Constraint Satisfaction through GBP-Guided Deliberate Bit Flipping

Mohsen Bahrami and Bane Vasić

Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721, USA
bahrami@email.arizona.edu and vasic@ece.arizona.edu

Abstract. In this paper, we consider the problem of transmitting binary messages over data-dependent two-dimensional channels. We propose a *deliberate bit flipping* coding scheme that removes channel harmful configurations prior to transmission. In this method, user messages are encoded with an error correction code, and therefore the number of bit flips should be kept small not to overburden the decoder. We formulate the problem of minimizing the number of bit flips as a binary constraint satisfaction problem, and devise a *generalized belief propagation* guided method to find approximate solutions. Applied to a data-dependent binary channel with the set of 2-D isolated bit configurations as its harmful configurations, we evaluated the performance of our proposed method in terms of uncorrectable bit-error rate.

Keywords: Probabilistic inference · Graphical models · Generalized belief propagation.

1 Introduction

Many of probabilistic inference problems can be reformulated as the computation of marginal probabilities of a joint probability distribution over the set of solutions of a constraint satisfaction problem (CSP) [1, 2]. A CSP consists of a number of variables and a number of constraints, where each constraint specifies admissible values of a subset of variables. A solution to a CSP is an assignment of variables satisfying all the constraints. Message passing algorithms have been successfully used for solving hard CSPs [3]. Traditional low-complexity approximate algorithms for solving these problems are based on belief propagation (BP) [4, 5] which operate on factor graphs. BP, as an algorithm to compute marginals over a factor graph, has its roots in the broad class of Bayesian inference problems [6]. It is well known that the BP algorithm gives exact inference only on cycle-free graphs (trees). It has been also observed that in some applications BP surprisingly can provide close approximations to exact marginals on loopy graphs. However, an understanding of the behavior of BP in the latter case is far from complete. Moreover, it is known that BP does not perform well on graphs which contain a large number of short cycles. A new class of message-passing algorithm called generalized belief propagation (GBP) is introduced in [7] to solve the problem of computing marginal probability distributions on factor graphs with short cycles. The algorithm relies on the extension of cluster variation method [8, 9], which is called the region graph method. The GBP algorithm provides approximate marginals by minimizing the Gibbs free energy using region graph method. In GBP, messages are sent among clusters of variables nodes instead of the node-to-node message passing fashion in BP and SP. More recently GBP has been shown empirically to have good performance, in either accuracy or convergence properties, for certain applications [10, 11].

In this paper, we consider the problem of transmitting a binary message over a data-dependent communication channel and recovering it back at the receiver side. This problem is one of the most fundamental problems in communication theory, and can be considered as an instance of a CSP. Shannon in his seminal work [12] introduced two coding schemes for reliable transmission of information over a noisy channel, namely error correction coding and constrained coding. The first method protects user messages against random errors, which are independent of input data, by introducing redundancy in the messages prior to transmission. On the other hand, a constrained coding method assumes that channel solely introduces errors in response to specific patterns in input messages, and removing these problematic patterns makes the channel noiseless. Recent advances in emerging data storage technologies like magnetic recording systems [13, 14], optical recording devices [15] and flash memory drives [16] necessitate to study two-dimensional coding (2-D) techniques for reliable storage of information. In these systems, user information bits are arranged into 2-D arrays for storing over the recording channel, and occurrences of specific patterns in input arrays are the significant cause of errors during read-back process. These systems require the use of some form of error-correction coding in addition to constrained coding of the input data or symbol sequences. It is therefore natural to investigate the interplay between these two forms of coding and the possibilities for efficiently combining their functions into a single coding operation. For this purpose, we introduce a generic 2-D channel with a set of harmful configurations to model patterning effects on an information bit from its neighboring bits in a 2-D channel input array. In this model, information bits contained in the harmful configurations are more vulnerable to errors than the other bits. Different 2-D constrained coding methods have been proposed to remedy the patterning effects in data-dependent 2-D channels, e.g., [17–22]. The goal of most of these methods is to achieve tighter bounds on the Shannon noiseless channel capacity of constraint. However, these schemes are non-linear in nature, and their encoder/decoder has a memory. Therefore, combinations of these methods with an error-correction coding scheme are challenging, and even a small number of bit errors can result multiple errors and severely degrade the performance of an error correction decoder. As an alternative coding scheme to address the non-linear effects of conventional 2-D constrained coding schemes, we present a *deliberate bit flipping* (DBF) coding scheme for data-dependent 2-D channels, where passing through channel specific patterns in inputs are the main cause of errors. The user message is first encoded by an error correction code, and is arranged into a 2-D array as an input to the channel. The idea is to completely eliminate a constrained encoder and, instead, to remove the harmful configurations by deliberately flipping the selected bits prior to transmission. The DBF method relies on the error correction capability of the error correction code (ECC) being used so that it should be able to correct both deliberate errors and channel errors. Therefore, it is crucial to keep the number of flipped bits small in order not to overburden the error correction decoder.

The problem of minimizing the number of deliberate bit flips for removing a set of configurations from a 2-D array is an instance of a CSP, where variables are arranged into a 2-D array, and constraints are defined locally over a set of neighboring variables. Assignments to variables are chosen from encoded messages of information bits (the codewords of ECC being used), and a constraint is violated if the realization of the neighboring variables involved in the constraint belongs to the given set of configurations. An initial realization of variables may violate some of constraints, and the goal is to change values of minimum number of variables to make all the constraints satisfied. This is equivalent to removing the forbidden configurations entirely from the 2-D array by flipping minimum number of bits. Using a factor graph representation, we devise a constrained combinatorial formulation for minimizing the number of bit flips in the DBF scheme for removing a given set of configurations. We find an approximate solution by reformulating the minimization problem as a 2-D maximum *a posteriori* (MAP) problem using a probabilistic graphical model. In this framework, patterns which do not contain harmful configurations are assumed to be uniformly distributed, and each pattern containing a harmful configuration has zero probability. The GBP algorithm, as a MAP inference method, is used to find the approximate solution for the 2-D MAP problem. Applied to a data-dependent 2-D channel with 2-D *isolated bit patterns* as the set of harmful patterns for the channel, we have shown the performance of DBF method in terms of uncorrectable bit-error rate.

The organization of the paper is as follows. Section 2 introduces the data-dependent 2-D channel model. The DBF coding scheme is presented in Section 3. Section 4 explains the probabilistic formulation devised for minimizing the number of bit flips in DBF coding scheme. Numerical results are given in Section 5.

2 Channel Model

In this section, we present a data-dependent 2-D communication channel which transmits binary rectangular patterns and produces as an output a binary pattern. Passing through the channel, information bits belong to a predefined set of configurations are more prone to errors than the other bits. The channel is characterized by this set of binary configurations, which is called the set of harmful configurations and is denoted by $\mathcal{F}$.

The set of channel input patterns and the set of channel output patterns are denoted by $\mathcal{X}$ and $\mathcal{Y}$. An input pattern $\mathbf{x} = [x_{i,j}]$ is chosen uniformly and randomly from $\mathcal{X}$, and is transmitted through the channel. A pattern $\mathbf{y} = [y_{i,j}] \in \mathcal{Y}$ is observed through the channel. The input pattern $\mathbf{x}$ can be considered as a

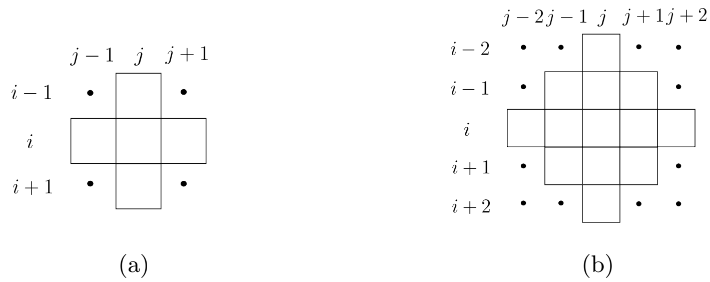

Fig. 1. Fig. shows (a) $Q^+(i,j)$ and (b) $\mathcal{P}_{i,j}$ over the lattice $\mathbb{Z}^2$ for the case of cross-shaped polyomino.

square binary tiling of a rectangle, where each information bit $x_{i,j}$ on the 2-D input pattern represents a colored tile (0 (1) refers to a white (black) tile). The channel is data-dependent, and for each tile $x_{i,j}$, error is characterized by a Bernoulli random variable which depends on the realization of polyominoes having intersection with this tile. A polyomino of order k is constructed by joining k square tiles. Here we consider cross-shaped polyominoes of order 5 which are defined over the 2-D lattice $\mathbb{Z}^2$ as the following

$$Q^+(i,j) = \{(i,j-1),(i-1,j),(i,j),(i,j+1),(i+1,j)\}\,. \tag{1}$$

The set of cross-shaped polyominoes that have intersection with tile $x_{i,j}$ over an $m \times n$ rectangle is identified by

$$\mathcal{P}_{i,j} = \bigcup_{(i',j') \in Q^+(i,j)} Q^+(i',j'). \tag{2}$$

Fig. 1 shows $Q^+(i,j)$ and $\mathcal{P}_{i,j}$ on a 2-D lattice $\mathbb{Z}^2$.

The received tile $y_{i,j}$ is characterized by

$$y_{i,j} = x_{i,j} \oplus z_{i,j}, \tag{3}$$

where $z_{i,j}$ is a Bernoulli random variable which depends on the realization of $\mathcal{P}_{i,j}$, $\mathbf{x}_{\mathcal{P}_{i,j}}$, and is defined by

$$z_{i,j} \sim \begin{cases} \text{Bern}(\alpha_b), & \mathbf{x}_{\mathcal{P}_{i,j}} \in \mathcal{F}, \\ \text{Bern}(\alpha_g), & \mathbf{x}_{\mathcal{P}_{i,j}} \notin \mathcal{F}. \end{cases} \tag{4}$$

Passing through the channel, colors of input tiles belong to $\mathcal{F}$ invert with probability α_b, while colors of other tiles invert with probability α_g. Since patterns belong to the set $\mathcal{F}$ are the main source of errors for this communication channel, we have $\alpha_b \gg \alpha_g$.

The introduced channel has two states where in each state acts as a binary symmetric channel with a different cross-over probability, and can be considered as an instance of the Gilbert-Elliot channel [23]. However, the state transitions in the introduced channel depends on input data which makes the problem of designing capacity achieving codes difficult. As we explain in the following section, we introduce a deliberate bit flipping coding strategy for this communication channel to overcome the effects of harmful configurations.

3 Deliberate Bit Flipping Coding Method

In this section, we characterize the deliberate bit flipping coding strategy for removing harmful configurations from 2-D channel input patterns before transmission through a data-dependent 2-D channel.

A user binary message $\mathbf{m}$ of length K is given. The message $\mathbf{m}$ is first encoded by an error correction code with rate $R = \frac{K}{N}$, and we have the codeword $\mathbf{c}$ of length N. The codeword is arranged into a 2-D array $\mathbf{x} = [x_{i,j}]$ of size $m \times n$, where $x_{i,j} = c_{(i-1)m+j}$ and $N = m \times n$. For each tile $x_{i,j}$, a 2-D constraint is defined over polyominoes having intersection with this tile. The 2-D constraint $\mathbb{S}$ forbids some of the configurations of $\mathcal{P}_{i,j}$, where the set of these configurations are denoted by $\mathcal{F}$. These configurations are essentially harmful configurations for the channel, and they must be removed before transmission. We use a deliberate error insertion approach to remove the harmful configurations from the input pattern $\mathbf{x}$ before transmission through the channel. Whenever there is a configuration from the list $\mathcal{F}$ in the input pattern $\mathbf{x}$, the color of selected tiles in $\mathbf{x}$ are inverted to remove the forbidden configurations. In the following, we present an example to highlight the basic ideas behind the DBF method for removing a set of predefined configurations from a 7×7 random binary pattern. In this example, the set of 2-D isolated bit patterns are required to be removed form the given random pattern.

Example 1. A 7×7 random binary pattern $\mathbf{x}$ as shown in Fig. 2 is given. The goal is to use the DBF scheme to remove the 2-D isolated bit configurations. We assume zero entries (white tiles) outside of $\mathbf{x}$, i.e., $x_{i,j} = 0$, while $i < 1$, $j < 1$, $i > 7$, or $j > 7$. There are two isolated bit patterns in $\mathbf{x}$, which are $\mathbf{x}_{Q^+(3,6)}$ and $\mathbf{x}_{Q^+(7,7)}$. Passing through the channel, the tiles whose belong to these two patterns are more prone to errors than the other tiles. These tiles are $(2,6)$, $(3,5)$, $(3,6)$, $(3,7)$, $(4,6)$, $(6,7)$, $(7,6)$ and $(7,7)$. For instance, for the tile $(2,6)$,

$$\mathcal{P}_{2,6} = \bigcup_{(i',j') \in Q^+(2,6)} Q^+(i',j'). \tag{5}$$

Since $Q^+(3,6) \subset \mathcal{P}_{2,6}$ and $\mathbf{x}_{Q^+(3,6)}$ is a 2-D isolated bit pattern, we have $\mathbf{x}_{\mathcal{P}_{2,6}}$ contains a 2-D isolated bit pattern. Similarly, we can verify this for the rest of tiles in $\mathbf{x}$. 2-D isolated bit configurations can be removed form $\mathbf{x}$ by inverting the colors of tiles $(3,6)$ and $(7,7)$.

In the DBF method, the main role is to select tiles whose colors need to be inverted for removing the harmful configurations. We define a *tile-selection* function to determine these tiles.

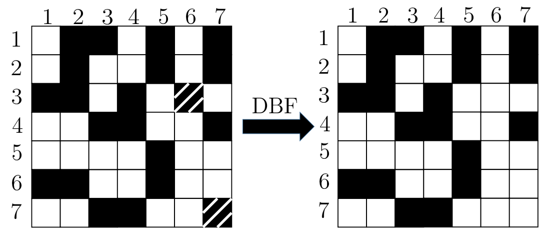

Fig. 2. In order to remove the 2-D isolated bit patterns from the given 7×7 binary pattern, the colors of tiles $(3, 6)$ and $(7, 7)$ are inverted.

Definition 1 (Tile-Selection Function). *The tile-selection function* $\theta : \mathcal{X} \to \{0,1\}^{m \times n}$ *selects tiles whose colors need to be inverted for removing the harmful configurations from the pattern* $\mathbf{x}$.

Using θ, $\mathbf{e}^{\mathrm{DBF}}$ is defined to identify the locations of tiles whose colors are inverted,

$$\mathbf{e}^{\mathrm{DBF}} = \theta(\mathbf{x}) = [e_{i,j}^{\mathrm{DBF}}], \tag{6}$$

where $e_{i,j}^{\mathrm{DBF}} = 1$ if the color of (i, j)-th tile is inverted, otherwise, $e_{i,j}^{\mathrm{DBF}} = 0$. Therefore, $\mathbf{x} \oplus \mathbf{e}^{\mathrm{DBF}}$ does not contain any harmful configurations from the list $\mathcal{F}$. Furthermore, the number of tiles whose colors are inverted is $w_H(\mathbf{e}_{\mathrm{DBF}})$. Now, instead of $\mathbf{x}$, we send $\mathbf{x} \oplus \mathbf{e}^{\mathrm{DBF}}$ over the channel, and the received pattern is $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}^{\mathrm{DBF}} \oplus \mathbf{e}^{\mathrm{CH}}$, where $\mathbf{e}^{\mathrm{CH}}$ indicates the locations of tiles whose colors are inverted due to channel errors. A decoder $\psi : \{0,1\}^{m \times n} \to \mathcal{X}$ maps a received pattern $\mathbf{y}$ to a pattern $\hat{\mathbf{x}}$ in the input set $\mathcal{X}$. In the following, we define the average probability of error and the capacity of the method.

Definition 2 (Average Probability of Error). $\lambda_{\mathbf{m}} = p(\hat{\mathbf{m}} \neq \mathbf{m}|\mathbf{m})$ *is the probability that the decoded message* $\hat{\mathbf{m}}$ *is different from the actual message* $\mathbf{m}$. *The average probability of error is defined by*

$$p_e^{(N)} = p(\hat{\mathbf{m}} \neq \mathbf{m}) = \sum_{\mathbf{m} \in \mathcal{M}} \lambda_{\mathbf{m}} p(\mathbf{m}) = \frac{1}{2^{\lceil NR \rceil}} \sum_{\mathbf{m}} \lambda_{\mathbf{m}}. \tag{7}$$

Definition 3 (Achievable Rate and Capacity). *A rate R is said to be achievable if for some N and $\epsilon_N > 0$, $p_e^{(N)} \leq \epsilon_N$. The capacity is defined as the supremum over all achievable rates.*

In this communication system with DBF method, there are two types of error. The first type of error is the deliberate errors which are introduced before transmission through the channel, and the second type is the random channel errors. If we assume that the main cause of errors are the presence of harmful patterns in input patterns, removing the harmful configurations makes the channel almost noiseless. Therefore, for the Hamming distance between the input and received patterns, we have

$$d_H(\mathbf{x}, \mathbf{y}) = w_H(\mathbf{x} \oplus \mathbf{y}) \simeq w_H(\mathbf{e}^{\text{DBF}}). \tag{8}$$

Without loss of generality, if we use a bounded-distance decoder, it should be ideally decode all the messages that

$$d_H\left(\mathbf{x}(\mathbf{m}), \mathbf{y}\right) \simeq w_H(\mathbf{e}^{\text{DBF}}) \leq \lfloor \frac{d_{\min} - 1}{2} \rfloor, \tag{9}$$

where $d_{\min}$ is the minimum distance of the code. Therefore, the main obstacle for using the DBF method for removing harmful configurations is to keep the number of deliberate errors small enough not to overburden the decoder. For a given binary user message and a set of forbidden configurations, we are interested in finding $\hat{\mathbf{x}}$, that minimizes $w_H(\hat{\mathbf{x}} \oplus \mathbf{x})$ and $\hat{\mathbf{x}} \in \mathbb{S}$. This minimization problem can be considered as a constrained combinatorial optimization problem. Finding a binary pattern which satisfies a certain local constraints (which do not contain a predefined set of 2-D configurations), and has the minimum Hamming distance with the input binary pattern $\mathbf{x}$ via an exhaustive search can be computationally prohibitive for large patterns. This problem can be regarded as an instance of the Levenshtine distance problem [24], which is known to be a hard combinatorial problem. In the following section, we present a probabilistic graphical model, and reformulate the problem as a maximum *a-posetriori* (MAP) problem to find an approximation solution for the problem.

4 A Probabilistic Formulation for DBF Method

In this section, we present a probabilistic formulation for the problem of minimizing the number of bit flips in the DBF scheme. In this framework, the set of input patterns which do not contain any harmful configurations has uniform distribution, while the patterns containing harmful configurations have zero probability. For a given random input pattern, the problem originally is to find the pattern which does not contain any harmful configurations, and has the minimum Hamming distance with the given input pattern. We translate this problem into the problem of finding the most likely pattern (that does not contain any harmful configurations) to the given pattern using a binomial expression.

An input pattern $\mathbf{x}$ is given. For each tile $x_{i,j}$ over $\mathbf{x}$, existence of harmful configurations is determined based on the configuration of $\mathcal{P}_{i,j}$, $\mathbf{x}_{\mathcal{P}_{i,j}}$. Therefore, the problem of finding $\hat{\mathbf{x}} \in \mathbb{S}$ which has the minimum $w_H(\hat{\mathbf{x}} \oplus \mathbf{x})$ can be break down locally over each $\mathcal{P}_{i,j}$. We define a *local distortion function* D over $\mathcal{P}_{i,j}$'s to determine the Hamming distance between $\hat{\mathbf{x}}_{\mathcal{P}_{i,j}}$ and $\mathbf{x}_{\mathcal{P}_{i,j}}$. The function $D : \{0,1\}^{|\mathcal{P}_{i,j}|} \times \{0,1\}^{|\mathcal{P}_{i,j}|} \rightarrow \mathbb{N}$ is defined over the tiles indexed by $\mathcal{P}_{i,j}$ as follows

$$D\left(\hat{\mathbf{x}}_{\mathcal{P}_{i,j}}, \mathbf{x}_{\mathcal{P}_{i,j}}\right) = \begin{cases} w_H\left(\hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \oplus \mathbf{x}_{\mathcal{P}_{i,j}}\right), & \hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \notin \mathcal{F}, \\ \infty, & \hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \in \mathcal{F}, \end{cases} \tag{10}$$

where the patterns belonging to $\mathcal{F}$ are specified by ∞. One may use the outputs of this function over the tiles in $\mathbf{x}$ to find $\mathbf{x}^\star \in \mathbb{S}$ which has the minimum Hamming distance with $\mathbf{x}$. This process can be intractable for large patterns as it needs to compute the output of D for each tile, which has $2^{|\mathcal{P}_{i,j}|}$ different configurations, and take exponentially large memory just to store. In the following, we present a probabilistic formulation to find approximate solution for this problem using GBP algorithm.

We use a binomial probability expression to reformulate the distortion indicator function defined in Eq. (10), and present a probabilistic formulation for the problem of minimizing the Hamming distance. We assume that the color of each tile contained in a harmful configuration is inverted with the probability $0 < \lambda \leq 1$. For each tile $x_{i,j}$, we define a function $D_p : \{0,1\}^{\mathcal{P}_{i,j}} \times \{0,1\}^{\mathcal{P}_{i,j}} \to \mathbb{R}^{[0,1]}$ over the tiles tiles indexed by $\mathcal{P}_{i,j}$,

$$D_p(\mathbf{x}_{\mathcal{P}_{i,j}}, \hat{\mathbf{x}}_{\mathcal{P}_{i,j}}) = \begin{cases} \lambda^{w_H(\mathbf{e}_{\mathcal{P}_{i,j}})}(1-\lambda)^{|\mathcal{P}_{i,j}|-w_H(\mathbf{e}_{\mathcal{P}_{i,j}})}, & \hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \notin \mathcal{F}, \\ 0, & \hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \in \mathcal{F}, \end{cases} \tag{11}$$

where $\mathbf{e}_{\mathcal{P}_{i,j}} = \hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \oplus \mathbf{x}_{\mathcal{P}_{i,j}}$, and $|\mathcal{P}_{i,j}|$ indicates the number of tiles in $\mathcal{P}_{i,j}$. This function is called as the *local probabilistic distortion* function. For each tile $(i,j) \in \mathcal{A}_{m,n}$, the distortion now is defined as the probability of having a distorted pattern $\mathbf{x}_{\mathcal{P}_{i,j}}$ which has the Hamming distance $w_H(\hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \oplus \mathbf{x}_{\mathcal{P}_{i,j}})$ with $\hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \notin \mathcal{F}$. When $\hat{\mathbf{x}}_{\mathcal{P}_{i,j}} \in \mathcal{F}$, this probability is set to be zero, as the first constraint is to find $\hat{\mathbf{x}} \in \mathbb{S}$.

For a given input pattern $\mathbf{x}$ and a set of harmful configurations $\mathcal{F}$, the goal is now to find $\hat{\mathbf{x}} \in \mathbb{S}$ that maximizes $p\left(\hat{\mathbf{x}}|\mathbf{x}\right)$, which is equivalent to finding $\hat{\mathbf{x}}$ that minimizes $w_H\left(\hat{\mathbf{x}} \oplus \mathbf{x}\right)$. In another word, we are interested in finding

$$\hat{\mathbf{x}} = \arg\max_{\hat{\mathbf{x}} \in \mathbb{S}} \left\{ p(\hat{\mathbf{x}}|\mathbf{x}) \right\}. \tag{12}$$

The *a-posteriori* probability $p\left(\hat{\mathbf{x}}|\mathbf{x}\right)$ for a fixed λ is factored into

$$\begin{aligned} p\left(\hat{\mathbf{x}}|\mathbf{x}\right) = \frac{p\left(\mathbf{x}|\hat{\mathbf{x}}\right) p\left(\hat{\mathbf{x}}\right)}{p\left(\mathbf{x}\right)} \overset{(a)}{\propto} p\left(\mathbf{x}|\hat{\mathbf{x}}\right) &\overset{(b)}{=} \prod_{(i,j)} p\left(\mathbf{x}_{\mathcal{P}_{i,j}} | \hat{\mathbf{x}}_{\mathcal{P}_{i,j}}\right), \\ &\overset{(c)}{=} \prod_{(i,j)} D_p(\mathbf{x}_{\mathcal{P}_{i,j}}, \hat{\mathbf{x}}_{\mathcal{P}_{i,j}}), \end{aligned} \tag{13}$$

where (a) comes from this assumption that the set of patterns which do not contain harmful configurations has uniform distribution, (b) is established as harmful configurations can be determined locally over $\mathcal{P}_{i,j}$'s, and (c) is obtained according to the local probabilistic distortion function, Eq. (11). Therefore, we have

$$p\left(\hat{\mathbf{x}}|\mathbf{x}\right) = \frac{1}{Z(\mathbf{x})} \prod_{(i,j) \in \mathcal{A}_{m,n}} D_p(\mathbf{x}_{\mathcal{P}_{i,j}}, \hat{\mathbf{x}}_{\mathcal{P}_{i,j}}), \tag{14}$$

where $Z(\mathbf{x})$ is the partition function and defined by

$$Z(\mathbf{x}) = \sum_{\mathbf{x}} \prod_{(i,j)\in\mathcal{A}_{m,n}} D_p(\mathbf{x}_{\mathcal{P}_{i,j}}, \hat{\mathbf{x}}_{\mathcal{P}_{i,j}}). \tag{15}$$

Providing either exact or approximate solutions for the marginal probabilities in general is a NP-hard problem [3], as we need to take sum over exponential number of variables. In [7, 25], it is shown that region-based approximation (RBA) method provides an approximate solution for the partition function by minimizing the region-based free energy (as an approximation to the variational free energy). Therefore, GBP as a method for finding approximate solution for region-based free energy can be used to solve the problem of minimizing the number of bit flips in the DBF scheme.

5 Numerical Results

In this section, we present the numerical results, and explain how the DBF method relies on the error correction capability of the code being used. We first provide an example of a short BCH code with incorporating DBF method.

5.1 Example of BCH-[15,5,7] Code

Consider the user messages of length 5, $\mathbf{m}_1 = (0,1,0,0,0)$, $\mathbf{m}_2 = (1,0,0,0,0)$, $\mathbf{m}_3 = (0,1,1,1,1)$ and $\mathbf{m}_4 = (0,1,1,0,1)$. The messages are encoded by the BCH-[15, 5, 7] code, and the codewords are

$$\mathbf{c}_1 = (0,1,0,0,0,1,1,1,1,0,1,0,1,1,0)\,,\ \mathbf{c}_2 = (1,0,0,0,0,1,0,1,0,0,1,1,0,1,1)\,,$$
$$\mathbf{c}_3 = (0,1,1,1,1,0,1,0,1,1,0,0,1,0,0)\,,\ \mathbf{c}_4 = (0,1,1,0,1,1,1,0,0,0,0,1,0,1,0)\,.$$

The codewords are arranged into 3×5 arrays as four different binary patterns. These patterns are shown in Fig. 3. We want to remove forbidden configurations by the 2-D n.i.b. constraint entirely from the patterns with flipping minimum number of bits. We only focus on these four patterns out of 32 possible binary patterns with BCH-[15, 5, 7] code as they present all different possible bit flipping scenarios for removing 2-D isolated bit patterns.

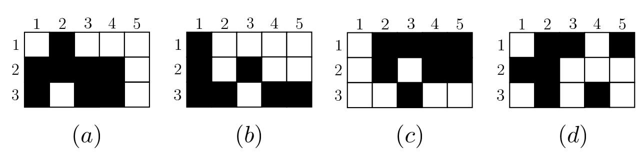

Fig. 3. The input patterns for this example. Outside of these patterns are filled with white tiles (zero entries).

In Fig. 3(a), the pattern does not contain any of the 2-D isolated bit configurations, therefore there is no need to invert the tile colors, and $w_H(\mathbf{e}_{(a)}) = 0$. The pattern in Fig. 3(b) contains single 2-D isolated bit pattern, which is $\mathbf{x}_{Q^+(2,3)}$. This 2-D isolated bit pattern can be removed by inverting the color of any one of the tiles in $Q^+(2,3)$, and therefore the minimum $w_H(\mathbf{e}_{(b)}) = 1$. For the pattern in Fig. 3(c), there are two overlapping 2-D isolated bit patterns, which are $\mathbf{x}_{Q^+(2,3)}$ and $\mathbf{x}_{Q^+(3,3)}$. These two isolated bit patterns can be removed simultaneously by inverting either the color of tile $(2,3)$ or $(3,3)$, and therefore for this case also the minimum $w_H(\mathbf{e}) = 1$. In Fig. 3(d), the pattern contains two non-overlapping 2-D isolated bit patterns, which are $\mathbf{x}_{Q^+(1,5)}$ and $\mathbf{x}_{Q^+(3,4)}$. At least colors of two tiles over this input pattern should be inverted, and for this case the minimum $w_H(\mathbf{e}_{(d)}) = 2$. For the systematic BCH-$[15,5,7]$ code (where the codewords are arranged into 3×5 arrays and the first row is equipped with the user bits), in average it needs to flip 0.6563 bits/pattern to remove the forbidden configurations by the 2-D n.i.b. constraint.

5.2 Uncorrectable Bit-Error Rate

In this section, we present the statistics of the number of bit flips required for removing 2-D isolated bit configurations from a random 2-D pattern of size 32×32, and also compute the uncorrectable bit-error rate (UBER) using these statistics.

The statistics of the number of bit flips are obtained by applying GBP-based DBF method for removing 2-D isolate bit patterns from a set of 8000 random 2-D patterns of size 32×32. Similar to the other examples, we assume white tiles or zero entries outside of patterns. Fig. 4(a) shows the occurrence probability of the number of flipped bits.

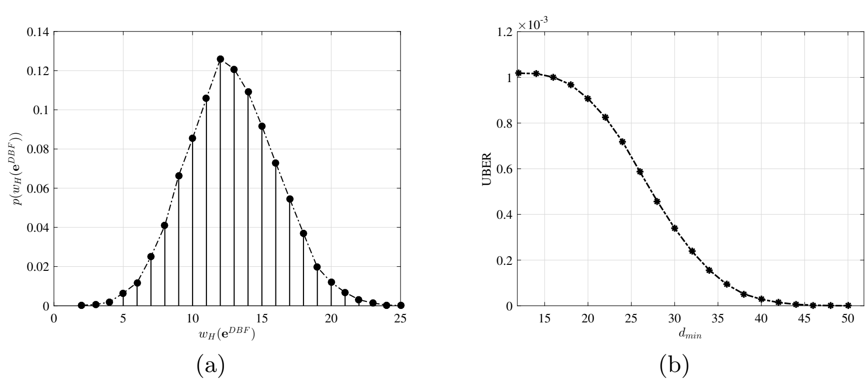

Fig. 4. (a) Fig. presents an occurrence probability of the number of bit flips for removing 2-D isolated bit configurations from a random 2-D pattern of size 32×32. (b) The UBER for the DBF scheme with BCH codes of length 1024 is given.

Using these occurrence probabilities, we can compute the UBER for an ECC being used, as follows

$$\text{UBER} = \left[\sum_{w_H(\mathbf{e}^{\text{DBF}}) > \lfloor \frac{d_{\min}-1}{2} \rfloor} p\left(w_H(\mathbf{e}^{\text{DBF}})\right) \right] / NR, \tag{16}$$

where $d_{\min}$ is the minimum distance of code, $N = m \times n$ is the size of the pattern (length of the code), and R is the rate of the ECC. In fact, we compute the UBER under the assumptions that the channel only introduces errors in response to presences of 2-D isolated bit configurations, and removing these configurations make the channel noiseless. In our introduced channel, this is the case when $\alpha_g = 0$ and $\alpha_b \neq 0$. As an example, we use BCH codes of length 1024 with different code rates, and draw the UBER for these codes in Fig. 4(b).

6 Conclusions

We have presented a deliberate bit flipping coding scheme for data-dependent 2-D channels. For this method, we have shown that the main obstacle is the number of deliberate errors which are introduced for removing harmful configurations before transmission through the channel. We have devised a combinatorial optimization formulation for minimizing the number of bit flips, and have explained how this problem can be related to a binary constraint satisfaction problem. Finally, through an example, we have presented uncorrectable bit-error rate results of incorporating DBF for removing 2-D isolated-bit configurations from 2-D patterns of certain size.

Acknowledgment

This work is funded by the NSF under grants ECCS-1500170 and SaTC-1813401. A comprehensive version of this paper has been submitted to IEEE Transactions on Communications [26].

References

1. M. Jordan, *Learning in Graphical Models.* MIT Press, 1999.
2. F. Kschischang, B. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498 –519, Feb. 2001.
3. M. J. Wainwright and M. I. Jordan, “Graphical models, exponential families, and variational inference,” *Foundations and Trends in Machine Learning*, vol. 1, pp. 1–305, Nov. 2008.
4. J. Pearl, *Probablisitic Reasoning in Intelligent Systems.* San Francisco, CA: Kaufmann, 1988.
5. R. G. Gallager, “Low density parity check codes,” Ph.D. dissertation, Cambridge, MA, 1963.
6. B. J. Frey, *Graphical models for machine learning and digital communication.* Cambridge, MA, USA: MIT Press, 1998.
7. J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Constructing free energy approximations and generalized belief propagation algorithms,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 2282–2312, July 2005.

8. R. Kikuchi, “A Theory of Cooperative Phenomena,” *Physical Review Online Archive (Prola)*, vol. 81, no. 6, p. 988, Mar. 1951.
9. T. Morita, *Foundations and applications of cluster variation method and path probability method*. Publication Office, Progress of Theoretical Physics, 1994.
10. S. Shamai, L. H. Ozarow, and A. D. Wyner, “Information rates for a discrete-time gaussian channel with intersymbol interference and stationary inputs,” *IEEE Trans. on Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov 1991.
11. G. Sabato and M. Molkaraie, “Generalized belief propagation for the noiseless capacity and information rates of run-length limited constraints,” *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 669–675, Mar. 2012.
12. C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
13. R. Wood, M. Williams, A. Kavcic, and J. Miles, “The feasibility of magnetic recording at 10 terabits per square inch on conventional media,” *IEEE Trans. Magn.*, vol. 45, no. 2, pp. 917 –923, Feb. 2009.
14. S. S. Garani, L. Dolecek, J. Barry, F. Sala, and B. Vasić, “Signal processing and coding techniques for 2-D magnetic recording: An overview,” *Proc. of the IEEE*, vol. 106, no. 2, pp. 286–318, Feb. 2018.
15. K. A. S. Immink, “A survey of codes for optical disk recording,” *IEEE J. on Selected Areas in Commun.*, vol. 19, no. 4, pp. 756–764, Apr. 2001.
16. L. Dolecek and Y. Cassuto, “Channel coding for nonvolatile memory technologies: Theoretical advances and practical considerations,” *Proc. of the IEEE*, vol. 105, no. 9, pp. 1705–1724, Sept. 2017.
17. S. Halevy, J. Chen, R. M. Roth, P. H. Siegel, and J. K. Wolf, “Improved bit-stuffing bounds on two-dimensional constraints,” *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 824–838, May 2004.
18. R. M. Roth, P. H. Siegel, and J. K. Wolf, “Efficient coding schemes for the hard-square model,” *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1166–1176, Mar. 2001.
19. S. Forchhammer and T. V. Laursen, “Entropy of bit-stuffing-induced measures for two-dimensional checkerboard constraints,” *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1537–1546, Apr. 2007.
20. I. Tal and R. M. Roth, “Bounds on the rate of 2-D bit-stuffing encoders,” *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2561–2567, Jun. 2010.
21. A. Sharov and R. M. Roth, “Two dimensional constrained coding based on tiling,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1800–1807, Apr. 2010.
22. A. R. Krishnan and B. Vasić, “Lozenge tiling constrained codes,” *Facta Universitatis, Series: Electronics and Energetics*, no. 4, pp. 1–9, October 2014.
23. M. Mushkin and I. Bar-David, “Capacity and coding for the Gilbert-Elliott channels,” *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.
24. V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” *Soviet Physics-Doklady*, vol. 10, no. 8, Feb. 1966.
25. P. Pakzad and V. Anantharam, “Kikuchi approximation method for joint decoding of LDPC codes and partial-response channels.” *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1149–1153, 2006.
26. M. Bahrami and B. Vasić, “A deliberate bit flipping coding scheme for data-dependent two-dimensional channels (under review),” *IEEE Trans. Commun.*, March 2019.