

# A Novel Stochastic Decoding of LDPC Codes with Quantitative Guarantees

Nima Noorshams                      Aravind Iyengar  
 nnoorshams@cal.berkeley.edu      ariyenga@qti.qualcomm.com

Qualcomm Research Silicon Valley  
 Santa Clara, CA, USA

May 27, 2014

## Abstract

Low-density parity-check codes, a class of capacity-approaching linear codes, are particularly recognized for their efficient decoding scheme. The decoding scheme, known as the sum-product, is an iterative algorithm consisting of passing messages between variable and check nodes of the factor graph. The sum-product algorithm is fully parallelizable, owing to the fact that all messages can be update concurrently. However, since it requires extensive number of highly interconnected wires, the fully-parallel implementation of the sum-product on chips is exceedingly challenging. Stochastic decoding algorithms, which exchange binary messages, are of great interest for mitigating this challenge and have been the focus of extensive research over the past decade. They significantly reduce the required wiring and computational complexity of the message-passing algorithm. Even though stochastic decoders have been shown extremely effective in practice, the theoretical aspect and understanding of such algorithms remains limited at large. Our main objective in this paper is to address this issue. We first propose a novel algorithm referred to as the Markov based stochastic decoding. Then, we provide concrete quantitative guarantees on its performance for tree-structured as well as general factor graphs. More specifically, we provide upper-bounds on the first and second moments of the error, illustrating that the proposed algorithm is an asymptotically consistent estimate of the sum-product algorithm. We also validate our theoretical predictions with experimental results, showing we achieve comparable performance to other practical stochastic decoders.

## 1 Introduction

Sparse graph codes, most notably low-density parity-check (LDPC), have been adopted by the latest wireless communication standards [16, 1, 11, 27]. They are known to approach the channel capacity [30, 21, 29, 28]. What makes them even more appealing for practical purposes is their simple decoding scheme [2, 18]. More specifically, LDPC codes are decoded via a message-passing algorithm called the sum-product (SP). It is an iterative algorithm consisting of passing messages between variable and check nodes in the factor graph [2, 18]. The fact that all messages in the SP algorithm can be updated concurrently, makes the fully-parallel implementation—where the factor graph is directly mapped onto the chip—most efficient. However, due to complex and seemingly random connections between check and variable nodes in the factor graph, fully-parallel implementation of the SP is challenging. The wiring complexity has a big impact on the circuit area and power consumption. Also longer, more inter-connected wires can create more parasitic capacitance and limit the clock rate.

Various solutions have been suggested by researchers in order to reduce the implementation complexity of the fully-parallel SP algorithm. Analog circuits have been designed for short LDPC codes [14, 36]. Bit serial algorithms, where messages are transmitted serially over single wires, have been proposed [6, 7, 3, 5]. Splitting row-modules by partitioning check

node operations has been shown to provide substantial gains in the required area and power efficiency [22, 23]. In another prominent line of work, researchers have proposed various stochastic decoding algorithms [10, 35, 33, 34, 24, 19, 20, 31]. They are all based on stochastic representation of the SP messages. More precisely, messages are encoded via Bernoulli sequences with correct marginal probabilities. As a result, the structure of check and variable nodes are substantially simplified and the wiring complexity is significantly reduced. (Benefits of such decoders are discussed in more details in Section 3.2.) Stochastic message-passing have also been used in other contexts, among which are distributed convex optimization and learning [17, 13], efficient belief propagation algorithms [25, 26], and efficient learning of distributions [32].

Although experimental results have proved stochastic decoding extremely beneficial, to date mathematical understanding of such decoders are very limited and largely missing from the literature. Since the output of stochastic decoders are random by construction, it is natural to ask the following questions: how does the stochastic decoder behave on average? can it be tuned to approach the performance of SP and if so how fast? is the average performance typical or do we have a measure of concentration around average? The main contribution of this paper is answering these questions by providing *theoretical analysis* for stochastic decoders. To that end, we propose a novel algorithm, referred to as Markov based stochastic decoding (MbSD), which is amenable to theoretical analysis. We provide quantitative bounds on the first and second moments of the error in terms of the underlying parameters for tree-structured (cycle free) as well as general factor graphs, showing that the performance of MbSD converges to that of SP.

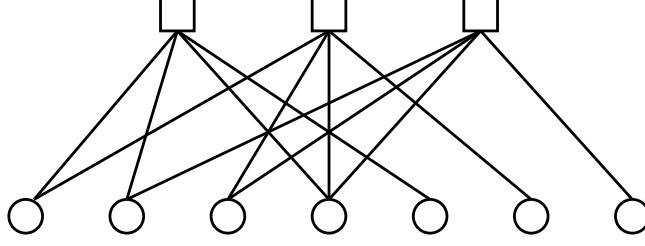
The remainder of this paper is organized as follows. We begin in Section 2 with some background on factor graph representation of LDPC codes, the sum-product algorithm, and stochastic decoding. In Section 3, we turn to our main results by introducing the MbSD algorithm followed by some discussion on its hardware implementation and statements of our main theoretical results (Theorems 1, and 2). Section 4 is devoted to proofs, with some technical aspects deferred to appendices. Finally in Section 5, we provide some experimental results, confirming our theoretical predictions.

## 2 Background and Problem Setup

In this section, we setup the problem and provide the necessary background.

### 2.1 Factor Graph Representation of LDPC Codes

A low-density parity-check code is a linear error-correcting code, satisfying a number of parity check constraints. These constraints are encoded by a sparse parity-check matrix  $H \in \{0, 1\}^{m \times n}$ . More specifically, a binary sequence  $x \in \{0, 1\}^n$  is a valid codeword if and only if  $Hx \stackrel{2}{=} 0$ , where all operations are module two [30]. A popular approach for modeling LDPC codes is via the notion of factor graphs [18]. A factor graph representing an LDPC code with the parity-check matrix  $H$  is a bipartite graph  $\mathcal{G} = (\mathcal{V}, \mathcal{C}, \mathcal{E})$ , consisting of a set of variable nodes  $\mathcal{V} := \{1, 2, \dots, n\}$ , a set of check nodes  $\mathcal{C} := \{1, 2, \dots, m\}$ , and a set of edges connecting variable and check nodes  $\mathcal{E} := \{(i, a) \mid i \in \mathcal{V}, a \in \mathcal{C}, \text{ and } H(a, i) = 1\}$ . (In this paper, we use letters  $i, j, \dots$ , and  $a, b, \dots$  to denote variable and check nodes respectively.) A typical factor graph representing an LDPC code (the Hamming code) is illustrated in Figure 1.



**Figure 1.** Factor graph of the Hamming code. Variable nodes are represented by circles, whereas check nodes are represented by squares.

## 2.2 The Sum-Product Algorithm

Suppose a transmitter sends the codeword  $x$  to a receiver over a memory-less, noisy communication channel. Some channel models that are commonly used in practice include the additive white Gaussian noise (AWGN), the binary symmetric channel, and the binary erasure channel. Having received the impaired signal  $y$ , the receiver attempts to recover the original signal by finding either the global maximum a posteriori (MAP) estimate  $\hat{x} = \arg \max_{Hx=0} \mathbb{P}(x | y)$ , or the bit-wise MAP estimates  $\hat{x}_i = \arg \max_{Hx=0} \mathbb{P}(x_i | y)$ , for  $i = 1, 2, \dots, n$ .

Without exploiting the underlying structure of the code, or equivalently its factor graph, the MAP estimation is intractable and requires an exponential number of operations in the code length. However, this problem can be circumvented using an algorithm called the sum-product (SP), also known as the belief propagation algorithm. The SP is an iterative algorithm consisting of passing messages, in the form of probability distributions, between nodes of the factor graph [2, 18]. It is known to converge to the correct bit-wise MAP estimates for cycle-free factor graphs; however, on loopy graphs, which includes almost all practical LDPC codes, such a guarantee no longer exists. Nonetheless, the SP algorithm has been shown to be extremely accurate and effective in practice [21, 30].

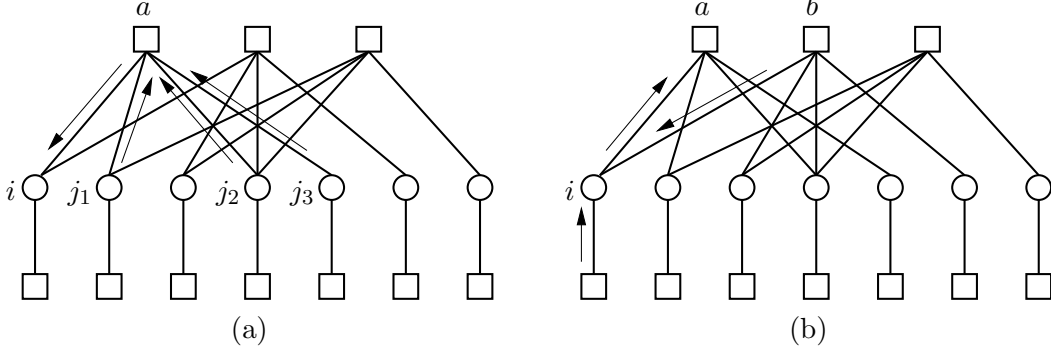
We now turn to the description of the SP algorithm. For every variable node  $i \in \mathcal{V}$  let  $\mathcal{N}(i) := \{a \mid (i, a) \in \mathcal{E}\}$  denote the set of its neighboring check nodes. Similarly define  $\mathcal{N}(a) := \{i \mid (i, a) \in \mathcal{E}\}$ , the set of neighboring variable nodes for every check node  $a \in \mathcal{C}$ . The SP algorithm allocates two messages to every edge  $(i, a) \in \mathcal{E}$ , one for each direction. At each iteration  $t = 0, 1, \dots$ , every variable node  $i \in \mathcal{V}$  (check node  $a \in \mathcal{C}$ ), calculates a message  $0 < \alpha_{i \rightarrow a}^{t+1} < 1$  (message  $0 < \alpha_{a \rightarrow i}^{t+1} < 1$ ) and transmit it to its neighboring check node  $a \in \mathcal{N}(i)$  (variable node  $i \in \mathcal{N}(a)$ ). In updating the messages, every variable node takes into account the incoming messages from its neighboring check nodes as well as the information from the channel, namely  $\alpha_i := \mathbb{P}(x_i = 1 \mid y_i)$ . With this notation at hand, the description of the SP algorithm is as follows: initialize messages from variable to check nodes,  $\alpha_{i \rightarrow a}^0 = \alpha_i$ , and update messages for each edge  $(i, a) \in \mathcal{E}$  and iteration  $t = 0, 1, \dots$  according to

$$\alpha_{a \rightarrow i}^t = \frac{1}{2} - \frac{1}{2} \prod_{j \in \mathcal{N}(a) \setminus \{i\}} (1 - 2 \alpha_{j \rightarrow a}^t), \quad (1)$$

and

$$\alpha_{i \rightarrow a}^{t+1} = \frac{\alpha_i \prod_{b \in \mathcal{N}(i) \setminus \{a\}} \alpha_{b \rightarrow i}^t}{\alpha_i \prod_{b \in \mathcal{N}(i) \setminus \{a\}} \alpha_{b \rightarrow i}^t + (1 - \alpha_i) \prod_{b \in \mathcal{N}(i) \setminus \{a\}} (1 - \alpha_{b \rightarrow i}^t)}. \quad (2)$$

Information flow on a factor graph is shown in Figure 2. Upon receiving all the incoming



**Figure 2.** Graphical representation of message-passing on factor graphs (a) check to variable node (b) variable to check node.

messages, variable node  $i \in \mathcal{V}$  update its marginal probability

$$\mu_i^{t+1} = \frac{\alpha_i \prod_{b \in \mathcal{N}(i)} \alpha_{b \rightarrow i}^t}{\alpha_i \prod_{b \in \mathcal{N}(i)} \alpha_{b \rightarrow i}^t + (1 - \alpha_i) \prod_{b \in \mathcal{N}(i)} (1 - \alpha_{b \rightarrow i}^t)}.$$

Accordingly, the receiver estimates the  $i$ -th bit by  $\hat{x}_i^{t+1} = \mathbb{I}(\mu_i^{t+1} > 0.5)$ , where  $\mathbb{I}(\cdot)$  is the indicator function. It should also be mentioned that in practice, in order to reduce the quantization error, log-likelihood ratios are mostly used as messages. Moreover, to further simplify the SP algorithm, the check node operation (1) is approximated. The resultant is known as the Min-Sum algorithm [8].

### 2.3 Stochastic Decoding of LDPC Codes

Stochastic computation in the context of LDPC decoding was first introduced in 2003 by Gaudet and Rapley [10]. Ever since, much research has been conducted in this field and numerous stochastic decoders have been proposed [35, 33, 34, 24, 20, 31]. For instance, Tehrani et al. [33] introduced and exploited the notions of edge memory and noise dependent scaling in order to make the stochastic decoding a viable method for long, practical, LDPC codes. Estimating the probability distributions via a successive relaxation method, Leduc-Primeau et al. [20] proposed a scheme with improved decoding gain. More recently, Sarkis et al. [31] extended the stochastic decoding to the case of non-binary LDPC codes.

The underlying structure of all these methods and most relevant to our work, however, are the following: they all encode messages by Bernoulli sequences, they all consist of ‘decoding cycles’ which should not be confused with SP iterations (roughly speaking, multiple decoding cycles correspond to one SP iteration.), the check node operation is the module-two sum (i.e. the message transmitted from a check node to a variable node is equal to the module-two sum of the incoming bits.), and finally the variable node operation is the equality (i.e. the message transmitted from a variable node to a check node is equal to one if all incoming bits are one, it is equal to zero if all incoming bits are zero, and it is equal to the previous decoding cycle’s bit in case incoming messages do not agree.). The intuition behind the stochastic variable and check node operations can be obtained from the inspection of SP message updates (1) and (2). More specifically, suppose  $Z_{j \rightarrow a}$ , for  $j \in \mathcal{N}(a) \setminus \{i\}$ , are independent Bernoulli random variables with distributions  $\mathbb{P}(Z_{j \rightarrow a} = 1) = \alpha_{j \rightarrow a}^t$ . Then  $\alpha_{a \rightarrow i}^t$ , derived from equation (1), becomes the probability of having odd number of ones in the sequence  $\{Z_{j \rightarrow a}\}_{j \in \mathcal{N}(a) \setminus \{i\}}$  (see Lemma 1

in the paper [9]). Therefore, the statistically consistent estimate of the check to variable node message is the module-two summation of the incoming bits. Similarly, to understand the stochastic variable node operation, let  $Z_i$  and  $Z_{b \rightarrow i}$ , for  $b \in \mathcal{N}(i) \setminus \{a\}$ , be independent Bernoulli random variables with probability distributions  $\mathbb{P}(Z_i = 1) = \alpha_i$ , and  $\mathbb{P}(Z_{b \rightarrow i} = 1) = \alpha_{b \rightarrow i}^t$ . Then  $\alpha_{i \rightarrow a}^{t+1}$ , derived from equation (2), becomes the probability of the event  $\{Z_i = 1, Z_{b \rightarrow i} = 1, \forall b \in \mathcal{N}(i) \setminus \{a\}\}$ , conditioned on the event  $\{Z_i = Z_{b \rightarrow i}, \forall b \in \mathcal{N}(i) \setminus \{a\}\}$ , thus supporting the intuition that one must transmit the common value from variable to check nodes in case all incoming bits are equal.

### 3 Algorithm and Main Results

In this section, we introduce the MbSD algorithm, discuss its hardware design aspect, and state some theoretical guarantees regarding its performance.

#### 3.1 The Proposed Stochastic Algorithm

The MbSD algorithm consists of passing messages between variable and check nodes of the factor graph. These messages are  $2k$ -dimensional binary vectors, for a fixed  $k$  (design parameter). However, variable and check node updates are element-wise, bit operations. Before stating the algorithm we need to define some notation.

Suppose  $(y_1, y_2, \dots, y_n)$  is the received codeword with the likelihood  $\alpha_i = \mathbb{P}(x_i = 1 | y_i)$ , for  $i = 1, 2, \dots, n$ . Our algorithm, involves messages from the channel to variable nodes at every iteration  $t = 0, 1, 2, \dots$ . More specifically, let  $Z_i^t \in \{0, 1\}^{2k}$  be the  $2k$ -dimensional binary message from the channel to the variable node  $i$  at time  $t$ , with independent and identically distributed (i.i.d.) entries

$$\mathbb{P}(Z_i^t(\ell) = 1) = \alpha_i, \quad \text{for all } \ell = 1, 2, \dots, 2k.$$

Moreover, let  $Z_{i \rightarrow a}^t \in \{0, 1\}^{2k}$  denote the  $2k$ -dimensional binary message from the variable node  $i$  to the check node  $a$  at time  $t = 0, 1, \dots$ . Similarly, let  $Z_{a \rightarrow i}^t \in \{0, 1\}^{2k}$  be the message from the check node  $a$  to the variable node  $i$  at time  $t$ .

We also need to define the element-wise, module-two summation operator  $\oplus$ , as well as the “equality” operator  $\odot$ . Suppose  $X_1, X_2, \dots, X_d$  are arbitrary  $2k$ -dimensional binary vectors. Then, the vector  $Y = \bigoplus_{i=1}^d X_i$  denotes the module-two summation of the vectors  $\{X_i\}_{i=1}^d$ , if and only if

$$Y(\ell) \stackrel{2}{=} X_1(\ell) + X_2(\ell) + \dots + X_d(\ell),$$

for all  $\ell = 1, 2, \dots, 2k$ . Furthermore, by  $Y = \bigodot_{i=1}^d X_i$  we mean

$$Y(\ell) = \begin{cases} 1 & \text{if } X_i(\ell) = 1, \text{ for all } i = 1, 2, \dots, d \\ 0 & \text{if } X_i(\ell) = 0, \text{ for all } i = 1, 2, \dots, d \\ Y(\ell - 1) & \text{otherwise} \end{cases}, \quad (3)$$

for all  $\ell = 1, 2, \dots, 2k$ . Here, we assume  $Y(0)$  is either zero or one, equally likely.

Now, the precise description of the MbSD algorithm is as follows:

1. Initialize messages from variable nodes to check nodes at time  $t = 0$  by  $Z_{i \rightarrow a}^0 = Z_i^0$ .

2. For iterations  $t = 0, 1, 2, \dots$ , and every edge  $(i, a) \in \mathcal{E}$ :

a) Update messages from check nodes to variable nodes

$$Z_{a \rightarrow i}^t = \bigoplus_{j \in \mathcal{N}(a) \setminus \{i\}} Z_{j \rightarrow a}^t.$$

b) Update messages from variable nodes to check nodes by following these steps:

– compute the auxiliary variable

$$W_{i \rightarrow a}^{t+1} = \bigodot_{b \in \mathcal{N}(i) \setminus \{a\}} Z_{b \rightarrow i}^t \bigodot Z_i^{t+1}.$$

– update the message entries  $Z_{i \rightarrow a}^{t+1}(\ell)$ , for  $\ell = 1, 2, \dots, 2k$ , by drawing i.i.d. samples from the set

$$\{W_{i \rightarrow a}^{t+1}(k+1), W_{i \rightarrow a}^{t+1}(k+2), \dots, W_{i \rightarrow a}^{t+1}(2k)\}.$$

c) Compute the binary vector  $U_i^{t+1} = \bigodot_{a \in \mathcal{N}(i)} Z_{a \rightarrow i}^t \bigodot Z_i^{t+1}$  and update the marginal estimates according to

$$\hat{\eta}_i^{t+1} = \frac{1}{k} \sum_{\ell=k+1}^{2k} U_i^{t+1}(\ell), \quad (4)$$

for all  $i = 1, 2, \dots, n$ .

Few comments, regarding the interpretation of the algorithm, are worth mentioning at this point. The check to variable node message update (step (a)) is a statistically consistent estimate of the actual check to variable BP message update. However, same can not be stated about the variable to check node update (step (b)). As will be shown in Section 4, the equality operator  $\bigodot$  generates Markov chains with desirable properties, thereby, justifying the “Markov based stochastic decoding” terminology. More specifically, the sequence  $\{W_{i \rightarrow a}^{t+1}(\ell)\}_{\ell=1}^{2k}$  is a Markov chain with the actual variable to check BP message as its stationary distribution. Our objective in step (b) is to estimate this stationary distribution. From basic Markov chain theory, we know that the marginal distribution of a chain converges to its stationary distribution. Therefore, for large enough  $k$ , the empirical distribution of the set  $\{W_{i \rightarrow a}^{t+1}(k+1), W_{i \rightarrow a}^{t+1}(k+2), \dots, W_{i \rightarrow a}^{t+1}(2k)\}$  becomes an accurate enough estimate of the stationary distribution of the Markov chain  $\{W_{i \rightarrow a}^{t+1}(\ell)\}_{\ell=1}^{2k}$ .

### 3.2 Discussion on Hardware Implementation

The proposed decoding scheme enjoys all the benefits of traditional stochastic decoders [10, 33]. Since messages between variable and check nodes are binary, stochastic decoding requires a substantially lower wiring complexity compared to fully-parallel sum-product or min-sum implementations. Shorter wires yield smaller circuit area, and smaller parasitic capacitance which in turn lead to higher clock frequencies and less power consumption. Another advantage of stochastic decoding algorithms is the very simple structure of check and variable nodes. As a matter of fact, check nodes can be carried out with simple XOR gates, and variable nodes can be implemented using a combination of a random number generator, a JK flip flop, and AND gates [10]. Finally, a very beneficial property of stochastic decoding is the

fact that the check node operation (XOR) is associative, i.e., can be partitioned arbitrarily without introducing any additional error. Mohsenin et al. [23], illustrated that partitioning check nodes can provide significant improvements (by a factor of four) in area, throughput, and energy efficiency.

It should be noted that in this paper and for mathematical convenience, the messages between check and variable nodes are represented by binary vectors. However, to implement the MbSD algorithm, there is no need to buffer all these vectors. We only need to count the number of ones between bits  $k + 1$ , and  $2k$  in each bulk, which can be accomplished by a simple counter. In that respect, MbSD has a great advantage compared to the algorithm proposed by Tehrani et al. [33], which requires buffering a substantial number of bits on each edge (edge memories). As will be discussed in Section 5, our algorithm has a superior bit error rate performance compared to [33], while maintaining the same order of maximum number of clocks, thereby achieving comparable if not better throughput. Moreover, MbSD is equipped with concrete theoretical guarantees, the subject to which we now turn.

### 3.3 Main Theoretical Results

Our results concern both cases of tree-structured (cycle free) as well as general factor graphs. Since factor graphs of randomly generated LDPC codes are locally tree-like [30], understanding the behavior of every decoding algorithm (stochastic as well as deterministic) on trees is of paramount importance. To that end, we first state some quantitative guarantees regarding the performance of the proposed stochastic decoder on tree-structured factor graphs.

Recalling the fact that there exists a unique path between every two variable nodes in a tree, we denote the largest path (also known as the graph diameter) by  $L$ . Moreover, we know that estimates generated by the SP algorithm on a tree converge to true marginals after  $L$  iterations [2], i.e., denoting true marginals by  $\{\mu_i^*\}_{i=1}^n$ , we have  $\mu_i^t = \mu_i^*$ , for all  $i = 1, 2, \dots, n$ , and  $t \geq L$ .

**Theorem 1 (Trees).** *Consider the sequence of marginals  $\{\hat{\eta}_i^{t+1}\}_{t=0}^\infty$ ,  $i = 1, 2, \dots, n$ , generated by the MbSD algorithm on a tree-structured factor graph. Then for arbitrarily small but fixed parameter  $\delta$ , and sufficiently large  $k = k(\delta, \mathcal{G})$  we have:*

(a) *The expected stochastic marginals become arbitrarily close to the true marginals, i.e.,*

$$\max_{1 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^t] - \mu_i^*| \leq \delta,$$

*for all  $t \geq L$ .*

(b) *Furthermore, we have*

$$\max_{1 \leq i \leq n} \max_{t \geq 0} \text{var}(\hat{\eta}_i^t) = \mathcal{O}\left(\frac{1}{k}\right).$$

**Remarks:** Theorem 1 provides quantitative bounds on the first and second moments of the MbSD marginal estimates. Combining parts (a), and (b), it can be easily observed that

$$\begin{aligned} \max_{1 \leq i \leq n} \mathbb{E}[(\hat{\eta}_i^t - \mu_i^*)^2] &\leq \max_{1 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^t] - \mu_i^*|^2 + \max_{1 \leq i \leq n} \text{var}(\hat{\eta}_i^t) \\ &\leq \delta^2 + \mathcal{O}\left(\frac{1}{k}\right), \end{aligned} \tag{5}$$

for all  $t \geq L$ . Therefore, as  $k \rightarrow \infty$  ( $\delta \rightarrow 0$ ), the sequence of estimates  $\hat{\eta}_i^L$  (ranging over  $k$ ) converges to the true marginal  $\mu_i^*$  in the  $\mathcal{L}^2$  sense. The rate of convergence, and its dependence on the underlying parameters, is fully characterized in expression (17). It is directly a function of the accuracy, and the factor graph structure (diameter, node degrees, etc.), and indirectly (through Lipschitz constants, etc.) a function of the signal to noise ratio (SNR).

We now turn to the statement of results for LDPC codes with general (loopy) factor graphs. Unlike tree-structured graphs, the existence and uniqueness of the SP fixed points on general graphs is not guaranteed, nor is the convergence of SP algorithm to such fixed points. Therefore, we have to make the assumption that the LDPC code of interest is well behaved. More precisely, we make the following assumptions:

**Assumption 1.** Suppose the SP message updates are consistent, that is  $\alpha_{i \rightarrow a}^t \rightarrow \alpha_{i \rightarrow a}^*$ , and  $\alpha_{a \rightarrow i}^t \rightarrow \alpha_{a \rightarrow i}^*$  as  $t \rightarrow \infty$  for all directed edges  $(i \rightarrow a)$ , and  $(a \rightarrow i)$ . Equivalently, there exists a sequence  $\{\mu_i^*\}_{i=1}^n$  such that  $\mu_i^t \rightarrow \mu_i^*$ , for all  $i = 1, 2, \dots, n$ .

For an accuracy parameter  $\delta > 0$ , arbitrarily small, we define the stopping time

$$T = T(\delta) := \inf \{t \mid \max_{1 \leq i \leq n} |\mu_i^t - \mu_i^*| \leq \delta\}. \quad (6)$$

According to assumption 1, the stopping time  $T$  is always finite.

**Theorem 2** (General factor graphs). *Consider the marginals  $\{\hat{\eta}_i^{t+1}\}_{t=0}^\infty$  generated by the MbSD algorithm on an LDPC code that satisfies Assumption 1. Then for arbitrarily small but fixed parameter  $\delta$ , and sufficiently large  $k = k(\delta, T, \mathcal{G})$  we have:*

(a) *The expected stochastic marginals become arbitrarily close to the SP marginals, i.e.,*

$$|\mathbb{E}[\hat{\eta}_i^t] - \mu_i^*| \leq \delta,$$

*for all  $i = 1, 2, \dots, n$ , and  $t = 0, 1, \dots, T$ .*

(b) *Furthermore, we have*

$$\max_{1 \leq i \leq n} \max_{0 \leq t \leq T} \text{var}(\hat{\eta}_i^t) = \mathcal{O}\left(\frac{1}{k}\right).$$

**Remarks:** Theorem 2, in contrast to Theorem 1, provides quantitative bounds on the error over a finite horizon specified by the stopping time (6). After  $T = T(\delta)$  iterations, the marginal estimates become arbitrarily close to the true marginals on average; in particular, we have

$$\max_{0 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^T] - \mu_i^*| \leq \max_{0 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^T] - \mu_i^T| + \max_{0 \leq i \leq n} |\mu_i^T - \mu_i^*| \leq 2\delta.$$

Moreover, since  $\text{var}(\hat{\eta}_i^T) = \mathcal{O}(1/k)$ , as  $k \rightarrow \infty$ , the random variables  $\{\hat{\eta}_i^T\}_{i=1}^n$ , become more and more concentrated around their means. Specifically, a very crude bound<sup>1</sup> using Chebyshev inequality [12] yields

$$\mathbb{P}\left(\max_{0 \leq i \leq n} |\hat{\eta}_i^T - \mathbb{E}[\hat{\eta}_i^T]| \geq \epsilon\right) \leq n \max_{0 \leq i \leq n} \mathbb{P}(|\hat{\eta}_i^T - \mathbb{E}[\hat{\eta}_i^T]| \geq \epsilon) = \mathcal{O}\left(\frac{n}{k \epsilon^2}\right).$$

Therefore, it is expected that the performance of the proposed stochastic decoding converges to that of SP, as  $k \rightarrow \infty$ .

---

<sup>1</sup>Tightening this bound exploiting Chernoff inequality and concentration of measure [4], can be further explored.



## 4 Proof of the Main Results

Conceptually, proofs of Theorems 1 and 2 are very similar. Therefore, in this section, we only prove Theorem 1 and highlight its important differences with Theorem 2 in Appendix D.

Poofs make use of basic probability and Markov chain theory. At a high level, the argument consists of two parts: characterizing the expected messages and controlling the error propagation in the factor graph. As it turns out, the check node operations (module two summation  $\oplus$ ) are consistent on average, that is expected messages from check to variable nodes are the same as SP messages. In contrast, the variable node operations (equality operator  $\odot$ ) are asymptotically consistent (as  $k \rightarrow \infty$ ). Therefore, for a finite message dimension  $k$ , variable node operations introduce error terms which become propagated throughout the factor graph. The main challenge is to characterize and control these errors.

### 4.1 Proof of Part (a) of Theorem 1

We start by stating a lemma which plays a key role in the sequel. Recall the definition of the equality operator  $\odot$  from (3).

**Lemma 1.** *Suppose  $Z_i = \{Z_i(\ell)\}_{\ell=1}^\infty$ , for  $i = 1, 2, \dots, d$ , are stationary, independent, and identically distributed binary sequence with  $\mathbb{P}(Z_i(\ell) = 1) = \beta_i$ . Then assuming  $\prod_{i=1}^d \beta_i + \prod_{i=1}^d (1 - \beta_i) > 0$ , the binary sequence  $U = \odot_{i=1}^d Z_i$  forms a time-reversible Markov chain with the following properties:*

(a) *The transition probabilities are*

$$\mathbb{P}(U(\ell) = 1 \mid U(\ell - 1) = 0) = \prod_{i=1}^d \beta_i, \quad \text{and} \quad (7)$$

$$\mathbb{P}(U(\ell) = 0 \mid U(\ell - 1) = 1) = \prod_{i=1}^d (1 - \beta_i). \quad (8)$$

(b) *The stationary distribution is equal to*

$$\lim_{\ell \rightarrow \infty} \mathbb{P}(U(\ell) = 1) = \frac{\prod_{i=1}^d \beta_i}{\prod_{i=1}^d \beta_i + \prod_{i=1}^d (1 - \beta_i)}.$$

The proof of this lemma is straight forward and is deferred to Appendix A. Now let  $\beta_{i \rightarrow a}^t = \mathbb{E}[Z_{i \rightarrow a}^t(\ell)] = \mathbb{P}(Z_{i \rightarrow a}^t(\ell) = 1)$  be the expected message from the variable node  $i$  to the check node  $a$ . By construction and the fact that the variables  $\{Z_{i \rightarrow a}^t(\ell)\}_{\ell=1}^{2k}$  are i.i.d., the expected value  $\beta_{i \rightarrow a}^t$  is independent of  $\ell$ . Similarly define  $\beta_{a \rightarrow i}^t = \mathbb{E}[Z_{a \rightarrow i}^t(\ell)] = \mathbb{P}(Z_{a \rightarrow i}^t(\ell) = 1)$ , the expected message from the check node  $a$  to variable node  $i$ . Taking expectation on both sides of the equation (4), we obtain

$$\mathbb{E}[\hat{\eta}_i^{t+1}] = \frac{1}{k} \sum_{\ell=k+1}^{2k} \mathbb{P}(U_i^{t+1}(\ell) = 1). \quad (9)$$

Therefore, in order to upper-bound the expected marginal  $\mathbb{E}[\hat{\eta}_i^{t+1}]$ , we need to calculate the probabilities  $\mathbb{P}(U_i^{t+1}(\ell) = 1)$ , for  $\ell = k + 1, k + 2, \dots, 2k$ . From Lemma 1, we know that the

sequence  $\{U_i^{t+1}(\ell)\}_{\ell=1}^{2k}$  is a Markov chain with the following transition probabilities:

$$f_i^t(\beta) := \alpha_i \prod_{a \in \mathcal{N}(i)} \beta_{a \rightarrow i}^t, \quad \text{and} \quad g_i^t(\beta) := (1 - \alpha_i) \prod_{a \in \mathcal{N}(i)} (1 - \beta_{a \rightarrow i}^t), \quad (10)$$

where  $f_i^t(\cdot)$ , and  $g_i^t(\cdot)$  are multivariate functions, taking values in the space  $[0, 1]^{|\mathcal{N}(i)|}$ . Recalling the basic Markov chain theory, we can calculate the probability  $\mathbb{P}(U_i^{t+1}(\ell) = 1)$  in terms of the stationary distribution, the iteration number, and the second eigenvalue<sup>2</sup> of the transition matrix [12]. Doing some algebra, we obtain

$$\begin{aligned} \mathbb{P}(U_i^{t+1}(\ell) = 1) &= \mathbb{P}(U_i^{t+1}(0) = 0) \mathbb{P}(U_i^{t+1}(\ell) = 1 \mid U_i^{t+1}(0) = 0) \\ &\quad + \mathbb{P}(U_i^{t+1}(0) = 1) \mathbb{P}(U_i^{t+1}(\ell) = 1 \mid U_i^{t+1}(0) = 1) \\ &= \left[ \frac{g_i^t(\beta) \mathbb{P}(U_i^{t+1}(0) = 1)}{f_i^t(\beta) + g_i^t(\beta)} - \frac{f_i^t(\beta) \mathbb{P}(U_i^{t+1}(0) = 0)}{f_i^t(\beta) + g_i^t(\beta)} \right] (1 - f_i^t(\beta) - g_i^t(\beta))^\ell \\ &\quad + \frac{f_i^t(\beta)}{f_i^t(\beta) + g_i^t(\beta)}. \end{aligned} \quad (11)$$

Substituting equation (11) into (9), doing some algebra simplifying the expression, and exploiting the facts

$$g_i^t(\beta) \mathbb{P}(U_i^{t+1}(0) = 1) - f_i^t(\beta) \mathbb{P}(U_i^{t+1}(0) = 0) \leq f_i^t(\beta) + g_i^t(\beta)$$

and

$$f_i^t(\beta) + g_i^t(\beta) \leq \alpha_i + (1 - \alpha_i) = 1 \quad (12)$$

yields

$$\mathbb{E}[\hat{\eta}_i^{t+1}] \leq \frac{f_i^t(\beta)}{f_i^t(\beta) + g_i^t(\beta)} + \frac{1}{k} \frac{1}{f_i^t(\beta) + g_i^t(\beta)} (1 - f_i^t(\beta) - g_i^t(\beta))^{k+1}.$$

On the other hand, denoting

$$f_i^t(\alpha) := \alpha_i \prod_{a \in \mathcal{N}(i)} \alpha_{a \rightarrow i}^t \quad \text{and} \quad g_i^t(\alpha) := (1 - \alpha_i) \prod_{a \in \mathcal{N}(i)} (1 - \alpha_{a \rightarrow i}^t), \quad (13)$$

by definition we have

$$\mu_i^{t+1} = \frac{f_i^t(\alpha)}{f_i^t(\alpha) + g_i^t(\alpha)}.$$

Since the multivariate function  $f_i^t(\cdot)/(f_i^t(\cdot) + g_i^t(\cdot))$  is Lipschitz, assuming

$$\min\{f_i^t(\alpha) + g_i^t(\alpha), f_i^t(\beta) + g_i^t(\beta)\} \geq c^*,$$

for some positive constant  $c^* > 0$ , there exists a constant  $M = M(c^*)$  such that

$$|\mathbb{E}[\hat{\eta}_i^{t+1}] - \mu_i^{t+1}| \leq M \sum_{a \in \mathcal{N}(i)} |\beta_{a \rightarrow i}^t - \alpha_{a \rightarrow i}^t| + \frac{1}{k} \frac{(1 - c^*)^{k+1}}{c^*}. \quad (14)$$

Subsequently, in order to upper-bound the error we need to bound the difference between expected stochastic messages and SP messages, i.e.  $|\beta_{a \rightarrow i}^t - \alpha_{a \rightarrow i}^t|$ . The following lemma, proved in Appendix B, addresses this problem.

---

<sup>2</sup>It is not hard to see that the second eigenvalue of the transition matrix of the Markov chain  $\{U_i^{t+1}(\ell)\}_{\ell=0}^\infty$  is equal to  $(1 - f_i^t(\beta) - g_i^t(\beta))$ .

**Lemma 2.** *On a tree-structured factor graph and for sufficiently large  $k$ , there exists a fixed positive constant  $c^* < 1$  such that*

$$\min \{f_i^t(\alpha) + g_i^t(\alpha), f_i^t(\beta) + g_i^t(\beta)\} \geq c^*, \quad (15)$$

for all  $t = 0, 1, 2, \dots$ , and  $i = 1, 2, \dots, n$ . Furthermore, denoting the maximum check and variable node degrees by  $d_c$ , and  $d_v$ , respectively, we have

$$\max_{(a \rightarrow i)} |\beta_{a \rightarrow i}^t - \alpha_{a \rightarrow i}^t| \leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \frac{[M(d_c - 1)(d_v - 1)]^L}{M(d_c - 1)(d_v - 1) - 1}, \quad (16)$$

for all  $t = 0, 1, 2, \dots$ .

Now substituting inequality (16) into (14), we obtain

$$\max_{0 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^{t+1}] - \mu_i^{t+1}| \leq \frac{(1 - c^*)^{k+1}}{k c^*} \left\{ M(d_c - 1)d_v \frac{[M(d_c - 1)(d_v - 1)]^L}{M(d_c - 1)(d_v - 1) - 1} + 1 \right\},$$

for all  $t = 0, 1, 2, \dots$ . Therefore, setting

$$k = \max \left\{ \frac{\log \delta - L \log(M(d_c - 1)(d_v - 1))}{\log(1 - c^*)}, \frac{3}{c^*} \right\}, \quad (17)$$

we obtain

$$\max_{0 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^t] - \mu_i^t| \leq \delta,$$

for all  $t = 0, 1, 2, \dots$ .

## 4.2 Proof of Part (b) of Theorem 1

To streamline the exposition, let  $U(\ell) := U_i^{t+1}(\ell)$ , for fixed  $i$ , and  $t$ . As previously stated, the sequence  $\{U(\ell)\}_{\ell=1}^\infty$  is a Markov chain with initial state  $p_0 := \mathbb{P}(U(0) = 0)$ ,  $p_1 := \mathbb{P}(U(0) = 1)$ , and transition probabilities  $f := f_i^t(\beta)$ , and  $g := g_i^t(\beta)$ ; more specifically we have

$$f = \mathbb{P}(U(\ell) = 1 | U(\ell - 1) = 0), \quad \text{and} \quad g = \mathbb{P}(U(\ell) = 0 | U(\ell - 1) = 1).$$

Since  $\mathbb{E}[(U(\ell) - \mathbb{E}[U(\ell)])^2] \leq 1$ , in order to upper-bound the variance

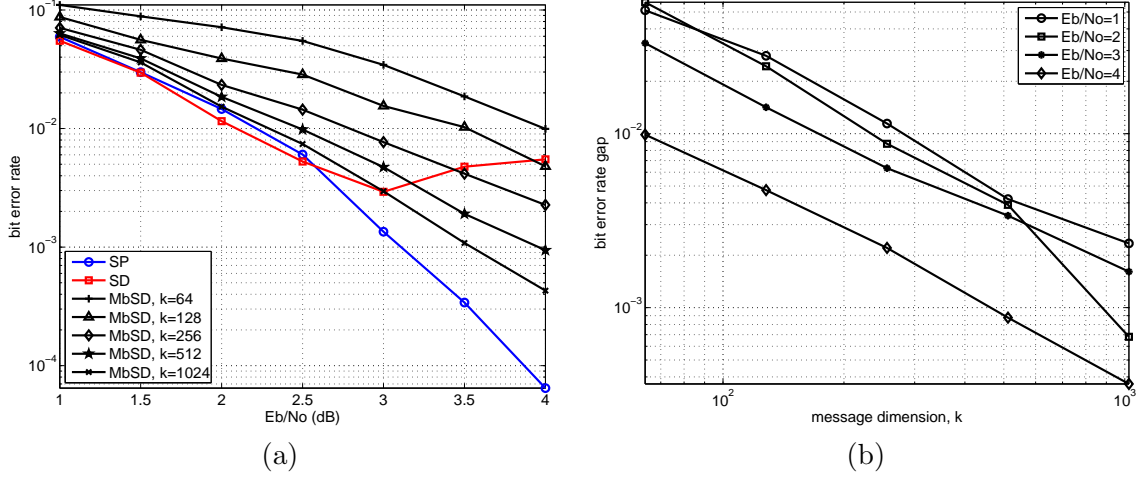
$$\text{var}(\hat{\eta}_i^{t+1}) = \frac{1}{k^2} \sum_{\ell=k+1}^{2k} \mathbb{E}[(U(\ell) - \mathbb{E}[U(\ell)])^2] + \frac{2}{k^2} \sum_{\ell' < \ell} \mathbb{E}[(U(\ell) - \mathbb{E}[U(\ell)])(U(\ell') - \mathbb{E}[U(\ell')])],$$

we only need to upper-bound the cross-product terms. Doing so, for  $\ell > \ell'$ , we have

$$\begin{aligned} \mathbb{E}[(U(\ell) - \mathbb{E}[U(\ell)])(U(\ell') - \mathbb{E}[U(\ell')])] &= \mathbb{E}[U(\ell)U(\ell')] - \mathbb{E}[U(\ell)]\mathbb{E}[U(\ell')] \\ &= \mathbb{P}(U(\ell') = 1) [\mathbb{P}(U(\ell) = 1 | U(\ell') = 1) - \mathbb{P}(U(\ell) = 1)] \\ &\leq \mathbb{P}(U(\ell) = 1 | U(\ell') = 1) - \mathbb{P}(U(\ell) = 1). \end{aligned}$$

Now, exploiting the Markov property and equation (11), we can further simplify the aforementioned inequality

$$\begin{aligned} \mathbb{E}[(U(\ell) - \mathbb{E}[U(\ell)])(U(\ell') - \mathbb{E}[U(\ell')])] &\leq \frac{g}{f + g} (1 - f - g)^{(\ell - \ell')} + \frac{f p_0}{f + g} (1 - f - g)^\ell - \frac{g p_1}{f + g} (1 - f - g)^\ell \\ &\stackrel{(i)}{\leq} (1 - f - g)^{(\ell - \ell')}, \end{aligned}$$



**Figure 3.** Performance of the MbSD algorithm on a (3,6)-LDPC code with  $n = 200$  variable nodes and  $m = 100$  check nodes. (a) Bit error rate versus the energy per bit to noise power spectral density ( $E_b/N_0$ ) for different decoders, namely, the SP, the SD (stochastic decoding without noise dependent scaling [33]), and the MbSD for different message dimensions  $k \in \{64, 128, 256, 512, 1024\}$ . As predicted by the theory, the performance of the MbSD converges to that of SP. Moreover, MbSD does not suffer from error floor, in contrast to the SD algorithm. (b) Bit error rate gap, i.e., the difference between the SP and the MbSD bit error rates, versus the message dimension. The rate of convergence is upper bounded by  $\mathcal{O}(1/k)$ , manifested as linear curves in the log-log domain plot.

where inequality (i) follows from (12) and the fact that  $g/(f+g) + p_0 f/(f+g) \leq 1$ . According to Lemma 2, for sufficiently large  $k$ , we have  $f+g \geq c^*$ . Therefore, putting the pieces together doing some algebra, we obtain

$$\begin{aligned}
 \text{var}(\hat{\eta}_i^{t+1}) &\leq \frac{1}{k} + \frac{2}{k^2} \sum_{k+1 \leq \ell' < \ell \leq 2k} (1 - c^*)^{(\ell - \ell')} \\
 &= \frac{1}{k} + \frac{2}{k^2} \sum_{i=1}^{k-1} (k-i)(1 - c^*)^i \\
 &\leq \frac{1}{k} + \frac{2}{k} \sum_{i=1}^{k-1} (1 - c^*)^i \leq \frac{1 + 2/c^*}{k},
 \end{aligned}$$

for all  $i = 1, 2, \dots, n$ , and  $t = 0, 1, 2, \dots$ .

## 5 Experimental Results

To confirm our theoretical predictions, we test the MbSD algorithm on a simple LDPC code. In our experiments, we set the block-length (variable nodes) and number of parity checks (check nodes) to be  $n = 200$  and  $m = 100$ , respectively. Using Gallager's construction [30], we first generate a regular (3, 6)-LDPC code, that is, all variable nodes have degree three, whereas all check nodes have degree six. Then, considering a binary pulse-amplitude modulation (BPAM) system over an AWGN channel, we run MbSD, for  $t = 60$  iterations, on several simulated signals in order to compute the bit error rate for different values of normalized

signal to noise ratio.<sup>3</sup> The test is carried out for a number of message dimensions  $k$ , and results are compared with the SP algorithm and the stochastic decoding (SD) proposed by Tehrani et al. [33] (see Figure 3 (a)).<sup>4</sup> As predicted by our theorems, the performance of the MbSD converges to that of SP as  $k$  grows. Therefore, in contrast to the SD algorithm, MbSD is an asymptotically consistent estimate of the SP and does not suffer from error floor. The rate of convergence, on the other hand, can be further explored in Figure 3 (b), wherein the bit error rate gap (i.e., the difference between SP and MbSD bit error rates) versus the message dimension is illustrated. As can be observed, the error curves in the log-log domain plot are roughly linear with slope one. This observation is consistent with equation (5), suggesting the upper-bound of  $\mathcal{O}(1/k)$  for the rate of convergence.

To improve upon the seemingly slow rate of convergence, we make use of the notion of noise dependent scaling (NDS). Sensitivity to random switching activities, referred to as ‘latching’, has been observed to be a major challenge in stochastic decoders [35, 33]. To circumvent this issue, the notion of NDS, in which the received log-likelihoods are down-scaled by a factor proportional to the SNR, was proposed and shown to be extremely effective [33]. The MbSD algorithm suffers from the latching problem too, especially for high SNR values. Intuitively, sequences generated by Markov chains (recall Lemma 1) are likely to be the all-one or the all-zero sequences when the SNR is sufficiently high. As a consequence, in such cases, the positive constant  $c^*$ , defined in (15), is more likely to be close to zero. The rate of convergence of the expectation, specified in equation (17), is inversely proportional to  $\log(1 - c^*)$ , therefore, the smaller the  $c^*$ , the slower the rate of convergence. Resolving this issue requires increasing the switching activities of Markov chain, which is accomplished by the NDS. Figure 4, illustrates the bit error rate versus  $E_b/N_0$  for the SP, the SD using NDS, and the MbSD using NDS.<sup>5</sup> As is evident, the rate of convergence of the MbSD algorithm, and thus its performance, is significantly improved. Moreover, having the same number of decoding cycles, MbSD outperforms the SD for high SNRs.

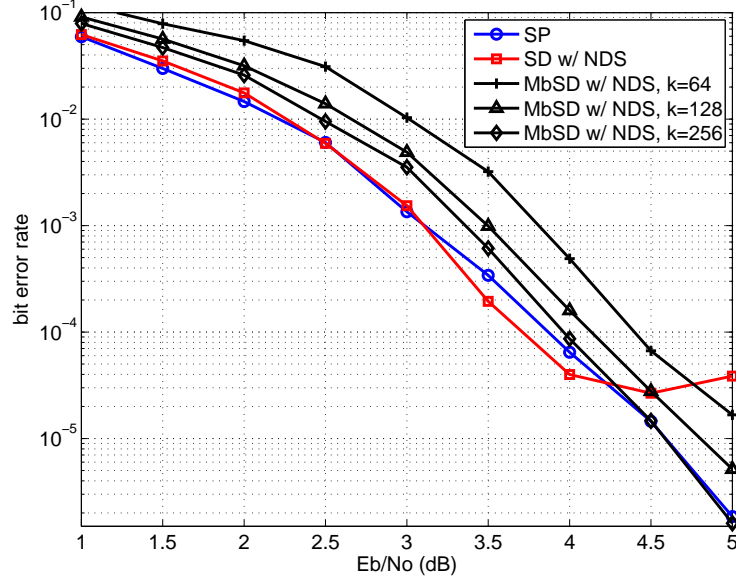
## 6 Conclusion

In this paper, we studied the theoretical aspect of stochastic decoders, a widely studied solution for fully-parallel implementation of LDPC decoding on chips. Generally speaking, encoding messages by binary sequences, stochastic decoders simplify check and node message updates by modulo-two summation and the equality operator, respectively. As it turns out, the check node operation is statistically consistent on average, whereas, the variable node equality operation generates a Markov chain with the desired quantity as its stationary distribution. Therefore, for a finite message dimension  $k$ , the stochastic message updates introduce error terms which become propagated in the factor graph. Controlling these errors is the main challenge in the theoretical analysis of stochastic decoders. To formalize these notions, we introduced a novel stochastic algorithm, referred to as the Markov based stochastic decoding, and provided concrete theoretical guarantees on its performance. More precisely, we showed that expected marginals produced by the MbSD become arbitrarily close to marginals generated by the SP algorithm on tree-structured as well as general factor graphs. The rate

<sup>3</sup>A BPAM system with transmit power one over an AWGN channel with noise variance  $\sigma^2$  has the energy per bit to noise power spectral density ( $E_b/N_0$ ) of  $1/(2R\sigma^2)$ , where  $R$  is the code rate.

<sup>4</sup>In simulating the SD algorithm, we used 30000 ‘decoding cycles’ and edge memory of length 25 without noise dependent scaling.

<sup>5</sup>In our simulations we set the NDS scaling parameter to be  $\sigma^2$ , the optimum choice as suggested in the paper [33].



**Figure 4.** Effect of the noise dependent scaling (NDS) on the performance of the MbSD algorithm. The panel contains several plots illustrating the bit error rate versus the energy per bit to noise power spectral density ( $E_b/N_0$ ) for different decoders, namely, the SP, the SD (stochastic decoding with noise dependent scaling [33]), and the MbSD for different message dimensions  $k \in \{64, 128, 256\}$ . As expected, NDS has significantly improved the performance of the MbSD. Simulations were conducted on a (3,6)-LDPC code with  $n = 200$  variable nodes and  $m = 100$  check nodes.

of convergence is governed by the message dimension, the graph structure, and the Lipschitz constant, formally specified in equation (17). Moreover, we proved that the variance of MbSD marginals are upper-bounded by  $\mathcal{O}(1/k)$ . These theoretical predictions were also supported by experimental results. We showed that, maintaining the same order of decoding cycles, our algorithm does not suffer from error floor; therefore, it achieves better bit error rate performance compared to other competing methods.

## Acknowledgements

Authors would like to thank Aman Bahtia for providing the C++ code, simulating sum-product and stochastic decoding algorithms.

## A Proof of Lemma 1

By definition we have

$$U(\ell) = \begin{cases} 1 & \text{if } Z_i(\ell) = 1, \text{ for all } i = 1, 2, \dots, d \\ 0 & \text{if } Z_i(\ell) = 1, \text{ for all } i = 1, 2, \dots, d \\ U(\ell - 1) & \text{otherwise} \end{cases}.$$

Therefore, given  $U(\ell - 1) = 0$  and regardless of the sequence  $\{U(\ell - 2), \dots, U(0)\}$ , the event  $\{U(\ell) = 1\}$  is equivalent to  $\{Z_i(\ell) = 1, \forall i = 1, \dots, d\}$ . Therefore, we have

$$\begin{aligned} \mathbb{P}(U(\ell) = 1 | U(\ell - 1) = 0, U(\ell - 2), \dots) &= \mathbb{P}(Z_i(\ell) = 1, \forall i = 1, \dots, d, ) \\ &\stackrel{(i)}{=} \prod_{a=1}^d \mathbb{P}(Z_i(\ell) = 1) = \prod_{i=1}^d \beta_i, \end{aligned}$$

where equality (i) follows from the i.i.d. nature of the sequence. The exact same argument yields the equation (8). Finally, the stationary distribution can be obtained from the detailed balance condition [12]

$$\lim_{\ell \rightarrow \infty} \mathbb{P}(U_i(\ell) = 1) \prod_{i=1}^d (1 - \beta_i) = \lim_{\ell \rightarrow \infty} \mathbb{P}(U_i(\ell) = 0) \prod_{i=1}^d \beta_i.$$

## B Proof of Lemma 2

Recall binary messages  $Z_{a \rightarrow i}^t$ , and  $Z_{i \rightarrow a}^t$  from steps 2(a) and 2(b) of the main algorithm. Also recall the definition of expected messages  $\beta_{a \rightarrow i}^t = \mathbb{P}(Z_{a \rightarrow i}^t(\ell) = 1)$ , and  $\beta_{i \rightarrow a}^t = \mathbb{P}(Z_{i \rightarrow a}^t(\ell) = 1)$ . From Lemma 1 of the paper [9], we know that

$$\beta_{a \rightarrow i}^t = \frac{1}{2} - \frac{1}{2} \prod_{j \in \mathcal{N}(a) \setminus \{i\}} (1 - 2\beta_{j \rightarrow a}^t). \quad (18)$$

On the other hand, by construction we have

$$\beta_{i \rightarrow a}^{t+1} = \frac{1}{k} \sum_{\ell=k+1}^{2k} \mathbb{P}(W_{i \rightarrow a}^{t+1}(\ell) = 1),$$

where  $W_{i \rightarrow a}^{t+1} = \odot_{b \in \mathcal{N}(i) \setminus \{a\}} Z_{b \rightarrow i}^t \odot Z_i^{t+1}$ . Since, according to Lemma 1, the sequence  $\{W_{i \rightarrow a}^{t+1}(\ell)\}$  forms a Markov chain with transition probabilities

$$f_{i \rightarrow a}^t(\beta) := \alpha_i \prod_{b \in \mathcal{N}(i) \setminus \{a\}} \beta_{b \rightarrow i}^t, \quad \text{and} \quad g_{i \rightarrow a}^t(\beta) := (1 - \alpha_i) \prod_{b \in \mathcal{N}(i) \setminus \{a\}} (1 - \beta_{b \rightarrow i}^t),$$

basic Markov chain theory yields

$$\begin{aligned} \mathbb{P}(W_{i \rightarrow a}^{t+1}(\ell) = 1) &= \frac{f_{i \rightarrow a}^t(\beta)}{f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta)} \\ &+ \left[ \frac{g_{i \rightarrow a}^t(\beta) \mathbb{P}(W_{i \rightarrow a}^{t+1}(0) = 1)}{f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta)} - \frac{f_{i \rightarrow a}^t(\beta) \mathbb{P}(W_{i \rightarrow a}^{t+1}(0) = 0)}{f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta)} \right] (1 - f_{i \rightarrow a}^t(\beta) - g_{i \rightarrow a}^t(\beta))^\ell. \end{aligned}$$

Therefore, doing some algebra, noticing the facts that

$$\frac{g_{i \rightarrow a}^t(\beta) \mathbb{P}(W_{i \rightarrow a}^{t+1}(0) = 1)}{f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta)} - \frac{f_{i \rightarrow a}^t(\beta) \mathbb{P}(W_{i \rightarrow a}^{t+1}(0) = 0)}{f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta)} \leq 1$$

and

$$f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta) \leq \alpha_i + (1 - \alpha_i) = 1,$$

we have

$$\beta_{i \rightarrow a}^{t+1} \leq \frac{f_{i \rightarrow a}^t(\beta)}{f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta)} + \frac{1}{k(f_{i \rightarrow a}^t(\beta) + g_{i \rightarrow a}^t(\beta))} (1 - f_{i \rightarrow a}^t(\beta) - g_{i \rightarrow a}^t(\beta))^{k+1}. \quad (19)$$

Equations (18), and (19) characterize the stochastic decoding message updates. Similarly, we have the SP update equations for all directed edges

$$\alpha_{a \rightarrow i}^t = \frac{1}{2} - \frac{1}{2} \prod_{j \in \mathcal{N}(a) \setminus \{i\}} (1 - 2\alpha_{j \rightarrow a}^t), \quad \text{and} \quad \alpha_{i \rightarrow a}^{t+1} = \frac{f_{i \rightarrow a}^t(\alpha)}{f_{i \rightarrow a}^t(\alpha) + g_{i \rightarrow a}^t(\alpha)}, \quad (20)$$

where we have denoted

$$f_{i \rightarrow a}^t(\alpha) := \alpha_i \prod_{b \in \mathcal{N}(i) \setminus \{a\}} \alpha_{b \rightarrow i}^t, \quad \text{and} \quad g_{i \rightarrow a}^t(\alpha) := (1 - \alpha_i) \prod_{b \in \mathcal{N}(i) \setminus \{a\}} (1 - \alpha_{b \rightarrow i}^t).$$

Since  $0 \leq \alpha_{j \rightarrow a}^t \leq 1$ , and  $0 \leq \beta_{j \rightarrow a}^t \leq 1$ , for all  $t$ , and  $(j \rightarrow a)$ , we have<sup>6</sup>

$$|\beta_{a \rightarrow i}^{t+1} - \alpha_{a \rightarrow i}^{t+1}| \leq \sum_{j \in \mathcal{N}(a) \setminus \{i\}} |\beta_{j \rightarrow a}^{t+1} - \alpha_{j \rightarrow a}^{t+1}|. \quad (21)$$

We now turn to upper-bounding the term

$$\begin{aligned} |\beta_{j \rightarrow a}^{t+1} - \alpha_{j \rightarrow a}^{t+1}| &\leq \left| \frac{f_{j \rightarrow a}^t(\beta)}{f_{j \rightarrow a}^t(\beta) + g_{j \rightarrow a}^t(\beta)} - \frac{f_{j \rightarrow a}^t(\alpha)}{f_{j \rightarrow a}^t(\alpha) + g_{j \rightarrow a}^t(\alpha)} \right| \\ &\quad + \frac{1}{k(f_{j \rightarrow a}^t(\beta) + g_{j \rightarrow a}^t(\beta))} (1 - f_{j \rightarrow a}^t(\beta) - g_{j \rightarrow a}^t(\beta))^{k+1}. \end{aligned}$$

Since  $0 < \alpha_{j \rightarrow a}^0 = \alpha_j < 1$ . Then by inspection of the SP updates, it is easy to see that  $0 < \alpha_{j \rightarrow a}^t < 1$ , and  $0 < \alpha_{a \rightarrow j}^t < 1$ , for all  $t \geq 0$ , and all directed edges  $(a \rightarrow j)$ , and  $(j \rightarrow a)$ . Therefore, recalling the definition (13), and the fact that  $\alpha_{a \rightarrow i}^t = \alpha_{a \rightarrow i}^*$ , for  $t$  larger than the graph diameter, there exists a positive constant  $c^* < 1$  such that

$$f_j^t(\alpha) + g_j^t(\alpha) \geq 2c^*, \quad \text{for all } j = 1, 2, \dots, n, \text{ and } t = 0, 1, 2, \dots$$

Now we show that for sufficiently large  $k$ , we have

$$f_j^t(\beta) + g_j^t(\beta) \geq c^*, \quad \text{for all } j = 1, 2, \dots, n, \text{ and } t = 0, 1, 2, \dots \quad (22)$$

Suppose for a fixed  $t$ , we have

$$\sum_{a \in \mathcal{N}(j)} |\beta_{a \rightarrow j}^\tau - \alpha_{a \rightarrow j}^\tau| \leq c^*, \quad \text{for all } j = 1, 2, \dots, n, \text{ and } \tau = 0, 1, \dots, t. \quad (23)$$

Since  $\beta_{j \rightarrow a}^0 = \alpha_{j \rightarrow a}^0 = \alpha_j$ , the left hand side of the above inequality is initially equal to zero; thus the condition (23) is satisfied for  $t = 0$ . Now making use of the mean-value theorem, we obtain<sup>7</sup>

$$|f_j^\tau(\beta) + g_j^\tau(\beta) - f_j^\tau(\alpha) - g_j^\tau(\alpha)| \leq \sum_{a \in \mathcal{N}(j)} |\beta_{a \rightarrow j}^\tau - \alpha_{a \rightarrow j}^\tau|.$$

<sup>6</sup>The inequality follows from the mean-value theorem and the fact that for the function  $h(x_1, \dots, x_d) = \prod_{i=1}^d (1 - 2x_i)/2$ , we have  $|\partial h / \partial x_i| \leq 1$ , if  $0 < x_i \leq 1$  for all  $i = 1, \dots, d$ .

<sup>7</sup>Let  $h(x_1, \dots, x_d) = \alpha \prod_{i=1}^d x_i + (1 - \alpha) \prod_{i=1}^d (1 - x_i)$  be an arbitrary function for  $0 \leq x_i \leq 1$ ,  $i = 1, \dots, d$ . Then we have  $|\partial h / \partial x_i| \leq \alpha \prod_{j \neq i} x_j + (1 - \alpha) \prod_{j \neq i} (1 - x_j) \leq 1$ .



Putting the pieces together, assuming (23), yields

$$f_j^\tau(\beta) + g_j^\tau(\beta) \geq c^* \quad \text{for all } j = 1, 2, \dots, n, \text{ and } \tau = 0, 1, \dots, t.$$

Since  $0 \leq \beta_{a \rightarrow j}^t, \alpha_{a \rightarrow j}^t \leq 1$ , we have

$$\begin{aligned} f_{j \rightarrow a}^\tau(\beta) + g_{j \rightarrow a}^\tau(\beta) &\geq f_j^\tau(\beta) + g_j^\tau(\beta) \geq c^*, \quad \text{and} \\ f_{j \rightarrow a}^\tau(\alpha) + g_{j \rightarrow a}^\tau(\alpha) &\geq f_j^\tau(\alpha) + g_j^\tau(\alpha) \geq 2c^*; \end{aligned}$$

therefore, there exist a constant  $M = M(c^*)$  such that

$$|\beta_{j \rightarrow a}^{\tau+1} - \alpha_{j \rightarrow a}^{\tau+1}| \leq M \sum_{b \in \mathcal{N}(j) \setminus \{a\}} |\beta_{b \rightarrow j}^\tau - \alpha_{b \rightarrow j}^\tau| + \frac{1}{k} \frac{(1 - c^*)^{k+1}}{c^*}, \quad (24)$$

for all  $\tau = 0, 1, \dots, t$ . Substituting the inequality (24) into (21), we obtain

$$|\beta_{a \rightarrow i}^{\tau+1} - \alpha_{a \rightarrow i}^{\tau+1}| \leq M \sum_{j \in \mathcal{N}(a) \setminus \{i\}} \sum_{b \in \mathcal{N}(j) \setminus \{a\}} |\beta_{b \rightarrow j}^\tau - \alpha_{b \rightarrow j}^\tau| + \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*}, \quad (25)$$

where we have denoted  $d_c := \max_{0 \leq a \leq m} |\mathcal{N}(a)|$ . Let  $e^t := \{|\beta_{a \rightarrow i}^t - \alpha_{a \rightarrow i}^t|\}_{(a \rightarrow i)}$  be the  $r := \sum_{a=1}^m |\mathcal{N}(a)|$  dimensional error vector. Now define a matrix  $A \in \mathbb{R}^{r \times r}$  with entries indexed by pairs of directed edges  $(a \rightarrow i)$ ; in particular, we have

$$A(a \rightarrow i, b \rightarrow j) := \begin{cases} M & \text{if } j \in \mathcal{N}(a) \setminus \{i\} \text{ and } b \in \mathcal{N}(j) \setminus \{a\} \\ 0 & \text{o.w.} \end{cases}. \quad (26)$$

Then by stacking the scalar inequalities (25), we obtain the vector inequality

$$e^{\tau+1} \leq A e^\tau + \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \vec{1}, \quad (27)$$

for all  $\tau = 0, 1, \dots, t$ . Here,  $\leq$  denotes the vector inequality, i.e., for  $r$ -dimensional vectors  $x$ , and  $y$  we say  $x \leq y$  if and only if  $x(i) \leq y(i)$ , for all  $i = 1, 2, \dots, r$ . Unwrapping the recursion (27), noticing that  $e^0 = \vec{0}$ , we obtain

$$e^{t+1} \leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} (I + A + \dots + A^t) \vec{1}. \quad (28)$$

The right hand side sequence of the previous inequality have, seemingly, a growing number of terms as  $t \rightarrow \infty$ . However, according to the following lemma, proved in Appendix C, the graph-respecting matrix  $A$  is nilpotent, i.e., there exists a positive integer  $\ell$  such that  $A^\ell = 0$ . (A similar statement regarding nilpotence of the tree-structured Markov random field have been shown in Lemma 1 [26].) Recall the definition of the factor graph diameter  $L$ , the largest path between any pair of variable nodes.

**Lemma 3.** *The graph-respecting matrix  $A$ , defined in (26), is nilpotent with degree at most the diameter of the factor graph  $L$ , that is  $A^L = 0$ .*

Exploiting the result of Lemma 3, we can further simplify the vector inequality (28)

$$e^t \leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} (I + A + \dots + A^{L-1}) \vec{1}, \quad (29)$$

for all  $t = 0, 1, 2, \dots$ . Now we take  $\|\cdot\|_\infty$  on the both sides of the inequality (29). Recalling the definition of the matrix norm infinity<sup>8</sup>,  $\|\cdot\|_\infty$ , triangle inequality, and using the fact that  $\|A^\ell\|_\infty \leq \|A\|_\infty^\ell$  [15], simple algebra yields

$$\begin{aligned} \max_{(a \rightarrow i)} |\beta_{a \rightarrow i}^{t+1} - \alpha_{a \rightarrow i}^{t+1}| &\leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \sum_{\ell=0}^{L-1} \|A\|_\infty^\ell \\ &\leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \sum_{\ell=0}^{L-1} [M(d_c - 1)(d_v - 1)]^\ell \\ &\leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \frac{[M(d_c - 1)(d_v - 1)]^L}{M(d_c - 1)(d_v - 1) - 1}, \end{aligned}$$

where we have denoted  $d_v = \max_{0 \leq i \leq n} |\mathcal{N}(i)|$ . For sufficiently large  $k$ , specifically when

$$\frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \frac{[M(d_c - 1)(d_v - 1)]^L}{M(d_c - 1)(d_v - 1) - 1} \leq \frac{c^*}{d_v},$$

we have

$$\sum_{a \in \mathcal{N}(j)} |\beta_{a \rightarrow j}^{t+1} - \alpha_{a \rightarrow j}^{t+1}| \leq c^*, \quad \text{for all } j = 1, 2, \dots, n,$$

and hence  $f_j^{t+1}(\beta) + g_j^{t+1}(\beta) \geq c^*$ , which proves the claim (22) and concludes the lemma.

## C Proof of Lemma 3

We first show, via induction, that for any positive integer  $\ell$  and edges  $(a \rightarrow i)$ , and  $(b \rightarrow j)$ , the entry  $A^\ell(a \rightarrow i, b \rightarrow j) \neq 0$  if and only if there exists a directed path of length<sup>9</sup>  $\ell$  between the edges  $(b \rightarrow j)$ , and  $(a \rightarrow i)$ . More specifically, there must exist a sequence of non-overlapping, directed (check-variable) edges  $\{(a_1 \rightarrow i_1), \dots, (a_{\ell-1} \rightarrow i_{\ell-1})\}$  such that

$$b \in \mathcal{N}(j) \setminus \{a_{\ell-1}\}, j \in \mathcal{N}(a_{\ell-1}) \setminus \{i_{\ell-1}\}, \dots, a_1 \in \mathcal{N}(i_1) \setminus \{a\}, i_1 \in \mathcal{N}(a) \setminus \{i\}.$$

The base case for  $\ell = 1$ , is obvious from construction (26). Suppose the claim is correct for  $\ell$ ; the goal is to prove it for  $\ell + 1$ . By definition, we have

$$A^{\ell+1}(a \rightarrow i, b \rightarrow j) = \sum_{(c \rightarrow k)} A^\ell(a \rightarrow i, c \rightarrow k) A(c \rightarrow k, b \rightarrow j).$$

Since the matrix is non-negative,  $A^{\ell+1}(a \rightarrow i, b \rightarrow j) \neq 0$ , if and only if there exists an edge  $(c \rightarrow k)$  such that  $A^\ell(a \rightarrow i, c \rightarrow k) \neq 0$ , and  $A(c \rightarrow k, b \rightarrow j) \neq 0$ . Therefore, according to the induction hypothesis, there exist a sequence of non-overlapping edges  $\{(a_1 \rightarrow i_1), \dots, (a_{\ell-1} \rightarrow i_{\ell-1})\}$  such that

$$c \in \mathcal{N}(k) \setminus \{a_{\ell-1}\}, k \in \mathcal{N}(a_{\ell-1}) \setminus \{i_{\ell-1}\}, \dots, a_1 \in \mathcal{N}(i_1) \setminus \{a\}, i_1 \in \mathcal{N}(a) \setminus \{i\}.$$

Moreover, we should have  $b \in \mathcal{N}(j) \setminus \{c\}$ , and  $j \in \mathcal{N}(c) \setminus \{k\}$ . Putting the pieces together, there must exist a directed path of length  $\ell + 1$ , consisting of  $\{(a_1 \rightarrow i_1), \dots, (a_{\ell-1} \rightarrow i_{\ell-1}), (c \rightarrow k)\}$ , between edges  $(b \rightarrow j)$ , and  $(a \rightarrow i)$ , which yields the claim. Finally, since there is no directed path longer than  $L$  (diameter) between any pair of edges in a tree-structured factor graph, we must have  $A^L = 0$  that concludes the proof.

<sup>8</sup>Norm infinity of a matrix is the maximum absolute row sum of the matrix [15]

<sup>9</sup>Here the length of the path is equal to the number of intermediate variable nodes plus one.

## D Proof of Theorem 2

As stated previously, proof of Theorem 2 is similar to that of Theorem 1. The major difference lie in the fact that due to its cycle-free structure, tree-respecting matrix  $A$ , defined in (26), is nilpotent (recall the result of Lemma 3). However, the same may not necessarily be true for general graphs. As a consequence, for a non-tree factor graph, the right hand side of the inequality (28) has indeed a growing number of terms as  $t \rightarrow \infty$ . However, we can upper-bound the error over a finite horizon provided by the stopping time  $T$ , defined in (6). More precisely, unwrapping the recursion (27) for  $\tau = 0, 1, \dots, t$ , taking norm-infinity on both sides of the outcome, and doing some algebra yields

$$\max_{(a \rightarrow i)} |\beta_{a \rightarrow i}^{t+1} - \alpha_{a \rightarrow i}^{t+1}| \leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \frac{[M(d_c - 1)(d_v - 1)]^{t+1}}{M(d_c - 1)(d_v - 1) - 1},$$

for all  $t = 0, 1, \dots, T - 1$ . Therefore, for  $k$  sufficiently large, specifically when

$$\frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \frac{[M(d_c - 1)(d_v - 1)]^T}{M(d_c - 1)(d_v - 1) - 1} \leq \frac{c^*}{d_v},$$

we have

$$\sum_{a \in \mathcal{N}(j)} |\beta_{a \rightarrow j}^{t+1} - \alpha_{a \rightarrow j}^{t+1}| \leq c^*,$$

for all  $j = 1, 2, \dots, n$ , and  $t = 0, 1, \dots, T - 1$ , thereby completing the proof of a slightly different version of the Lemma 2 over a finite horizon.

**Lemma 4.** *For sufficiently large  $k$ , there exists a fixed positive constant  $c^* = c^*(T) < 1$  such that*

$$\min \{f_i^t(\alpha) + g_i^t(\alpha), f_i^t(\beta) + g_i^t(\beta)\} \geq c^*,$$

for all  $t = 0, 1, \dots, T$ , and  $i = 1, 2, \dots, n$ . Furthermore, we have

$$\max_{(a \rightarrow i)} |\beta_{a \rightarrow i}^t - \alpha_{a \rightarrow i}^t| \leq \frac{(d_c - 1)(1 - c^*)^{k+1}}{k c^*} \frac{[M(d_c - 1)(d_v - 1)]^t}{M(d_c - 1)(d_v - 1) - 1}, \quad (30)$$

for all  $t = 0, 1, \dots, T$ .

Now substituting the inequality (30) into (14), setting

$$k = \max \left\{ \frac{\log \delta - T \log(M(d_c - 1)(d_v - 1))}{\log(1 - c^*)}, \frac{3}{c^*} \right\},$$

we obtain

$$\max_{0 \leq i \leq n} |\mathbb{E}[\hat{\eta}_i^t] - \mu_i^t| \leq \delta,$$

for all  $t = 0, 1, \dots, T$ , which concludes proof of part (a) of Theorem 2. Proof of part (b) follows the exact same steps outlined in the Section 4.2

## References

- [1] IEEE 802.16e. Air interface for fixed and mobile broadband wireless access systems, October 2005. IEEE p802.16e/d12 Draft.
- [2] S. M. Aji and R. J. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, March 2000.
- [3] Tyler Brandon, Robert Hang, Gary Block, Vincent C. Gaudet, Bruce Cockburn, Sheryl Howard, Christian Giasson, Keith Boyle, Paul Goud, Siavash Sheikh Zeinoddin, Anthony Rapley, Stephen Bates, Duncan Elliott, and Christian Schlegel. A scalable LDPC decoder ASIC architecture with bit-serial message exchange. *the VLSI Journal of Integration*, 41(3):385 – 398, 2008.
- [4] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics*, 3(1):79–127, 2006.
- [5] K. Cushon, C. Leroux, S. Hemati, S. Mannor, and W.J. Gross. A min-sum iterative decoder based on pulsewidth message encoding. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 57(11):893–897, 2010.
- [6] A. Darabiha, A. C. Carusone, and F. R. Kschischang. A bit-serial approximate min-sum ldpc decoder and FPGA implementation. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, May 2006.
- [7] A. Darabiha, A. C. Carusone, and F. R. Kschischang. Power reduction techniques for LDPC decoders. *IEEE Journal of Solid-State Circuits*, 43(8):1835–1845, 2008.
- [8] M. P. C. Fossorier, M. Mihaljevic, and H. Imai. Reduced complexity iterative decoding of low-density parity check codes based on belief propagation. *IEEE Transactions on Communications*, 47(5):673–680, May 1999.
- [9] R. G. Gallager. Low-density parity check codes. *IRE Transactions on Information Theory*, 8(1):21–28, 1962.
- [10] V. C. Gaudet and A. C. Rapley. Iterative decoding using stochastic computation. *Electronics Letters*, 39(3):299–301, 2003.
- [11] G.hn/G.9960. Next generation standard for wired home network. <http://www.itu.int/ITU-T>.
- [12] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford Science Publications, Clarendon Press, Oxford, 1992.
- [13] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Journal of Machine Learning*, 69(2-3):169–192, 2007.
- [14] S. Hemati, A. H. Banihashemi, and C. Plett. A 0.18-  $\mu$ m CMOS analog min-sum iterative decoder for a (32,8) low-density parity-check (LDPC) code. *IEEE Journal of Solid-State Circuits*, 41(11):2531–2540, 2006.
- [15] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

- [16] T. T. S. I. Digital video broadcasting (DVB) second generation framing structure for broadband satellite applications. <http://www.dvb.org>.
- [17] A. Juditsky, G. Lan, A. Nemirovski, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.
- [18] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transaction on Information Theory*, 47(2):498–519, 2001.
- [19] H. Kuo-Lun, V. Gaudet, and M. Salehi. A Markov chain model for edge memories in stochastic decoding of ldpc codes. In *Proceedings of the 45th Annual Conference on Information Sciences and Systems*, pages 1–4, 2011.
- [20] F. Leduc-Primeau, S. Hemati, S. Mannor, and W. J. Gross. Relaxed half-stochastic belief propagation. *IEEE Transactions on Communications*, 61(5):1648–1659, 2013.
- [21] D. MacKay. Good error correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45(2):399–431, 1999.
- [22] T. Mohsenin and B. M. Baas. Split-row: A reduced complexity, high throughput LDPC decoder architecture. In *Proceedings of International Conference on Computer Design*, pages 320–325, 2006.
- [23] T. Mohsenin, D. N. Truong, and B. M. Baas. A low-complexity message-passing algorithm for reduced routing congestion in LDPC decoders. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 57(5):1048–1061, 2010.
- [24] A. Naderi, S. Mannor, M. Sawan, and W. J. Gross. Delayed stochastic decoding of LDPC codes. *IEEE Transactions on Signal Processing*, 59(11):5617–5626, 2011.
- [25] N. Noorshams and M. J. Wainwright. Belief propagation for continuous state spaces: Stochastic message-passing with quantitative guarantees. *Journal of Machine Learning Research*, 14:2799–2835, September 2013.
- [26] N. Noorshams and M. J. Wainwright. Stochastic belief propagation: A low-complexity alternative to the sum-product algorithms. *IEEE Transactions on Information Theory*, 59(4):1981–2000, April 2013.
- [27] IEEE P802.3an. 10GBASE-T Task force. <http://www.ieee802.org/3/an>.
- [28] T. Richardson, A. Shokrollahi, and R. Urbanke. Design of capacity-approaching irregular low-density parity check codes. *IEEE Transactions on Information Theory*, 47:619–637, February 2001.
- [29] T. Richardson and R. Urbanke. The capacity of low-density parity check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47:599–618, February 2001.
- [30] T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.
- [31] G. Sarkis, S. Hemati, S. Mannor, and W. J. Gross. Stochastic decoding of LDPC codes over  $\text{GF}(q)$ . *IEEE Transactions on Communications*, 61(3):939–950, 2013.

- [32] A. D. Sarwate and T. Javidi. Opinion dynamics and distributed learning of distributions. In *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing*, pages 1151–1158, September 2011.
- [33] S. S. Tehrani, W. J. Gross, and S. Mannor. Stochastic decoding of LDPC codes. *IEEE Communications Letters*, 10(10):716–718, 2006.
- [34] S. S. Tehrani, A. Naderi, G. A. Kamendje, S. Hemati, S. Mannor, and W. J. Gross. Majority-based tracking forecast memories for stochastic LDPC decoding. *IEEE Transactions on Signal Processing*, 58(9):4883–4896, 2010.
- [35] C. Winstead, V. C. Gaudet, A. Rapley, and C. Schlegel. Stochastic iterative decoders. In *Proceedings of the International Symposium on Information Theory*, pages 1116–1120, February 2005.
- [36] C. Winstead, N. Nguyen, V.C. Gaudet, and C. Schlegel. Low-voltage CMOS circuits for analog iterative decoders. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 53(4):829–841, 2006.