# Batched Sparse Codes

Shenghao Yang and Raymond W. Yeung, *Fellow, IEEE*

**Abstract**

BATched Sparse codes (BATS codes) are proposed for transmitting a collection of packets through a communication network with packet loss. A BATS code consists of an inner code and an outer code over a finite field. The outer code is a matrix generalization of a fountain code that preserves desirable properties of the latter such as ratelessness and low encoding/decoding complexity. The outer code encodes the file to be transmitted into batches, each of which containing $M$ packets. When the batch size $M$ is equal to 1, the outer code reduces to a fountain code. The inner code is comprised of the linear network coding performed by the intermediate network nodes. With the constraint that linear network coding is applied only to packets within the same batch, the structure of the outer code is preserved. Furthermore, the computational capability of the intermediate network nodes required to apply BATS codes is independent of the number of packets for transmission. For tree networks, the size of the buffer required at the intermediate nodes is also independent of the number of packets for transmission. It is verified theoretically for certain cases and demonstrated numerically for some general cases that BATS codes asymptotically achieve rates very close to the capacity of the underlying networks.

**Index Terms**

Network coding, fountain codes, sparse graph codes, belief propagation.

## I. Introduction

One fundamental task of communication networks is to distribute a bulk of digital data, called a *file*, from a source node to a set of destination nodes. We consider this file distribution problem, called *multicast*, in *packet networks*, in which data packets transmitted on the network links can be lost due to channel noise, congestion, faulty network hardware, and so on.

Existing network protocols, for example TCP, mostly use retransmission to guarantee reliable transmission of individual packets. Retransmission relies on feedback and is not scalable for multicast transmission. On the other

The work of S. Yang was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00302. The work of S. Yang and R. W. Yeung was supported in part by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08).

This paper was presented in part at the IEEE International Symposium on Information Theory, Saint Petersburg, Russia, August 2011.

S. Yang is with the Institute for Theoretical Computer Science, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, P. R. China. This work was done when he was with the Institute of Network Coding, The Chinese University of Hong Kong (email: shyang@tsinghua.edu.cn).

R. W. Yeung is with the Department of Information Engineering and the Institute of Network Coding, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong Special Administrative Region, P. R. China (email: whyeung@ie.cuhk.edu.hk).

hand, fountain codes, including LT codes [1], Raptor codes [2] and online codes [3], provide a good solution for routing networks without relying on feedback, where the intermediate nodes apply store-and-forward. When using fountain codes, the source node keeps transmitting coded packets generated by a fountain code encoder and a destination node can decode the original file after receiving $n$ coded packets, where $n$ typically is only slightly larger than the number of the input packets, regardless of which $n$ packets are received. Fountain codes have the advantages of ratelessness, universality, and low encoding/decoding complexity. Taking Raptor codes as an example, both the encoding and decoding of a packet has constant complexity.

Routing, however, is not an optimal operation at the intermediate nodes in the presence of packet loss from the throughput point of view. For example, the routing capacity of the network in Fig. 1 is $0.64$ packet per use. If we allow decoding and encoding operations at the intermediate node and treat the network as a concatenation of two erasure channels, we can achieve the rate $0.8$ packet per use by using erasure codes on both links. For a general network, the maximum multicast rate can be achieved only by *network coding* [4]. Network coding allows an intermediate node to generate and transmit new packets using the packets it has received. Linear network coding [5] was proved to be sufficient for multicast communications and can be realized distributedly by random linear network coding [6]–[9].

The following network coding method has been proved to achieve the multicast capacity for networks with packet loss in a wide range of scenarios [10], [11]. The source node transmits random linear combinations of the input packets and an intermediate node transmits random linear combinations of the packets it has received. Note that no erasure codes are required for each link though packet loss is allowed. Network coding itself plays the role of end-to-end erasure codes. A destination node can decode the input packets when it receives enough coded packets with linearly independent coding vectors.

The above scheme, referred to as the baseline random linear network coding scheme, has been implemented in wireline peer-to-peer (P2P) networks [12], [13] (see [14] for a network coding analysis), in which every node in the network is required to decode the file. However, the computational and storage complexities of this scheme are not suitable for many other practical applications, in particular wireless applications. Consider transmitting $K$ packets where each packet consists of $T$ symbols in a finite field. The computational complexity of encoding in the source node is $\mathcal{O}(TK)$ per packet. An intermediate node needs to buffer all the packets it has received for network coding, so in the worst case, the storage cost is $K$ packets, and the computational complexity of encoding is $\mathcal{O}(TK)$ per packet. Decoding using Gaussian elimination has complexity $\mathcal{O}(K^2+TK)$ per packet. Though these complexities are polynomials in $K$, the baseline random linear network coding scheme is still difficult to implement for large $K$. In particular, the intermediate nodes, like network routers, usually have limited buffer capability. Since the size of the required buffer at the intermediate node depends on the file size, such an implementation cannot handle an arbitrarily large file.

In practice, we hope to build network coding enabled devices with limited storage and computational capabilities. Accordingly, it is desirable for a network coding scheme to have i) low encoding complexity in the source node and low decoding complexity in the destination nodes, ii) constant computational complexity of encoding a packet

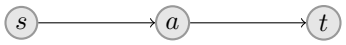


Fig. 1. In this network, $s$ is the source node, $t$ is the destination node, and $a$ is the intermediate node that does not demand the file. Both links are capable of transmitting one packet per use and have a packet loss rate 0.2.

in an intermediate node and constant buffer requirement in an intermediate node, iii) small protocol overhead, and iv) high transmission rate.

### *A. Some Previous Works*

There are roughly two classes of works for designing efficient file transmission schemes in networks with coding at the intermediate nodes, but they either cannot meet our requirement at the intermediate nodes or have other drawbacks.

The first class of works tries to extend fountain codes to networks with coding at the intermediate nodes. Since coding in the intermediate nodes changes the degrees of the packets, it is difficult to guarantee that the degrees of the received packets follow a specific distribution. Solutions have been proposed for special network topologies (e.g., line networks [15], [16]) and special communication scenarios (e.g., peer-to-peer file sharing [17], [18]), but those solutions are difficult to be extended to general network settings and cannot meet our requirement for the intermediate nodes. For example, all the schemes proposed in [15]–[18] require the intermediate nodes to have a buffer size that increases linearly with the number of packets for transmission.

The second class of works try to simplify the complexity of linear network coding using chunks [7]. A chunk (also called generation or class) is a subset of the packets for transmission. Encoding, recoding and decoding are all performed within one chunk. It reduces the encoding and decoding complexity to $\mathcal{O}(TL)$ and $\mathcal{O}(L^2+TL)$ per packet, respectively, where chunks are disjoint and have size $L$, but at the same time introduces the scheduling issues of chunks. Specifically, sequential scheduling of chunks requires feedback and is not scalable for multicast, while random scheduling of chunks requires the intermediate nodes to cache all the chunks [19]–[22]. For a detailed discussion on the scheduling issues, we refer the reader to [23].

### *B. Our Solution*

To address the issues of the existing schemes, we propose a solution called *BATched Sparse codes (BATS codes)*, which extends fountain codes to the realm of networks and at the same time incorporates random linear network coding. A BATS code consists of an inner code and an outer code over a finite field. The outer code is a matrix generalization of a fountain code, and hence rateless. The outer code encodes the file to be transmitted into *batches*, each containing $M$ packets. When the batch size $M$ is equal to 1, the outer code reduces to a fountain code. The inner code is comprised of the linear network coding performed by the intermediate network nodes. The only constraint on the linear network coding scheme (other than causality) is that only packets belonging to the same batch can be combined. Since the inner code does not change the structure of the outer code, an efficient belief

propagation (BP) decoding algorithm can be used to decode BATS codes. When the batch size $M$ is equal to $K$, a BATS code can become the baseline random linear network coding scheme (see a discussion in Section II-E).

When applying BATS codes, the encoding complexity of the outer code is $\mathcal{O}(TM)$ per packet and the corresponding decoding complexity is $\mathcal{O}(M^2 + TM)$ per packet. An intermediate node uses $\mathcal{O}(TM)$ time to recode a packet, and an intermediate node is required to buffer only $\mathcal{O}(M)$ packets for tree networks, including the three-node network in Fig. 1. Note that all these requirements for BATS codes do not depend on $K$, the total number of packets for transmission.

BATS codes are suitable for any network that allows linear network coding at the intermediate nodes. These codes are robust against dynamical network topology and packet loss since the end-to-end operation remains linear. Moreover, BATS codes can operate with small finite fields. In contrast, most existing random linear network coding schemes require a large field size to guarantee a full rank for the transfer matrix. For BATS codes, as we will see, the transfer matrices of the batches are allowed to have arbitrary rank deficiency.

Even though the underlying network can vary, the performance of a BATS code can be evaluated independent of the details of the intermediate operations and network topologies given the ranks of the transfer matrices applied on the batches. We use density evolution to analyze the BP decoding process of BATS codes, and obtain a sufficient and a necessary condition for the BP decoding succeeding with high probability.

A near optimal degree distribution for a BATS code can be obtained by solving an optimization problem induced by the sufficient condition. The optimization problem can be approximately solved by linear programming. When the empirical distribution of the transfer matrix ranks converges to a probability vector $(h_0, h_1, \ldots, h_M)$, we verify theoretically for certain cases and demonstrate numerically for some general cases that BATS codes can achieve rates very close to $\sum_i i h_i$, the maximum achievable rate in term of packet per use for such transfer matrices.

In the rest of this paper, BATS codes are formally introduced in Section II. The belief propagation decoding of BATS codes is analyzed in Section III. An optimization of the degree distribution is discussed in Section IV. An example of how to use BATS codes in the three-node network is illustrated in Section V.

## II. BATS Codes

In this section, we discuss the encoding and decoding of BATS codes. Consider encoding $K$ input packets, each of which has $T$ symbols in a finite field $\mathbb{F}$ with size $q$. A packet is denoted by a column vector in $\mathbb{F}^T$. The rank of a matrix is denoted by $\text{rk}(\mathbf{A})$. In the following discussion, we equate a set of packets to a matrix formed by juxtaposing the packets in this set. For example, we denote the set of the input packets by the matrix

$$\mathbf{B} = \Big[b_1, b_2, \cdots, b_K\Big],$$

where $b_i$ is the $i$th input packet. When treating the packets as a set, with an abuse of notation, we also write $b_i \in \mathbf{B}$, $\mathbf{B}' \subset \mathbf{B}$, etc.

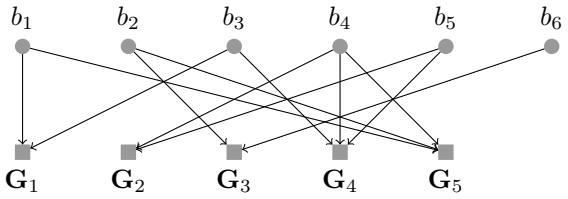


Fig. 2. Tanner graph for encoding and transmitting of the first five batches. Nodes in the first row are the variable nodes representing the input packets. Nodes in the second row are the check nodes representing the batches.

### A. Encoding of Batches

A *batch* is a set of $M$ coded packets generated from a subset of the $K$ input packets. For $i = 1, 2, \ldots$, the $i$th batch $\mathbf{X}_i$ is generated from a subset $\mathbf{B}_i \subset \mathbf{B}$ of the input packets by the operation

$$\mathbf{X}_i = \mathbf{B}_i \mathbf{G}_i,$$

where $\mathbf{G}_i$, a matrix with $M$ columns, is called the *generator matrix* of the $i$th batch. We call the packets in $\mathbf{B}_i$ the contributors of the $i$th batch. The formation of $\mathbf{B}_i$ is specified by a *degree distribution* $\Psi = (\Psi_0, \Psi_1, \cdots, \Psi_K)$: 1) sample the distribution $\Psi$ which returns a *degree* $d_i$ with probability $\Psi_{d_i}$; 2) uniformly at random choose $d_i$ input packets to form $\mathbf{B}_i$. The design of $\Psi$ is discussed later in Section IV.

The dimension of $\mathbf{G}_i$ is $d_i \times M$. In this paper, we analyze BATS codes with random generator matrices. Specifically, all the components of $\mathbf{G}_i$ are independently and uniformly chosen at random by the encoder. Such a random matrix is also called a *totally random matrix*. Random generator matrices do not only facilitate analysis but are also readily implementable. For example, $\mathbf{G}_i$, $i = 1, 2, \cdots$ can be generated by a pseudorandom number generator and can be recovered at the destinations by the same pseudorandom number generator.

The code described above, called the *outer code* of the BATS code, can be described by a Tanner graph. A Tanner graph has $K$ *variable nodes*, where variable node $i$ corresponds to the $i$th input packet $b_i$, and $n$ *check nodes*, where check node $j$ corresponds to the $j$th batch $\mathbf{X}_j$. Check node $j$ is connected to variable node $i$ if $b_i$ is a contributor of $\mathbf{X}_j$. Fig. 2 illustrates an example of a Tanner graph for encoding.

### B. Transmission of Batches

To transmit a batch, the source node transmits the packets in the batch, not necessarily in the order they are generated. No feedback is required to stop the transmission of each batch. A BATS code can be used as a rateless code, i.e., the number of batches transmitted is not fixed and is potentially unlimited. An intermediate node encodes the received packets within the same batch into new packets by taking random linear combinations and transmits these new packets on the outgoing links, i.e., random linear network coding is applied to packets belonging to the same batch. These new packets so generated are regarded as belonging to the same batch. The rule is that packets belonging to different batches are not mixed inside the network. BATS codes are robust against dynamical network

topology and packet loss since the end-to-end operation remains linear. The random linear network coding applied on batches is referred to as the *inner code* of the BATS code.

To apply BATS codes, we further need to consider how to schedule the transmission of batches at the source node and at the intermediate nodes, and how to manage the buffers at the intermediate nodes. The design of these network operations varies for different scenarios. For the file distribution in a P2P network, since all the network nodes request the file, random scheduling of the batches can reduce the protocol overhead. In contrast, since the intermediate node in the three-node network in Fig. 1 does not require the file, sequential scheduling of the batches at both the source node and the intermediate node can minimize the buffer requirement at the intermediate node. As we will show in Section V, caching one batch in the intermediate node is asymptotically optimal. The point is that the intermediate node always receives the packets of the same batch consecutively. Since only the packets of the same batch can be combined by network coding, it is not necessary to keep the batches whose transmission has been completed by the intermediate node. Note that the completion of the transmission of a batch at the intermediate node is signaled by the reception of the first packet of the next batch. Similarly, for tree networks with the root being the source node, sequential scheduling of the batches can also minimize the buffer requirement at the intermediate nodes.

Given the end-to-end transformations applied to the batches, the design of the outer code does not depend on the details of the network operations. So we will not discuss the detailed network operations on the batches in general networks. Nevertheless, we demonstrate how a BATS code works in the three-node network in Section V, where some general guidelines on the design of the intermediate operation are given. Note that though the three-node network is simple, it models many situations that arise in multiple hop transmissions in wireline and wireless communications.

Let $\mathbf{Y}_i$ be the received packets at a destination node that belong to the $i$th batch. We write

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{H}_i = \mathbf{B}_i\mathbf{G}_i\mathbf{H}_i, \tag{1}$$

where $\mathbf{H}_i$ is the transfer matrix incurred by the linear network coding operation of the network [5], [24] for the $i$th batch. The number of rows of $\mathbf{H}_i$ is $M$, while the number of columns varies for different batches and is finite. We assume that $\mathbf{H}_i$ is known by the destination node through the coding vectors in the packet headers. When the packet length $T$ is sufficiently large, this overhead is negligible. See an introduction of linear network coding in [25] for more details.

The operation of the network on the batches in (1) can be modeled as a *linear operator channel (LOC)*, which has been studied for linear network coding [26]–[28]. The outer code of a BATS code can be regarded as a channel code for the LOC. In the analysis of BATS codes, we assume that the empirical rank distribution of the transfer matrices converges in probability to a probability vector. This is a mild assumption since it does not require the ranks of the transfer matrices to be i.i.d. as in [27], [28]. See Appendix I for more discussion and a characterization of the capacity of such LOCs.

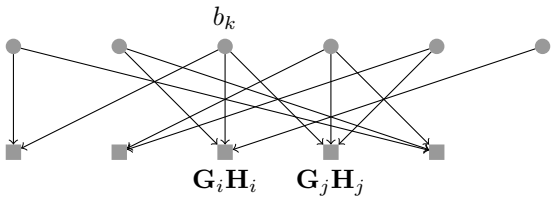


Fig. 3. A decoding graph. Nodes in the first row are the variable nodes representing the input packets. Nodes in the second row are the check nodes representing the batches.

### C. Belief Propagation Decoding

A destination tries to decode the input packets using $\mathbf{Y}_i$ and the knowledge of $\mathbf{G}_i\mathbf{H}_i$ for $i = 1, 2, \ldots, n$. The decoding process is better described using the bipartite graph in Fig. 3, which is the same as the encoding graph in Fig. 2 except that associated with each check node $i$ is the matrix $\mathbf{G}_i\mathbf{H}_i$.

A check node $i$ is called decodable if $\mathbf{G}_i\mathbf{H}_i$ has rank $d_i$, the degree of the $i$th batch. If so, then $\mathbf{B}_i$ is recovered by solving the linear system of equations $\mathbf{Y}_i = \mathbf{B}_i\mathbf{G}_i\mathbf{H}_i$, which has a unique solution since $\text{rk}(\mathbf{G}_i\mathbf{H}_i) = d_i$. After decoding the $i$th batch, we recover the $d_i$ input packets in $\mathbf{B}_i$. Then substitute the values of these input packets in the undecoded batches. Consider that $b_k$ is in $\mathbf{B}_i$. If variable node $k$ has only one edge that connects with check node $i$, just remove variable node $k$. If variable node $k$ also connects check node $j \neq i$, then besides removing the variable node, also remove the row in $\mathbf{G}_j\mathbf{H}_j$ corresponding to variable node $k$. In the decoding graph, this is equivalent to first removing check node $i$ and its neighboring variable nodes, and then for each removed variable node update its neighboring check nodes. We repeat this decoding-substitution procedure on the new graph until no more check nodes are decodable.

The degree distribution is the crucial parameter that affects the performance of the BP decoding. We want to design a degree distribution such that i) the BP decoding succeeds with high probability, ii) the encoding/decoding complexity is low, and iii) the coding rate is high. Based on the analysis of the decoding process in Section III, an optimization of the degree distribution will be provided in Section IV.

### D. Precoding of BATS Codes

The same technique of Raptor codes is applied here to reduce the encoding/decoding complexity of BATS codes. The input packets are first encoded using a traditional erasure code (precode), and then encoded by a BATS code. We require that the belief propagation decoding of the BATS code recovers a given fraction of its input packets. The traditional erasure code is capable of recovering the original input packets in face of a fixed fraction of erasures. Fig. 4 demonstrates a systematic precode together with a BATS code.

### E. Computation Complexity

The complexity of encoding a batch with degree $d$ is $\mathcal{O}(TMd)$. For a encoding graph with $n$ check nodes, i.e., $n$ batches, the encoding complexity is $\mathcal{O}(TM\sum_{i=1}^{n} d_i)$, which converges to $\mathcal{O}(TMn\,\mathbb{E}[\Psi])$ when $n$ is large, where

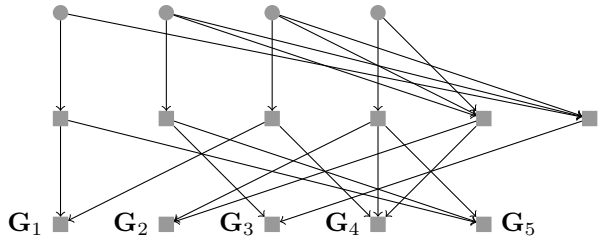


Fig. 4. Precoding of BATS codes. Nodes in the first row represent the input packets. Nodes in the second row represent the intermediate packets generated by the precode. Nodes in the third row represent the batches generated by the encoding of a BATS code.

$\mathbb{E}[\Psi] = \sum_d d\Psi_d$.

Let $k_i = \text{rk}(\mathbf{H}_i)$ and let $k_i'$ be the rank of $\mathbf{G}_i\mathbf{H}_i$ when check node $i$ is decodable. It is clear that $k_i' \leq k_i \leq M$. The decoding processing involves two parts: the first part is the decoding of the decodable check nodes, which has complexity $\mathcal{O}(\sum_i k_i'^3 + T\sum_i k_i'^2)$; the second part is updating the decoding graph, which has complexity $\mathcal{O}(T\sum_i(d_i - k_i')M)$. So the total complexity is $\mathcal{O}(\sum_i k_i'^3 + T\sum_i k_i'^2 + T\sum_i(d_i - k_i')M)$, which can be simplified to $\mathcal{O}(nM^3 + TM\sum_i d_i)$. When $n$ is large, the complexity converges to $\mathcal{O}(M^3 n + TMn\,\mathbb{E}[\Psi])$. Usually, $T$ and $\mathbb{E}[\Psi]$ is larger than $M$ and the second term is dominant.

We will see from Section IV that we can find a degree distribution with $\mathbb{E}[\Psi] = \mathcal{O}(M)$. In the design of BATS codes, $M$ is a parameter independent of $K$. The rate of the code is $\frac{K}{nM}$ packets per transmission. When the rate of the code converges to a constant value, we see that the encoding and decoding complexity are $\mathcal{O}(TKM)$ and $\mathcal{O}(KM^2 + TKM)$, respectively.

The batch size $M$ determines the tradeoff between the complexity and the maximum achievable rate. When $M = 1$, a BATS code degenerates to a Raptor code, which has the lowest computation complexity but cannot get the benefit of network coding. When $M = K$ and the degrees of all batches are $K$, a BATS code becomes the baseline random linear network coding scheme. In the second case, though the complexity is high, the potential of network coding can be fully realized.

## III. Decoding Analysis

Some existing methods for analyzing the BP decoding of erasure codes can be modified to analyze the BP decoding of BATS codes. In this paper, we adopt the differential equation approach [29] that has been used in [30] (see also [31]).

Compared with the analysis of fountain codes, BATS codes have a relatively complex decoding criteria that involves both the degree and the rank value of a check node. In addition to the evolution of the degrees of the check nodes, the evolution of the ranks of the check nodes also needs to be tracked in the decoding analysis.

### A. Random Decoding Graph

Consider a random decoding graph with $K$ variable nodes and $n$ check nodes. Fix a degree distribution $\Psi = (\Psi_0, \Psi_1, \cdots, \Psi_D)$, where $D$ is the maximum integer such that $\Psi_D$ is nonzero. Assume that $D = \mathcal{O}(M)$. The feasibility of this assumption will be justified later. The degree $d_i$ of check node $i$ is obtained by sampling the degree distribution $\Psi$. The $d_i$ neighbors of check node $i$ is uniformly chosen and the generator matrix $G_i$ of check node $i$ is a $d_i \times M$ totally random matrix, i.e., its components are uniformly i.i.d.

Let $H_i$ be the transfer matrix associated with check node $i$. Assume that the empirical distribution of the transfer matrix ranks converges in probability to a probability vector $h = (h_0, \ldots, h_M)$. Specifically, for $k = 0, \ldots, M$ let

$$\pi_k \triangleq \frac{|\{i : \text{rk}(H_i) = k\}|}{n}.$$

Note that $\pi_k$ depends on $n$. We assume that the convergence of the matrix ranks satisfies

$$|\pi_k - h_k| = \mathcal{O}(n^{-1/6}), \quad 0 \le k \le M, \tag{2}$$

with probability at least $1 - \gamma(n)$, where $\gamma(n) = o(1)$, i.e., there exists a constant $c$ such that for all sufficiently large $n$,

$$\Pr\{|\pi_k - h_k| < cn^{-1/6},\ 0 \le k \le M\} < 1 - \gamma(n),$$

and

$$\lim_{n\to\infty} \gamma(n) = 0.$$

Note that the above assumption on the convergence of $\{\pi_k\}$ is valid when $\{H_i\}$ are i.i.d. and $\text{rk}(H_i)$ follows the distribution $h$. We also assume that the transfer matrices are independent of the generation of batches. The random decoding graph of a BATS code described above is denoted by $\text{BATS}(K, n, \Psi, h)$.

We call $r_i = \text{rk}(G_i H_i)$ the *rank* of check node $i$. Define the following two regions of the degree-rank pair:

$$\bar{\mathcal{F}} \triangleq \{(d, r) : 1 \le r \le M, r \le d \le D\},$$

$$\mathcal{F} \triangleq \{(d, r) : 1 \le r \le M, r < d \le D\}.$$

We see that $\bar{\mathcal{F}} = \mathcal{F} \cup \{(r, r), r = 1, \ldots, M\}$. A check node with rank zero does not help the decoding, so we do not include $(d, 0)$ in $\bar{\mathcal{F}}$ and $\mathcal{F}$. To analyze the decoding process, we use the degree-rank distribution of the edges defined as follows. An edge is said to be of degree $d$ and rank $r$ if it is connected to a check node with degree $d$ and rank $r$. Let $R_{d,r}$ be the number of edges of degree $d$ and rank $r$. Define the *degree-rank distribution of the edges* as

$$\bar{R} \triangleq (R_{d,r}, (d, r) \in \bar{\mathcal{F}}).$$

Note that $R_{d,r}/d$ gives the number of nodes with degree $d$ and rank $r$.

For a check node with degree $d$ and transfer matrix rank $k$, the probability that it has rank $r$ is denoted by $\zeta_r^{d,k}$. The details can be found in Appendix II-A, but for the purpose of the discussion here, an explicit form of $\zeta_r^{d,k}$ is

not needed. Let

$$h_{d,r} \triangleq \sum_{k=r}^{M} \zeta_r^{d,k} h_k \tag{3}$$

be the probability that a check node with degree $d$ has rank $r$ when the rank of the transfer matrix is chosen according to the probability vector $h$. Let

$$\rho_{d,r} \triangleq d\Psi_d h_{d,r}, \tag{4}$$

where $n\rho_{d,r}$ is the expected number of edges of degree $d$ and rank $r$ in the decoding graph when the rank of a transfer matrix is chosen according to the probability vector $h$ independently. The following lemma shows that $R_{d,r}/n$ converges in probability to $\rho_{d,r}$ as $n$ goes to infinity.

*Lemma 1:* With probability at least $1 - (\gamma(n) + 2MD\exp(-2n^{2/3}))$,

$$\left|\frac{R_{d,r}}{n} - \rho_{d,r}\right| = \mathcal{O}(n^{-1/6}), \quad (d,r) \in \bar{\mathcal{F}}.$$

*Proof:* Consider the instances of $\{\pi_k\}$ satisfying (2). By the assumption on $\{\pi_k\}$, this will decrease the bound by at most $\gamma(n)$. With an abuse of notation, we treat $\{\pi_k\}$ as an instance satisfying (2) in the following of this proof.

The decoding graph has $n\pi_k$ check nodes with transfer matrix rank $k$. For a check node with degree $d$ and transfer matrix rank $k$, the probability that it has rank $r$ is $\zeta_r^{d,k}$ when $r \leq \min\{d,k\}$, and is zero otherwise. Thus the expected number of check nodes with degree $d$ and rank $r$ is

$$\sum_{k=r}^{M} n\pi_k \Psi_d \zeta_r^{d,k} = n\Psi_d \sum_{k=r}^{M} \pi_k \zeta_r^{d,k}.$$

By Hoeffding's inequality, with probability at least $1 - 2MD\exp(-2n^{2/3})$,

$$\left|\frac{R_{d,r}}{dn} - \Psi_d \sum_{k=r}^{M} \pi_k \zeta_r^{d,k}\right| < n^{-1/6}, \quad (d,r) \in \bar{\mathcal{F}}. \tag{5}$$

Then,

$$\begin{aligned}\left|\frac{R_{d,r}}{dn} - \Psi_d h_{d,r}\right| &= \left|\frac{R_{d,r}}{dn} - \Psi_d \sum_{k=r}^{M} \pi_k \zeta_r^{d,k} + \Psi_d \sum_{k=r}^{M} \pi_k \zeta_r^{d,k} - \Psi_d h_{d,r}\right| \\ &\leq \left|\frac{R_{d,r}}{dn} - \Psi_d \sum_{k=r}^{M} \pi_k \zeta_r^{d,k}\right| + \Psi_d \sum_{k=r}^{M} |\pi_k - h_k| \zeta_r^{d,k},\end{aligned}$$

where the last inequality follows from the triangle inequality and the definition of $h_{d,r}$ in (3). By (5), under the condition of (2), we have

$$\left|\frac{R_{d,r}}{dn} - \Psi_d h_{d,r}\right| = \mathcal{O}(n^{-1/6})$$

with probability at least $1 - 2MD\exp(-2n^{2/3})$. The proof is completed by considering $\{\pi_k\}$ not satisfying (2). ■

We will analyze the average decoding performance of BATS$(K, n, \Psi, h)$ with a random decoding strategy. In each decoding step, an edge $(U, V)$ with degree equal to the rank is uniformly chosen, where $U$ is a check node and $V$ is a variable node. Since check node $U$ has degree equal to the rank, variable node $V$ is decodable. Variable node $V$, as well as all the edges connected to it, are removed in the decoding graph. For each check node connected

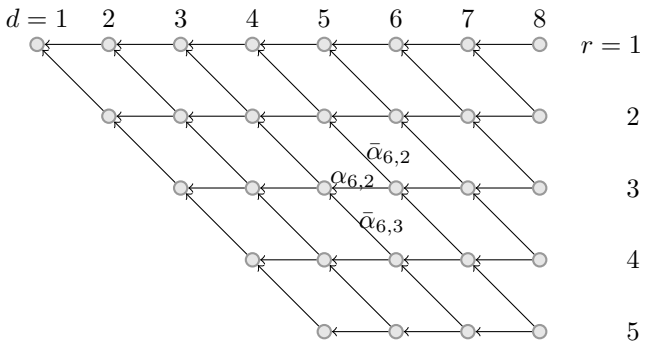


Fig. 5. State transition diagram for $M = 5$ and $D = 8$. Each node in the graph represent a degree-rank pair. In each step, if the check node connects to the decoded variable node, its state changes according to the direction of the outgoing edges of its current state. The label on an edge shows the probability that a direction is chosen.

to variable node $V$, three operations are applied: 1) the degree is reduced by 1; 2) the row in the generator matrix corresponding to the variable node $V$ is removed; and 3) the rank is updated accordingly. The decoding process stops when there is no edge with degree equal to the rank. The following decoding analysis is based on this random decoding strategy. In the decoding process described in the last section, decoding a check node with degree equal to the rank can recover several variable nodes simultaneously. Note that for a given instance of the decoding graph, both strategies will reduce the decoding graph to the same residual graph when they stop (see the discussion in Appendix III).

### *B. Density Evolution*

Consider the evolution of BATS$(K, n, \Psi, h)$ during the decoding process. Time $t$ starts at zero and increases by one for each variable node removed by the decoder. Let $R_{d,r}(t)$ denote the number of edges in the residual graph of degree $d$ and rank $r$ at time $t \geq 0$ with $R_{d,r}(0) = R_{d,r}$.

Upon removing a neighboring variable node of a check node with degree $d$ and rank $r$, the degree of the check node will change to $d-1$. The rank of the check node may remain unchanged with probability

$$\alpha_{d,r} = \frac{1 - q^{-d+r}}{1 - q^{-d}}, \quad (d, r) \in \bar{\mathcal{F}} \tag{6}$$

(see the derivation in Appendix II-A), or may change to $r-1$ with probability $\bar{\alpha}_{d,r} = 1 - \alpha_{d,r}$. Regarding a degree-rank pair as a state, the state transition of a check node during the decoding process is illustrated in Fig. 5.

Assume that the process has not stopped. At time $t$, we have $K - t$ variable nodes left in the residual graph, and an edge with degree equal to the rank is uniformly chosen to be removed. Let

$$\bar{R}(t) \triangleq (R_{d,r}(t) : (d, r) \in \bar{\mathcal{F}}).$$

As we will show in the following lemma, the random process $\{\bar{R}(t)\}$ is a Markov chain. This suggests a straightforward approach to compute all the transition probabilities in the Markov chain, but as discussed in [29], this

approach may lead to a complicated formula. Instead of taking this approach, we work out the expected change $R_{d,r}(t+1) - R_{d,r}(t)$ explicitly for all $t \geq 0$. Let

$$R_0(t) = \sum_{r=1}^{M} R_{r,r}(t).$$

We do not need to study the behavior of $R_{r,r}(t)$ for the individual value of $r$ since $R_0(t)$ is sufficient to determine when the decoding process stops. Specifically, the decoding process stops as soon as $R_0(t)$ becomes zero.

*Lemma 2:* The random process $\{\bar{R}(t)\}$ is a Markov chain, and for any constant $c \in (0,1)$, as long as $t \leq cK$ and $R_0(t) > 0$, we have

$$\mathbb{E}[R_{d,r}(t+1) - R_{d,r}(t)|\bar{R}(t)] = (\alpha_{d+1,r} R_{d+1,r}(t) + \bar{\alpha}_{d+1,r+1} R_{d+1,r+1}(t) - R_{d,r}(t)) \frac{d}{K-t}, \ (d,r) \in \mathcal{F}, \quad (7)$$

and

$$\mathbb{E}[R_0(t+1) - R_0(t)|\bar{R}(t)] = \frac{\sum_r r\alpha_{r+1,r} R_{r+1,r}(t)}{K-t} - \frac{R_0(t)}{K-t} - 1 + \mathcal{O}(1/K). \quad (8)$$

*Proof:* Fix a time $t \geq 0$. With an abuse of notation, we treat $\bar{R}(0), \ldots, \bar{R}(t)$ as instances in the proof, i.e., the values of these random vectors are fixed. Let $(U,V)$ be the edge chosen to be removed at time $t$, where $V$ is the variable node and $U$ is the check node, according to the random decoding algorithm described in Section III-A. Note that $V$ is uniformly distributed among all variable nodes and $U$ must be a check node with degree equal to the rank at time $t$.

Define indicator random variables $\iota_{d,r}(i)$, $i = 1, \ldots, \frac{R_{d,r}(t)}{d}$, where $\iota_{d,r}(i) = 1$ if the $i$th check node with degree $d$ and rank $r$ becomes degree $d-1$ and rank $r$ at time $t+1$. Define indicator random variables $\mu_{d,r}(i)$, $i = 1, \ldots, \frac{R_{d,r}(t)}{d}$, where $\mu_{d,r}(i) = 1$ if the $i$th check node with degree $d$ and rank $r$ becomes degree $d-1$ and rank $r-1$ at time $t+1$. The difference $R_{d,r}(t+1) - R_{d,r}(t)$ can then be expressed as

$$\begin{aligned} &R_{d,r}(t+1) - R_{d,r}(t) \\ &= \sum_{i=1}^{R_{d+1,r}(t)/(d+1)} d \cdot \iota_{d+1,r}(i) + \sum_{i=1}^{R_{d+1,r+1}(t)/(d+1)} d \cdot \mu_{d+1,r+1}(i) - \sum_{i=1}^{R_{d,r}(t)/d} d(\iota_{d,r}(i) + \mu_{d,r}(i)). \end{aligned} \quad (9)$$

Let us look at the joint distribution of $\iota_{d,r}(i), \mu_{d,r}(i)$, $(d,r) \in \bar{\mathcal{F}}$, $1 \leq i \leq \frac{R_{d,r}(t)}{d}$. Let $A_r(i)$ be the event that $U$ is the $i$th check node with degree $r$ and rank $r$. Since $(U,V)$ is uniformly distributed among all edges with degree equal to rank, we have that $A_r(i)$ for all $r$ and $i$ are mutually exclusive and

$$\Pr\{A_r(i)\} = \frac{r}{R_0(t)}.$$

Define indicator random variable $\beta_{d,r}(i)$ with $\beta_{d,r}(i) = 1$ if $V$ is a neighbor of the $i$th check node with degree $d$ and rank $r$. Conditioning on $A_{r'}(i')$, by the construction of the random decoding graph, we know that $\beta_{d,r}(i)$, $(d,r) \in \bar{\mathcal{F}}$, $i = 1, \ldots, \frac{R_{d,r}(t)}{d}$ are independent and

$$\Pr\{\beta_{d,r}(i) = 1 | A_{r'}(i')\} = \begin{cases} 1 & d = r = r', i = i', \\ \frac{d}{K-t} & \text{otherwise.} \end{cases} \quad (10)$$

Conditioning on $\beta_{d,r}(i)$, $(d,r) \in \bar{\mathcal{F}}$, $i = 1, \ldots, \frac{R_{d,r}(t)}{d}$, by the construction of the random decoding graph, we know that $(\iota_{d,r}(i), \mu_{d,r}(i))$, $(d,r) \in \bar{\mathcal{F}}$, $1 \le i \le \frac{R_{d,r}(t)}{d}$ are independent and $(\iota_{d,r}(i), \mu_{d,r}(i))$ only depends on $\beta_{d,r}(i)$. Specifically, we have

$$\Pr\{\iota_{d,r}(i) = 0, \mu_{d,r}(i) = 0 | \beta_{d,r}(i) = 0\} = 1$$

since the degree of a check node will not change if $V$ is not a neighbor,

$$\Pr\{\iota_{d,r}(i) = 1, \mu_{d,r}(i) = 1 | \beta_{d,r}(i) = 1\} = 0$$

since $\iota_{d,r}(i)$ and $\mu_{d,r}(i)$ cannot both be 1,

$$\Pr\{\iota_{d,r}(i) = 1, \mu_{d,r}(i) = 0 | \beta_{d,r}(i) = 1\} = \alpha_{d,r} \tag{11}$$

(cf. (6)), and

$$\Pr\{\iota_{d,r}(i) = 0, \mu_{d,r}(i) = 1 | \beta_{d,r}(i) = 1\} = \bar{\alpha}_{d,r}. \tag{12}$$

By (9), $\bar{R}(t+1)$ is a deterministic function of $\bar{R}(t)$ and $\iota_{d,r}(i), \mu_{d,r}(i)$, $(d,r) \in \bar{\mathcal{F}}$, $i = 1, \ldots, \frac{R_{d,r}(t)}{d}$, where the distribution of the latter part is determined by $\bar{R}(t)$ independent of $\bar{R}(t'), t' < t$. Thus, the random process $\{\bar{R}(t)\}$ is a Markov chain.

Now we calculate the marginal distribution of $\iota_{d,r}(i)$ and $\mu_{d,r}(i)$ for all $(d,r)$ and $i$. When $d \neq r$, we have

$$\begin{aligned}
&\Pr\{\iota_{d,r}(i) = 1\} \\
&= \sum_{a,b} \sum_{r',i'} \Pr\{\iota_{d,r}(i) = 1, \mu_{d,r}(i) = b | \beta_{d,r}(i) = a\} \Pr\{\beta_{d,r}(i) = a | A_{r'}(i')\} \Pr\{A_{r'}(i')\} \\
&= \sum_{r',i'} \alpha_{d,r} \Pr\{\beta_{d,r}(i) = 1 | A_{r'}(i')\} \Pr\{A_{r'}(i')\} && (13) \\
&= \sum_{r',i'} \alpha_{d,r} \frac{d}{K-t} \Pr\{A_{r'}(i')\} && (14) \\
&= \alpha_{d,r} \frac{d}{K-t},
\end{aligned}$$

where (13) follows from (11), and (14) follows from (12) with $d \neq r$; and similarly

$$\Pr\{\mu_{d,r}(i) = 1\} = \bar{\alpha}_{d,r} \frac{d}{K-t}.$$

When $d = r$, we have

$$\begin{aligned}
&\Pr\{\iota_{r,r}(i) = 1\} \\
&= \sum_{a,b} \sum_{r',i'} \Pr\{\iota_{r,r}(i) = 1, \mu_{r,r}(i) = b | \beta_{r,r}(i) = a\} \Pr\{\beta_{r,r}(i) = a | A_{r'}(i')\} \Pr\{A_{r'}(i')\} \\
&= \sum_{r',i'} \alpha_{r,r} \Pr\{\beta_{r,r}(i) = 1 | A_{r'}(i')\} \Pr\{A_{r'}(i')\} && (15) \\
&= 0, && (16)
\end{aligned}$$

where (15) follows from (11), and (16) follows from $\alpha_{r,r} = 0$ (cf. (6)); and

$$\begin{aligned}
&\Pr\{\mu_{r,r}(i) = 1\} \\
&= \sum_{a,b} \sum_{r',i'} \Pr\{\iota_{r,r}(i) = b, \mu_{r,r}(i) = 1 | \beta_{r,r}(i) = a\} \Pr\{\beta_{r,r}(i) = a | A_{r'}(i')\} \Pr\{A_{r'}(i')\} \\
&= \sum_{r',i'} \Pr\{\beta_{r,r}(i) = 1 | A_{r'}(i')\} \Pr\{A_{r'}(i')\} \qquad (17) \\
&= \Pr\{A_r(i)\} + \sum_{r',i'} \frac{r}{K-t} \Pr\{A_{r'}(i')\} - \frac{r}{K-t} \Pr\{A_r(i)\} \qquad (18) \\
&= \frac{r}{R_0(t)} + \frac{r}{K-t} - \frac{r}{R_0(t)} \frac{r}{K-t},
\end{aligned}$$

where (17) follows from (12) and $\bar{\alpha}_{r,r} = 1$, and (18) follows from (10).

The expectation in (7) is obtained by taking expectation on (9). To verify (8), note that when $d = r$ in (9), $\iota_{r,r} = 0$ in the last term. Then we have

$$\begin{aligned}
R_0(t+1) - R_0(t) &= \sum_r \left( R_{r,r}(t+1) - R_{r,r}(t) \right) \\
&= \sum_r r \sum_{i=1}^{R_{r+1,r}(t)/(r+1)} \iota_{r+1,r}(i) - \sum_r \sum_{i=1}^{R_{r,r}(t)/r} \mu_{r,r}(i). \qquad (19)
\end{aligned}$$

Taking expectation on (19), we have

$$\begin{aligned}
\mathbb{E}[R_0(t+1) - R_0(t)] &= \sum_r r\alpha_{r+1,r} \frac{R_{r+1,r}(t)}{K-t} - \sum_r \left( \frac{R_{r,r}(t)}{R_0(t)} + \left(1 - \frac{r}{R_0(t)}\right) \frac{R_{r,r}(t)}{K-t} \right) \\
&= \sum_r r\alpha_{r+1,r} \frac{R_{r+1,r}(t)}{K-t} - \frac{R_0(t)}{K-t} - 1 + \sum_r \frac{r}{R_0(t)} \frac{R_{r,r}(t)}{K-t}.
\end{aligned}$$

The expectation in (8) is obtained by noting that $\sum_r \frac{r}{R_0(t)} \frac{R_{r,r}(t)}{K-t} < \frac{M^2}{K(1-c)}$ since $t \le cK$ by assumption. ■

### C. Sufficient and Necessary Conditions

We care about when $R_0(t)$ goes to zero for the first time. The evolution of $R_0(t)$ depends on that of $R_{d,r}(t)$, $(d,r) \in \mathcal{F}$. To study the trend of $R_0(t)$, the differential equation approach [29] leads us to consider the system of differential equations

$$\begin{aligned}
\frac{d\rho_{d,r}(\tau)}{d\tau} &= \big( \alpha_{d+1,r} \rho_{d+1,r}(\tau) + \bar{\alpha}_{d+1,r+1} \rho_{d+1,r+1}(\tau) \\
&\qquad - \rho_{d,r}(\tau) \big) \frac{d}{\theta - \tau}, \quad (d,r) \in \mathcal{F}, \qquad (20) \\
\frac{d\rho_0(\tau)}{d\tau} &= \frac{\sum_{r=1}^{D-1} r\alpha_{r+1,r} \rho_{r+1,r}(\tau) - \rho_0(\tau)}{\theta - \tau} - 1 \qquad (21)
\end{aligned}$$

with initial values $\rho_{d,r}(0) = \rho_{d,r}$, $(d,r) \in \mathcal{F}$, and $\rho_0(0) = \sum_r \rho_{r,r}$, where $\theta = K/n$ is the design rate of the BATS code.

We can get some intuition about how the system of differential equations is obtained by replacing $R_{d,r}(t)$ and $R_0(t)$ with $n\rho_{d,r}(t/n)$ and $n\rho_0(t/n)$, respectively, in (7) and (8). Defining $\tau = t/n$ and letting $n \to \infty$, we obtain

the system of differential equations in (20) and (21). The expectation is ignored because $\rho_{d,r}(\tau)$ and $\rho_0(\tau)$ are deterministic functions. Theorem 5 in Appendix IV makes the above intuition rigorous.

The system of differential equations in (20) and (21) is solved in Appendix V for $0 \leq \tau < \theta$. The solution of (21) is

$$\rho_0(\tau) = \left(1 - \frac{\tau}{\theta}\right)\left(\sum_{r=1}^{M} \alpha_{r+1,r} \sum_{d=r+1}^{D} \rho_{d,r}^{(d-r-1)} \, \mathrm{I}_{d-r,r}\left(\frac{\tau}{\theta}\right) + \sum_{r=1}^{M} \rho_{r,r} + \theta \ln(1 - \tau/\theta)\right), \tag{22}$$

where $\rho_{d,r}^{(d-r-1)}$ is defined by the recursive formula

$$\rho_{d,r}^{(0)} \triangleq \rho_{d,r}, \tag{23}$$

$$\rho_{d,r}^{(i+1)} \triangleq \alpha_{d-i,r}\rho_{d,r}^{(i)} + \bar{\alpha}_{d-i,r+1}\rho_{d,r+1}^{(i)}; \tag{24}$$

and

$$\mathrm{I}_{a,b}(x) \triangleq \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} x^j (1-x)^{a+b-1-j}$$

is called the *regularized incomplete beta function*. For $\bar{\eta} \in (0,1)$, the following theorem shows that if $\rho_0(\tau) > 0$ for $\tau \in [0, \bar{\eta}]$, then the decoding does not stop until $t > \bar{\eta}K$ with high probability, and $R_{d,r}(t)$ and $R_0(t)$ can be approximated by $n\rho_{d,r}(t/n)$ and $n\rho_0(t/n)$, respectively.

*Theorem 1:* Consider a sequence of decoding graphs BATS$(K, n, \Psi, h)$, $n = 1, 2, \ldots$ with fixed $\theta = K/n$, and the empirical rank distribution of transfer matrices $(\pi_0, \ldots, \pi_M)$ satisfying

$$|\pi_i - h_i| = \mathcal{O}(n^{-1/6}), \quad 0 \leq i \leq M, \tag{25}$$

with probability at least $1 - \gamma(n)$, where $\gamma(n) = o(1)$. For $\bar{\eta} \in (0,1)$,

(i) if $\rho_0(\tau) > 0$ for $\tau \in [0, \bar{\eta}\theta]$, then for sufficiently large $K$, with probability $1 - \mathcal{O}(n^{7/24}\exp(-n^{1/8})) - \gamma(n)$, the decoding terminates with at least $\bar{\eta}K$ variable nodes decoded, and

$$|R_{d,r}(t) - n\rho_{d,r}(t/n)| = \mathcal{O}(n^{5/6}), \ (d,r) \in \mathcal{F}$$

$$|R_0(t) - n\rho_0(t/n)| = \mathcal{O}(n^{5/6})$$

uniformly for $t \in [0, \bar{\eta}K]$;

(ii) if $\rho_0(\tau) < 0$ for some $\tau \in [0, \bar{\eta}\theta]$, then for sufficiently large $K$, with probability $1 - \mathcal{O}(n^{7/24}\exp(-n^{1/8})) - \gamma(n)$, the decoding terminates before $\bar{\eta}K$ variable nodes are decoded.

*Proof:* See Appendix IV. ■

## IV. Optimization of Degree Distribution

Theorem 1 gives a sufficient and a necessary condition such that the BP decoding succeeds with high probability. These conditions induce an optimization problem that generates a degree distribution that meets our requirement in Section II-C.

### A. Optimization

We first define some new notations to help the formulation of the optimization of the degree distribution. Let

$$h_r^* \triangleq \alpha_{r+1,r} h_{r+1,r}. \tag{26}$$

Since $h_{r+1,r}$ is a linear function of $h$ (ref. (3)), $h_r^*$ is also a linear function of $h$. Define

$$\Omega(x; h, \Psi) \triangleq \sum_{r=1}^{M} h_r^* \sum_{d=r+1}^{D} d\Psi_d \, \mathrm{I}_{d-r,r}(x) + \sum_{r=1}^{M} h_{r,r} r \Psi_r. \tag{27}$$

When the context is clear, we also write $\Omega(x; \Psi)$, $\Omega(x; h)$ or $\Omega(x)$ to simplify the notation. The expression of $\rho_0$ in (22) can be simplified as follows.

*Lemma 3:* $\rho_0(\tau) = (1 - \tau/\theta)\,(\Omega(\tau/\theta) + \theta \ln(1 - \tau/\theta))$.

*Proof:* Define

$$h_{d,r}^{(0)} \triangleq h_{d,r}, \quad 1 \le r \le M, r \le d \le D. \tag{28}$$

For $d > r$ and $0 \le i \le d - r - 1$, define

$$h_{d,r}^{(i+1)} \triangleq \alpha_{d-i,r} h_{d,r}^{(i)} + \bar{\alpha}_{d-i,r+1} h_{d,r+1}^{(i)}. \tag{29}$$

By Lemma 4 in Appendix II-A, $h_{d,r}^{(d-r-1)} = h_{r+1,r}$. By the definitions in (4), (23) and (28), we have $\rho_{d,r}^{(0)} = d\Psi_d h_{d,r}^{(0)}$. Since the recursive formulas in (24) and (29) are the same, we have for $i = 1, \ldots, d - r - 1$,

$$\rho_{d,r}^{(i)} = d\Psi_d h_{d,r}^{(i)}.$$

Substitute $\rho_{d,r}^{(d-r-1)} = d\Psi_d h_{d,r}^{(d-r-1)} = d\Psi_d h_{r+1,r}$ in (22). Last, using the definition of $h_r^*$ in (26), we obtain the formula in the lemma. ■

For $\bar{\eta} \in (0, 1)$, we say a rate $\theta$ is *$\bar{\eta}$-achievable* by BATS codes if for every $\epsilon > 0$ and every sufficiently large $K$ there exists a BATS code with $K$ input packets such that for $n \le K/\theta$ received batches, the BP decoding recovers at least $\bar{\eta}K$ input packets with probability at least $1 - \epsilon$. Define an optimization problem

$$\max \quad \theta \tag{30a}$$

$$\text{s.t.} \quad \Omega(x) + \theta \ln(1 - x) \ge 0, \qquad 0 \le x \le \bar{\eta}, \tag{30b}$$

$$\sum_d \Psi_d = 1 \quad \text{and} \quad \Psi_d \ge 0, \; d = 1, \cdots, D. \tag{30c}$$

Let $\hat{\theta}$ be the optimal value in (30).

*Proposition 1:* When the empirical rank distribution of the transfer matrices converges to $h = (h_0, \ldots, h_M)$ (in the sense of (25)), for any $\epsilon > 0$, the rate $\hat{\theta} - \epsilon$ is $\bar{\eta}$-achievable by BATS codes.

*Proof:* To show that $\hat{\theta} - \epsilon$ is $\bar{\eta}$-achievable, by Theorem 1 and Lemma 3, we only need to show that there exists a degree distribution such that

$$\Omega(x) + (\hat{\theta} - \epsilon) \ln(1 - x) > 0, \quad 0 \le x \le \bar{\eta}. \tag{31}$$

For the degree distribution $\Psi$ that achieves $\hat{\theta}$, we have from (30b)

$$\Omega(x;\Psi) + \hat{\theta}\ln(1-x) \geq 0, \quad 0 \leq x \leq \bar{\eta}.$$

Multiplying by $\frac{\hat{\theta}-\epsilon}{\hat{\theta}}$, we have

$$\frac{\hat{\theta}-\epsilon}{\hat{\theta}}\Omega(x;\Psi) + (\hat{\theta}-\epsilon)\ln(1-x) \geq 0, \quad 0 \leq x \leq \bar{\eta}. \tag{32}$$

Since $\Omega(x;\Psi) > 0$ for $x > 0$, (32) implies that $\Psi$ satisfies (31) except possibly for $x = 0$. Checking the definition of $\Omega$ in (27), we have $\Omega(0;\Psi) = \sum_{r=1}^{M} h_{r,r} r \Psi_r$. If $\sum_{r=1}^{M} h_{r,r} r \Psi_r > 0$, which implies $\Psi$ satisfies (31), we are done. In the following, we consider $\sum_{r=1}^{M} h_{r,r} r \Psi_r = 0$.

Let $r^*$ be the largest integer $r$ such that $h_r > 0$. We can characterize that $h_{r,r} = 0$ for $r > r^*$ and $h_{r,r} > 0$ for $r \leq r^*$ (cf. (3) and (45) in Appendix II-A). Since $\sum_{r=1}^{M} h_{r,r} r \Psi_r = 0$, we know that $\sum_{d \leq r^*} \Psi_d = 0$. Define a new degree distribution $\Psi'$ by $\Psi'_d = \Psi_d \frac{\hat{\theta}-\epsilon}{\hat{\theta}}$ for $d > r^*$ and $\Psi'_d = \Delta$ for $d \leq r^*$, where $\Delta > 0$ can be determined by the constraint $\sum_d \Psi'_d = 1$. The formulation of $\Omega$ in (27) can be rewritten as

$$\Omega(x;\Psi) = \sum_{d=1}^{D} \Psi_d f_d(x)$$

for certain functions $f_d(x)$, $d = 1, \ldots, D$ not related to $\Psi$. Using the above formulation, we have for $0 \leq x \leq \bar{\eta}$,

$$\begin{aligned}
\Omega(x,\Psi') - \frac{\hat{\theta}-\epsilon}{\hat{\theta}}\Omega(x;\Psi) &= \sum_{d=1}^{r^*} \Psi'_d f_d(x) + \sum_{d=r^*+1}^{D} \frac{\hat{\theta}-\epsilon}{\hat{\theta}} \Psi_d f_d(x) - \frac{\hat{\theta}-\epsilon}{\hat{\theta}} \sum_{d=r^*+1}^{D} \Psi_d f_d(x) \\
&= \sum_{d=1}^{r^*} \Delta f_d(x) \\
&\geq \Delta \sum_{d=1}^{r^*} d h_{d,d} \qquad (33) \\
&> 0,
\end{aligned}$$

where (33) follows from $f_d(x) \geq d h_{d,d}$. By (32), $\Psi'$ satisfies (31). ■

For many cases, we can directly use the degree distribution $\Psi$ obtained by solving (30). But when $\Omega(0;\Psi) = 0$, by Lemma 3, $\rho_0(0) = \sum_r \rho_{r,r} = 0$, and hence $\Psi_d = 0$, $d \leq M$ (cf. (4)). Thus, $\Psi$ does not guarantee that the decoding can start. We can then modify $\Psi$ as we do in the proof of Proposition 1 by increasing the probability masses $\Psi_d$, $d \leq M$ a little bit to make sure that the decoding can start.

The maximum degree $D$ in (30c) affects the encoding/decoding complexity. In Section III-A, we have assumed that $D = \mathcal{O}(M)$. The next theorem shows that it is optimal to choose $D \leq \lceil M/\eta \rceil - 1$, where $\eta = 1 - \bar{\eta}$.

*Theorem 2:* Using $D > \lceil M/\eta \rceil - 1$ does not give a better optimal value in (30), where $\eta = 1 - \bar{\eta}$.

*Proof:* Consider an integer $\Delta$ such that $\eta \geq \frac{M}{\Delta+1}$. Let $\Psi$ be a degree distribution with $\sum_{d>\Delta} \Psi_d > 0$. Construct a new degree distribution $\tilde{\Psi}$ as follows:

$$\tilde{\Psi}_d = \Psi_d, \quad d < \Delta,$$

$$\tilde{\Psi}_\Delta = \sum_{d \geq \Delta} \Psi_d,$$

and

$$\tilde{\Psi}_d = 0, \quad d > \Delta.$$

We can show that $\Omega(x;\tilde{\Psi}) > \Omega(x;\Psi)$ for all $0 < x \leq 1-\eta$. Write

$$\begin{aligned}&\Omega(x;\tilde{\Psi}) - \Omega(x;\Psi)\\ &\quad= \sum_{d=\Delta+1}^{\infty} \Psi_d \sum_{r=1}^{M} h_r^*(\Delta\, \mathrm{I}_{\Delta-r,r}(x) - d\, \mathrm{I}_{d-r,r}(x))\end{aligned}$$

For $d \geq \Delta + 1$,

$$\frac{r-1}{d-r} \leq \frac{M-1}{d-M} < \frac{M}{\Delta - M + 1} \leq \frac{\eta}{1-\eta}.$$

So we can apply Lemma 8 in Appendix II-B to show that, for $0 < x \leq 1-\eta$,

$$\begin{aligned}\frac{d\, \mathrm{I}_{d-r,r}(x)}{(d-1)\, \mathrm{I}_{d-1-r,r}(x)} &< \frac{d}{d-1}\left(1 - \frac{\eta}{r}\right)\\ &\leq \frac{d}{d-1}\left(1 - \frac{\eta}{M}\right)\\ &\leq \frac{\Delta+1}{\Delta}\left(1 - \frac{1}{\Delta+1}\right)\\ &= 1,\end{aligned}$$

which gives $\Omega(x;\tilde{\Psi}) > \Omega(x;\Psi)$ for $0 < x \leq 1-\eta$.

Thus, for certain $\theta$ such that

$$\Omega(x;\Psi) + \theta \ln(1-x) \geq 0, \quad 0 \leq x \leq 1-\eta,$$

we have

$$\Omega(x;\tilde{\Psi}) + \theta \ln(1-x) \geq 0, \quad 0 \leq x \leq 1-\eta.$$

This means that $\tilde{\Psi}$ is potentially better than $\Psi$. So we do not need to consider a degree distribution $\Psi$ with $\sum_{d>\Delta} \Psi_d > 0$. Thus, it is sufficient to take the maximum degree $D \leq \min_{\eta \geq \frac{M}{\Delta+1}} \Delta = \lceil M/\eta \rceil - 1$. ■

The converse of Proposition 1 is that "a rate larger than $\hat{\theta}$ is not $\bar{\eta}$-achievable". Intuitively, for any $\epsilon > 0$, we cannot have a degree distribution such that

$$\Omega(x) + (\hat{\theta} + \epsilon) \ln(1-x) \geq 0, \quad 0 \leq x \leq \bar{\eta},$$

since otherwise, $\hat{\theta}$ is not the optimal value in (30). Thus, for any degree distribution, $\Omega(x) + (\hat{\theta} + \epsilon) \ln(1-x) < 0$ for some $x \in [0, \bar{\eta}]$. Taking $\hat{\theta} + \epsilon$ in place of $\theta$ in Lemma 3, for any degree distribution, $\rho_0(\tau) < 0$ for some $\tau \in [0, \bar{\eta}(\hat{\theta} + \epsilon)]$. Hence, we can apply the second part of Theorem 1 to show that $\hat{\theta} + \epsilon$ is not $\bar{\eta}$-achievable for any degree distribution, since for any degree distribution there exists $K_0$ such that when the number of input packets $K \geq K_0$, with probability approaching 1 the BATS code cannot recover $\bar{\eta}K$ input packets. To rigorously prove this converse, however, we need a uniform bound $K_0$ for all degree distributions such that the second part of Theorem 1 holds, which is very tedious if not impossible. Instead of taking this approach, we demonstrate in the rest of the section that $\hat{\theta}$ is close to the capacity of the underlying LOC (cf. Section II-B and Appendix I).

### B. Upper Bounds on the Achievable Rates

The first upper bound on the optimal value $\hat{\theta}$ of (30) is given by the capacity of LOCs. When the empirical rank distribution of the transfer matrices converging to $h = (h_0, \ldots, h_M)$, the capacity of a LOC, in terms of packets per use, is $\mathbb{E}[h] = \sum_r r h_r$ (see Appendix I). The BP decoding algorithm recovers at least a fraction of $\bar{\eta}$ of all the input packets with high probability. So asymptotically BATS codes under BP decoding can recover at least $\bar{\eta}\hat{\theta}$ fraction of the input packets. Thus, we have $\bar{\eta}\hat{\theta} \leq \mathbb{E}[h]$.

A tighter upper bound can be obtained by analyzing (30) directly. Using Lemma 5 in Appendix II-A, rewrite

$$\begin{aligned}\Omega(x;\Psi) &= \sum_{r=1}^{M} h_r^* \sum_{d=r+1}^{D} d\Psi_d \, \mathrm{I}_{d-r,r}(x) + \sum_{r=1}^{M} h_r^* \sum_{d=1}^{r} d\Psi_d, \\ &= \sum_{r=1}^{M} h_r^* S_r(x;\Psi), \end{aligned} \tag{34}$$

where

$$S_r(x;\Psi) = S_r(x) \triangleq \sum_{d=r+1}^{D} d\Psi_d \, \mathrm{I}_{d-r,r}(x) + \sum_{d=1}^{r} d\Psi_d. \tag{35}$$

This form of $\Omega(x;\Psi)$ will be used in subsequent proofs.

*Theorem 3:* The optimal value $\hat{\theta}$ of (30) satisfies

$$\bar{\eta}\hat{\theta} \leq \sum_{r=1}^{M} r h_r^*,$$

where $h_r^*$ is defined in (26).

Note that by Lemma 6 in Appendix II-A, we have $\sum_r r h_r^* \leq \sum_r r h_r$, i.e., Theorem 3 gives a strictly better upper bound than $\mathbb{E}[h]$. When $q \to \infty$, $\sum_r r h_r^* \to \sum_r r h_r$ (cf. (56) in Appendix II-A). So when the field size is large, these two upper bounds are very close.

*Proof:* Using (58) in Appendix II-B, we have

$$\begin{aligned}\int_0^1 S_r(x)\,dx &= \sum_{d=r+1}^{D} d\Psi_d \int_0^1 \mathrm{I}_{d-r,r}(x)\,dx + \sum_{d=1}^{r} d\Psi_d \\ &= \sum_{d=r+1}^{D} r\Psi_d + \sum_{d=1}^{r} d\Psi_d \\ &\leq r \sum_{d=1}^{D} \Psi_d \\ &= r.\end{aligned}$$

Hence,

$$\int_0^1 \Omega(x)\,dx = \int_0^1 \sum_{r=1}^{M} h_r^* S_r(x)\,dx \leq \sum_{r=1}^{M} r h_r^*. \tag{36}$$

Since $\Omega(x)$ is an increasing function and the inequality in (30b) holds for $x = 1 - \eta = \bar{\eta}$,

$$\int_{1-\eta}^{1} \Omega(x)\,dx \geq \eta\Omega(1-\eta) \geq -\eta\hat{\theta}\ln\eta. \tag{37}$$

Since $\Omega(x)+\hat{\theta}\ln(1-x)\geq 0$ for $0<x\leq 1-\eta$,

$$\begin{aligned}&\int_0^{1-\eta}\Omega(x)dx-\hat{\theta}(\eta\ln\eta+1-\eta)\\&=\int_0^{1-\eta}\Omega(x)dx+\hat{\theta}\int_0^{1-\eta}\ln(1-x)dx\\&\geq 0.\end{aligned}\tag{38}$$

Therefore, by (36)-(38), we have

$$\begin{aligned}\sum_{r=1}^{M}rh_r^*&\geq\int_0^1\Omega(x)dx\\&=\int_0^{1-\eta}\Omega(x)dx+\int_{1-\eta}^1\Omega(x)dx\\&\geq\hat{\theta}(\eta\ln\eta+1-\eta)-\eta\hat{\theta}\ln\eta\\&=\hat{\theta}(1-\eta).\end{aligned}$$

■

### C. Lower Bound on the Achievable Rates

We prove for a special case and demonstrate by simulation for general cases that the optimal value $\hat{\theta}$ of (30) is very close to $\sum_r rh_r^*$.

*Theorem 4:* The optimal value $\hat{\theta}$ of (30) satisfies

$$\hat{\theta}\geq\max_{r=1,2,\cdots,M}r\sum_{i=r}^{M}h_i^*.$$

Before proving Theorem 4, we first explain why Theorem 3 and Theorem 4 together demonstrate that $\hat{\theta}$ is close to the capacity of the underlying LOC for a special case. Consider the case with $h_\kappa=1$ for some $1\leq\kappa\leq M$. Theorem 4 implies that $\hat{\theta}\geq\kappa h_\kappa^*$. On the other hand, Theorem 3 says that $(1-\eta)\hat{\theta}\leq\sum_r rh_r^*=\kappa h_\kappa^*+\sum_{r<\kappa}rh_r^*$ (cf. (56) in Appendix II-A). Note that $\eta$ can be arbitrarily small, and $\sum_{r<\kappa}rh_r^*\to 0$ and $h_\kappa^*\to h_\kappa$ when the field size goes to infinity (again cf. (56)). Thus, the upper bound in Theorem 3 and the lower bound in Theorem 4 match $\mathbb{E}[h]=\kappa h_\kappa$ asymptotically when $h_\kappa=1$ for some $1\leq\kappa\leq M$.

*Proof:* Define degree distribution $\Psi^r$ as

$$\Psi_d^r=\begin{cases}0 & d\leq r,\\ \frac{r}{d(d-1)} & d=r+1,\cdots,D-1,\\ \frac{r}{D-1} & d=D.\end{cases}\tag{39}$$

Recall the definition of $S_r(x;\Psi)$ in (35). For $M\geq r'\geq r$, we will show that

$$S_{r'}(x;\Psi^r)+r\ln(1-x)>0,\quad 0\leq x\leq 1-\eta.\tag{40}$$

By Lemma 9 in Appendix II-B,

$$-r\ln(1-x)=r\sum_{d=r'+1}^{\infty}\frac{1}{d-1}\,\mathrm{I}_{d-r',r'}(x).$$

By (35) and (39),

$$\begin{aligned} S_{r'}(x;\Psi^r) + r\ln(1-x) &\geq \sum_{d=r'+1}^{D} d\Psi_d^r \, \mathrm{I}_{d-r',r'}(x) - r\sum_{d=r'+1}^{\infty} \frac{1}{d-1}\,\mathrm{I}_{d-r',r'}(x) \\ &\geq r\frac{D}{D-1}\,\mathrm{I}_{D-r',r'}(x) - r\sum_{d=D}^{\infty}\frac{1}{d-1}\,\mathrm{I}_{d-r',r'}(x) \\ &= r\,\mathrm{I}_{D-r',r'}(x) - r\sum_{d=D+1}^{\infty}\frac{1}{d-1}\,\mathrm{I}_{d-r',r'}(x). \end{aligned}$$

We will show that $\mathrm{I}_{D-r',r'}(x) > \sum_{d=D+1}^{\infty}\frac{1}{d-1}\,\mathrm{I}_{d-r',r'}(x)$ for $x \in [0, 1-\eta]$. This is equivalent to show that

$$\sum_{d=D+1}^{\infty}\frac{1}{d-1}\frac{\mathrm{I}_{d-r',r'}(x)}{\mathrm{I}_{D-r',r'}(x)} < 1 \quad \text{for } x \in [0,1-\eta]. \tag{41}$$

By Lemma 7 in Appendix II-B, $\frac{\mathrm{I}_{d-r',r'}(x)}{\mathrm{I}_{D-r',r'}(x)}$ is monotonically increasing, so we only need to prove the above inequality for $x = 1-\eta$. By Lemma 8 in Appendix II-B, $\frac{\mathrm{I}_{d-r',r'}(1-\eta)}{\mathrm{I}_{D-r',r'}(1-\eta)} < (1-\frac{\eta}{M})^{d-D}$. Therefore,

$$\begin{aligned} \sum_{d=D+1}^{\infty}\frac{1}{d-1}\frac{\mathrm{I}_{d-r',r'}(x)}{\mathrm{I}_{D-r',r'}(x)} &\leq \frac{1}{D}\sum_{d=D+1}^{\infty}\frac{\mathrm{I}_{d-r',r'}(1-\eta)}{\mathrm{I}_{D-r',r'}(1-\eta)} \\ &< \frac{1}{D}\sum_{d=D+1}^{\infty}(1-\frac{\eta}{M})^{d-D} \\ &= \frac{M-\eta}{D\eta} \\ &\leq 1, \end{aligned}$$

where the last inequality follows from $D = \lceil M/\eta \rceil - 1$. So we have established (41) and hence (40).

Last, by (34) and (40), we have for $0 \leq x \leq 1-\eta$,

$$\begin{aligned} \Omega(x;\Psi^r) &\geq \sum_{r'\geq r} h^*_{r'} S_{r'}(x;\Psi^r) \\ &> -\ln(1-x) r \sum_{r'\geq r} h^*_{r'}, \end{aligned}$$

or

$$\Omega(x;\Psi^r) + \left(r\sum_{r'\geq r} h^*_{r'}\right)\ln(1-x) > 0. \tag{42}$$

Comparing (42) and (30b), we conclude that $\hat{\theta} \geq r\sum_{r'\geq r} h^*_{r'}$. The proof is completed by considering all $r = 1, 2, \cdots, M$. ∎

### *D. Numerical Results*

To see the achievable rates for the general cases, we numerically solve (30) by taking discrete values of $x$. Let $x_i = (1-\eta)\frac{i}{N}$ for some integer $N$. We relax (30b) by considering only $x = x_i$, $i = 1, \ldots, N$. Let $\tilde{\theta}$ be the optimal value of (30) with this relaxation. Numerical results show that when $N$ is large, $\tilde{\theta}$ becomes small. When $N$ is reasonably large, e.g., 100, the optimal value becomes stable.

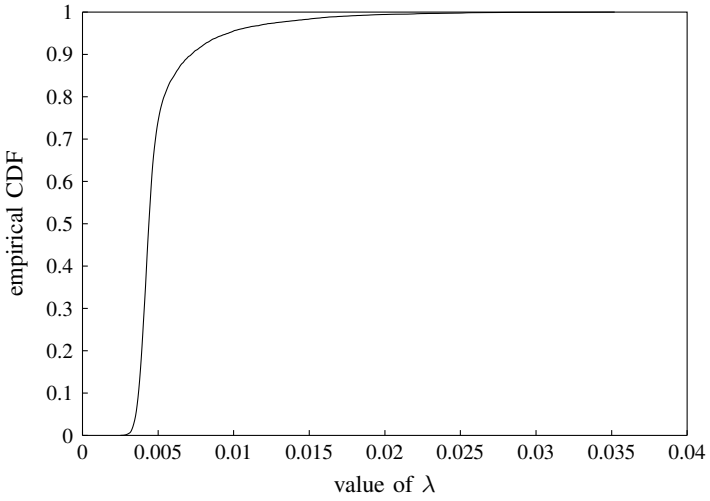


Fig. 6. The empirical cumulative distribution function (CDF) of $\lambda \triangleq (\sum_{r=1}^{M} rh_r^* - (1-\eta)\tilde{\theta})/\sum_{r=1}^{M} rh_r^*$ for 10000 rank distributions uniformly at random chosen. Here $q = 2^8$ and $M = 16$.

Set $M = 16$, $q = 2^8$ and $\eta = 0.01$. A rank distribution $(h_0, h_1, \ldots, h_M)$ is generated as follows: First, for $i = 0, \ldots, M$, $h_i$ is independently and uniformly chosen between zero and one; Second, normalize the rank distribution such that $\sum_i h_i = 1$. We compute $\tilde{\theta}$ for 10000 rank distributions independently generated and compare $(1-\eta)\tilde{\theta}$ with $\sum_r rh_r^*$ by computing $\lambda \triangleq (\sum_r rh_r^* - (1-\eta)\tilde{\theta})/\sum_r rh_r^*$. The results show that for more than $99\%$ of the rank distributions, $\lambda$ is smaller than $0.02$, and among all the samples the largest $\lambda$ is $0.0352$. Note that for these parameters, the difference $\sum_r rh_r - \sum_r rh_r^*$ is of the order $10^{-3}$, so the upper bound in Theorem 3 is indeed very close to the capacity. The empirical cumulative distribution function of $\lambda$ is drawn in Fig. 6.

## V. An Example of BATS Codes

We apply BATS codes in the network in Fig. 1. The source node $s$ performs BATS code encoding. In each time slot, node $s$ sends a packet to node $a$. Assume transmission is instantaneous and node $a$ receives the packet, if not erased, at the same time slot. No matter whether certain packets are received or not, node $a$ transmits at each time slot a linear combination of the packets it has received so far. After $M$ time slots, node $s$ switches to another batch and node $a$ clears its buffer for the last batch. These operations of the network minimize the transmission delay and are asymptotically optimal when $M$ goes to infinity.

The operation at node $a$ for a batch is given by a random matrix $\Phi$, an $M \times M$ upper unitriangular[1] matrix with all the upper triangular, off-diagonal entries being independent and uniformly distributed. Let $E$ be an $M \times M$ random diagonal matrix with independent components. A diagonal component of $E$ is $0$ with probability $0.2$ and is $1$ with

[1]A unitriangular matrix has unit entries on the main diagonal. The intermediate operation modelled by a unitriangular matrix becomes forwarding when $M = 1$.

TABLE I

DEGREE DISTRIBUTIONS FOR $M = 16$, $q = 2, 4, 8, 16$, RESPECTIVE. HERE ONLY THE DOMINANT PROBABILITY MASSES ARE LISTED. THE SUMMATION OF ALL OTHER PROBABILITY MASSES ARE LESS THAN $0.001$.

| $\Psi$ | $q = 2$ | $q = 4$ | $q = 8$ | $q = 16$ |
|---|---|---|---|---|
| $\Psi_{13}$ | 0.1500 | | | |
| $\Psi_{14}$ | 0.3262 | 0.2691 | 0.1850 | 0.1593 |
| $\Psi_{15}$ | | 0.1225 | 0.1903 | 0.2105 |
| $\Psi_{20}$ | | | | 0.0020 |
| $\Psi_{21}$ | 0.0491 | 0.2081 | 0.2078 | 0.2055 |
| $\Psi_{22}$ | 0.1546 | | | |
| $\Psi_{28}$ | | | | 0.0019 |
| $\Psi_{29}$ | | 0.0996 | 0.1172 | 0.1227 |
| $\Psi_{30}$ | | | 0.0019 | |
| $\Psi_{31}$ | 0.0310 | | | |
| $\Psi_{32}$ | 0.0848 | | | |
| $\Psi_{35}$ | | 0.0854 | | |
| $\Psi_{36}$ | | | 0.0155 | |
| $\Psi_{37}$ | | | 0.0732 | 0.0437 |
| $\Psi_{38}$ | | | | 0.0481 |
| $\Psi_{46}$ | 0.0986 | | | |
| $\Psi_{51}$ | | 0.0936 | | |
| $\Psi_{52}$ | | 0.0168 | | |
| $\Psi_{53}$ | | | 0.1073 | 0.0015 |
| $\Psi_{54}$ | | | | 0.1031 |
| $\Psi_{81}$ | 0.1058 | | | |
| $\Psi_{91}$ | | 0.1040 | | |
| $\Psi_{94}$ | | | 0.1026 | |
| $\Psi_{95}$ | | | | 0.1017 |

probability $0.8$. The matrix $E$ models the erasures in a link. The transfer matrix of the network is $H = E_1 \Phi E_2$, where $E_1$, $\Phi$ and $E_2$ are independent, and $E_1$ and $E_2$ follow the same distribution of $E$.

The rank distribution of $H$ is approximated by the empirical distribution obtained using $10^5$ independent samples of $H$. Using the (empirical) rank distribution, a degree distribution is obtained by solving (30) by taking discrete values of $x$. Table I lists some degree distributions for different parameters by setting $\eta = 0.08$.

BATS codes are rateless, i.e., the coding rate is not fixed. To see the performance of a BATS code, we use the average coding rate defined as follows. Consider that the source node encodes $K$ packets using a BATS code, and the decoder stops after recovering $\bar{\eta}K$ packets. Here we assume that a precode is used to first encode the original message into $K$ packets and any $\bar{\eta}K$ out of these $K$ packets are sufficient to recover the message by decoding the precode (cf. Section II-D). Repeat the above simulation $J$ times and let $n_j$ be the number of batches used when the decoder stops in the $j$th simulation. The average coding rate of the BATS code is defined as $\bar{\eta}KJ/(M\sum_j n_j)$, where the rate is normalized by $M$ for the sake of comparison of different value of $M$. In the following, we will

TABLE II

AVERAGE CODING RATES FOR $M = 16$.

| $K$ | $q = 2$ | $q = 4$ | $q = 8$ | $q = 16$ |
|---|---|---|---|---|
| 16000 | 0.5466 | 0.6208 | 0.6384 | 0.6484 |
| 32000 | 0.5589 | 0.6377 | **0.6563** | **0.6636** |
| 64000 | 0.5671 | **0.6484** | **0.6670** | **0.6755** |
| Capacity | 0.6948 | 0.7074 | 0.7115 | 0.7125 |

TABLE III

AVERAGE CODING RATES FOR $M = 32$.

| $K$ | $q = 2$ | $q = 4$ | $q = 8$ | $q = 16$ |
|---|---|---|---|---|
| 16000 | 0.5826 | 0.6145 | 0.6203 | 0.6248 |
| 32000 | 0.6087 | **0.6441** | **0.6524** | **0.6574** |
| 64000 | 0.6259 | **0.6655** | **0.6762** | **0.6818** |
| Capacity | 0.7178 | 0.7292 | 0.7325 | 0.7334 |

compare the average coding rates for different parameters, where the average coding rates are maximized over $\bar{\eta}$.

The first thing we want to show is that BATS codes outperform fountain codes. We know that when $M = 1$, BATS codes become Raptor codes and the intermediate operation $T$ becomes forwarding. BATS codes can achieve rates exceeding $0.64$ (see the rates in bold letters in Table II and III), the routing capacity, which serves as an upper bound on the maximum achievable rate for Raptor codes.

The capacity of the LOC formed by the network operation (normalized by $M$) is $\mathbb{E}[\text{rk}(H)]/M$, which takes the field size $q$ and the batch size $M$ as parameters. The rows labeled by Capacity in Table II and III show the numerical values of $\mathbb{E}[\text{rk}(H)]/M$. The simulation results demonstrate that for fixed $q$ and $M$, when $K$ becomes larger, the achievable rate approaches the capacity. For any fixed $q$, it is not difficult to show that

$$\frac{\mathbb{E}[\text{rk}(H)]}{M} \to 0.8, \quad M \to \infty.$$

So when $M$ is large, BATS codes can potentially achieve higher rates. Our simulation results also illustrate this trend. We observe that when $M$ becomes larger, capacity values are generally higher when the field sizes are the same.

Another trend we observe is that using large $q$ also increases the rates. A closer look at the simulations further reveals that the gain by increasing $q$ becomes smaller when $q$ is large. For example, when $M = 16$ and $K = 32000$, increasing $q$ from $2$ to $4$ gains $5.82\%$ in the rate, but increasing $q$ from $4$ to $8$ gains only $1.29\%$.

## VI. CONCLUDING REMARKS

Benefiting from network coding and the properties of fountain codes, BATS codes are ideal for transmitting files through communication networks. Besides low encoding/decoding complexity, BATS codes can be realized with

constant computation and storage complexity at the intermediate nodes. This desirable property makes BATS code a suitable candidate for the making of universal network coding based network devices that can potentially replace routers.

In this paper, we mainly discussed the design of BATS codes for one destination node. Given a rank distribution, we can design BATS codes that can achieve nearly optimal rates when the empirical distribution of the transfer matrix rank converges to rank distribution. For more practical applications, we need to consider BATS codes for multiple destination nodes which may have different empirical distributions of the transfer matrix ranks and to design BATS codes for unknown empirical distribution of the transfer matrix ranks. The sufficient condition of the degree distribution for successful decoding (Theorem 1) can be readily applied for multiple rank distributions. We leave the discussion of designing BATS codes for these scenarios to future works.

## ACKNOWLEDGEMENTS

We thank Prof. Rüdiger Urbanke for his insightful input to this work.

## APPENDIX I
## LINEAR OPERATOR CHANNELS

The network operation given by the linear network coding on batches described in Section II-B can be modelled by a *linear operator channel (LOC)* over finite fields. Let $X_i$ be a $T \times M$ random matrix representing the $i$th input, let $Y_i$ be a $T$-row random matrix representing the output of the network for the $i$th use, and let $H_i$ be an $M$-row random matrix representing the network operation. A LOC with input $X_i$, $i = 1, 2, \ldots$, and output $Y_i$, $i = 1, 2, \ldots$, is given by

$$Y_i = X_i H_i.$$

We assume that the instances of $H_i$ are unknown at the transmitter but known at the receiver. The number of columns of $H_i$ is arbitrary but finite.

Let $X^n = (X_1, \ldots, X_n)$. $Y^n$ and $H^n$ are defined similarly. We assume that $X^n$ and $H^n$ are independent for all $n$. When $H_i$, $i = 1, 2, \ldots$ are independent and follow the same but arbitrary distribution of a random matrix $H$, the LOC is a discrete memoryless channel (DMC) and its capacity is $\mathbb{E}[\text{rk}(H)]$ packet per use (see [28], [32]). Here we show that the capacity of the LOC can be similarly characterized when the transfer matrices change in an arbitrary way defined as follows.

Let

$$\pi_k \triangleq \frac{|\{i : 1 \le i \le n, \text{rk}(H_i) = k\}|}{n}.$$

Note that $\pi_k$ depends on $n$. Let $(h_0, \ldots, h_M)$ be a probability vector. We assume that the convergence of the matrix ranks satisfies

$$\Pr\{|\pi_k - \zeta_k| \le \sigma(n),\ k = 0, \ldots, M\} \ge 1 - \psi(n), \tag{43}$$

where $\sigma(n) = o(1)$ and $\psi = o(1)$. In other words, the empirical rank distribution of $H^n$ converges to $(h_0, \ldots, h_M)$ as $n$ goes to infinity. Note that the above assumption on the convergence of $\{\pi_k\}$ is valid when $\{H_i\}$ are i.i.d. and $\mathrm{rk}(H_i)$ follows the distribution $h$. Further, we assume that $X^n$ and $H^n$ are statistically independent for all $n$.

An LOC described above is an arbitrarily varying channel (AVC) with convergent state constraints (refer to [33, Chapter 6] and [34] for more information about AVCs). Here we are concerned about the capacity of the LOC for average error probability and randomized codes. For a randomized code, the encoder and the decoder can share a common randomness which is independent of the channel states. Randomized codes can potentially achieve higher rates than deterministic codes. Based on the results of the capacity of AVCs with state constraints [35], the capacity of AVCs with convergent state constraints can be obtained [36]. As a special case, the capacity of the LOC with constraint (43) is $\sum_{k=1}^{M} kh_k$ packet per use.

## APPENDIX II
## SOME LEMMAS

### *A. Rank of a Random Matrix*

All matrices discussed here are over the finite field with $q$ elements. Define

$$\zeta_r^m \triangleq \begin{cases} (1-q^{-m})(1-q^{-m+1})\cdots(1-q^{-m+r-1}) & r > 0, \\ 1 & r = 0. \end{cases}$$

For $1 \le r \le m$, we can count as follows (see also [37], [38]) that the number of full rank $r \times M$ matrices is

$$q^{rm}\zeta_r^m = (q^m - 1)(q^m - q)\cdots(q^m - q^{r-1}). \tag{44}$$

A full rank $m \times r$ matrix can be obtained by picking its $r$ columns from $\mathbb{F}^m$ one by one. The first column has $q^m - 1$ choices, and the $i$th column, $1 < i \le r$, cannot be picked in the subspaces spanned by the first $i-1$ columns, and hence has $q^m - q^{i-1}$ choices. So (44) is the number of full rank $r \times M$ matrices. We say a matrix is totally random if all its components are uniformly i.i.d. By the above counting problem, the probability that an $r \times m$ totally random matrix is full rank is exactly $\zeta_r^m$.

Let $G_d$ be a totally random matrix with $d$ rows and $M$ columns, and let $H$ be an arbitrary random matrix with $M$ rows. We show in the following that that for $i \le M$ and $r \le \min\{d, i\}$,

$$\Pr\{\mathrm{rk}(G_d H) = r \,|\, \mathrm{rk}(H) = i\} = \frac{\zeta_r^d \zeta_r^i}{\zeta_r^r q^{(d-r)(i-r)}} \triangleq \zeta_r^{d,i}. \tag{45}$$

Let $\mathbf{H}$ be any instance of $H$ with rank $i$. The $d$ rows of $G_d\mathbf{H}$ are i.i.d. and are uniformly distributed among the $i$-dimensional vector space spanned by the rows of $\mathbf{H}$. Thus, the probability of $\mathrm{rk}(G_d\mathbf{H}) = r$ is equal to the probability of a totally random $d \times i$ matrix being rank $r$. The latter is the ratio of the number of $d \times i$ matrices with rank $r$, which is $q^{dr}\zeta_r^d q^{ir}\zeta_r^i/(q^{rr}\zeta_r^r)$ (see [37], [39]). Thus,

$$\begin{aligned} \Pr\{\mathrm{rk}(G_d\mathbf{H}) = r\} &= \frac{q^{dr}\zeta_r^d q^{ir}\zeta_r^i/(q^{rr}\zeta_r^r)}{q^{di}} \\ &= \frac{\zeta_r^d \zeta_r^i}{\zeta_r^r q^{(d-r)(i-r)}}. \end{aligned}$$

Noting that the above probability is only related to the rank of $\mathbf{H}$, the verification of (45) is completed.

Let $h_i = p_{\mathrm{rk}(H)}(i)$. Using (45), we have for $r \leq \min\{d, M\}$,

$$\begin{aligned}\Pr\{\mathrm{rk}(G_d H) = r\} &= \sum_{i=r}^{M} \Pr\{\mathrm{rk}(G_d H) = r | \, \mathrm{rk}(H) = i\} p_{\mathrm{rk}(H)}(i) \\ &= \sum_{i=r}^{M} \zeta_r^{d,i} h_i. \end{aligned} \tag{46}$$

This gives another meaning of $h_{d,r}$ defined in (3), i.e.,

$$h_{d,r} = \Pr\{\mathrm{rk}(G_d H) = r\}. \tag{47}$$

Let $g$ be a row of $G_d$ and let $G'_d$ be the submatrix of $G_d$ without the row $g$. We see that $\alpha_{d,r}$ define in (6) is given by the conditional probability

$$\alpha_{d,r} = \Pr\{\mathrm{rk}(G'_d H) = r | \, \mathrm{rk}(G_d H) = r\}. \tag{48}$$

Clearly, $\alpha_{r,r} = 0$. When $d > r$,

$$\begin{aligned}\alpha_{d,r} &= \frac{\Pr\{\mathrm{rk}(G'_d H) = r, \mathrm{rk}(G_d H) = r\}}{\Pr\{\mathrm{rk}(G_d H) = r\}} \\ &= \frac{\sum_i \Pr\{\mathrm{rk}(G'_d H) = r, \mathrm{rk}(G_d H) = r | \, \mathrm{rk}(H) = i\} h_i}{\Pr\{\mathrm{rk}(G_d H) = r\}} \\ &= \frac{\sum_i \zeta_r^{d-1,i} q^{r-i} h_i}{\Pr\{\mathrm{rk}(G_d H) = r\}} \qquad (49) \\ &= \frac{\frac{1-q^{-d+r}}{1-q^{-d}} \sum_i \zeta_r^{d,i} h_i}{\Pr\{\mathrm{rk}(G_d H) = r\}} \\ &= \frac{1 - q^{-d+r}}{1 - q^{-d}}, \qquad (50)\end{aligned}$$

where (49) follows from

$$\begin{aligned}&\Pr\{\mathrm{rk}(G'_d H) = r, \mathrm{rk}(G_d H) = r | \, \mathrm{rk}(H) = i\} \\ &\quad = \Pr\{\mathrm{rk}(G'_d H) = r | \, \mathrm{rk}(H) = i\} \Pr\{\mathrm{rk}(G_d H) = r | \, \mathrm{rk}(H) = i, \mathrm{rk}(G'_d H) = r\} \\ &\quad = \zeta_r^{d-1,i} \Pr\{gH \in \langle G'_d H \rangle | \, \mathrm{rk}(H) = i, \mathrm{rk}(G'_d H) = r\} \\ &\quad = \zeta_r^{d-1,i} q^{r-i}, \end{aligned} \tag{51}$$

and (50) follows from (46).

As define in (28) and (29), for $1 \leq r \leq M$ and $r \leq d \leq D$,

$$h_{d,r}^{(0)} = h_{d,r},$$

and for $0 \leq i \leq d - r - 1$ and $d > r$,

$$h_{d,r}^{(i+1)} = \alpha_{d-i,r} h_{d,r}^{(i)} + \bar{\alpha}_{d-i,r+1} h_{d,r+1}^{(i)}.$$

*Lemma 4:* $h_{d,r}^{(i)} = h_{d-i,r}$ for $d > r$ and $1 \leq i \leq d - r$.

*Proof:* First we show that $h_{d,r}^{(1)} = h_{d-1,r}$ for $d > r$. Let $R = \text{rk}(G_d H)$ and $R' = \text{rk}(G'_d H)$. We have

$$\begin{aligned} h_{d,r}^{(1)} &= \alpha_{d,r} h_{d,r} + \bar{\alpha}_{d,r+1} h_{d,r+1} \\ &= \Pr\{R' = r | R = r\} \Pr\{R = r\} + \Pr\{R' = r | R = r+1\} \Pr\{R = r+1\} \quad (52) \\ &= \Pr\{R' = r, R = r\} + \Pr\{R' = r, R = r+1\} \\ &= \Pr\{R' = r, R \in \{r, r+1\}\} \\ &= \Pr\{R' = r\} \Pr\{R \in \{r, r+1\} | R' = r\} \\ &= \Pr\{R' = r\} \quad (53) \\ &= h_{d-1,r}, \quad (54) \end{aligned}$$

where (52) and (54) follow from (47) and (48); (53) follows because $\text{rk}(G_d H)$ must be either $r$ or $r+1$ under the condition that $\text{rk}(G'_d H) = r$.

The general case of the Lemma is proved by induction. Assume $h_{d,r}^{(i)} = h_{d-i,r}$ for $d > r$ and $i < d - r$. Then

$$\begin{aligned} h_{d,r}^{(i+1)} &= \alpha_{d-i,r} h_{d,r}^{(i)} + \bar{\alpha}_{d-i,r+1} h_{d,r+1}^{(i)} \\ &= \alpha_{d-i,r} h_{d-i,r} + \bar{\alpha}_{d-i,r+1} h_{d-i,r+1} \\ &= h_{d-i,r}^{(1)} \\ &= h_{d-i-1,r}, \end{aligned}$$

completing the proof. ■

By (47) and (48), $h_r^*$ defined in (26) can be written as

$$h_r^* = \Pr\{\text{rk}(G'_{r+1} H) = r, \text{rk}(G_{r+1} H) = r\}. \quad (55)$$

*Lemma 5:*

$$\sum_{k=r}^{M} h_k^* = h_{r,r}.$$

*Proof:* By (51) and (55),

$$h_r^* = \sum_i \Pr\{\text{rk}(G'_{r+1} H) = r, \text{rk}(G_{r+1} H) = r | \, \text{rk}(H) = i\} p_{\text{rk}(H)}(i) = \sum_{i=r}^{M} \frac{\zeta_r^i}{q^{i-r}} h_i. \quad (56)$$

Therefore,

$$\begin{aligned} \sum_{k=r}^{M} h_k^* &= \sum_{k=r}^{M} \sum_{i=k}^{M} \frac{\zeta_k^i}{q^{(i-k)}} h_i \\ &= \sum_{i=r}^{M} h_i \sum_{k=r}^{i} \frac{\zeta_k^i}{q^{(i-k)}} \\ &= \sum_{i=r}^{M} h_i \zeta_r^i = h_{r,r}. \end{aligned}$$

■

*Lemma 6:* $\sum_r r h_r^* \le \sum_r r h_r$.

*Proof:* By Lemma 5, $\sum_{k=r}^{M} h_k^* = h_{r,r} = \sum_{i=r}^{M} h_i \zeta_r^i \le \sum_{k=r}^{M} h_k$, where the last inequality follows from $\zeta_r^i < 1$. Hence,

$$\begin{aligned}\sum_r r h_r^* &= \sum_{r=1}^{M}\sum_{k=r}^{M} h_k^* \\ &\le \sum_{r=1}^{M}\sum_{k=r}^{M} h_k \\ &= \sum_r r h_r.\end{aligned}$$

■

### *B. Incomplete Beta Function*

Beta function with integer parameters is used extensively in this work. Related results are summarized here. For positive integer $a$ and $b$, the *beta function* is defined by

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{(a-1)!(b-1)!}{(a+b-1)!}.$$

The *(regularized) incomplete beta function* is defined as

$$\begin{aligned}\mathrm{I}_{a,b}(x) &= \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{B(a,b)} \\ &= \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} x^j (1-x)^{a+b-1-j}.\end{aligned} \tag{57}$$

For more general discussion of beta functions, as well as incomplete beta functions, please refer to [40].

Using the above definitions, we can easily show that

$$\int_0^1 \mathrm{I}_{a,b}(x)dx = \frac{b}{a+b}, \tag{58}$$

and

$$\mathrm{I}_{a+1,b}(x) = \mathrm{I}_{a,b}(x) - \frac{x^a(1-x)^b}{aB(a,b)}. \tag{59}$$

*Lemma 7:* $\frac{\mathrm{I}_{a+1,b}(x)}{\mathrm{I}_{a,b}(x)}$ is monotonically increasing in $x$.

*Proof:* By (59),

$$\begin{aligned}\frac{\mathrm{I}_{a+1,b}(x)}{\mathrm{I}_{a,b}(x)} &= 1 - \frac{x^a(1-x)^b}{aB(a,b)\,\mathrm{I}_{a,b}(x)} \\ &= 1 - \frac{1}{aB(a,b)\sum_{j=a}^{a+b-1}\binom{a+b-1}{j}x^{j-a}(1-x)^{a-1-j}} \\ &= 1 - \frac{1}{aB(a,b)\sum_{j=0}^{b-1}\binom{a+b-1}{j+a}x^{j}(1-x)^{-1-j}},\end{aligned}$$

in which $x^j(1-x)^{-1-j}$ is monotonically increasing. ■

*Lemma 8:* When $\frac{b-1}{a+1} \leq \frac{\eta}{1-\eta}$ where $0 < \eta < 1$, $\frac{\mathrm{I}_{a+1,b}(x)}{\mathrm{I}_{a,b}(x)} \leq 1 - \frac{\eta}{b}$ for $0 < x \leq 1-\eta$ with equality when $b = 1$ and $x = 1 - \eta$.

*Proof:* Since $\frac{\mathrm{I}_{a+1,b}(x)}{\mathrm{I}_{a,b}(x)}$ is monotonically increasing in $x$ (cf. Lemma 7), it is sufficient to show $\frac{\mathrm{I}_{a+1,b}(1-\eta)}{\mathrm{I}_{a,b}(1-\eta)} \leq 1 - \frac{\eta}{b}$. Since $a + 1 \geq (b-1)\frac{1-\eta}{\eta}$,

$$\begin{aligned}\mathrm{I}_{a,b}(1-\eta) &= \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} (1-\eta)^j \eta^{a+b-1-j} \\ &\leq b\binom{a+b-1}{a}(1-\eta)^a \eta^{b-1},\end{aligned}$$

where the equality holds for $b = 1$. Thus,

$$\begin{aligned}\frac{\mathrm{I}_{a+1,b}(1-\eta)}{\mathrm{I}_{a,b}(1-\eta)} &= 1 - \frac{(1-\eta)^a \eta^b}{aB(a,b)\,\mathrm{I}_{a,b}(1-\eta)} \\ &\leq 1 - \frac{(1-\eta)^a \eta^b}{abB(a,b)\binom{a+b-1}{a}(1-\eta)^a \eta^{b-1}} \\ &= 1 - \frac{\eta}{b}.\end{aligned}$$

■

We will use the following result about the summation of binomial coefficients:

$$\sum_{j=0}^{n} (-1)^{j-n} \binom{j+m}{n} \binom{n}{j} = 1, \quad m \geq n. \tag{60}$$

The above equality can be verified as follows

$$\begin{aligned}\sum_{j=0}^{n} (-1)^{j-n} \binom{j+m}{n} \binom{n}{j} &= \sum_{j=0}^{n} (-1)^{j-n} \binom{j+m}{j+m-n} \binom{n}{j} \\ &= \sum_{j=0}^{n} (-1)^{j-n}(-1)^{j+m-n} \binom{-j-m+j+m-n-1}{j+m-n} \binom{n}{j} \qquad (61) \\ &= \sum_{j=0}^{n} (-1)^{m} \binom{-n-1}{j+m-n} \binom{n}{n-j} \\ &= (-1)^m \binom{-1}{m} \qquad (62) \\ &= 1, \qquad (63)\end{aligned}$$

where (62) follows from Vandermonde's identity; (61) and (63) use the relation between binomial coefficients with negative integers and positive integers.

*Lemma 9:* For $r \geq 1$,

$$\sum_{d=r+1}^{\infty} \frac{1}{d-1} \mathrm{I}_{d-r,r}(x) = -\ln(1-x), \quad x \in [0,\ 1).$$

*Proof:* As a special case, when $r = 1$, the equality becomes

$$\sum_{d=2}^{\infty} \frac{x^{d-1}}{d-1} = -\ln(1-x), \tag{64}$$

which is the Taylor expansion of $-\ln(1-x)$ for $x \in [0, 1)$.

To prove the general case, let us first derive an alternative form of $\mathrm{I}_{d-r,r}(x)$. For $a>0$,

$$\begin{aligned}
\mathrm{I}_{a,b}(x) &= \sum_{j=a}^{a+b-1}\binom{a+b-1}{j}x^j\sum_{i=0}^{a+b-1-j}(-1)^i\binom{a+b-1-j}{i}x^i\\
&= \sum_{m=a}^{a+b-1}x^m\sum_{j=a}^{m}\binom{a+b-1}{j}(-1)^{m-j}\binom{a+b-1-j}{m-j}\\
&= \sum_{m=a}^{a+b-1}(-x)^m\binom{a+b-1}{m}\sum_{j=a}^{m}\binom{m}{j}(-1)^j\\
&= \sum_{m=a}^{a+b-1}(-x)^m\binom{a+b-1}{m}\binom{m-1}{a-1}(-1)^a\\
&= b\binom{a+b-1}{b}(-1)^a\sum_{m=a}^{a+b-1}\frac{(-x)^m}{m}\binom{b-1}{m-a}.
\end{aligned}$$

Using this form for $\mathrm{I}_{d-r,r}(x)$, we have

$$\begin{aligned}
\sum_{d=r+1}^{\infty}\frac{1}{d-1}\mathrm{I}_{d-r,r}(x) &= \sum_{d=r+1}^{\infty}\frac{r}{d-1}\binom{d-1}{r}\sum_{m=d-r}^{d-1}\binom{r-1}{m-d+r}(-1)^{m-d+r}\frac{x^m}{m}\\
&= \sum_{m=1}^{\infty}\frac{x^m}{m}\sum_{d=\max\{m,r\}+1}^{m+r}\frac{r}{d-1}\binom{d-1}{r}\binom{r-1}{m-d+r}(-1)^{m-d+r}\\
&= \sum_{m=1}^{\infty}\frac{x^m}{m}A_m, \qquad (65)
\end{aligned}$$

where

$$A_m \triangleq \sum_{d=\max\{m,r\}+1}^{m+r}\frac{r}{d-1}\binom{d-1}{r}\binom{r-1}{m-d+r}(-1)^{m-d+r}.$$

For $m\le r$,

$$\begin{aligned}
A_m &= \sum_{d=r+1}^{m+r}\frac{r}{d-1}\binom{d-1}{r}\binom{r-1}{m-d+r}(-1)^{m-d+r}\\
&= \sum_{d=r+1}^{m+r}\binom{d-2}{r-1}\binom{r-1}{m-d+r}(-1)^{m-d+r}\\
&= \sum_{j=0}^{m-1}\binom{j+r-1}{r-1}\binom{r-1}{m-j-1}(-1)^{m-j-1}\\
&= \sum_{j=0}^{m-1}\binom{j+r-1}{m-1}\binom{m-1}{m-j-1}(-1)^{m-j-1}\\
&= 1,
\end{aligned}$$

where the last equality follows from (60). Similarly, for $m>r$,

$$\begin{aligned}
A_m &= \sum_{d=m+1}^{m+r}\frac{r}{d-1}\binom{d-1}{r}\binom{r-1}{m-d+r}(-1)^{m-d+r}\\
&= \sum_{d=m+1}^{m+r}\binom{d-2}{r-1}\binom{r-1}{m-d+r}(-1)^{m-d+r}
\end{aligned}$$

$$= \sum_{j=0}^{r-1} \binom{j+m-1}{r-1} \binom{r-1}{r-j-1} (-1)^{r-j-1}$$

$$= 1.$$

The proof is completed by referring to (64) and (65) with $A_m = 1$. ■

## APPENDIX III
## LAYERED DECODING GRAPH

We have discussed different decoding strategies under the rule that a check node is decodable if and only if its rank equals its degree. We say a variable node is decodable if it is connected to a decodable check node. In Section II-C, a decodable check node is chosen and all its neighbors (variable nodes) are recovered simultaneously, while in Section III-A, a decodable variable node is uniformly chosen to be recovered. Here we show that under the decoding rule that a check node is decodable if and only if its rank equals to its degree, both strategies stop with the same subset of the variable nodes undecoded.

For a given decoding graph $\mathcal{G}$, let $\mathcal{G}^0 = \mathcal{G}$. Label by $L_1$ all the decodable check nodes in $\mathcal{G}^0$ and label by $L_2$ all the variable nodes in $\mathcal{G}^0$ connected to the check nodes with label $L_1$. We repeat the above procedure as follows. For $i = 1, 2, \ldots$, let $\mathcal{G}^i$ be the subgraph of $\mathcal{G}$ obtained by removing all the nodes with labels $L_j$ for $j \leq 2i$, as well as the related edges. (The generator matrices of the check nodes are also updated.) Label by $L_{2i+1}$ all the decodable check nodes in $\mathcal{G}^i$ and label by $L_{2i+2}$ all the variable nodes in $\mathcal{G}^i$ connected to the check nodes with label $L_{2i+1}$. This procedure stops when $\mathcal{G}^i$ has no more decodable check nodes. Let $i_0$ be the index where the procedure stops. The above labelling procedure is deterministic and generates unique labels for each decodable variable nodes and check nodes.

With the labels, we can generate a layered subgraph $\mathcal{G}'$ of $\mathcal{G}$. In $\mathcal{G}'$, layer $j$, $j = 1, 2, \ldots, 2i_0$, contains all the check/variable nodes with label $L_j$. Only the edges connecting two nodes belonging two consecutive layers are preserved in $\mathcal{G}'$. By the assigning rule of the labels, it is clear that a variable node on layer $2i$ must connect to one check node on layer $2i-1$, $i = 1, \ldots, i_0$, since otherwise, the variable node is not decodable. Further, a check node on layer $2i+1$ must connect to some variable nodes on layer $2i$, $i = 1, \ldots, i_0 - 1$, since otherwise, the check node should be on layer $2i-1$.

By the definition of decodability, a decoding strategy must process the variable/check nodes in $\mathcal{G}'$ following an order such that a variable/check node is processed after all its lower layer descendant variable/check nodes have been processed. The two random decoding strategies we have discussed in Section II-C and Section III-A both can process all the nodes in $\mathcal{G}'$ before stopping.

## APPENDIX IV
## CONCENTRATION

Theorem 1 is proved by applying a general theorem by Wormald [29], [41].

*A. Wormald's Differential Equation*

The statement of the next theorem follows that of [29, Theorem 5.1] with an extra initial condition. A similar version is provided in [31, Theorem C.28] with the boundedness condition holding deterministically.

We say a function $f(u_1,\dots,u_j)$ satisfies a *Lipschitz condition* on $\mathcal{D}\subset\mathbb{R}^j$ if there exists a constant $C_L$ such that

$$|f(u_1,\cdots,u_j)-f(v_1,\cdots,v_j)|\le C_L\max_{1\le i\le j}|u_i-v_i|$$

for all $(u_1,\cdots,u_j)$ and $(v_1,\cdots,v_j)$ in $\mathcal{D}$. We call $C_L$ the Lipschitz constant for $f$. Note that $\max_{1\le i\le j}|u_i-v_i|$ is the distance between $(u_1,\cdots,u_j)$ and $(v_1,\cdots,v_j)$ in the $l^\infty$-norm.

*Theorem 5:* Let $\mathcal{G}_0,\mathcal{G}_1,\dots$ be a random process with a positive integer parameter $n$, and let $(Y_l(t))_{l=0}^L$ be a random vector determined by $\mathcal{G}_0,\dots,\mathcal{G}_t$. For some constant $C_0$ and all $l$, $|Y_l(t)|<C_0n$ for $t\ge0$ and all $n$. Let $\mathcal{D}$ be some bounded connected open set containing the closure of

$$\{(0,z_1,\dots,z_L):\exists n,\Pr\{Y_l(0)=z_ln,1\le l\le L\}\ne0\}.$$

Define the *stopping time* $T_{\mathcal{D}}$ to be the minimum $t$ such that $(t/n,Y_1(t)/n,\dots,Y_L(t)/n)\notin\mathcal{D}$. Assume the following conditions hold.

(i) (Boundedness) For some functions $\beta=\beta(n)\ge1$ and $\gamma=\gamma(n)$, the probability that

$$\max_l|Y_l(t+1)-Y_l(t)|\le\beta,$$

is at least $1-\gamma$ for $t<T_{\mathcal{D}}$.

(ii) (Trend) For some function $\lambda_1=\lambda_1(n)=o(1)$, if $t<T_{\mathcal{D}}$,

$$\mathbb{E}[Y_l(t+1)-Y_l(t)|\mathcal{G}_1,\dots,\mathcal{G}_t]=f_l\left(\frac{t}{n},\left(\frac{Y_i(t)}{n}\right)_{i=0}^L\right)+\mathcal{O}(\lambda_1),$$

for $1\le l\le L$.

(iii) (Lipschitz) Each function $f_l$ satisfies a Lipschitz condition on $\mathcal{D}\cap\{(t,z_1,\dots,z_L),t\ge0\}$ with the same Lipschitz constant $C_L$ for each $l$.

(iv) (Initial condition) For some point $(0,z_1^0,\dots,z_l^0)\in\mathcal{D}$,

$$|Y_l(0)/n-z_l^0|\le\sigma=o(1),0\le l\le L.$$

Then the following are true.

(a) For $(0,(\hat z_l)_{l=1}^L)\in\mathcal{D}$, the system of differential equations

$$\frac{dz_l(\tau)}{d\tau}=f_l(\tau,(z_{l'}(\tau))_{l'=1}^L),\quad l=1,\dots,L,$$

has a unique solution in $\mathcal{D}$ for $z_l:\mathbb{R}\to\mathbb{R}$ passing through $z_l(0)=\hat z_l$, $l=1,\dots,L$, and this solution extends to points arbitrarily close to the boundary of $\mathcal{D}$.

(b) Let $\lambda>\max\{\sigma,\lambda_1+C_0n\gamma\}$ with $\lambda=o(1)$. There exists a sufficiently large constant $C_1$ such that when $n$ is sufficiently large, with probability $1-\mathcal{O}(n\gamma+\frac{\beta}{\lambda}\exp(-\frac{n\lambda^3}{\beta^3}))$,

$$|Y_l(t)-nz_l(t/n)|=\mathcal{O}(\lambda n)\tag{66}$$

uniformly for $0 \le t \le \bar{\tau} n$ and for each $l$, where $\hat{z}_l = z_l^0$, and $\bar{\tau} = \bar{\tau}(n)$ is the supremum of those $\tau$ to which the solution of the system of differential equations in (a) can be extended before reaching within $l^\infty$-distance $C_1\lambda$ of the boundary of $\mathcal{D}$.

*Proof:* The proof follows exactly the proof of [29, Theorem 5.1] except for the place where we should handle the initial condition (iv). We only need to modify the definition of $B_j$ (below (5.9) in [29]) in the original proof to

$$B_j = (n\lambda + \omega)\left(\left(1 + \frac{B\omega}{n}\right)^j - 1\right) + B_0\left(1 + \frac{B\omega}{n}\right)^j,$$

where $B_0 = n\lambda$. The induction in the original proof now begins by the fact that $|z_l(0) - Y_l(0)/n| \le \sigma < \mathcal{O}(\lambda)$. The other part of the proof stays the same as that of [29, Theorem 5.1]. ■

### *B. Proof of Theorem 1*

We first prove two technical lemmas. For BATS$(K, n, \Psi, h)$, the degrees of the variable nodes are not independent but follow the same distribution. The following lemma shows that the degree of a variable node is not likely to be much larger than its expectation.

*Lemma 10:* Let $V$ be the degree of a variable node of BATS$(K, n, \Psi, h)$. For any $\alpha > 0$,

$$\Pr\{V > (1+\alpha)\,\mathbb{E}[\Psi]/\theta\} < \left(\frac{e^\alpha}{(1+\alpha)^{(1+\alpha)}}\right)^{\mathbb{E}[\Psi]/\theta},$$

where $\theta = K/n$.

*Proof:* Fix a variable node. Let $X_i$ be the indicator random variable of the $i$th check node being the neighbor of the specific variable node. Then $V = \sum_i X_i$. We have $\mathbb{E}[V] = \sum_i \mathbb{E}[X_i] = \sum_i \sum_d \frac{d}{K}\Psi_d = \frac{n}{K}\,\mathbb{E}[\Psi] = \frac{\mathbb{E}[\Psi]}{\theta}$. Since $X_i$, $i = 1, \ldots, n$, are mutually independent, the lemma is proved by applying the Chernoff bound. ■

The following lemma verifies the boundedness condition of Theorem 5.

*Lemma 11:* When $\beta/D > \mathbb{E}[\Psi]/\theta$, the probability that

$$\max_{\iota \in \mathcal{F} \cup \{0\}} |R_\iota(t+1) - R_\iota(t)| \le \beta,$$

is at least

$$1 - \theta n \exp\left(-\frac{\beta}{D}(\ln(\beta/D) - \ln(\mathbb{E}[\Psi]/\theta) - 1) - \frac{\mathbb{E}[\Psi]}{\theta}\right).$$

*Proof:* Let $V$ be the degree of the variable node to be removed at the beginning of time $t+1$. By (9), we have for $(d, r) \in \mathcal{F}$,

$$|R_{d,r}(t+1) - R_{d,r}(t)| \le DV,$$

and by (19), we have

$$|R_0(t+1) - R_0(t)| \le DV.$$

Hence when $\beta/D > \mathbb{E}[\Psi]/\theta$,

$$\begin{aligned}
&\Pr\left\{\max_{\iota\in\mathcal{F}\cup\{0\}} |R_\iota(t+1) - R_\iota(t)| \le \beta\right\} \\
&\ge \Pr\{VD \le \beta\} \\
&\ge \Pr\{\text{the degrees of all variable nodes at time zero} \le \beta/D\} \\
&> 1 - \theta n \exp\left(-\frac{\beta}{D}(\ln(\beta/D) - \ln(\mathbb{E}[\Psi]/\theta) - 1) - \frac{\mathbb{E}[\Psi]}{\theta}\right),
\end{aligned}$$

where the last inequality follows from Lemma 10 and the union bound. ■

*Proof of Theorem 1:* We consider in the proof only the instances of BATS$(K, n, \Psi, h)$ satisfying

$$\left|\frac{R_{d,r}}{n} - \rho_{d,r}\right| = \mathcal{O}(n^{-1/6}), \quad (d,r) \in \bar{\mathcal{F}}. \tag{67}$$

By Lemma 1 this will decrease the probability bounds we will obtained by at most $\gamma(n) + 2MD\exp(-2n^{2/3})$.

Define the stopping time $T_0$ as the first time $t$ such that $R_0(t) = 0$. By defining proper functions $f_{d,r}$, $f_0$ we can rewrite (7) and (8) as

$$\begin{aligned}
&\mathbb{E}[R_{d,r}(t+1) - R_{d,r}(t)|\bar{R}(t)] \\
&= f_{d,r}\left(\frac{t}{n}, \left(\frac{R_0(t)}{n}\right), \left(\frac{R_{d',r'}(t)}{n}\right)_{(d',r')\in\mathcal{F}}\right), \ (d,r) \in \mathcal{F} \\
&\mathbb{E}[R_0(t+1) - R_0(t)|\bar{R}(t)] \\
&= f_0\left(\frac{t}{n}, \left(\frac{R_0(t)}{n}\right), \left(\frac{R_{d',r'}(t)}{n}\right)_{(d',r')\in\mathcal{F}}\right) + \mathcal{O}\left(\frac{1}{n}\right),
\end{aligned}$$

for $t < T_0$. For $\iota \in \mathcal{F}\cup\{0\}$, define random variable $\hat{R}_\iota$ as $\hat{R}_\iota(0) = R_\iota(0)$ and for $t \ge 0$,

$$\hat{R}_\iota(t+1) = \begin{cases} R_\iota(t+1) & t < T_0 \\ \hat{R}_\iota(t) + f_\iota\left(\frac{t}{n}, \left(\frac{R_0(t)}{n}\right), \left(\frac{R_{d',r'}(t)}{n}\right)_{(d',r')\in\mathcal{F}}\right) & t \ge T_0. \end{cases}$$

Note that $T_0$ is also the first time that $\hat{R}_0(t)$ becomes zeros.

We apply Theorem 5 on $(\hat{R}_0(t), (\hat{R}_{d,r}(t))_{(d,r)\in\mathcal{F}})$ in place of $(Y_l(t))_{l=1}^L$. The region $\mathcal{D}$ is defined as

$$\mathcal{D} = (-\eta, (1-\eta/2)\theta) \times (-M, M+\eta) \times (-\eta, d)^{|\mathcal{F}|}.$$

So 1) $t/n$ is in the interval $(-\eta, (1-\eta/2)\theta)$; 2) $\hat{R}_0(t)/n$ is in the interval $(-M, M+\eta)$; and 3) $\hat{R}_{d,r}(t)/n$, $(d,r) \in \mathcal{F}$, is in the interval $(-\eta, d)$. As required, $\mathcal{D}$ is a bounded connected open set and containing all the possible initial state $(0, \hat{R}_0(0)/n, (\hat{R}_{d,r}(0)/n)_{(d,r)\in\mathcal{F}})$.

The conditions of Theorem 5 are ready to be verified. When $t \ge T_0$, the change $|\hat{R}_\iota(t+1) - \hat{R}_\iota(t)|$ for $\iota \in \mathcal{F}\cup\{0\}$ is deterministic and upper bounded. When $t < T_0$, by Lemma 11 with $\beta = n^{1/8}$, the boundedness condition (i) holds with

$$\gamma = n\exp\left(-n^{1/8}\left(c_{1,3}\ln n - c_{1,1}\right) - c_{1,2}\right),$$

where $c_{1,1}$, $c_{1,2}$, and $c_{1,3}$ are only related to $\mathbb{E}[\Psi]$ and $\theta$. The trend condition (ii) is satisfied with $\lambda_1 = \mathcal{O}(1/n)$. By definition, it can be verified that $f_\iota$, $\iota \in \mathcal{F} \cup \{0\}$ satisfy the Lipschitz condition (iii). The initial condition (iv) holds with $\sigma = \mathcal{O}(n^{-1/6})$.

Wormald's method leads us to consider the system of differential equations

$$\frac{d\rho_{d,r}(\tau)}{d\tau} = f_{d,r}(\tau, \rho_0(\tau), (\rho_{d',r'}(\tau))_{(d',r')\in\mathcal{F}}), \quad (d,r) \in \mathcal{F}$$
$$\frac{d\rho_0(\tau)}{d\tau} = f_0(\tau, \rho_0(\tau), (\rho_{d',r'}(\tau))_{(d',r')\in\mathcal{F}})$$

with the initial condition $\rho_{d,r}(0) = \rho_{d,r}$, $(d,r) \in \mathcal{F}$, and $\rho_0(0) = \sum_r \rho_{r,r}$. The conclusion (a) of Theorem 5 shows the existence and uniqueness of the solution of the above system of differential equations. We solve the system of differential equations explicitly in Appendix V.

Let $\lambda = \mathcal{O}(n^{-1/6})$. By the conclusion (b) of Theorem 5, we know that for a sufficiently large constant $C_1$, with probability $1 - \mathcal{O}(n\gamma + \frac{\beta}{\lambda}\exp(-\frac{n\lambda^3}{\beta^3}))$,

$$|\hat{R}_{d,r}(t) - n\rho_{d,r}(t/n)| = \mathcal{O}(n^{5/6}), \ (d,r) \in \mathcal{F},$$
$$|\hat{R}_0(t) - n\rho_0(t/n)| = \mathcal{O}(n^{5/6})$$

uniformly for $0 \le t \le \bar{\tau}n$, where $\bar{\tau}$ is defined in Theorem 5. Increase $n$ if necessary so that $\frac{\beta}{\lambda}\exp(-\frac{n\lambda^3}{\beta^3}) = n^{7/24}\exp(-n^{-1/8}) > n\gamma$ and $C_1\lambda < \frac{\eta}{2}\theta$, which implies $\bar{\tau} \ge (1-\eta)\theta$. So there exists constants $c_0$ and $c_0'$ such that the event

$$E_0 = \{|\hat{R}_0(t)/n - \rho_0(t/n)| \le c_0 n^{-1/6}, \ 0 \le t \le (1-\eta)K\}$$

holds with probability at least $1 - c_0' n^{7/24}\exp(-n^{-1/8})$.

Now we consider the two cases in the theorem to prove. (i) If $\rho_0(\tau) > 0$ for $\tau \in [0, (1-\eta)\theta]$, then there exists $\epsilon > 0$ such that $\rho_0(\tau) \ge \epsilon$ for $\tau \in [0, (1-\eta)\theta]$. Increase $n$ if necessary so that $c_0 n^{-1/6} < \epsilon$. Then, we have

$$\begin{aligned}\Pr\{T_0 > (1-\eta)K\} &= \Pr\{\hat{R}(t) > 0, 0 \le t \le (1-\eta)K\} \\ &\ge \Pr\{E_0\} \qquad (68) \\ &\ge 1 - c_0' n^{7/24}\exp(-n^{-1/8}),\end{aligned}$$

where (68) follows that under the condition $E_0$, for all $t \in [0, (1-\eta)K]$, $\hat{R}_0(t)/n \ge \rho_0(t/n) - c_0 n^{-1/6} > 0$. Since $\hat{R}_\iota = R_\iota$, $\iota \in \mathcal{F} \cup \{0\}$, when $t < T_0$, the first part of the theorem is proved.

(ii) Consider $\rho_0(\tau_0) < 0$ for $\tau_0 \in [0, (1-\eta)\theta]$. There exists $\epsilon > 0$ such that $\rho_0(\tau) \le -\epsilon$ for all $\tau \in [\tau_0 - \epsilon, \tau_0 + \epsilon] \cap [0, (1-\eta)\theta]$. Increase $n$ if necessary so that $c_0 n^{-1/6} < \epsilon$ and $n\epsilon > 1$. Then, we have

$$\begin{aligned}&\Pr\{T_0 \le (1-\eta)K\} \\ &\quad = \Pr\{\hat{R}_0(t) < 0, \text{ for some } t \in [0, (1-\eta)K]\} \\ &\quad \ge \Pr\{E_0\} \qquad (69) \\ &\quad \ge 1 - c_0' n^{7/24}\exp(-n^{-1/8}),\end{aligned}$$

where (69) can be shown as follows. Since $n\epsilon > 1$, there exists $t_0$ such that $t_0/n \in [\tau_0 - \epsilon, \tau_0 + \epsilon] \cap [0, (1-\eta)\theta]$. Hence, under the condition $E_0$, $\hat{R}_0(t_0)/n \le c_0 n^{-1/6} + \rho_0(t_0/n) < 0$.

The proof of the theorem is completed by subtracting the probability that (67) does not hold. ■

## APPENDIX V
## SOLVE THE SYSTEM OF DIFFERENTIAL EQUATIONS

We solve the following system of differential equations given in (20) and (21), which is repeated as follows:

$$\begin{aligned}\frac{d\rho_{d,r}(\tau)}{d\tau} &= (\alpha_{d+1,r}\rho_{d+1,r}(\tau) + \bar{\alpha}_{d+1,r+1}\rho_{d+1,r+1}(\tau) - \rho_{d,r}(\tau))\times \\ &\qquad \times \frac{d}{\theta - \tau}, \quad 1 \le r \le M,\ r < d \le D \\ \frac{d\rho_0(\tau)}{d\tau} &= \frac{\sum_{r=1}^{D-1} r\alpha_{r+1,r}\rho_{r+1,r}(\tau) - \rho_0(\tau)}{\theta - \tau} - 1\end{aligned}$$

with $\rho_{d,r}(0) = \hat{\rho}_{d,r}$ and $\rho_0(0) = \sum_{r=1}^{D} \hat{\rho}_{1,1}$.

Let $y_{d,r}(\tau) = (1 - \tau/\theta)^{-d}\rho_{d,r}(\tau)$. We have

$$\frac{dy_{d,r}(\tau)}{d\tau} = \frac{d}{\theta}(\alpha_{d+1,r}y_{d+1,r}(\tau) + \bar{\alpha}_{d+1,r+1}y_{d+1,r+1}(\tau)).$$

We see that $y_{d,r}(0) = \hat{\rho}_{d,r}$. Define

$$\begin{aligned}\hat{\rho}_{d,r}^{(0)} &= \hat{\rho}_{d,r} \\ \hat{\rho}_{d,r}^{(i+1)} &= \alpha_{d-i,r}\hat{\rho}_{d,r}^{(i)} + \bar{\alpha}_{d-i,r+1}\hat{\rho}_{d,r+1}^{(i)}.\end{aligned}$$

We can verify that

$$y_{d,r}(\tau) = \sum_{j=d}^{D} \binom{j-1}{d-1} (\tau/\theta)^{j-d} \hat{\rho}_{j,r}^{(j-d)}.$$

Thus

$$\rho_{d,r}(\tau) = (1 - \tau/\theta)^d \sum_{j=d}^{D} \binom{j-1}{d-1} (\tau/\theta)^{j-d} \hat{\rho}_{j,r}^{(j-d)}. \tag{70}$$

Using the general solution of linear differential equations, we obtain that

$$\begin{aligned}\rho_0(\tau) &= (1 - \tau/\theta)\left( \int_0^\tau \frac{\sum_{r=1}^{M} r\alpha_{r+1,r}\rho_{r+1,r}(t)}{\theta - t}(1 - t/\theta)^{-1}dt + \theta \ln(1 - \tau/\theta) + \sum_{r\ge 1} \hat{\rho}_{r,r} \right) \\ &= (1 - \tau/\theta)\left( \sum_{r=1}^{M} r\alpha_{r+1,r} \int_0^\tau \frac{\rho_{r+1,r}(t)}{\theta - t}(1 - t/\theta)^{-1}dt + \theta \ln(1 - \tau/\theta) + \sum_{r\ge 1} \hat{\rho}_{r,r} \right).\end{aligned} \tag{71}$$

The integral in (71) can be further calculated as follows:

$$\begin{aligned}
&\int_0^{\tau} \frac{\rho_{r+1,r}(t)}{\theta - t}(1-t/\theta)^{-1}dt \\
&= \int_0^{\tau} \frac{\sum_{j=r+1}^{D} \hat{\rho}_{j,r}^{(j-r-1)} \binom{j-1}{r} (1-t/\theta)^{r+1}(t/\theta)^{j-r-1}}{(\theta-t)(1-t/\theta)} dt \\
&= \int_0^{\tau} \sum_{j=r+1}^{D} \hat{\rho}_{j,r}^{(j-r-1)} \binom{j-1}{r} (1-t/\theta)^{r-1}(t/\theta)^{j-r-1} \frac{dt}{\theta} \\
&= \sum_{j=r+1}^{D} \hat{\rho}_{j,r}^{(j-r-1)} \binom{j-1}{r} \int_0^{\tau/\theta} (1-t)^{r-1} t^{j-r-1} dt \\
&= \sum_{j=r+1}^{D} \hat{\rho}_{j,r}^{(j-r-1)} \binom{j-1}{r} \frac{(j-r-1)!(r-1)!}{(j-1)!} \mathrm{I}_{j-r,r}(\tau/\theta) \qquad (72) \\
&= 1/r \sum_{j=r+1}^{D} \hat{\rho}_{j,r}^{(j-r-1)} \mathrm{I}_{j-r,r}(\tau/\theta),
\end{aligned}$$

where (72) is obtained by substituting $\rho_{r+1,r}(t)$ in (70), and (72) is obtained by the definition of incomplete beta function (cf. (57)).

## REFERENCES


[1] M. Luby, "LT codes," in *Proc. 43rd Ann. IEEE Symp. on Foundations of Computer Science*, Nov. 2002, pp. 271–282.

[2] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.

[3] P. Maymounkov, "Online codes," NYU, Tech. Rep., Nov. 2002.

[4] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[5] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[6] T. Ho, B. Leong, M. Medard, R. Koetter, Y. Chang, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. IEEE ISIT '03*, Jun. 2003.

[7] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Allerton Conf. Comm., Control, and Computing*, Oct. 2003.

[8] S. Jaggi, P. A. Chou, and K. Jain, "Low complexity optimal algebraic multicast codes," in *Proc. IEEE ISIT'03*, Jun. 2003.

[9] P. Sanders, S. Egner, and L. Tolhuizen, "Polynomial time algorithms for network information flow," in *Proceedings of the fifteenth annual ACM symposium on Parallel algorithms and architectures*, ser. SPAA '03. New York, NY, USA: ACM, 2003, pp. 286–294.

[10] D. S. Lun, M. Médard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," *Physical Communication*, vol. 1, no. 1, pp. 3–20, 2008.

[11] Y. Wu, "A trellis connectivity analysis of random linear network coding with buffering," in *Proc. IEEE ISIT '06*, Seattle, USA, Jul. 2006.

[12] C. Gkantsidis and P. Rodriguez, "Network coding for large scale content distribution," in *Proc. IEEE Infocom'05*, 2005.

[13] M.Wang and B.Li, "Lava: A reality check of network coding in peer-to-peer live streaming," in *Proc. INFOCOM*, 2007.

[14] R. W. Yeung, "Avalanche: a network coding analysis," *Comm. Info. and Syst.*, vol. 7, 2007.

[15] P. Pakzad, C. Fragouli, and A. Shokrollahi, "Coding schemes for line networks," in *Proc. IEEE ISIT '05*, 2005, pp. 1853–1857.

[16] R. Gummadi and R. Sreenivas, "Relaying a fountain code across multiple nodes," in *Proc. IEEE ITW '08*, May 2008, pp. 149 –153.

[17] M.-L. Champel, K. Huguenin, A.-M. Kermarrec, and N. Le Scouarnec, "LT network codes," in *Proc. IEEE ICDCS '10*, Jun. 2010.

[18] N. Thomos and P. Frossard, "Degree distribution optimization in Raptor network coding," in *Proc. IEEE ISIT '11*, Aug. 2011.

[19] P. Maymounkov, N. J. A. Harvey, and D. S. Lun, "Methods for efficient network coding," in *Proc. Allerton Conf. Comm., Control, and Computing*, Sep. 2006.

[20] D. Silva, W. Zeng, and F. R. Kschischang, "Sparse network coding with overlapping classes," in *Proc. NetCod '09*, 2009, pp. 74–79.

[21] A. Heidarzadeh and A. H. Banihashemi, "Overlapped chunked network coding," in *Proc. ITW '10*, 2010, pp. 1–5.

[22] Y. Li, E. Soljanin, and P. Spasojevic, "Effects of the generation size and overlap on throughput and complexity in randomized linear network coding," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 1111–1123, Feb. 2011.

[23] S. Yang and R. W. Yeung, "Large file transmission in network-coded networks with packet loss – A performance perspective," in *Proc. ISABEL 2011*, Barcelona, Spain, 2011.

[24] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[25] R. W. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[26] D. Silva, F. R. Kschischang, and R. Koetter, "Communication over finite-field matrix channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1296–1305, Mar. 2010.

[27] M. Jafari, S. Mohajer, C. Fragouli, and S. Diggavi, "On the capacity of non-coherent network coding," in *Proc. IEEE ISIT'09*, Jul. 2009.

[28] S. Yang, J. Meng, and E.-h. Yang, "Coding for linear operator channels over finite fields," in *Proc. IEEE Inte. Symp. on Information Theory ISIT '10*, Austin, USA, Jun. 2010.

[29] N. C. Wormald, "The differential equation method for random graph processes and greedy algorithms," *Karonsky and Proemel, eds., Lectures on Approximation and Randomized Algorithms PWN, Warsaw*, pp. 73–155, 1999.

[30] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, 2001.

[31] T. Richardsan and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, Cambridge, England, 2008.

[32] S. Yang, S.-W. Ho, J. Meng, and E.-h. Yang, "Linear operator channels over finite fields," 2010, online, http://arxiv.org/abs/1002.2293v1.

[33] I. Csiszár and J. Körner, *Information theory*. Cambridge University Press, 2011.

[34] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[35] I. Csiszar and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inform. Theory*, vol. 34, no. 1, pp. 27–34, 1988.

[36] S. Yang, "Arbitrarily varying channels with convergent state constraints," 2012, under preparation.

[37] G. E. Andrews, *The theory of partitions*, ser. vol. 2, Encyclopedia of mathematics and its applications. Addison-Wesley Pub. Co., 1976.

[38] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3579–3591, Aug. 2008.

[39] M. Gadouleau and Z. Yan, "Packing and covering properties of rank metric codes," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 3873–3883, Sep. 2008.

[40] M. Zelen and N. C. Severo, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, Eds. New York: Dover, 1972.

[41] N. C. Wormald, "Differential equations for random processes and random graphs," *Ann. Appl. Probab.*, no. 5, pp. 1217–1235, 1995.