f\left(y^{(\ell+1)}\right) = \varepsilon x^2 \end{aligned}$$

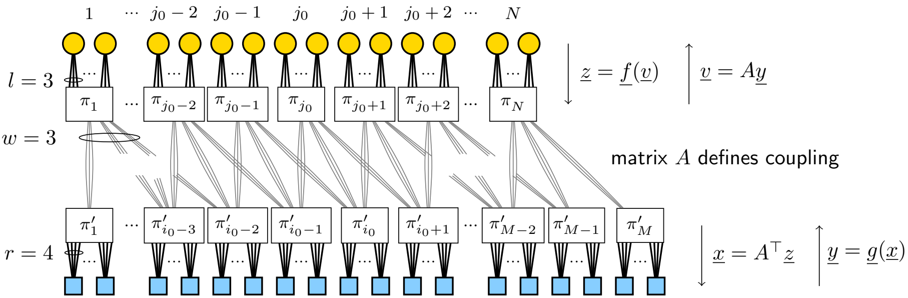
Ex. (3,4) LDPC

characterizes fixed point of the coupled recursion ($\underline{x}^{(0)}=\underline{1}$)

$$\begin{aligned} \underline{y}^{(\ell+1)} &= \underline{g}\left(\underline{x}^{(\ell)}\right) \\ \underline{x}^{(\ell+1)} &= A^\top \underline{f}\left(A\,\underline{y}^{(\ell+1)}\right) \end{aligned}$$

$$A = \frac{1}{w}\begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \ddots & 1 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

Example: DE for Spatially-Coupled LDPC Codes



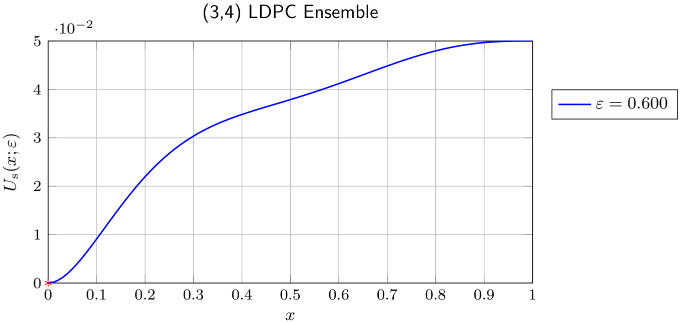
$\underline{z} = \underline{f}(\underline{v})$ $\underline{v} = A\underline{y}$

matrix A defines coupling

$\underline{x} = A^{\top}\underline{z}$ $\underline{y} = \underline{g}(\underline{x})$

- N bit-node and $M \triangleq N+w-1$ check-node sections
  - Erasure probability by section: $\underline{x}, \underline{y} \in [0,1]^M$ and $\underline{v}, \underline{z} \in [0,1]^N$
  - Symmetry points: $j_0 = \lceil N/2 \rceil$ and $i_0 = \lceil M/2 \rceil$
  - Coupling matrix: $A_{j,i}$ is the fraction of π_j edges attached to π'_i
  - Vector updates: $[\underline{f}(\underline{y})]_j = f(y_j),\ j \in [N]$ and $[\underline{g}(\underline{x})]_i = g(x_i),\ i \in [M]$

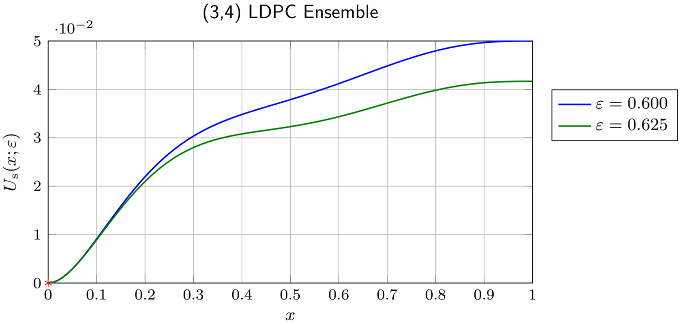
The Potential Function and Threshold Saturation



Let the potential function $U_{\mathrm{s}} \colon \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

$$U_{\mathrm{s}}(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z) \mathrm{d}z.$$

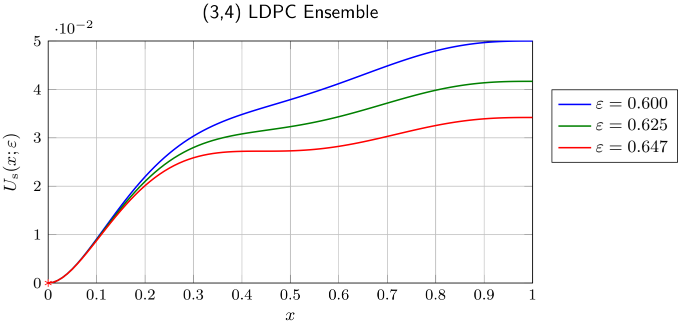
The Potential Function and Threshold Saturation



Let the potential function $U_{\mathrm{s}} \colon \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

$$U_{\mathrm{s}}(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z) \mathrm{d}z.$$

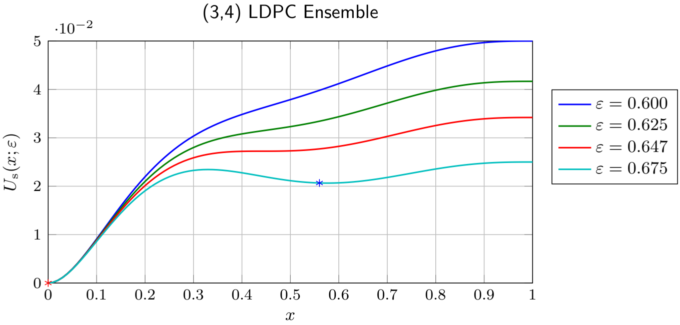
The Potential Function and Threshold Saturation



Let the potential function $U_{\mathrm{s}} \colon \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

$$U_{\mathrm{s}}(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z) \mathrm{d}z.$$

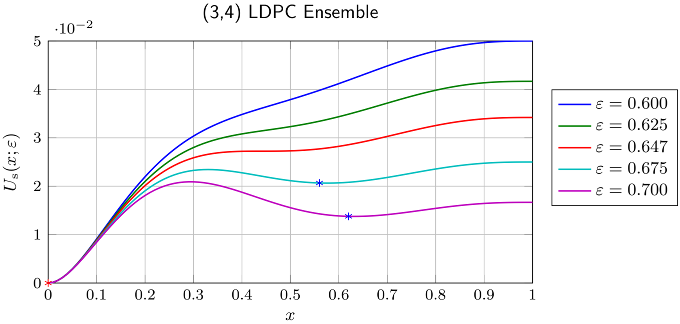
The Potential Function and Threshold Saturation



Let the potential function $U_\mathrm{s} \colon \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

$$U_\mathrm{s}(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z) \mathrm{d}z.$$

The Potential Function and Threshold Saturation



Let the potential function $U_s\colon \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

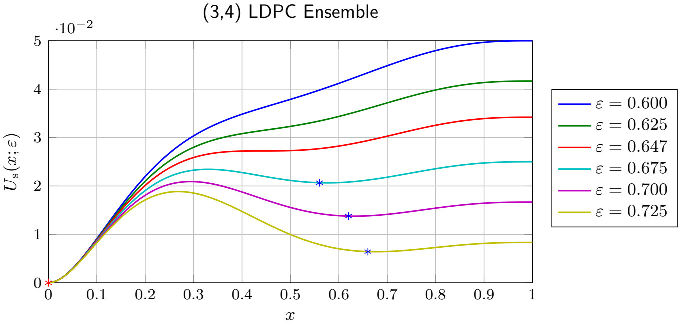
$$U_s(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z)\mathrm{d}z.$$

Threshold Saturation

If $f(g(x)) < x$ for $x \in (0,\delta)$ and $\min_{x\in[\delta,1]} U_s(x) > 0$, then $\exists w_0 < \infty$:

for $w > w_0$, only fixed point of coupled recursion is $\underline{x}^{(\infty)} = \mathbf{0}$

The Potential Function and Threshold Saturation



Let the potential function $U_s : \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

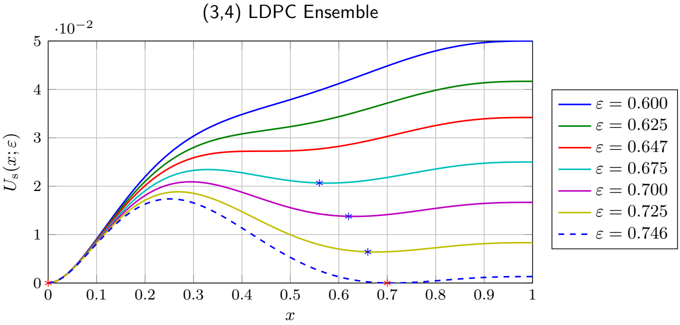
$$U_s(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z) \mathrm{d}z.$$

Threshold Saturation

If $f(g(x)) < x$ for $x \in (0, \delta)$ and $\min_{x \in [\delta, 1]} U_s(x) > 0$, then $\exists w_0 < \infty$:

for $w > w_0$, only fixed point of coupled recursion is $\underline{x}^{(\infty)} = \mathbf{0}$

The Potential Function and Threshold Saturation



Let the potential function $U_{\mathrm{s}} \colon \mathcal{X} \to \mathbb{R}$ of the scalar recursion be

$$U_{\mathrm{s}}(x) \triangleq \int_0^x \big(z - f\left(g(z)\right)\big) g'(z) \mathrm{d}z.$$

Threshold Saturation

If $f(g(x)) < x$ for $x \in (0, \delta)$ and $\min_{x \in [\delta, 1]} U_{\mathrm{s}}(x) > 0$, then $\exists w_0 < \infty$:

for $w > w_0$, only fixed point of coupled recursion is $\underline{x}^{(\infty)} = \mathbf{0}$

A Little History: Threshold Saturation Proofs

For:

- the BEC by KRU in 2010
  - Established many properties and tools used by later approaches
- the Curie-Weiss model in physics by HMU in 2010
- CDMA using a GA by TTK in 2011
- CDMA with outer code via GA by Truhachev in 2011
- compressive sensing using a GA by DJM in 2011
- regular codes on BMS channels by KRU in 2012
- increasing scalar and vector recursions by YJNP in 2012
- irregular LDPC codes on BMS channels by KYMP in 2012
- non-decreasing scalar recursions by KRU in 2012

Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

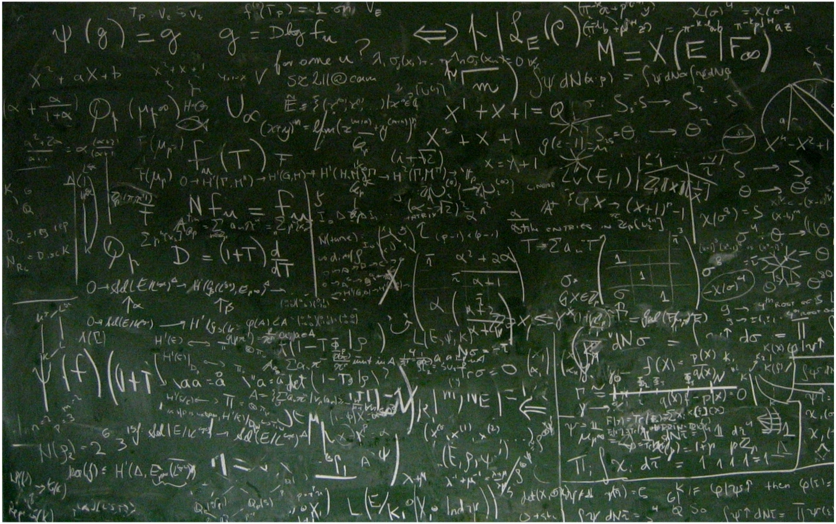
Chalkboard Proof

General Factor Graphs

Wyner-Ziv and Gelfand-Pinsker

Conclusions

Thus, it follows naturally that...



Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

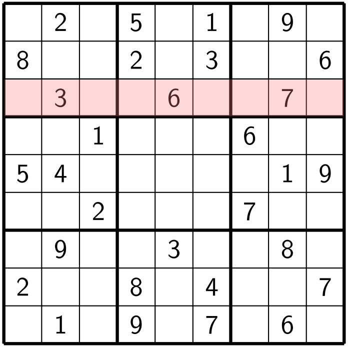
Chalkboard Proof

General Factor Graphs

Wyner-Ziv and Gelfand-Pinsker

Conclusions

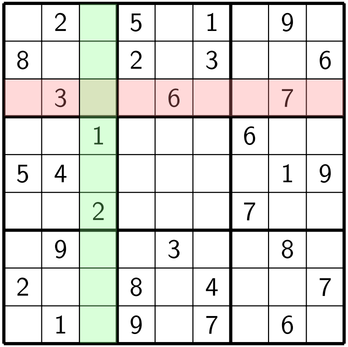
Sudoku: A Well-Known Example



	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	

rows are permutations of $\{1, 2, \ldots, 9\}$

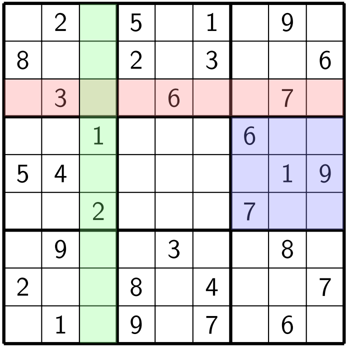
Sudoku: A Well-Known Example



	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	

rows are permutations of $\{1, 2, \ldots, 9\}$
columns are permutations of $\{1, 2, \ldots, 9\}$

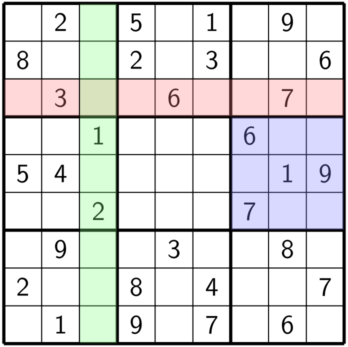
Sudoku: A Well-Known Example



	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	

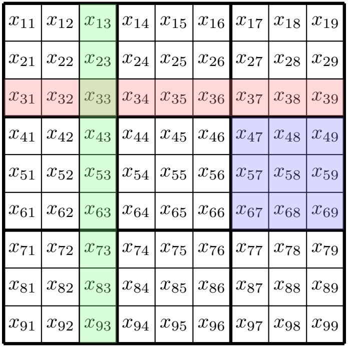
rows are permutations of $\{1, 2, \ldots, 9\}$
columns are permutations of $\{1, 2, \ldots, 9\}$
subblocks are permutations of $\{1, 2, \ldots, 9\}$

Sudoku: A Well-Known Example



	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	

rows are permutations of $\{1, 2, \ldots, 9\}$
columns are permutations of $\{1, 2, \ldots, 9\}$
subblocks are permutations of $\{1, 2, \ldots, 9\}$



x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	x_{46}	x_{47}	x_{48}	x_{49}
x_{51}	x_{52}	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	x_{58}	x_{59}
x_{61}	x_{62}	x_{63}	x_{64}	x_{65}	x_{66}	x_{67}	x_{68}	x_{69}
x_{71}	x_{72}	x_{73}	x_{74}	x_{75}	x_{76}	x_{77}	x_{78}	x_{79}
x_{81}	x_{82}	x_{83}	x_{84}	x_{85}	x_{86}	x_{87}	x_{88}	x_{89}
x_{91}	x_{92}	x_{93}	x_{94}	x_{95}	x_{96}	x_{97}	x_{98}	x_{99}

implied factor graph has
81 variable and 27 factor nodes

$$f(\underline{x}) = \left(\prod_{i=1}^{9} f_\sigma(x_{i*})\right) \left(\prod_{j=1}^{9} f_\sigma(x_{*j})\right) \left(\prod_{k=1}^{9} f_\sigma(x_{B(k)})\right) \prod_{(i,j)\in O} \mathbb{I}(x_{ij} = y_{ij})$$

Solving Sudoku via Marginalization

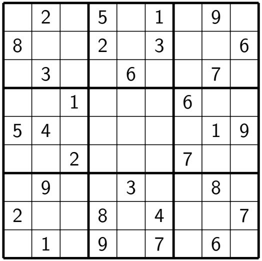
- Consider any constraint satisfaction problem with erased entries
  - One can write $f(\underline{x})$ as the product of indicator functions
  - Some factors force $\underline{x}$ to be valid (i.e., satisfy constraints)
  - Other factors force $\underline{x}$ to be compatible with observed values
  - Summing over $\underline{x}$ counts the $\#$ of valid compatible sequences

Solving Sudoku via Marginalization

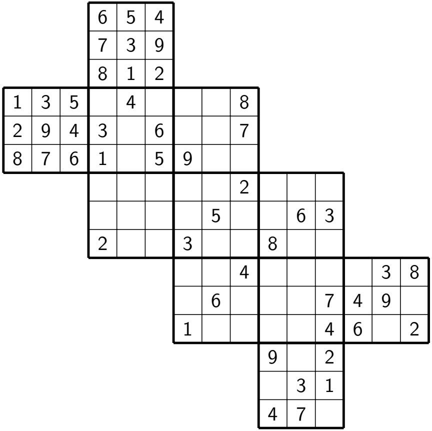
- Consider any constraint satisfaction problem with erased entries
  - One can write $f(\underline{x})$ as the product of indicator functions
  - Some factors force $\underline{x}$ to be valid (i.e., satisfy constraints)
  - Other factors force $\underline{x}$ to be compatible with observed values
  - Summing over $\underline{x}$ counts the # of valid compatible sequences

- Marginalization allows uniform sampling from valid compatible set
  - Sample $x_1' \sim g_1(\cdot)$, fix $x_1 = x_1'$, sample $x_2' \sim g_2(\cdot|x_1)$, etc...
    - For Sudoku, this always works because only one solution!
  - Low complexity if factor graph forms a tree
  - If not a tree, low-complexity approximation sometimes possible
  - But, in general, marginalization is #P-complete

Spatial Coupling: Beyond LDPC

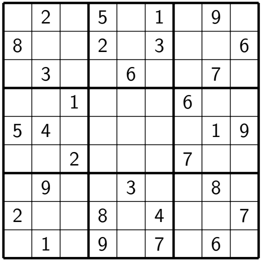


	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	

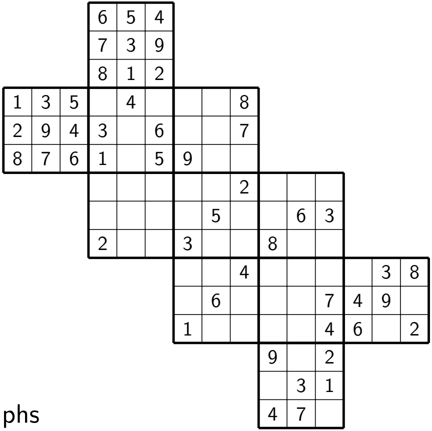


			6	5	4									
			7	3	9									
			8	1	2									
1	3	5		4				8						
2	9	4	3		6			7						
8	7	6	1		5	9								
								2						
							5			6	3			
			2			3			8					
								4					3	8
							6				7	4	9	
						1					4	6		2
									9		2			
										3	1			
									4	7				

Spatial Coupling: Beyond LDPC



	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	



			6	5	4									
			7	3	9									
			8	1	2									
1	3	5		4				8						
2	9	4	3		6			7						
8	7	6	1		5	9								
								2						
							5			6	3			
			2			3			8					
								4					3	8
							6				7	4	9	
						1					4	6		2
									9		2			
										3	1			
									4	7				

- Spatially-Coupled Factor Graphs
  - Variable nodes have a natural global orientation
  - Boundaries help variables to be recovered in an ordered fashion

Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

Chalkboard Proof

General Factor Graphs

Wyner-Ziv and Gelfand-Pinsker

Conclusions

Rate-Distortion, Wyner-Ziv, and Gelfand-Pinsker

- Rate Distortion (RD) Problem
  - What is the minimum data rate to transmit a source with average distortion less than D?

Rate-Distortion, Wyner-Ziv, and Gelfand-Pinsker

- Rate Distortion (RD) Problem
  - What is the minimum data rate to transmit a source with average distortion less than D?
- Wyner-Ziv (WZ) Problem
  - WZ extends RD to the case of side-information at the decoder

Rate-Distortion, Wyner-Ziv, and Gelfand-Pinsker

- Rate Distortion (RD) Problem
  - What is the minimum data rate to transmit a source with average distortion less than D?
- Wyner-Ziv (WZ) Problem
  - WZ extends RD to the case of side-information at the decoder
- Gelfand-Pinsker (GP) Problem
  - Channel coding with non-causal side-information at transmitter

Rate-Distortion, Wyner-Ziv, and Gelfand-Pinsker

- Rate Distortion (RD) Problem
  - What is the minimum data rate to transmit a source with average distortion less than D?
- Wyner-Ziv (WZ) Problem
  - WZ extends RD to the case of side-information at the decoder
- Gelfand-Pinsker (GP) Problem
  - Channel coding with non-causal side-information at transmitter
- WZ and GP problems arise naturally in network information theory

Belief-Propagation Guided Decimation (BPGD)

- RD-type problems are challenging for graph codes with BP decoding
  - They require quantization of an arbitrary sequence to a codebook
  - BP converges only if received sequence is “close” to a codeword
  - But, vanishing fraction of total space is “close” to codewords

Belief-Propagation Guided Decimation (BPGD)

- RD-type problems are challenging for graph codes with BP decoding
  - They require quantization of an arbitrary sequence to a codebook
  - BP converges only if received sequence is “close” to a codeword
  - But, vanishing fraction of total space is “close” to codewords

- When the received vector is not “close” to a codeword
  - BP decoder typically converges to a non-informative fixed point
  - There are exponentially many codewords with low distortion
  - But, the decoder just cannot pick one
  - The bias of a bit is defined to be $|LLR| = \left|\ln \frac{P(X=0)}{P(X=1)}\right|$

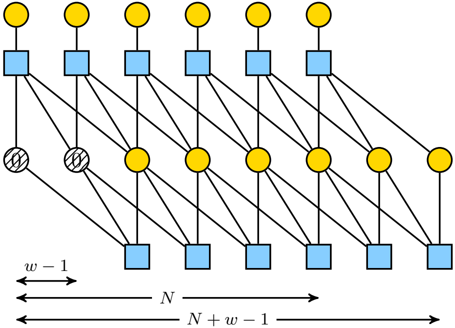
Belief-Propagation Guided Decimation (BPGD)

- RD-type problems are challenging for graph codes with BP decoding
  - They require quantization of an arbitrary sequence to a codebook
  - BP converges only if received sequence is "close" to a codeword
  - But, vanishing fraction of total space is "close" to codewords

- When the received vector is not "close" to a codeword
  - BP decoder typically converges to a non-informative fixed point
  - There are exponentially many codewords with low distortion
  - But, the decoder just cannot pick one
  - The bias of a bit is defined to be $|LLR| = \left| \ln \frac{P(X=0)}{P(X=1)} \right|$

- To force convergence, bits are sequentially "decimated":
  1. The BP decoder is run for a fixed number of iterations
  2. A bit with large bias is sampled and "decimated"

Once Again, Spatial-Coupling Comes to the Rescue



- Rate Distortion
  - SC low-density generator matrix (LDGM) codes can approach the RD limit with BPGD [AMUV12]
- Wyner-Ziv and Gelfand-Pinsker
  - SC compound LDGM/LDPC codes can approach the WZ/GP limits with BPGD [KVNP14]

Outline

Review of LDPC Codes and Density Evolution

Spatially-Coupled Graphical Models

Universality for Multiuser Scenarios

Abstract Formulation of Threshold Saturation

Chalkboard Proof

General Factor Graphs

Wyner-Ziv and Gelfand-Pinsker

Conclusions

Summary and Open Problems

- Spatial Coupling
  - Powerful technique for designing and understanding factor graphs
  - Related to the statistical physics of supercooled liquids
  - General proof of threshold saturation for scalar recursions
- Interesting Open Problems
  - Clever constructions to reduce the rate-loss due to termination
  - Finding new problems where SC gives real benefits
  - Proving SC codes with decimation achieve the rate-distortion limit

Thanks for your attention