On Achievable Rates of Line Networks with Generalized Batched Network Coding

Jie Wang, Shenghao Yang, Yanyan Dong and Yiheng Zhang

Abstract

Towards better understanding of the wireless network design with a large number of hops, we study the achievable rate of a line network formed by general discrete memoryless channels (DMCs), which are not necessarily identical. To bound the communication latency and computation cost, we take the buffer size and inner blocklength constraints into consideration. We focus a class of codes called Generalized Batched Network Codes (GBNCs), which include most existing schemes as special cases and achieve the min-cut upper bounds when certain parameters go infinity. Using a new "bottleneck status" technique, we provide upper bounds on the achievable rates of GBNCs that match the achievable rates in the order of the network length. Using a "channel reduction" technique, we generalize the existing achievability results for a line network of identical DMCs, and obtain achievability results with the buffer size of the sub-linear order of the inner blocklength. Stronger results for various special DMCs are obtained. For line networks of canonical channels, certain upper bounds holds even with relaxed inner blocklength constraints. For line networks of packet erasure channels, we provide refined upper and lower bounds and showcase their proximity through numerical evaluations.

Index Terms

multi-hop network, line network, batched network code, capacity bound, buffer size, latency

I. Introduction

Discrete memoryless channels (DMCs) are basic channel models that capture the essential issues in many communication systems. Motivated by the recent development in communication

J. Wang is with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, USA (jwang3163@gatech.edu). S. Yang is with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China (shyang@cuhk.edu.cn). Y. Dong is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117597, Singapore (yan.dong@nus.edu.sg). Y. Zhang is with the Information Networking Institute, Carnegie Mellon University, Pittsburgh, USA (yihengzh@andrew.cmu.edu).

networks, we study multi-hop, line topology networks formed by the concatenation of DMCs. In a line network, the first node is called *source node*, the last node is called the *destination node*, and all the other nodes are called the *intermediate nodes*. When there is no limits on the storage and computation cost at the intermediate nodes, the network capacity is the *min-cut* from the source node to the destination node, achievable by implementing capacity achieving channel codes hop-by-hop [3]. However, when the number of hops becomes large, the hop-by-hop coding approach introduces significant communication latency, computational overhead, and storage expenses at the intermediate nodes. In this paper, we study this network with a large number of hops while considering constraints on the intermediate node buffer size and latency.

A. Related Work

Multi-hop wireless communication networks are emerging in wireless LAN [4], underwater acoustic networks [5], and deep space communication networks [6]. Cellular networks introduce integrated access and backhaul (IAB) which enables the interconnection of multiple base stations through wireless communications to form a multi-hop wireless network [7], [8]. With IAB, some base stations do not need to have wireline backhaul links. One trend of these multi-hop wireless networks is that the number of hops is expected to be large. Notably, a Low Earth Orbit (LEO) satellite network often necessitates a considerable number of wireless hops [9]. Additionally, due to the expansive communication cost in the ocean and the relatively limited range of sound communication, wireless underwater acoustic networks may greatly benefit from networks of hundreds of hops [10]. Furthermore, applications such as V2X autonomous driving highly demand wireless networks with a substantial number of hops [11].

In addition to rate, communication latency is a major concern when utilizing multi-hop wireless networks [12]. In the case of the hop-by-hop coding approach, as the number of hops increases, it becomes necessary to increase the code block length to minimize hop-by-hop error probability. Consequently, the end-to-end communication latency experiences a growth rate higher than linear with the number of hops. In general, we may question how the end-to-end latency and communication rate are affected by an increase in the number of hops.

In [13], Niesen, Fragouli, and Tuninetti studied the line network capacity with a fixed intermediate node processing blocklength N (called the *inner blocklength* in this paper) that can affect the delay, buffer size, and the computation cost. Suppose all the channels in the line network are identical, denoted as Q. When the zero-error capacity of Q is nonzero, using a constant N can achieve any constant rate below the zero-error capacity for any given number of hops L [13]. When the zero-error capacity of Q is zero, they showed that a class of codes with a constant N can achieve rates $\Omega(e^{-cL})$, where c is a constant. They also showed that if the inner blocklength N is of order $\ln L$, any rate below the capacity of Q can be achieved.

In contrast to the aforementioned achievability results, the min-cut is still the best upper bound. It is still an open question whether these diminishing rates with increasing network length are necessary. Furthermore, although not explicitly examined in [13], the processing latency and buffer size at an intermediate node are both $O(N)$. However, we are curious to explore the possibility of these factors being of a lower order than N. With these inquiries in mind, we embark on a more comprehensive investigation of line networks formed by DMCs.

This paper is a refined and extended version of our previous conference papers [1], [2]. In the following, we clarify the difference between the previous two papers and this paper.

1) Regarding the achievable rate of GBNC, our previous work [1] only discussed the repetition recoding using binary symmetric channel (BSC), and how to reduced a general channel to BSC of size 2×2. In this paper, we provide a detailed analysis on the achievable rates of more advanced recoding schemes (including decode-and-forward, repetition recoding, and channel reduction) in Section IV. We also analyze the repetition recoding using a general channel, which is not necessarily a BSC. For the channel reduction part, we outline how to reduce a general channel to an uniform channel (which includes BSC as a special case).
2) Regarding the upper bound of the achievable rate of GBNC, our previous work [2] only focused on the case where all channels have 0 zero-error capacity. For this scenario, we adpot a new techinque that obtains a tighter upper bound (Theorem 4). We also generalize this result to channels that may also have positive zero-error capacity (Section VI).
3) Our previous work [1] briefly discussed the achievable rate of GBNC for packet erasure channels. In this paper, we perform a detailed theoretical and numerical analysis on the lower bound and upper bound of the achievable rate for this type of channels (Section V). For theoretical part, we give the order of the achievable rate for a broader choices of channel paramerers. For numerical simulation part, we illustrate the upper bound and achievable rates of GBNC in Fig. 6. We also illustrate those results using BSC in Fig. 2, 4, 5, respectively.

Remark 1. We do not take feedback messages into consideration in this paper. In certain applications of multi-hop networks with a significant number of hops, the feedback may experience long delays and be unreliable, as seen in underwater acoustic networks and deep space networks. Therefore, incorporating feedback in these networks would further increase communication latency. Besides, feedback does not increase the capacity of the network, i.e., the min-cut. For line networks formed by packet erasure channels with a constant buffer size at the intermediate nodes, the capacity of the line network is preserved when perfect hop-by-hop feedback is allowed [14], and batched network codes achieve rate very close to the min-cut with a small batch size [15].

B. Paper Contributions

For a line network formed by general DMCs, which are not necessarily identical, we study a class of codes called Generalized Batched Network Codes (GBNCs) (see Sec. II). A GBNC has an outer code and an inner code. The outer code encodes the information messages into *batches*, each of which is a sequence of coded symbols, while the inner code performs a general recoding for the symbols belonging to the same batch. There are a couple of reasons why we choose GBNCs.

First, it is challenging to characterize a non-trivial upper bound for a line network formed by general DMCs with buffer size or inner blocklength constraints, as hinted in the literature of network information theory [16]. GBNC is general enough to include a broad class of codes as special cases. The coding schemes studied in [13] are GBNCs with the inner blocklength the same as the batch size. GBNCs extend the batched network codes studied for networks with packet loss [17]–[25] by allowing general DMCs. Furthermore, if both the batch size and the inner blocklength can be arbitrarily large, GBNCs can achieve the min-cut. In this paper, we derive upper bounds on the achievable rates of GBNCs for line networks, considering various constraints on the batch size and the inner blocklength (Theorem 6), and demonstrate that these upper bounds align with the achievable rates established in [13] for certain cases.

Second, GBNCs enable us to to characterize latency and buffer size explicitly. In our formulation, we show that the recoding latency and buffer size at an intermediate node are bounded above by a linear order of the inner blocklength. However, we also provide codes that exhibit buffer sizes of a sub-linear order of the inner blocklength.

In this paper, we characterize the achievable rates of GBNCs with the network length L. A GBNC has three parameters: batch size M, buffer size B, and inner blocklength N, which also represents the recoding latency per batch at each intermediate node. Consequently, the end-to-end recoding latency for a batch is $N(L-1)$.

TABLE I

SUMMARIZATION OF THE ACHIEVABLE RATE SCALABILITY FOR THE CHANNELS WITH 0 ZERO-ERROR CAPACITY USING BATCHED CODES. HERE, c AND c' HAVE CONSTANT VALUES THAT DO NOT CHANGE WITH L. THE UPPER/LOWER BOUND MARKED WITH $*$ IS OBTAINED IN THIS PAPER.

(a) upper bound

batch size M	inner blk-length N	buffer size B	upper bound
unbounded	$O(1)$	unbounded	$O(e^{-c'L})^*$
$O(1)$	$\Omega(\ln L)$	unbounded	$O(1/\ln L)^*$
unbounded	unbounded	unbounded	$O(1)$ [3]

(b) lower bound

batch size M	inner blk-length N	buffer size B	lower bound
$O(1)$	$O(1)$	$O(1)$	$\Omega(e^{-cL})$ [13],*
$O(1)$	$O(\ln L)$	$O(\ln \ln L)$	$\Omega(1/\ln L)^*$
$O(\ln L)$	$O(\ln L)$	$O(\ln L)$	$\Omega(1)$ [13],*

1) Upper Bound: For line networks of channels with 0 zero-error capacity, we derive a general upper bound on the achievable rates of batched codes (Theorem 6), which allows us to obtain the following upper bound scalability results in Table I. When $N = O(1)$, the achievable rate must be exponentially decaying with L, which aligns with the achievable rate obtained in [13]. When $N = O(\ln L)$ and $M = O(1)$, the achievable rate is $O(1/\ln L)$, which again matches the achievable rate obtained in [13].

We use a new "bottleneck status" technique to prove the upper bound of a line network. For a packet erasure channel, the bottleneck status occurs when all the N channel uses for a batch output erasure. The end-to-end transition matrix induced by the inner code can be decomposed as the linear combination of two parts, where one part occurs under the bottleneck status and the other part otherwise. An upper bound on the achievable rate of GBNCs can then be obtained following the convexity of mutual information.

We first prove the converses for a class of DMCs called *canonical channels*, which have an output symbol that occurs with a positive probability for all possible input symbols. Many channels extensively studied in the literature, such as packet erasure channels and binary symmetric channels, are canonical. For canonical channels, stronger upper bounds are obtained: the achievable rate is $O(1/\ln L)$ when $M = O(1)$ and N is unbounded. To solve the general case, we show that the concatenation of a number of non-canonical channels can form a canonical channel.

2) Lower Bound: We construct several coding schemes whose achievable rates match the upper bound in order with respect to the hop length L (see Sec. IV). When all the links in the line networks are identical DMCs, it has been shown in [13] that a constant rate can be achieved using batched codes with $M, N = O(\ln L)$, while exponentially decreasing rates can be achieved using batched codes with $M, N = O(1)$. We generalize these achievability results to a line network where the DMCs are not necessarily identical (see Sec. IV-A and Sec. IV-C). To prove the general result, we use a "channel reduction" technique to find proper inner coding that allows the line network to be transformed into an equivalent network with identical channels.

We also show that the rate $\Omega(1/\ln L)$ can be achieved using $M = O(1)$ and $N = O(\ln L)$. In a general decode-and-forward approach, a buffer size of $O(\ln L)$ is required. However, specific codes can achieve a reduced buffer size of $O(\ln \ln L)$ (see Sec. IV-B). To showcase this result, we employ a repetition coding scheme, which inspires us to explore simpler schemes like convolutional codes with Viterbi decoding for line networks comprising a large number of hops.

3) Extensions: In the case of line networks with packet erasure channels, we improve upon both the upper and lower bounds. Through numerical evaluations, we demonstrate the proximity of the upper and lower bounds (see Sec. V). Additionally, we extend our results to networks where some channels have a *positive* zero-error capacity (see Sec. VI).

Throughout this paper, we use $\log$ to denote the logarithm of base 2 and $\ln$ to denote the logarithm of base e. All omitted proofs and experiment codes can be found in the supplementary material at https://github.com/WalterBabyRudin/xxxxxxx. Most of the notations used throughout this manuscript are defined in Table II of Appendix A in the supplementary material. Others are defined following their first appearances, as needed.

II. Line Networks and Generalized Batched Network Coding

In this section, we describe the line network model and introduce batched network coding.

A. Line Network Model

A line network of length L consists of nodes labeled as $0, 1, \ldots, L$, with directed communication links from node $\ell - 1$ to node ℓ. Each link is a discrete memoryless channel (DMC) with fixed finite input and output alphabets $\mathcal{Q}_\mathrm{i}$ and $\mathcal{Q}_\mathrm{o}$ respectively. The transition matrix for link ℓ is denoted as Q_ℓ. The line network is formed by concatenating $Q_1, Q_2, \ldots, Q_L$. This study focuses on communication between the first node, referred to as the *source node*, and the last node, known as the *destination node*. The nodes numbered $1, 2, \ldots, L-1$ are referred to as the *intermediate nodes*.

Let $C(Q)$ and $C_0(Q)$ denote the channel capacity and zero-error capacity of a DMC with transition matrix Q respectively. Without any constraints at the network nodes, the capacity of the network is given by $\min_{\ell=1}^{L} C(Q_\ell)$, which is also known as the *min-cut*. Achieving the min-cut involves using a capacity achieving code at each hop, where intermediate nodes decode the previous link's code and encode the message using the next link's code. This scheme is commonly referred to as *decode-and-forward*. However, as we will discuss later, decode-and-forward is not always the optimal solution when considering both latency and buffer size at the intermediate nodes. Next, we present a general coding scheme for the line network and examine the relationship between the coding parameters and latency as well as buffer size.

B. Generalized Batched Network Coding

A *Generalized Batched Network Code* (GBNC) comprises an outer code and an inner code. The outer code, executed at the source node, encodes a message from a finite set and generates multiple batches, each containing M symbols from a finite set $\mathcal{A}$. The parameter M is known as the *batch size*. The inner code operates on individual batches separately, employing recoding operations at nodes $0, 1, \ldots, L-1$.

Let's define the recoding process for a generic batch $\mathbf{X} \in \mathcal{A}^M$. At the source node, the recoding transforms the original M symbols of $\mathbf{X}$ into N recoded symbols $\mathbf{U}_1$ in $\mathcal{Q}_\mathrm{i}$, where N is a positive integer referred to as the *inner blocklength*. The recoding at the source node is represented by the function $\phi_0 : \mathcal{A}^M \to \mathcal{Q}_\mathrm{i}^N$, such that $\mathbf{U}_1 = \phi_0(\mathbf{X})$.

At an intermediate node ℓ, recoding is performed on the N received symbols $\mathbf{Y}_\ell \in \mathcal{Q}_\mathrm{o}^N$ to generate N recoded symbols $U_{\ell+1} \in \mathcal{Q}_\mathrm{i}^N$ for transmission on the outgoing link of node ℓ. Due to the memoryless property of Q_ℓ, the conditional probability of $\mathbf{Y}_\ell = \mathbf{y}$ given $\mathbf{U}_\ell = \mathbf{u}$ is

$$P(\mathbf{Y}_\ell = \mathbf{y} | \mathbf{U}_\ell = \mathbf{u}) = Q_\ell^{\otimes N}(\mathbf{y}|\mathbf{u}) \triangleq \prod_{i=1}^{N} Q_\ell(\mathbf{y}[i]|\mathbf{u}[i]), \tag{1}$$

where $\mathbf{y}[k]$ $(1 \leq k \leq M)$ represents the kth entry in $\mathbf{y}$. The recoding at node ℓ is represented by the function $\phi_\ell : \mathcal{A}^M \to \mathcal{Q}_\mathrm{i}^N$, such that $\mathbf{U}_{\ell+1} = \phi_\ell(\mathbf{Y}_\ell)$. In general, the number of recoded symbols transmitted by different nodes can vary [26], [27]. However, for simplicity, we assume they are all the same for the analysis.

At the destination node, all received symbols, which may belong to different batches, are jointly decoded. The inner code's end-to-end operation, with the given recoding function ϕ_ℓ at all nodes, can be viewed as a memoryless channel referred to as a *batch channel*, which takes

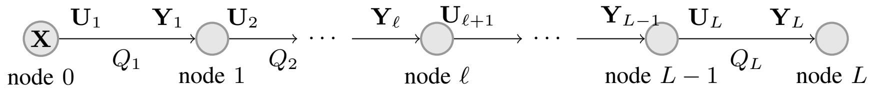

Fig. 1. A line network with the random variables involved in recoding.

$\mathbf{X}$ as the input and produces $\mathbf{Y}_L$ as the output. Fig. 1 illustrates the variables involved in the recoding process, forming the Markov chain:

$$\mathbf{X} \to \mathbf{U}_1 \to \mathbf{Y}_1 \to \cdots \to \mathbf{U}_L \to \mathbf{Y}_L. \tag{2}$$

The end-to-end transition matrix W_L of the batch channel can be derived using ϕ_ℓ and Q_ℓ.

The outer code serves as a channel code for the batch channel W_L to ensure end-to-end reliability. Given a recoding scheme $\{\phi_\ell\}$, the maximum achievable rate of the outer code is $\max_{p_\mathbf{X}} I(\mathbf{X};\mathbf{Y}_L)$ for N channel uses, where $p_\mathbf{X}$ represents the distribution of $\mathbf{X}$. The objective of designing a recoding scheme, given parameters M and N, is to maximize $\frac{1}{N}\max_{p_\mathbf{X}} I(\mathbf{X};\mathbf{Y}_L)$. Let $C_L(M,N)$ denote the maximum achievable rate among all recoding schemes with batch size M and inner blocklength N, defined as:

$$C_L(M,N) = \frac{1}{N}\max_{\{\phi_\ell\}}\max_{p_\mathbf{X}} I(\mathbf{X};\mathbf{Y}_L) = \frac{1}{N}\max_{\{\phi_\ell\}}\max_{p_\mathbf{X}} I(p_\mathbf{X}, W_L). \tag{3}$$

$C_L(M,N)$ is also referred to as the capacity of GBNCs with parameters M and N. We can then maximize $C_L(M,N)$ while considering constraints on M and N, which impact both the recoding latency and the buffer size.

Recoding functions $\{\phi_\ell\}$ can generally be random. However, the convexity of $I(p_\mathbf{X}, W_L)$ for a fixed $p_\mathbf{X}$ with respect to W_L implies the existence of a deterministic recoding scheme that achieves $C_L(M,N)$. In particular, the coding scheme analyzed in [13] considers the case where $M = N$. A special inner code known as decode-and-forward will be discussed in §IV. GBNCs generalize the batched network codes studied for networks with packet erasure channels in literature (see discussion in §V).

C. Buffer Size and Latency at Intermediate Nodes

Let's now delve into the buffer size requirement and latency at the intermediate nodes in GBNCs. In this discussion, we consider a sequential transmission model where symbols of a batch are transmitted consecutively. We will disregard the space and time costs associated with executing recoding ϕ_ℓ, but instead focus on the buffer size needed for caching received symbols and the corresponding latency. Specifically, we will discuss the buffer size required for caching the received symbols for recoding at an intermediate node, as well as the latency between receiving the first symbol of a batch and transmitting the first symbol of the same batch.

The key principle of GBNCs is the independent application of recoding to each batch. In the worst case scenario, an intermediate node begins transmitting the first recoded symbol of a batch only after receiving all N symbols of that batch. Consequently, the latency of a batch at an intermediate node is upper bounded by N. Since an intermediate node can only transmit symbols of a batch after receiving at least one symbol from that batch, the lower bound on the latency at an intermediate node is 1. The accumulated end-to-end recoding latency across all intermediate nodes falls within the range of $L-1$ to $(L-1)N$.

Similarly, in the worst-case scenario, an intermediate node starts transmitting the first recoded symbol of a batch only after receiving all N symbols of that batch. Additionally, these received symbols need to be cached for N more channel uses. Therefore, an intermediate node needs to cache at most $2N$ symbols: N symbols of the batch for transmitting and N symbols of the same batch for receiving. This indicates that the buffer size required for caching symbols at an intermediate node is $O(N)$.

III. Converse for Line Networks of Channels with 0 Zero-error Capacity

In this section, we study upper bounds on $C_L(M,N)$ for line networks of channels with 0 zero-error capacity, i.e., $C_0(Q_\ell)=0$. One known upper bound of $C_L(M,N)$ is the min-cut $\min_{\ell=1}^{L} C(Q_\ell)$. However, this bound may not be sufficient for small values of M and N. In this section, we introduce a technique called a "bottleneck status" to derive a tighter bound on $C_L(M,N)$ when M and N are small.

The bottleneck status refers to an event E_0 that is associated with the channel W_L. Let

$$W_L^{(0)}(\mathbf{y}\mid\mathbf{x}) = P(\mathbf{Y}_L=\mathbf{y}\mid\mathbf{X}=\mathbf{x},E_0),\quad W_L^{(1)}(\mathbf{y}\mid\mathbf{x}) = P(\mathbf{Y}_L=\mathbf{y}\mid\mathbf{X}=\mathbf{x},\overline{E_0}). \tag{4}$$

The channel W_L can be expressed as $W_L = W_L^{(0)}p_0 + W_L^{(1)}p_1$, where $p_0=P(E_0), p_1=P(\overline{E_0})$. As mutual information $I(p_\mathbf{X},W_L)$ is convex w.r.t. W_L for given $p_\mathbf{X}$, we can establish the upper bound as follows:

$$I(p_\mathbf{X},W_L) \le p_0 I(p_\mathbf{X},W_L^{(0)}) + p_1 I(p_\mathbf{X},W_L^{(1)}). \tag{5}$$

The crucial step is to design the event E_0 in order to obtain the desired upper bound. We first introduce our technique for canonical channels, and then discuss the general channels.

A. *Line Network of Canonical Channels*

For $0 < \varepsilon \leq 1$, we call a channel $Q : \mathcal{Q}_\mathrm{i} \to \mathcal{Q}_\mathrm{o}$ an *ε-canonical channel* if there exists $y^* \in \mathcal{Q}_\mathrm{o}$ such that for every $x \in \mathcal{Q}_\mathrm{i}$, $Q(y^*|x) \geq \varepsilon$. In other words, for a canonical channel, there exists an output symbol y^* that occurs with a positive probability for all the inputs. The binary erasure channel (BEC) and binary symmetric channel (BSC) are both canonical channels, but a typewriter channel is non-canonical. Note that a canonical Q has $C_0(Q) = 0$. In this subsection, we study a line network consisting of ε-canonical channels $Q_\ell, \ell = 1, \dots, L$.

To design the bottleneck status E_0, we adopt a formulation of DMCs in [28, Sec. 7.1]. The relation between the input X and output Y of a DMC Q can be modeled as a function α as $Y = \alpha(X, Z)$, where Z is a random variable independent of X. In particular, we have

$$Y = \alpha(X, Z = (Z[x], x \in \mathcal{Q}_\mathrm{i})) = \sum_{x \in \mathcal{Q}_\mathrm{i}} \mathbf{1}\{X = x\} Z[x], \tag{6}$$

where $\mathbf{1}$ denotes the indicator function, and $Z[x], x \in \mathcal{Q}_\mathrm{i}$ are independent random variables on $\mathcal{Q}_\mathrm{o}$ with the distribution $P(Z[x] = y) = Q(y|x)$. Here $Z = (Z[x], x \in \mathcal{Q}_\mathrm{i})$ is also called *channel status variable*, and α is called the *channel function*. We denote by α_ℓ the channel function of Q_ℓ.

Consider a GBNC with inner blocklength N for the line network. With the alternative channel formulation (6), we can write for $\ell = 1, \dots, L$, and $i = 1, \dots, N$, $\mathbf{Y}_\ell[i] = \alpha_\ell(\mathbf{U}_\ell[i], \mathbf{Z}_\ell[i])$. Here $\mathbf{Z}_\ell[i] = (\mathbf{Z}_\ell[i, x], x \in \mathcal{Q}_\mathrm{i})$ is the channel status variable for the i-th use of the channel Q_ℓ, where

$$P(\mathbf{Z}_\ell[i, x] = y) = Q_\ell(y|x). \tag{7}$$

Define $\mathbf{Z}_\ell = (\mathbf{Z}_\ell[i], i = 1, \dots, N)$. For notation simplicity, we rewrite the channel relation as

$$\mathbf{Y}_\ell = \alpha_\ell^{(N)}(\mathbf{U}_\ell, \mathbf{Z}_\ell). \tag{8}$$

Given that Q_ℓ is ε-canonical, there exists an output denoted as y_ℓ^* satisfying

$$Q_\ell(y_\ell^*|x) \geq \varepsilon \text{ for all } x \in \mathcal{Q}_\mathrm{i}. \tag{9}$$

Let's define $E_{0,\ell} = \{\mathbf{Z}_\ell[i,x] = y_\ell^*, i \in \{1,\dots,N\}, x \in \mathcal{Q}_{\mathrm{i}}\}$. Under the condition $E_{0,\ell}$, all N outputs of Q_ℓ are equal to y_ℓ^* for any possible channel input, rendering the channel useless. We can quantify the probability of $E_{0,\ell}$ as follows:

$$P(E_{0,\ell}) = \prod_{i\in\{1,\dots,N\},x\in\mathcal{Q}_{\mathrm{i}}} P(\mathbf{Z}_\ell[i,x] = y_\ell^*) = \prod_{i\in\{1,\dots,N\},x\in\mathcal{Q}_{\mathrm{i}}} Q_\ell(y_\ell^*|x) \geq \varepsilon^{|\mathcal{Q}_{\mathrm{i}}|N}, \tag{10}$$

where the second equality follows from (7), and the inequality follows from (9). Now consider the event

$$E_0 = \vee_{\ell=1}^{L} E_{0,\ell}. \tag{11}$$

This event implies the existence of at least one link ℓ in the network that is deemed useless. To establish this rigorously, we show the following lemma.

Lemma 1. *When Q_ℓ, $\ell = 1,\dots,L$ are all ε-canonical channels, for $W_L^{(0)}$ defined in* (4) *and E_0 defined in* (11)*, $I(p_{\mathbf{X}}, W_L^{(0)}) = 0$.*

Proof: Write $E_0 = \vee_{\ell=1}^{L}(E_{0,\ell} \wedge_{\ell'>\ell} \overline{E_{0,\ell'}})$, where $(E_{0,\ell} \wedge_{\ell'>\ell} \overline{E_{0,\ell'}})$, $\ell = 1,\dots,L$ are disjoint. Hence,

$$P(\mathbf{y}_L, \mathbf{x}, E_0) = \sum_{\ell=1}^{L} P(\mathbf{y}_L, \mathbf{x}, E_{0,\ell}, \wedge_{\ell'>\ell}\overline{E_{0,\ell'}}) = \sum_{\ell,\mathbf{y}_\ell,\mathbf{u}_\ell} P(\mathbf{y}_L, \mathbf{x}, E_{0,\ell}, \wedge_{\ell'>\ell}\overline{E_{0,\ell'}}, \mathbf{y}_\ell, \mathbf{u}_\ell) \tag{12}$$

$$= \sum_{\ell,\mathbf{y}_\ell} P(\mathbf{y}_L, \wedge_{\ell'>\ell}\overline{E_{0,\ell'}} \mid \mathbf{y}_\ell) \sum_{\mathbf{u}_\ell} P(\mathbf{x}, E_{0,\ell}, \mathbf{y}_\ell, \mathbf{u}_\ell), \tag{13}$$

where the last equality follows from the Markov chain in (2). Further,

$$\sum_{\mathbf{u}_\ell} P(\mathbf{x}, E_{0,\ell}, \mathbf{y}_\ell, \mathbf{u}_\ell) = \sum_{\mathbf{u}_\ell} P(\mathbf{x}, \mathbf{u}_\ell) P(E_{0,\ell}) P(\mathbf{y}_\ell \mid \mathbf{u}_\ell, E_{0,\ell}) \tag{14}$$

$$= \sum_{\mathbf{u}_\ell} P(\mathbf{x}, \mathbf{u}_\ell) P(E_{0,\ell}) P(\mathbf{y}_\ell \mid E_{0,\ell}) = p_{\mathbf{X}}(\mathbf{x}) P(E_{0,\ell}) P(\mathbf{y}_\ell \mid E_{0,\ell}), \tag{15}$$

where (14) follows from $\mathbf{Y}_\ell = \alpha_\ell^{(N)}(\mathbf{U}_\ell, \mathbf{Z}_\ell | E_{0,\ell}) = y_\ell^*$ (ref. (8)). By (13) and (15),

$$P(\mathbf{y}_L, \mathbf{x}, E_0) = p_{\mathbf{X}}(\mathbf{x}) \sum_{\ell,\mathbf{y}_\ell} P(\mathbf{y}_L, \wedge_{\ell'>\ell}\overline{E_{0,\ell'}} \mid \mathbf{y}_\ell) P(\mathbf{y}_\ell, E_{0,\ell}) \tag{16}$$

$$= p_{\mathbf{X}}(\mathbf{x}) \sum_{\ell,\mathbf{y}_\ell} P(\mathbf{y}_L, \wedge_{\ell'>\ell}\overline{E_{0,\ell'}}, \mathbf{y}_\ell, E_{0,\ell}) = P(\mathbf{y}_L, \mathbf{x}, E_0), \tag{17}$$

which implies $I(p_{\mathbf{X}}, W_L^{(0)}) = 0$. ■

Lemma 2. *When Q_ℓ, $\ell = 1, \ldots, L$ are all ε-canonical channels, for $W_L^{(1)}$ defined in* (4) *and E_0 defined in* (11)*, we have 1) $P(\overline{E_0}) \leq (1 - \varepsilon^{|\mathcal{Q}_\mathrm{i}|N})^L$ and 2)*

$$I(p_\mathbf{X}, W_L^{(1)}) \leq \min\left\{H(\mathbf{X}), \max_{p_{\mathbf{U}_L}} I(\mathbf{U}_\ell; \mathbf{Y}_\ell \mid \overline{E_{0,\ell}}), \ell = 1, \ldots, L\right\}. \tag{18}$$

Proof: As $\overline{E_0} = \wedge_{\ell=1}^L \overline{E_{0,\ell}}$, by (10), $P(\overline{E_0}) = \prod_{\ell=1}^L (1 - P(E_{0,\ell})) \leq (1 - \varepsilon^{|\mathcal{Q}_\mathrm{i}|N})^L$. We first show that given $\overline{E_0}$, $\mathbf{Z}_1, \ldots, \mathbf{Z}_L$ are independent. Write

$$P(\overline{E_0}) = P(\overline{E_{0,\ell}}, \ell = 1, \ldots, L) = \prod_{\ell=1}^L P(\overline{E_{0,\ell}}) \tag{19}$$

$$P(\mathbf{Z}_\ell, \ell = 1, \ldots, L, \overline{E_0}) = P(\mathbf{Z}_\ell, \overline{E_{0,\ell}}, \ell = 1, \ldots, L) = \prod_{\ell=1}^L P(\mathbf{Z}_\ell \mid \overline{E_{0,\ell}}) P(\overline{E_{0,\ell}}) \tag{20}$$

Hence,

$$P(\mathbf{Z}_\ell, \ell = 1, \ldots, L \mid \overline{E_0}) = \prod_{\ell=1}^L P(\mathbf{Z}_\ell \mid \overline{E_{0,\ell}}) = \prod_{\ell=1}^L P(\mathbf{Z}_\ell \mid \overline{E_0}). \tag{21}$$

Under the condition of $\overline{E_0}$, as $\mathbf{Z}_1, \ldots, \mathbf{Z}_L$ are independent, we have the Markov chain in (2) holds and hence $I(p_\mathbf{X}, W_L^{(1)}) \leq I(\mathbf{U}_\ell; \mathbf{Y}_\ell \mid \overline{E_{0,\ell}})$ and $I(p_\mathbf{X}, W_L^{(1)}) \leq H(\mathbf{X})$. ■

In Lemma 2, $\max_{p_\mathbf{U}} I(\mathbf{U}_\ell; \mathbf{Y}_\ell \mid \overline{E_{0,\ell}})$ is the capacity of the channel Q_ℓ^N under the condition $\overline{E_{0,\ell}}$. One upper bound is $\frac{1}{N} \max_{p_{\mathbf{U}_L}} I(\mathbf{U}_\ell; \mathbf{Y}_\ell \mid \overline{E_{0,\ell}}) \leq \log \min(|\mathcal{Q}_\mathrm{i}|, |\mathcal{Q}_\mathrm{o}|)$. In the following lemma, we give a better upper bound that converges $C(Q_\ell)$ when N tends to infinity.

Lemma 3. *Consider a channel Q as defined in* (6) *by (α, Z). Fix an output y^* such that $Q(y^*|x) = P(Z[x] = y^*) \geq \epsilon$ for all input x, where $\epsilon > 0$. For N uses of the channel, let $Z[i, x]$ be the channel variable of the ith uses associated with the input x. Let E_0 be the event that $\{Z[i, x] = y^*, i = 1, \ldots, N, x \in \mathcal{Q}_\mathrm{i}\}$. Let W be the channel formed by N uses of Q under the condition of $\overline{E_0}$. Then*

$$\frac{1}{N} I(p, W) \leq C^*(Q, N) \triangleq \frac{1}{1 - p_0}\left(C(Q) + \frac{1}{N}\left((q^* + p_0)\log\frac{q^* - p_0}{\epsilon^N - p_0} + q^* \log\frac{\epsilon^N}{q^*}\right)\right), \tag{22}$$

where $p_0 = P(E_0) = [\prod_x Q(y^*|x)]^N$, and $q^* = \max_\mathbf{x} Q^{\otimes N}(\mathbf{y}^*|\mathbf{x})$.

Based on the relation (5) together with Lemmas 1, 2, and 3, we derive the following theorem. Its full proof is provided in Appendix B.

Theorem 4. *Consider a length-L line network of ε-canonical channels with finite input and output alphabets $\mathcal{Q}_\mathrm{i}$ and $\mathcal{Q}_\mathrm{o}$, respectively. The capacity of GBNCs with batch size M and inner blocklength N has the following upper bound:*

$$C_L(M, N) \leq (1 - \varepsilon^{|\mathcal{Q}_\mathrm{i}|N})^L \min\left\{C^*(Q_\ell, N), \log|\mathcal{Q}_\mathrm{i}|, \log|\mathcal{Q}_\mathrm{o}|, \frac{M \log|\mathcal{A}|}{N}\right\}. \tag{23}$$

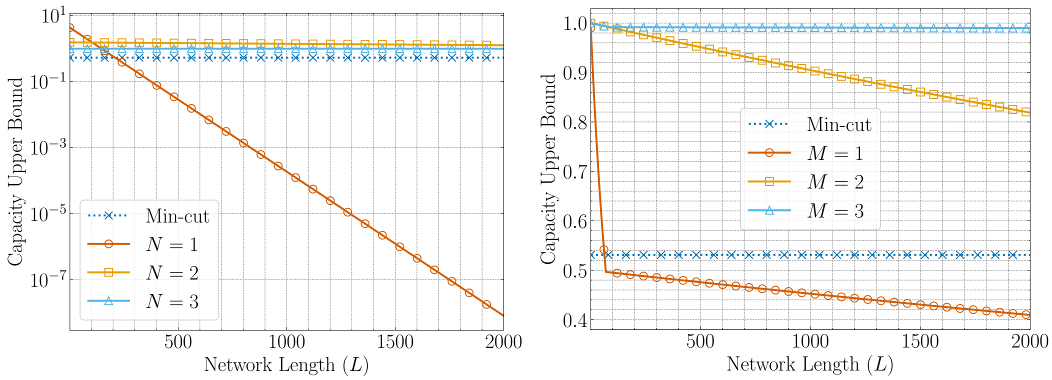

Fig. 2. Numerical illustrations of the capacity of GBNCs using BSC with crossover probability $\epsilon = 0.2, |\mathcal{A}| = 2$. The hop length L ranges from 1 to 2000. Figures from left to right correspond to (a) the upper bound (23) of the capacity with fixed $N \in \{1, 2, 3\}$ and optimized batch size M; and (b) the upper bound (23) of the capacity with fixed $M \in \{1, 2, 3\}$ and optimized inner blocklength N.

Moreover,

1) *when* $N = O(1)$*,* $\max_M C_L(M, N) = O((1 - \varepsilon^{|\mathcal{Q}_\mathrm{i}|N})^L)$*;*
2) *when* $M = O(1)$*,* $\max_N C_L(M, N) = O(1/\ln L)$*;*
3) *when* M *and* N *are arbitrary,* $\max_{M,N} C_L(M, N) = O(1)$.

To illustrate the capacity upper bound in Theorem 4, we evaluate it for the network formed by BSCs in Fig. 2, and use the min-cut for baseline comparison. Fig. 2(a) depicts, for each hop length L, the optimal value of the upper bound (23) in terms of batch size M, with a fixed inner blocklength N. For the case $N = 1$, it reveals the exponential decay of the capacity with respect to L, and the min-cut is in geneal a loose upper bound. For cases $N = 2, 3$, however, the hop length L may not be sufficiently large to reveal the exponential decaying trend of the capacity. Fig. 2(b) shows the optimal value of the upper bound (23) in terms of inner blocklength N, with a fixed batch size M. For the case $M = 1$, it indicates the capacity decays slowly as L increases, and the min-cut is a loose upper bound as well. Similarly, the hop length L may not be sufficiently large to illustrate the $1/\ln L$ decaying trend of the capacity.

B. General Channels

Consider a channel $Q : \mathcal{Q}_\mathrm{i} \to \mathcal{Q}_\mathrm{o}$ with $C_0(Q) = 0$, modeled as in (6). As Q may not be canonical, an output symbol that occurs with a positive probability for all the inputs may not exist. Moreover, if Q is non-canonical, $Q^{\otimes m}$ is non-canonical for any positive integer m. For example,

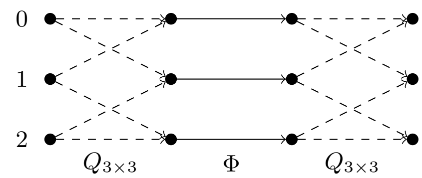

Fig. 3. Concatenation of two $Q_{3\times 3}$'s with recoding Φ. The end-to-end channel is given by $W = Q_{3\times 3}\Phi Q_{3\times 3}$. Here Φ is a deterministic transition matrix with $\Phi(i|i) = 1$. The transition from an input to an output connected by a dashed has probability $1/2$. Any output of W can occur with a positive probability for all inputs.

define channel $Q_{3\times 3}$ with $\mathcal{Q}_\mathrm{i} = \mathcal{Q}_\mathrm{o} = \{0,1,2\}$ and $Q_{3\times 3}(0|0) = Q_{3\times 3}(0|1) = Q_{3\times 3}(1|0) = Q_{3\times 3}(1|2) = Q_{3\times 3}(2|1) = Q_{3\times 3}(2|2) = 1/2$. For a non-canonical channel, we cannot find an event of the channel status with a positive probability such that the channel outputs are the same for all inputs.

To study the converse of general channels, our technique is to use the concatenation of multiple channels by recoding, which forms a new channel that can be canonical. We use $Q_{3\times 3}$ to illustrate the idea. Consider the concatenation of two copies of $Q_{3\times 3}$ with a 3×3 *deterministic* transition matrix Φ, which yields the new channel $W = Q_{3\times 3}\Phi Q_{3\times 3}$. In other words, Φ maps an output of the first channel as an input of the second channel. See an illustration in Fig. 3. For the first channel, at least one output of $\{0,1\}$ occurs for any channel input. If Φ maps these two outputs 0 and 1 of the first channel to the same input, say 2, of the second channel, then output 2 of W occurs with a positive probability for any input of W. If Φ maps these two outputs 0 and 1 of the first channel to two different inputs, say 1 and 2 respectively, of the second channel, then output 2 of W occurs with a positive probability for any input of W.

Now we discuss the general case. For a channel $Q: \ \mathcal{Q}_\mathrm{i} \to \mathcal{Q}_\mathrm{o}$, denote by ε_Q the maximum value such that for any $x, x' \in \mathcal{Q}_\mathrm{i}$, there exists $y \in \mathcal{Q}_\mathrm{o}$ so that $Q(y|x) \geq \varepsilon_Q$ and $Q(y|x') \geq \varepsilon_Q$. For $Q_{3\times 3}$, we have $\varepsilon_{Q_{3\times 3}} = 1/2$. Note that $\varepsilon_Q > 0$ if and only if $C_0(Q) = 0$ (see [29]). As $C_0(Q_\ell) = 0$, For any two channel inputs of Q_ℓ, it is possible to observe the same output. Based on this fact, we can further show that for any subset $\mathcal{S}_\mathrm{i}$ of $\mathcal{Q}_\mathrm{i}$, there exists a subset $\mathcal{S}_\mathrm{o}$ of $\mathcal{Q}_\mathrm{o}$ size less than half of $\mathcal{S}_\mathrm{i}$ such that for any input in $\mathcal{S}_\mathrm{i}$, it is possible to observe an output in $\mathcal{S}_\mathrm{o}$. Formally, we have the following lemma, proved in Appendix B:

Lemma 5. *Consider a DMC $Q : \mathcal{Q}_\mathrm{i} \to \mathcal{Q}_\mathrm{o}$ with $\varepsilon_Q > 0$, channel function α and channel status variable Z. For any non-empty set $\mathcal{S}_\mathrm{i} \subseteq \mathcal{Q}_\mathrm{i}$, there exist a subset $\mathcal{Z}$ of the range of Z and a* *subset* $\mathcal{S}_\mathrm{o} \subseteq \mathcal{Q}_\mathrm{o}$ *with* $|\mathcal{S}_\mathrm{o}| \le \lceil |\mathcal{S}_\mathrm{i}|/2 \rceil$ *such that* $\alpha(x,z) \in \mathcal{S}_\mathrm{o}$ *for any* $x \in \mathcal{S}_\mathrm{i}$ *and* $z \in \mathcal{Z}$*, and* $P(Z \in \mathcal{Z}) \ge \varepsilon_Q^{|\mathcal{S}_\mathrm{i}|}$.

Based on the above lemma, we can concatenate a number of consecutive channels in the line network to form a canonical channel, and hence the line network becomes one formed by canonical channels. Then similar to Theorem 4, we can prove the following result, where the proof is in Appendix B.

Theorem 6. *Consider a length-L line network of channels $\{Q_\ell\}_{\ell=1}^L$ with finite input and output alphabets and $\varepsilon_{Q_\ell} \ge \varepsilon > 0$ for all ℓ. The capacity of GBNCs with batch size M and inner blocklength N has the following upper bound:*

$$C_L(M,N) \le (1-\varepsilon^{N(2|\mathcal{Q}_\mathrm{i}|^N+K)})^{\lfloor L/K \rfloor} \min\{M/N \log|\mathcal{A}|, \log|\mathcal{Q}_\mathrm{i}|, \log|\mathcal{Q}_\mathrm{o}|\}, \tag{24}$$

where $K = \lceil N \log|\mathcal{Q}_\mathrm{i}| \rceil$. Moreover,

1) *when $N = O(1)$, $C_L^* = O((1-\varepsilon')^L)$ for certain $\varepsilon' \in (0,1)$;*
2) *when $M = O(1)$ and $N = \Omega(\ln L)$, $C_L^* = O(1/\ln L)$;*
3) *when M and N are arbitrary, $C_L^* = O(1)$.*

When applied to line networks of canonical channels, Theorem 4 gives stronger results than that of Theorem 6. The upper bound in (23) is strictly better than that in (24). For scalabilities, the first and the third cases in both theorems have the same condition and the same scalability order of L. For the second case, in Theorem 6, there is an extra condition that $N = \Omega(\ln L)$.

IV. Achievable Rates using Decode-and-Forward Recoding

In this section, we discuss the lower bounds of the achievable rates of line networks. We will first study the achievable rates when $N = O(\ln L)$ using two recoding schemes: decode-and-forward and repetition, which can achieve different scalability of the buffer size B. When $N = O(1)$, for a line network of identical channels, a rate that exponentially decays with L can be achieved as proved in [13]. We will extend their results for line networks where channels may not be identical.

A. Decode-and-forward Recoding

We discuss a class of GBNC recoding called *decode-and-forward*. When there is a trivial outer code, decode-and-forward has been extensively studied and widely applied in the existing communication systems [16]. We first describe decode-and-forward recoding in the GBNC framework, and then discuss the scalability of achievable rates.

Following the notations in Sec. II-B, we consider a GBNC with batch size M. Let (f_ℓ, g_ℓ) be an (N, M) channel code for Q_ℓ where $f_\ell : \mathcal{A}^M \to \mathcal{Q}_\mathrm{i}^N$ and $g_\ell : \mathcal{Q}_\mathrm{o}^N \to \mathcal{A}^M$ are the encoding and decoding functions, respectively. Consider the transmission of a generic batch $\mathbf{X}$. The source node transmits $\mathbf{U}_1 = f_1(\mathbf{X})$. Each intermediate node ℓ first catches the N received symbols of $\mathbf{Y}_\ell$ and then transmits $\mathbf{U}_{\ell+1} = f_{\ell+1}(g_\ell(\mathbf{Y}_\ell))$. In other words, the recoding function ϕ_ℓ behaves as follows:

- For $i = 1, \ldots, N$, the recoding just keeps the received symbols in the buffer, i.e., $\mathbf{B}_\ell[i] = \mathbf{Y}_\ell[1 : i]$. Therefore, $B = \Theta(N)$.
- After receiving the N symbols of $\mathbf{Y}_\ell$, the recoding generates $f_{\ell+1}(g_\ell(\mathbf{Y}_\ell))$. If the decoding is correct at nodes $1, \ldots, \ell$, then $g_\ell(\mathbf{Y}_\ell) = \mathbf{X}$ and $\mathbf{U}_{\ell+1} = f_{\ell+1}(\mathbf{X})$.

Let ϵ_ℓ denote the maximum decoding error probability of (f_ℓ, g_ℓ) for Q_ℓ. Due to the fact that if the decoding is correct at all the nodes $1, \ldots, L$, it holds that $g_L(\mathbf{Y}_L) = \mathbf{X}$, we have

$$P(g_L(\mathbf{Y}_L) = \mathbf{X}) \geq \prod_{\ell=1}^{L} (1 - \epsilon_\ell). \tag{25}$$

Let $C = \min_{\ell=1}^{L} C(Q_\ell)$ be the min-cut of the line network. When $\frac{M}{N} \log |\mathcal{A}| < C$ and N is sufficiently large, by the channel coding theorem of DMCs, there exists (f_ℓ, g_ℓ) such that ϵ_ℓ can be arbitrarily small. This gives us the well-known result that the min-cut C is achievable using decode-and-forward recoding when M, N and B are allowed to be arbitrarily large [3]. When all the channels are identical, it has been shown that if $M = \Theta(N)$ and $N = O(\ln L)$ (and hence $B = O(\ln L)$), a constant rate lower than C can be achieved by GBNC [13].

We briefly rephrase their discussion for the case where the channels of the line network are not identical. Consider a sequence of DMCs $Q_\ell, \ell = 1, 2, \ldots$ with $C = \inf\{C(Q_\ell), \ell \geq 1\} > 0$. Suppose parameters M and N are chosen to satisfy $r \triangleq \frac{M}{N} \log |\mathcal{A}| \in [0, C]$. Using random coding arguments [30], there exists (f_ℓ, g_ℓ) such that

$$\epsilon_\ell \leq \exp(-N \mathrm{Er}_\ell(r)), \tag{26}$$

where Er_ℓ is the random coding error exponent for Q_ℓ. For certain $0 < C' \leq C$, assume $\mathrm{Er}^*(r) \triangleq \inf\{\mathrm{Er}_\ell(r), \ell \geq 1\} > 0$ for all $0 \leq r < C'$. The following theorem shows the achievable rate of decode-and-forward recoding scheme.

Theorem 7. *For the line network of length L, where the ℓ-th link is Q_ℓ, the GBNC with decode-and-forward recoding scheme, batch size M, and inner blocklength N achieves rate*

$$C_L(M,N) \geq \frac{M \log|\mathcal{A}|}{N}\left(1-e^{-N\mathrm{Er}^*(M\log|\mathcal{A}|/N)}\right)^L - \frac{1}{N}. \tag{27}$$

The lower bound (27) *is non-trivial only if*

$$\left(1-e^{-N\mathrm{Er}^*(M\log|\mathcal{A}|/N)}\right)^L > |\mathcal{A}|^{-M}. \tag{28}$$

Moreover,

1) *when $M = \Theta(N)$ and $N = O(\ln L)$, $C_L^* = \Omega(1)$;*
2) *when $M = O(1)$ and $N = O(\ln L)$, $C_L^* = \Omega(1/\ln L)$.*

Proof of Theorem 7: Substituting the error bound of ϵ_ℓ in (26) into (25), we obtain the end-to-end decoding error bound:

$$P(g_L(\mathbf{Y}_L) \neq \mathbf{X}) \leq 1 - \prod_{\ell=1}^{L}(1-e^{-N\mathrm{Er}_\ell(r)}) \leq 1-\left(1-e^{-N\mathrm{Er}^*(r)}\right)^L. \tag{29}$$

Using a similar argument as in the proof of [13, Theorem V.3], the GBNC achieves rate $r\left(1-e^{-N\mathrm{Er}^*(r)}\right)^L - 1/N$. When (28) does not hold, this rate is bounded by $\frac{1}{N}[\log(|\mathcal{A}|^M)/|\mathcal{A}|^M - 1] \leq 0$, which is trivial. Next, we discuss the scalability of the rate for different scalings of M and N.

- When $M = \Theta(N)$, i.e., $r_1 \leq r \leq r_2$ for some $0 < r_1 < r_2 < C'$, as long as $N = O(\ln L)$ and $\left(1-e^{-N\mathrm{Er}^*(r_2)}\right)^L > |\mathcal{A}|^{-M}$, the achievable rate of GBNC is at least $r_2\left(1-e^{-N\mathrm{Er}^*(r_2)}\right)^L - \frac{1}{N} = \Theta(1)$. Note that in $\left(1-e^{-N\mathrm{Er}^*(r_2)}\right)^L > |\mathcal{A}|^{-r_1 N} > |\mathcal{A}|^{-M}$, the first inequality can be satisfied when $N = O(\ln L)$ and L is sufficiently large.
- When $M = O(1)$, i.e., $r \leq r_3$ for some $0 < r_3 < C$, as long as $N = O(\ln L)$ and $\left(1-e^{-N\mathrm{Er}^*(r)}\right)^L > |\mathcal{A}|^{-M}$, the achievable rate of GBNC is at least $r\left(1-e^{-N\mathrm{Er}^*(r)}\right)^L - \frac{1}{N} = \Omega(\ln L)$. Note that in $\left(1-e^{-N\mathrm{Er}^*(r)}\right)^L \geq \left(1-e^{-N\mathrm{Er}^*(r_3)}\right)^L > |\mathcal{A}|^{-M}$, the second inequality can be satisfied when $N = O(\ln L)$ and $M = O(1)$ are chosen properly. ■

We provide a special example showcasing the achievable rates of GBNC based on decode-and-forward in Fig. 4: we use a BSC with crossover error probability $\epsilon = 0.2$, $|\mathcal{A}| = 2$ and vary the hop length L from 1 to 2000. The solid lines correspond to the case where $M = \Theta(N), N = O(\ln L)$, from which we find the achievable rate remains to be a constant for increasing hop length L. The dash lines correspond to the case where $M = O(1), N = O(\ln L)$, whears in

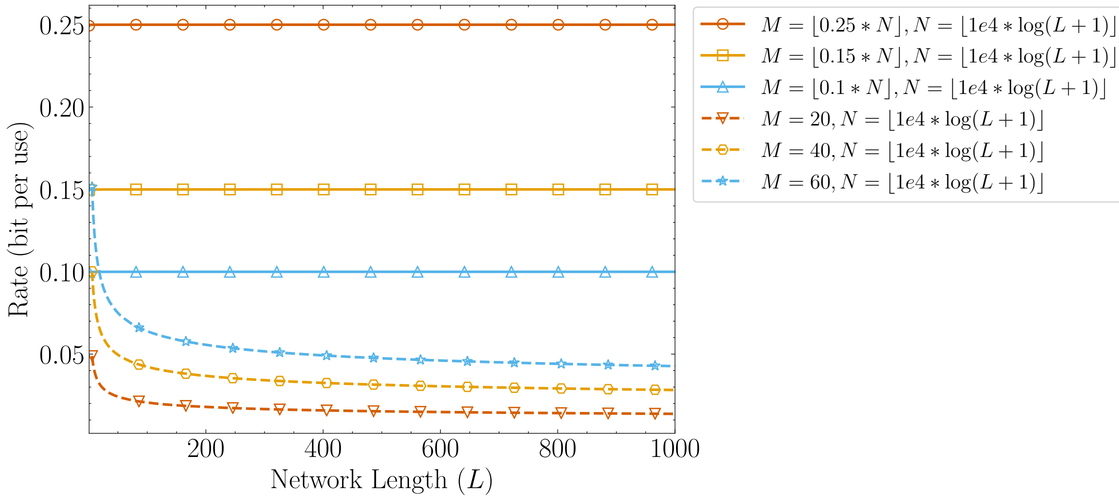

Fig. 4. Numerical illustrations of the achievable rates of GBNC based on decode-and-forward recoding (see (27)) using BSC with crossover probability $\epsilon = 0.2, |\mathcal{A}| = 2$. The hop length L ranges from 1 to 1000. The solid lines correspond to achievable rates when $M = \lfloor c_1 N \rfloor, N = \lfloor c_2 \log(L+1) \rfloor$, and the dashed lines correspond to rates when $M = c_1', N = \lfloor c_2 \log(L+1) \rfloor$. Here constants $c_1 \in \{0.1, 0.15, 0.25\}, c_2 = 1e4, c_1' \in \{20, 40, 60\}$.

this situation, the achievable rate decays slowly when L increases. As a summary, decode-and-forward recoding can achieve the same order of rate scalability as the upper bound in Theorem 6 for case 2) and 3), where the buffer size requirement is $B = O(N) = O(\ln L)$. In Sec. IV-B, we will show that it is possible to achieve $\Omega(1/\ln L)$ using $M = O(1)$, $N = O(\ln L)$ and $B = O(\ln \ln L)$.

The above approach, however, cannot be used to show the scalability with $M = O(1)$ and $N = O(1)$, since (28) does not hold when $N = O(1)$ and L is large. This case will be discussed in Sec.IV-C using another approach.

B. Special Decode-and-forward Recoding: Repetition Recoding

In this subsection, we provide a special recoding scheme that achieves rate $\Omega(1/\ln L)$ using $M = O(1), N = O(\ln L)$, and $B = O(\ln \ln L)$. Specifically, we discuss the repetition recoding scheme, which is a special decode-and-forward recoding scheme. In the following, we introduce this recoding scheme by specifying $f_\ell, g_\ell, \ell = 1, \dots, L$ defined in Sec. IV-A.

We first discuss the case $M = 1$. For any ℓ, let $\mathcal{Q}_\mathrm{i}^\ell$ be the maximal subset of $\mathcal{Q}_\mathrm{i}$ such that for any $x \neq x' \in \mathcal{Q}_\mathrm{i}^\ell$, $Q_\ell(\cdot|x) \neq Q_\ell(\cdot|x')$. For $\ell = 1, \dots, L$, assume $|\mathcal{Q}_\mathrm{i}^\ell| \geq |\mathcal{A}| \geq 2$, and let u_ℓ

be a one-to-one mapping from $\mathcal{A}$ to $\mathcal{Q}_{\mathrm{i}}^{\ell}$. For a generic batch $x \in \mathcal{A}$ with $M = 1$, node $\ell - 1$ transmits $u_\ell(x)$ for N times, i.e.,

$$f_\ell(x) = (u_\ell(x), \dots, u_\ell(x)). \tag{30}$$

Suppose $\mathbf{Y}_\ell = \mathbf{y}_\ell$, i.e., node ℓ receives $\mathbf{y}_\ell$ for the transmission $f_\ell(x)$. The decoding function g_ℓ is defined based on the maximum likelihood (ML) criterion:

$$g_\ell(\mathbf{y}_\ell) = \arg\max_{x \in \mathcal{A}} \prod_{i=1}^{N} Q_\ell(\mathbf{y}_\ell[i] \mid u_\ell(x)), \tag{31}$$

where a tie is broken arbitrarily. Let

$$\mathcal{L}_\ell(x; \mathbf{y}_\ell) = \sum_{i=1}^{N} \ln Q_\ell(\mathbf{y}_\ell[i] \mid u_\ell(x)) = \sum_{y \in \mathcal{Q}_{\mathrm{o}}} \mathcal{N}(y|\mathbf{y}_\ell) \ln Q_\ell(y \mid u_\ell(x)), \tag{32}$$

where $\mathcal{N}(y|\mathbf{y}_\ell)$ denote the number of times that y appears in $\mathbf{y}_\ell$. Then the ML decoding problem can be equivalently written as $g_\ell(\mathbf{y}_\ell) = \arg\max_{x\in\mathcal{A}} \ \mathcal{L}_\ell(x; \mathbf{y}_\ell)$.

To perform the ML decoding, node ℓ needs to count the frequencies of symbols y for any $y \in \mathcal{Q}_{\mathrm{o}}$ among N received symbols. As a result, the buffer size B_1 at each intermediate node satisfies $B_1 = O(|\mathcal{Q}_{\mathrm{o}}| \log N) = O(\ln N)$. The following lemma bounds the maximum decoding error probability ϵ_ℓ of (f_ℓ, g_ℓ) for Q_ℓ, whose proof is provided in Appendix C.

Lemma 8. *For any $\ell = 1, \dots, L$, under the condition $|\mathcal{Q}_{\mathrm{i}}^{\ell}| \geq |\mathcal{A}| \geq 2$, using the repetition encoding f_ℓ and the ML decoding g_ℓ in* (30) *and* (31)*, respectively, the maximum decoding error probability ϵ_ℓ for Q_ℓ satisfies $\epsilon_\ell \leq \big(|\mathcal{A}| - 1\big) \exp\left(-N E_\ell\right)$, where $E_\ell > 0$ is a constant depends only on the channel Q_ℓ.*

Based on the above lemma, the following theorem establishes the achievable rate of repetition recoding scheme. Consider a sequence of DMCs $Q_\ell, \ell = 1, 2, \dots$ with $C(Q_\ell) > 0, \ell \geq 1$ and

$$E^* \triangleq \inf\{E_\ell, \ell \geq 1\} > 0, \quad S^* \triangleq \inf\{|\mathcal{Q}_{\mathrm{i}}^{\ell}| : \ \ell \geq 1\} \geq 2. \tag{33}$$

We choose the alphabet $\mathcal{A}$ such that $|\mathcal{A}| \in [2, S^*]$.

Theorem 9. *Assume N is chosen such that $\left(1 - e^{-NE^*}\right)^L \geq |\mathcal{A}|^{-1}$. For the line network of length L, where the ℓ-th link is Q_ℓ, the GBNC with repetition recoding scheme, batch size $M = 1$, inner blocklength N, and batch alphabet $\mathcal{A}$ achieves rate*

$$C_L(1, N) \geq \frac{1}{N} \left\{ \log|\mathcal{A}| - \left(1 - \left(1 - e^{-NE^*}\right)^L\right) \log(|\mathcal{A}| - 1) - \mathcal{H}\left(\left(1 - e^{-NE^*}\right)^L\right) \right\}, \tag{34}$$

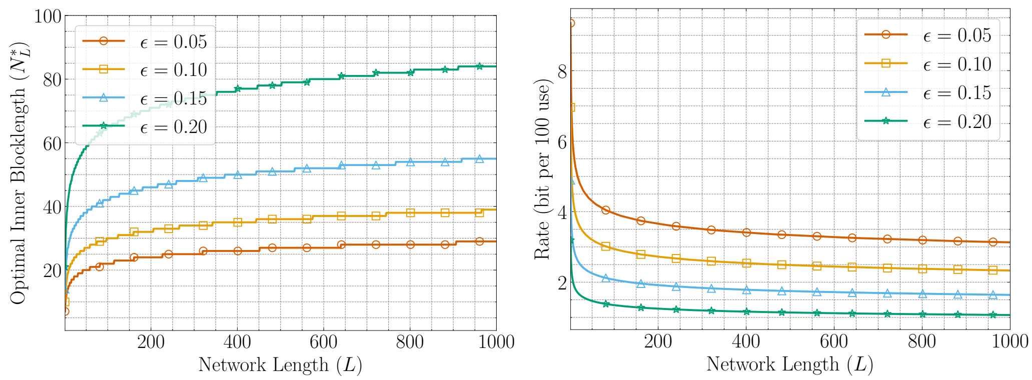

Fig. 5. Numerical illustrations of the achievable rates of GBNC based on repetition recoding (see (34)) using BSC with crossover probability $\epsilon = 0.2, |\mathcal{A}| = 2$. Figures from left to right correspond to (a) plot of the optimal value of N, denoted by N_L^*, that maximizes the lower bound of $C_L(1, N)$ in (34); and (b) plot of the achievable rates with optimized inner blocklength N and fixed batch size $M = 1$.

where $\mathcal{H}(\cdot)$ denotes the binary entropy function. When specifying $N = O(\ln L)$ and $B = O(\ln N) = O(\ln \ln L)$, it holds that $C_L^* = \Omega(1/\ln L)$.

We plot the rate of repetition recoding using BSC with crossover error probability $\epsilon \in \{0.05, 0.1, 0.15, 0.2\}$ and $|\mathcal{A}| = 2$ with respect to the hop length L in Fig. 5. In Fig.5(a), for each hop length L, we plot the optimal value of N maximizing the lower bound (34), which is denoted as N_L^*. This illustration highlights the observed trend of N_L^* increasing roughly in the order of $\ln L$. In Fig. 5(b), we plot the lower bound (34) for each hop length L, showcasing an approximate decrease rate in the order of $1/\ln L$.

Proof of Theorem 9: As $C(Q_\ell) > 0$, $|\mathcal{Q}_{\mathrm{i}}^\ell| \geq 2$. We consider a GBNC with $M = 1$ and $\mathcal{A}$ such that $\inf\{|\mathcal{Q}_{\mathrm{i}}^\ell|, \ell \geq 1\} \geq |\mathcal{A}| \geq 2$. Let $E^* = \inf\{E_\ell, \ell \geq 1\}$. By the condition of the theorem, $E^* > 0$, and hence $P(g_L(\mathbf{Y}_L) = \mathbf{X}) \geq \left(1 - e^{-NE^*}\right)^L$. Applying an argument in [13, Theorem V.3], we obtain the desired lower bound (34). When $N = O(\ln L)$ is chosen properly, we have $C_L(1, N) = \Omega(1/\ln L)$. ■

Though in the proof of the above theorem, we use GBNC with $M = 1$, the repetition coding scheme can be extended to $M > 1$. Fix integers $M = O(1)$ and $\tilde{N} = O(\ln L)$. For an integer m, let $\mathcal{Q}_{\mathrm{i}}^{m,\ell}$ be the maximum subset of $\mathcal{Q}_{\mathrm{i}}^m$ such that for any $\mathbf{x} \neq \mathbf{x}' \in \mathcal{Q}_{\mathrm{i}}^{m,\ell}$, $Q_\ell^{\otimes m}(\cdot|\mathbf{x}) \neq Q_\ell^{\otimes m}(\cdot|\mathbf{x}')$. Following the similar notations as in (33), we define

$$E^{m,*} = \inf\{E_\ell^m, \ell \geq 1\}, \quad S^{m,*} = \inf\{|\mathcal{Q}_{\mathrm{i}}^{m,\ell}| : \ \ell \geq 1\} \geq 2, \tag{35}$$

with E_ℓ^m being the random coding error exponent for $Q_\ell^{\otimes m}$. Assume that $(1-e^{-NE^{m,*}})^L \geq |\mathcal{A}|^{-M}$. Fix $m = O(1)$ and a finite alphabet $\mathcal{A}$ such that $|\mathcal{A}|^M \in [2, S^{m,*}]$. Conseuqently, we can view the line network of channels $Q_1, \ldots, Q_L$ as one of $Q_1^{\otimes m}, \ldots, Q_L^{\otimes m}$. For the latter, we can apply the repetition recoding with batch size 1, inner blocklength $\tilde{N}$ and the batch alphabet $\mathcal{A}^M$, which for the original line network of $Q_1, \ldots, Q_L$ is a GBNC with batch size M, inner blocklength $m\tilde{N}$ and the batch alphabet $\mathcal{A}$. Based on Theorem 9, such a coding scheme achieves rate

$$\sup_N \frac{1}{N}\left\{\log|\mathcal{A}|^M - \left(1-\left(1-e^{-NE^{m,*}}\right)^L\right)\log(|\mathcal{A}|^M-1) - \mathcal{H}\left(\left(1-e^{-NE^{m,*}}\right)^L\right)\right\}. \quad (36)$$

While the repetition code may appear straightforward, it serves as an illustrative example of how to reduce the buffer size at the intermediate node. Using convolutional codes with Viterbi decoding, due to their analogous encoding and decoding nature, can achieve the same order of the buffer size. However, the corresponding achievable rate is challenging to analyze.

C. Channel Reduction

When all the links in the line network are identical DMCs, it has been shown in [13] that an exponentially decreasing rate can be achieved using $N = O(1)$, which corresponds to the first case in Theorem 6. Here we discuss how to generalize this scalability result to line networks where the DMCs Q_ℓ are not necessarily identical. Our approach is to perform recoding so that the line network is reduced to one with identical channels.

We introduce the reduction of an $m \times n$ stochastic matrix Q with $C(Q) > 0$. Let $r = \mathrm{rank}(Q)$. Note that $C(Q) > 0$ if and only if $r \geq 2$. Let s be an integer such that $2 \leq s \leq r$. We would like to reduce Q by multiplying an $s \times m$ matrix R and an $n \times s$ matrix S before and after Q, respectively, so that RQS becomes an $s \times s$ matrix $U_s(\varrho)$ with $(U_s(\varrho))_{i,j} = \varrho$ if $i = j$ and otherwise $(U_s(\varrho))_{i,j} = \frac{1-\varrho}{s-1}$, where ϱ is a parameter in the range $(1/s, 1]$. When $1/s < \varrho \leq 1$, among all the $s \times s$ stochastic matrices with trace $s\varrho$, $U_s(\varrho)$ is the one that has the least mutual information for the uniform input distribution (ref. [13, Theorem V.3]). The reduction described above, if exists, is called *uniform reduction*.

We give an example of uniform reduction with $s = 2$. Choose R so that RQ is an s-row matrix formed by s linearly independent rows of Q. Let a_{ij} be the (i, j) entry of RQ, where

$i = 1,2$ and $1 \leq j \leq n$. Define an $n \times 2$ stochastic matrix $W = (w_{ij})$ as

$$w_{i1} = \begin{cases} \frac{a_{1i}}{a_{1i}+a_{2i}} & \text{if } a_{1i} + a_{2i} > 0, \\ 1 & \text{otherwise}, \end{cases} \tag{37}$$

and $w_{i2} = 1 - w_{i1}$, where $1 \leq i \leq n$. With the above R and W, we see that $RQW = U_2(\varrho)$, where $\varrho = \sum_{k:a_{1k}+a_{2k}>0} \frac{a_{1k}^2}{a_{1k}+a_{2k}}$. The following lemma, proved in Appendix C, states a range of ϱ such that the reduction to $U_2(\varrho)$ is feasible.

Lemma 10. *For a stochastic matrix Q such that $C(Q) > \epsilon$ for some $\epsilon > 0$, there exists a constant $B > 1/2$ depending only on ϵ such that Q has a uniform reduction to $U_2(\varrho)$ for all $1/2 < \varrho \leq B$.*

Fix any $\epsilon > 0$. Consider the line network formed by $Q_1, \ldots, Q_L$, where $C(Q_\ell) > \epsilon$ and hence $\operatorname{rank}(Q_\ell) \geq 2$. We discuss a GBNC with $|\mathcal{A}| = 2$ and $M = N = 1$. By Lemma 10, there exists $\varrho > 1/2$ such that for any ℓ, there exists stochastic matrices R_ℓ and S_ℓ such that $R_\ell Q_\ell S_\ell = U_2(\varrho)$. Define the recoding at the source node as R_1, and for $\ell = 1, \ldots, L-1$, define the recoding at node ℓ as $S_\ell R_{\ell+1}$. At the destination node, process all the received batches by R_L. The overall operation of a batch from the source node to the destination node is $W'_L \triangleq (U_2(\varrho))^L$. Applying the argument in [13, Theorem III.5], we get

$$\log\left(\frac{1}{2\varrho - 1}\right) \leq \liminf_{L\to\infty} -\frac{1}{L}\log C(W'_L) \leq \limsup_{L\to\infty} -\frac{1}{L}\log C(W'_L) \leq 2\log\left(\frac{1}{2\varrho-1}\right), \tag{38}$$

where $\frac{1}{2\varrho-1}$ is the second largest eigenvalue of $U_2(\varrho)$. Therefore, a channel code for the transition matrix W_L as the outer code can achieve the rate $\Omega(e^{-cL})$ as $L \to \infty$, where the constant c is between $\log\left(\frac{1}{2\varrho-1}\right)$ and $2\log\left(\frac{1}{2\varrho-1}\right)$. The above discussion is summarized as the following theorem:

Theorem 11. *Consider a sequence of DMCs $Q_\ell, \ell = 1, 2, \ldots$ with $\inf\{C(Q_\ell), \ell \geq 1\} > 0$. For the line network of length L, where the ℓ-th link is Q_ℓ, the GBNC with $M = O(1)$ and $N = O(1)$ achieves rate $C_L^* \geq c' \cdot e^{-cL}$, where c is a constant between $\log\left(\frac{1}{2\varrho-1}\right)$ and $2\log\left(\frac{1}{2\varrho-1}\right)$, and $c' > 0$ is a constant.*

The technique used in the proof of Theorem 11 can be generalized for $M, N \geq 1$. We first show that for an $m \times n$ stochastic matrix Q with $\operatorname{rank}(Q) \geq 2$, for any $2 \leq s \leq r$, the uniform reduction to $U_s(\varrho)$ exists if ϱ is sufficiently close to $1/s$. For an integer $2 \le s \le r$, let

$$\kappa_s(Q) = \max_{\substack{s\times m \text{ stochastic matrix } R\\ n\times s \text{ stochastic matrix } W}} \min \mathrm{inv}(RQW) \tag{39}$$

where $\min \mathrm{inv}(RQW)$ is the minimum value of $(RQW)^{-1}$ when RQW is invertible and is ∞ otherwise. We give an example of R and W such that RQW is invertible. Choose R so that RQ is an s-row matrix formed by s linearly independent rows of Q. Let a_{ij} be the (i,j) entry of RQ, where $1 \le i \le s$ and $1 \le j \le n$. To simplify the discussion, we assume all the columns of RQ are non-zero. Define $W = D(RQ)^\top$, where D is an $n \times n$ diagonal matrix with the (i,i) entry $1/\sum_{j'} a_{j'i}$. With the above R and W, we see that RQW is positive definite and hence invertible. Let $\rho_s(Q) = \frac{\min\{\kappa_s(Q),0\}-1}{s\min\{\kappa_s(Q),0\}-1}$. We see that $\rho_s(Q) > 1/s$. The following lemma, proved in Appendix C, states a range of ϱ such that the reduction to $U_s(\varrho)$ is feasible.

Lemma 12. *Consider an $m \times n$ stochastic matrix Q with rank $r \ge 2$. For any $2 \le s \le r$ and $1/s < \varrho \le \rho_s(Q)$, there exist an $s \times m$ stochastic matrix R and an $n \times s$ stochastic matrix S such that $RQS = U_s(\varrho)$.*

Remark 2. Lemma 10 is stronger than Lemma 12 for the case $s = 2$ as the former gives a uniform bound B that does not depend on Q as long as $C(Q) > \epsilon$.

Consider a line network formed by $Q_1, \dots, Q_L$, where $C(Q_\ell) > 0$ and hence $\mathrm{rank}(Q_\ell) \ge 2$. Let $r = \min_{\ell=1}^L \mathrm{rank}(Q_\ell)$. Assuming $r \ge |\mathcal{A}|$, we first discuss a recoding scheme with $M = N = 1$. Let $\varrho = \min_{\ell=1}^L \rho_r(Q_\ell)$. By Lemma 12, there exists stochastic matrices R_ℓ and S_ℓ such that $R_\ell Q_\ell S_\ell = U_r(\varrho)$. The following argument is similar as that of the proof of Theorem 11. Now we consider recoding with $M, N = O(1)$. Fix $M, N = O(1)$ and a finite alphabet $\mathcal{A}$ such that $r^N \ge |\mathcal{A}|^M$. Regarding the line network $\mathcal{L}$ as one formed by $Q_1^{\otimes N}, \dots, Q_L^{\otimes N}$, we can apply the above GBNC with batch size 1, inner blocklength 1 and the batch alphabet $\mathcal{A}^M$, which for the original line network $\mathcal{L}$ of $Q_1, \dots, Q_L$ is a GBNC with batch size M, inner blocklength N and the batch alphabet $\mathcal{A}$.

V. Line Networks of Packet Erasure Channels

Batched network codes for line networks of *packet erasure channels* have been studied as efficient variations of random linear network coding [17]–[24]. In this section, we discuss line networks with identical packet erasure channels, for which, we demonstrate stronger converse and achievability results than the general ones.

Fix the alphabet $\mathcal{A}$ with $|\mathcal{A}| \geq 2$. Suppose that the input alphabet $\mathcal{Q}_\mathrm{i}$ and the output alphabet $\mathcal{Q}_\mathrm{o}$ are both $\mathcal{A} \cup \{e\}$ where $e \notin \mathcal{A}$ is called the erasure. For example, we may use a sequence of bits to represent a packet so that $\mathcal{A} = \{0,1\}^T$, i.e., each packet is a sequence of T bits. Henceforth, a symbol in $\mathcal{A}$ is also called a packet in this section.

A packet erasure channel with erasure probability ϵ $(0 < \epsilon < 1)$ has the transition matrix Q_era: for each $x \in \mathcal{A}$, $Q_\text{era}(y|x) = 1 - \epsilon$ if $y = x$ and $Q_\text{era}(y|x) = \epsilon$ if $y = e$. The input e can be used to model the input when the channel is not used for transmission and we define $Q_\text{era}(e|e) = 1$. When the input e is not used for encoding information, erasure codes can achieve a rate of $1-\epsilon$ symbols (in $\mathcal{A}$) per use. It is also clear that $C_0(Q_\text{era}) = 0$.

A. Upper Bound

We follow the argument in Sec. III-A to obtain the upper bound. It is worth noting that for this special channel, the upper bound in Lemma 2 can be tightened as $p_1 = P(\overline{E_0}) = (1-\epsilon^N)^L$. Based on this observation and following a similar procedure as in the proof of Theorem 4, we have

$$C_L(M,N) \leq \frac{(1-\epsilon^N)^L}{N} \min\{M \log|\mathcal{A}|, N \log|\mathcal{Q}_\mathrm{o}|\}, \tag{40}$$

which is a tighter upper bound than (23).

B. Achievability by Random Linear Recoding

We now introduce a class of inner codes with batch size $M = O(1)$, which provides the achievability counterpart for the cases 1) and 2) in Theorem 4. Let $\mathbb{F}_q$ be the finite field of q symbols, and let $T > 0$ be an integer. Suppose $\mathcal{A} = \mathbb{F}_q^T$, i.e., each packet is a sequence of T symbols from the finite field $\mathbb{F}_q$. The outer code generates batches that consist of M packets in $\mathcal{A}$, and can be represented as a $T \times M$ matrix over $\mathbb{F}_q$. In each packet generated by the outer code, the first M symbols in $\mathbb{F}_q$ are called the *coefficient vector*. A batch $\mathbf{X}$ has the first M rows, called the *coefficient matrix*, forming the identity matrix. In the following discussion, we treat the erasure e as the all-zero vector $\mathbf{0}$ in $\mathbb{F}_q^T$, which is not used as a packet in the batches. In other words, when a packet is erased, an intermediate node assumes $\mathbf{0}$ is received.

The inner code is formed by *random linear recoding*, which have been studied in [17]–[24]. A random linear combination of vectors in $\mathcal{A}$ has the linear combination coefficients chosen uniformly at random from $\mathbb{F}_q$. The inner code includes the following operations:

- The source node generates N packets for a batch using random linear combinations of the M packets of the batch generated by the outer code.
- Each intermediate node generates N packets for a batch using random linear combinations of all packets of the received packets of the batch.

Note that for each batch, only the packets with linearly independent coefficient vectors are needed for random linear recoding. Therefore, the buffer size used to store batch content is $B_1 = O(MT \log q)$ bits. Also, the computational overlead for each intermediate node is $O(N^2 T \log q)$.

At each node, the rank of the coefficient matrix of a batch (i.e., the first M rows of the matrix formed by the generated/received packets of the batch) is also called the rank of the batch. At each node, the ranks of all the batches follow an identical and independent distribution. Denote by π_ℓ the rank distribution of a batch at node ℓ. As all the batches at the source node have rank M, we know that $\pi_0 = (0, 0, \ldots, 0, 1)$. Moreover, the rank distributions $\pi_0, \pi_1, \ldots, \pi_L$ form a Markov chain so that for $\ell = 1, \ldots, L$, it holds that

$$\pi_\ell = \pi_{\ell-1}\mathbf{P} \tag{41}$$

where $\mathbf{P}$ is the transition matrix characterized in [15, Lemma 4.2] (see (77) in Appendix C for the formula).

The maximum achievable rate of this class of GBNCs is $(1 - \frac{M}{T})\frac{\mathbf{E}[\pi_L]}{N}$ packets (in $\mathcal{A}$) per use, and can be achieved by BATS codes [15], [31], where the factor $1 - \frac{M}{T}$ comes from the overhead of M symbols in a packet used to transmit the coefficient vector. Denote

$$\text{BATS}_L(M, N) = \left(1 - \frac{M}{T}\right) \frac{\mathbf{E}[\pi_L]}{N} \log |\mathcal{A}|. \tag{42}$$

In Fig. 6, we compare numerically the upper bound and the achievable rates of BATS codes by evaluating (40) and (42), respectively. Throughout the experiment, we specify parameters $\epsilon = 0.2$, $q = 256$ and $T = 1024$ following the same setup as in [22, Fig. 10]. Specifying $q = 256, T = 1024$ leads to random linear network coding on a packet of T symbols on the finite field $\mathbb{F}_q$, which can be implemented efficiently. Specifying $\epsilon = 0.2$ ensures the multi-hop communication consisting of 1000 hops is a challenging task. Note that each packet has 8192 bits and the min-cut is 6553.6 bits per use.

- First, we consider fixed $M = N = 2, 3, 4$, and plot the calculation for L up to 1000 in Fig. 6(a). We see from the figure that for a fixed N, the achievable rates of BATS codes and the upper bound in (40) share the same exponential decreasing trend.

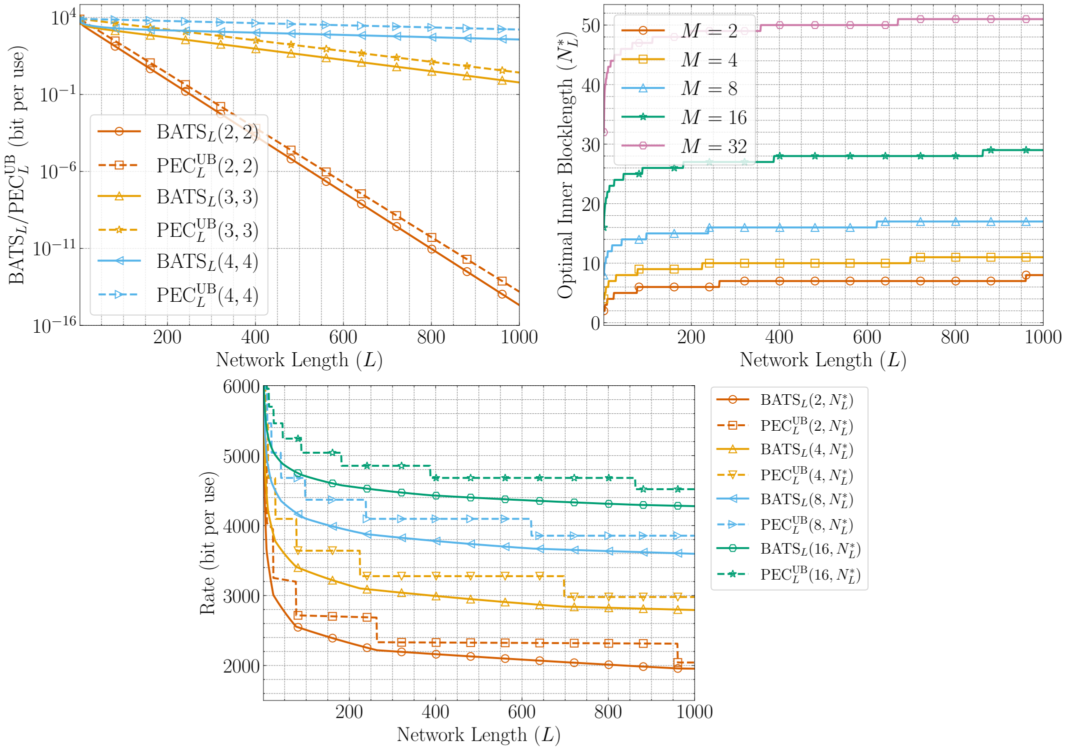

Fig. 6. Numerical illustrations of the upper bound and achievable rates of BATS codes. Figures from left to right correspond to (a) plot of $\mathrm{BATS}_L(M, N)$ and $\mathrm{PEC}_L^{\mathrm{UB}}(M, N)$ when L increases for the case $M = N$ fixed; (b) plot of the optimal value of N, denoted by N_L^*, that maximizes $\mathrm{BATS}_L(M, N)$ for a fixed value of M; and (c) plot of $\mathrm{BATS}_L(M, N)$ and $\mathrm{PEC}_L^{\mathrm{UB}}(M, N)$ when L increases for the case that M is fixed and $N = N_L^*$.

- Second, we consider fixed $M = 2, 4, 8, 16, 32$. For each value of M, we find the optimal value of N, denoted by N_L^*, that maximizes $\mathrm{BATS}_L(M, N)$. We see from Fig. 6(b) that N_L^* demonstrates a low increasing rate with L. We further illustrate $\mathrm{BATS}_L(M, N_L^*)$ and $\mathrm{PEC}_L^{\mathrm{UB}}(M, N_L^*)$ for each value of M in Fig. 6(c).

The following theorem, proved in Appendix C, justifies the scalability of $\mathrm{BATS}_L(M, N)$ when L is large, where the $M = 1$ case was proved in [15].

Theorem 13. *Consider a line network of L packet erasure channels with erasure probability ϵ. For GBNCs of fixed batch size $M < T$ and inner blocklength N using random linear recoding,*

$$\mathit{BATS}_L(M, N) = \Theta\left(\frac{\left(1 - (\epsilon + (1-\epsilon)/q)^N\right)^L}{N}\right).$$

When q is relatively large, $\mathrm{BATS}_L(M, N)$ has nearly the same scalability as $\mathrm{PEC}_L^{\mathrm{UB}}(M, N)$, as illustrated by Fig. 6(c). Consider two cases of N for the scalability of $\mathrm{BATS}_L(M, N)$ and buffer size:

- When N is a fixed number, $\text{BATS}_L(M,N)$ decreases exponentially with L, and the buffer size for the counter $B_2 = O(1)$.
- When M is a fixed number and N is unconstrained, based on the optimization theory (see, e.g., [1, Lemma 1]) we know that $\max_N \text{BATS}_L(M,N) = \Theta(1/\ln L)$, and the maximum is achieved by $N = \Theta(\ln L)$. Therefore, the buffer size for the counter $B_2 = O(\ln\ln L)$.

VI. Line Networks with Channels of Positive Zero-Error Capacity

In this section, we study the capacity scalability of line networks of channels that have positive capacity but may also have positive zero-error capacity. Denote by $\mathcal{L}(L)$ a line network of length L formed by channels $Q_1,\dots,Q_L$, where it is not necessary that $C_0(Q_\ell)=0$. Recall that for a GBNC with batch size M and inner blocklength N, the end-to-end transition matrix of a batch for $\mathcal{L}(L)$ with recoding operations F and $\{\Phi_i\}_{i=1}^{L-1}$ is

$$W_L = FQ_1^{\otimes N}\Phi_1 Q_2^{\otimes N}\Phi_2\cdots\Phi_{L-1}Q_L^{\otimes N}. \tag{43}$$

Denote the maximum achievable rate of all recoding schemes with batch size M and inner blocklength N for $\mathcal{L}(L)$ as $C_{\mathcal{L}(L)}(M,N)$. Let $L_0(L)$ be the number of channels in $\mathcal{L}(L)$ with 0 zero-error capacity, i.e., $L_0(L) = |\{1\le \ell\le L : C_0(Q_\ell)=0\}|$. Let $\{l_1,\dots,l_{L_0(L)}\} = \{1\le\ell\le L : C_0(Q_\ell)=0\}$ where $l_1<l_2<\cdots<l_{L_0(L)}$. In the following, we argue that $C_{\mathcal{L}(L)}(M,N)$ scales like a line network of length $L_0(L)$ formed by channels with 0 zero-error capacity.

A. Upper Bound

Denote by $\mathcal{L}'(L_0(L))$ the line network formed by the concatenation of $Q_{l_1},\dots,Q_{L_0(L)}$, each of which has 0 zero-error capacity in $\mathcal{L}(L)$. For a GBNC with batch size M and inner blocklength N, the end-to-end transition matrix of a batch for $\mathcal{L}'(L_0(L))$ with recoding operations F' and $\{\Phi'_i\}_{i=1}^{L_0(L)-1}$ is $W'_{L_0(L)} = F'Q_{l_1}^{\otimes N}\Phi'_1 Q_{l_2}^{\otimes N}\Phi'_2\cdots\Phi'_{L_0(L)-1}Q_{l_{L_0(L)}}^{\otimes N}$. Denote the maximum achievable rate of all recoding schemes with batch size M and inner blocklength N for $\mathcal{L}'(L_0(L))$ as $C_{\mathcal{L}'(L_0(L))}(M,N)$.

Notice that $W'_{L_0(L)} = W_L$ when proper recoding operations F' and $\{\Phi'_i\}_{i=1}^{L_0(L)-1}$ are selected, and hence $C_{\mathcal{L}(L)}(M,N)\le C_{\mathcal{L}'(L_0(L))}(M,N)$. For network $\mathcal{L}'(L_0(L))$, Sec. III provides the upper bounds on the achievable rates as functions of length $L_0(L)$ under certain coding parameter sets, which are also upper bounds for network $\mathcal{L}(L)$.

B. Lower Bound

We derive a lower bound of achievable rates of $\mathcal{L}(L)$ using the uniform reduction approach introduced in Sec. IV-C. Suppose $C = \inf\{C(Q_{l_i}) :\ i \geq 1\} > 0$. For Q_{l_i}, $i = 1, \dots, L_0(L)$, we know $C_0(Q_{l_i}) = 0$. By Lemma 10, there exists a constant $B \in (1/2, 1)$ depending only on C such that there exist stochastic matrices R_{l_i} and S_{l_i} with $R_{l_i} Q_{l_i} S_{l_i} = U_2(B)$ for all i. For Q_ℓ with $C_0(Q_\ell) > 0$, we can find R_ℓ and S_ℓ so that $R_\ell Q_\ell S_\ell$ equals the identity matrix I_2. The existence of R_ℓ and S_ℓ is guaranteed by the following lemma, whose proof is provided in Appendix C.

Lemma 14. *For an $m \times n$ stochastic matrix Q with $C_0(Q) > 0$, there exists a $2 \times m$ stochastic matrix R and a $n \times 2$ stochastic matrix S such that $RQS = I_2$, the 2×2 identity matrix.*

Denote by $\mathcal{L}''(L_0(L))$ the line network formed by the concatenation of $L_0(L)$ identical channels $U_2(B)$. For a GBNC with batch size M and inner blocklength N, the end-to-end transition matrix of a batch for $\mathcal{L}''(L_0(L))$ with recoding operations F'' and $\{\Phi''_i\}_{i=1}^{L_0(L)-1}$ is

$$\begin{aligned} W''_{L_0(L)} &= F'' U_2(B)^{\otimes N} \Phi''_{l_1} U_2(B)^{\otimes N} \Phi''_{l_2} \cdots \Phi''_{l_{L_0(L)-1}} U_2(B)^{\otimes N} \\ &= F'' R_{l_1}^{\otimes N} Q_{l_1}^{\otimes N} S_{l_1}^{\otimes N} \Phi''_{l_1} R_{l_2}^{\otimes N} Q_{l_2}^{\otimes N} S_{l_2}^{\otimes N} \Phi''_{l_2} \cdots \Phi''_{l_{L_0(L)-1}} R_{l_{L_0(L)}}^{\otimes N} Q_{l_{L_0(L)}}^{\otimes N} S_{l_{L_0(L)}}^{\otimes N}. \end{aligned}$$

Denote the maximum achievable rate of all recoding schemes with batch size M and inner blocklength N for $\mathcal{L}''(L_0(L))$ as $C_{\mathcal{L}''(L_0(L))}(M, N)$.

In addition to the recoding operations F and $\{\Phi_i\}_{i=1}^{L-1}$ as used in (43), deploying an extra recoding operation Φ_L at the destination node, the end-to-end transition matrix of a batch for $\mathcal{L}(L)$ becomes $W_L^* = W_L \Phi_L$. By properly choosing F, $\Phi_1, \dots, \Phi_L$, W_L^* becomes $W''_{L_0(L)}$:

- Let $F = F'' R_1^{\otimes N}$ and $\Phi_L = S_L^{\otimes N}$;
- If $L_0(L) \neq L$, let $\Phi_{L_0(L)} = S_{L_0(L)}^{\otimes N} R_{L_0(L)+1}^{\otimes N}$.
- For $\ell = 1, \dots, L-1$ and $\ell \neq L_0(L)$, if $C_0(Q_\ell) > 0$, let $\Phi_\ell = S_\ell^{\otimes N} R_{\ell+1}^{\otimes N}$; if $C_0(Q_\ell) = 0$, let $\Phi_\ell = S_\ell^{\otimes N} \Phi''_\ell R_{\ell+1}^{\otimes N}$.

Hence, we obtain $C_{\mathcal{L}(L)}(M, N) \geq C_{\mathcal{L}''(L_0(L))}(M, N)$, where the later can be lower bounded by the techniques in Sec. IV. In particular, the error exponent condition in Theorem 9 can be verified by checking the proof of Lemma 8 for the special case of BSCs.

VII. Concluding Remarks

This paper studied the achievable rates of generalized batched network codes (GBNCs) for line networks of general discrete memoryless channels (DMCs) with buffer size and latency constraints. It provides some new guidelines for designing multi-hop communication networks when the number of hops is large. First, it may not be necessary to consider only the capacity achieving codes for the DMCs to form the inner code. Some simple codes like repetition and convolutional codes have the advantage of lower buffer size while preserving the same achievable rate order. Second, hop-by-hop communication is not necessarily reliable for communications with buffer size constraints. GBNCs require that the inner code is designed to maximize the capacity of the end-to-end batch channel.

In the future, it is desirable to study better recoding schemes for line networks of special channels, such as BSCs. Generalization of our analysis for channels with infinite alphabets and continuous channels is of research interest. Last, we are curious whether our upper bound holds for more general codes than GBNCs.

References

[1] S. Yang, J. Wang, Y. Dong, and Y. Zhang, "On the capacity scalability of line networks with buffer size constraints," in *Proc. IEEE ISIT '19*. IEEE, 2019, pp. 1507–1511.

[2] S. Yang and J. Wang, "Upper bound scalability on achievable rates of batched codes for line networks," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1629–1634.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc, 2006.

[4] V. Sevani, B. Raman, and P. Joshi, "Implementation-based evaluation of a full-fledged multihop tdma-mac for wifi mesh networks," *IEEE transactions on mobile computing*, vol. 13, no. 2, pp. 392–406, 2012.

[5] W. Zhang, M. Stojanovic, and U. Mitra, "Analysis of a linear multihop underwater acoustic network," *IEEE Journal of Oceanic Engineering*, vol. 35, no. 4, pp. 961–970, 2010.

[6] Z. Huakai, D. Guangliang, and L. Haitao, "Simplified bats codes for deep space multihop networks," in *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*. IEEE, 2016, pp. 311–314.

[7] 3GPP, "Release 16," 2020.

[8] 3GPP, "Release 17," 2022.

[9] C. Duan, J. Feng, H. Chang, B. Song, and Z. Xu, "A novel handover control strategy combined with multi-hop routing in leo satellite networks," in *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE, 2018, pp. 845–851.

[10] S. Yang, J. Ma, and X. Huang, "Multi-hop underwater acoustic networks based on BATS codes," in *Proceedings of WUWNet '18*, Shenzhen, China, 2018.

[11] S. Yang, H. H. Yin, R. W. Yeung, X. Xiong, Y. Huang, L. Ma, M. Li, and C. Tang, "On scalable network communication for infrastructure-vehicle collaborative autonomous driving," *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 310–324, 2022.

[12] B. Soret, S. Ravikanti, and P. Popovski, "Latency and timeliness in multi-hop satellite networks," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[13] U. Niesen, C. Fragouli, and D. Tuninetti, "On capacity of line networks," *Information Theory, IEEE Transactions on*, vol. 53, no. 11, pp. 4039–4058, Nov. 2007.

[14] B. N. Vellambi, N. Torabkhani, and F. Fekri, "Throughput and latency in finite-buffer line networks," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3622–3643, June 2011.

[15] S. Yang and R. W. Yeung, *BATS Codes: Theory and Practice*, ser. Synthesis Lectures on Communication Networks. Morgan & Claypool Publishers, 2017.

[16] A. E. Gammal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

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

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

[19] 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.

[20] S. Yang and R. W. Yeung, "Coding for a network coded fountain," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, Saint Petersburg, Russia, July 31 - Aug. 5 2011, pp. 2647–2651.

[21] K. Mahdaviani, M. Ardakani, H. Bagheri, and C. Tellambura, "Gamma codes: A low-overhead linear-complexity network coding solution," in *Network Coding (NetCod), 2012 International Symposium on*, June 2012, pp. 125–130.

[22] S. Yang and R. W. Yeung, "Batched sparse codes," *Information Theory, IEEE Transactions on*, vol. 60, no. 9, pp. 5322–5346, Sep. 2014.

[23] B. Tang, S. Yang, B. Ye, Y. Yin, and S. Lu, "Expander chunked codes," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–13, 2015. [Online]. Available: http://dx.doi.org/10.1186/s13634-015-0297-8

[24] B. Tang and S. Yang, "An LDPC approach for chunked network codes," *IEEE/ACM Transactions on Networking*, vol. 26, no. 1, pp. 605–617, 2018.

[25] S. Yang and R. W. Yeung, "Network communication protocol design from the perspective of batched network coding," *IEEE Communications Magazine*, vol. 60, no. 1, pp. 89–93, 2022.

[26] J. Wang, Z. Jia, H. H. F. Yin, and S. Yang, "Small-sample inferred adaptive recoding for batched network coding," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 1427–1432.

[27] H. H. Yin, B. Tang, K. H. Ng, S. Yang, X. Wang, and Q. Zhou, "A unified adaptive recoding framework for batched network coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 4, pp. 1150–1164, 2021.

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

[29] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[30] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, Inc, 1968.

[31] S. Yang, J. Meng, and E.-h. Yang, "Coding for linear operator channels over finite fields," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, Austin, USA, June 13-18 2010, pp. 2413–2417.

[32] Z. Zhou, C. Li, S. Yang, and X. Guang, "Practical inner codes for batched sparse codes in wireless multi-hop networks," *IEEE Transactions on Vehicular Technology*, vol. 1, no. 1, pp. 1–1, 2019.

Supplementary Material for "*On Achievable Rates of Line Networks with Generalized Batched Network Coding*"

Appendix A
Nomenclature

TABLE II

Some notations used in the paper, listed in the alphabetical order.

Notation	Explanation
$\mathcal{A}$	Batch alphabet.
B	Buffer size.
$\mathbf{B}_\ell$	Buffer content at node ℓ.
$C(Q)$	Channel capacity of channel Q.
$C_0(Q)$	Zero-error capacity of channel Q.
$C_L(M,N)$	Maximum achievable rate of all recoding schemes with batch size M and inner blocklength N.
$E_{0,\ell}$	Event that all N outputs of Q_ℓ are equal to the same value regardless of channel input.
E_0	Event that there exists one link ℓ such that $E_{0,\ell}$ holds.
Er_ℓ	Coding error exponent for channel Q_ℓ
Er^*	Smallest coding error exponent among all $\ell \geq 1$.
ℓ	Index of node/channel.
L	Network length.
M	Batch size.
N	Inner blocklength.
Q	Discrete memoryless channel.
$Q^{\otimes N}$	Discrete memoryless channel with N channel uses.
$\mathcal{Q}_\mathrm{i}/\mathcal{Q}_\mathrm{o}$	Input/output alphabets of Q.
$\mathbf{U}_\ell/\mathbf{Y}_\ell$	The input/output of N uses of the ℓ-th communication link.
W_L	End-to-end transition matrix of the batch channel from $\mathbf{X}$ to $\mathbf{Y}_L$.
$\mathbf{X} \in \mathcal{A}^M$	A generic batch.
$\mathbf{X}[k]$	The k-th entry in $\mathbf{X}$.
$\mathbf{Z}_\ell$	Channel status of Q_ℓ.

Appendix B
Proofs about Converse

Proof of Lemma 3: Denote by $\mathbf{y}^* = (y^* \cdots y^*)$. We have

$$W(\mathbf{y}|\mathbf{x}) = \begin{cases} \frac{Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) - p_0}{1-p_0} & \mathbf{y} = \mathbf{y}^*, \\ \frac{Q^{\otimes N}(\mathbf{y}|\mathbf{x})}{1-p_0} & \text{otherwise.} \end{cases} \tag{44}$$

Let $P(\mathbf{y}) = \sum_{\mathbf{x}} Q^{\otimes N}(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ and $P'(\mathbf{y}) = \sum_{\mathbf{x}} W(\mathbf{y}|\mathbf{x})p(\mathbf{x})$. We have

$$P'(\mathbf{y}) = \begin{cases} \frac{1}{1-p_0}(P(\mathbf{y}) - p_0) & \mathbf{y} = \mathbf{y}^*, \\ \frac{1}{1-p_0}P(\mathbf{y}) & \text{otherwise.} \end{cases} \tag{45}$$

Substituting (44) and (45) into $I(p, W)$, we get

$$I(p, W) = \sum_{\mathbf{x}} p(\mathbf{x}) \sum_{\mathbf{y}} W(\mathbf{y}|\mathbf{x}) \log \frac{W(\mathbf{y}|\mathbf{x})}{P'(\mathbf{y})} \tag{46}$$

$$= \frac{1}{1-p_0} I(p, Q^{\otimes N}) + \frac{1}{1-p_0} U(\mathbf{y}^*), \tag{47}$$

where

$$U(\mathbf{y}^*) \triangleq \sum_{\mathbf{x}} p(\mathbf{x}) \left((Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) - p_0) \log \frac{Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) - p_0}{P(\mathbf{y}^*) - p_0} - Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) \log \frac{Q^{\otimes N}(\mathbf{y}^*|\mathbf{x})}{P(\mathbf{y}^*)} \right). \tag{48}$$

Using $P(\mathbf{y}^*) = \sum_{\mathbf{x}} Q^{\otimes N}(\mathbf{y}^*|\mathbf{x})p(\mathbf{x}) \geq \sum_{\mathbf{x}} \epsilon^N p(\mathbf{x}) = \epsilon^N$, we have

$$U(\mathbf{y}^*) = -p_0 \sum_{\mathbf{x}} p(\mathbf{x}) \log(Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) - p_0) + P(\mathbf{y}^*) \log \frac{P(\mathbf{y}^*)}{P(\mathbf{y}^*) - p_0} \tag{49}$$

$$+ p_0 \log(P(\mathbf{y}^*) - p_0) + \sum_{\mathbf{x}} p(\mathbf{x}) Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) \log \frac{Q^{\otimes N}(\mathbf{y}^*|\mathbf{x}) - p_0}{Q^{\otimes N}(\mathbf{y}^*|\mathbf{x})} \tag{50}$$

$$\leq -p_0 \log(\epsilon^N - p_0) + q^* \log \frac{\epsilon^N}{\epsilon^N - p_0} + p_0 \log(q^* - p_0) + q^* \log \frac{q^* - p_0}{q^*} \tag{51}$$

$$= (q^* + p_0) \log \frac{q^* - p_0}{\epsilon^N - p_0} + q^* \log \frac{\epsilon^N}{q^*} \tag{52}$$

The proof is completed by combining (47) and (52). ■

Lemma 15. *For fixed real number $0 < \epsilon < 1$ and integer $L > 1$, the function $F(N) = (1-\epsilon^N)^L/N$ of integer N is maximized when N is $\Theta(\ln L)$, and the optimal value of $F(N)$ is $\Theta\left(\frac{\ln(1/\epsilon)}{\ln L}\right)$.*

Proof: We relax N to a real number and solve $\frac{\mathrm{d}F(N)}{\mathrm{d}N} = 0$, i.e.,

$$1 - \epsilon^N + LN\epsilon^N \ln \epsilon = 0, \tag{53}$$

or

$$\epsilon^{-N} - 1 + LN \ln \epsilon = 0. \tag{54}$$

Let $t = -N \ln \epsilon$, and denote by $t^*(L)$ the solution of $g(t) \triangleq e^t - 1 - Lt = 0, t > 0$. Then the solution of (53) is $N^* = t^*(L)/\ln(1/\epsilon)$.

We know that $g(t) < 0$ for $0 < t < t^*(L)$; and $g(t) > 0$ for $t > t^*(L)$. Since $g(\ln L) = L - 1 - L \ln L < 0$ and $g(2 \ln L) = L^2 - 1 - 2L \ln L > 0$ when $L > 1$, we have $\ln l < t^*(L) < 2 \ln L$ when $L > 1$. Last, using $\epsilon^{N^*} = e^{-t^*(L)}$,

$$0.25 \leq (1 - 1/L)^L \leq (1 - \epsilon^{N^*})^L \leq \left(1 - 1/L^2\right)^L < 1, \tag{55}$$

and hence $F(N^*) = \frac{(1-\epsilon^{N^*})^L}{N^*} = \frac{\ln \frac{1}{\epsilon}(1-\epsilon^{N^*})^L}{t^*(L)} = \Theta(\frac{\ln \frac{1}{\epsilon}}{\ln L})$. ■

Proof of Theorem 4: Write

$$\begin{aligned}
I(p_{\mathbf{X}}, W_L) &\leq p_0 I(p_{\mathbf{X}}, W_L^{(0)}) + p_1 I(p_{\mathbf{X}}, W_L^{(1)}) && (56)\\
&= p_1 I(p_{\mathbf{X}}, W_L^{(1)}) && (57)\\
&\leq (1 - \varepsilon^{|\mathcal{Q}_\mathrm{i}|N})^L \min\left\{H(\mathbf{X}), \max_{p_{\mathbf{U}_L}} I(\mathbf{U}_\ell; \mathbf{Y}_\ell \mid \overline{E_{0,\ell}}), \ell = 1, \ldots, L\right\} && (58)\\
&\leq (1 - \varepsilon^{|\mathcal{Q}_\mathrm{i}|N})^L \min\{NC^*(Q_\ell, N), N \log |\mathcal{Q}_\mathrm{i}|, N \log |\mathcal{Q}_\mathrm{o}|, M \log |\mathcal{A}|\}, && (59)
\end{aligned}$$

where (56) follows from (5), (57) is obtained by applying Lemma 1, (58) follows from Lemma 2, and (59) holds due to $H(\mathbf{X}) \leq M \log |\mathcal{A}|$, $I(\mathbf{U}_\ell; \mathbf{Y}_\ell \mid \overline{E_{0,\ell}}) \leq \min(\log |\mathcal{Q}_\mathrm{i}^N|, \log |\mathcal{Q}_\mathrm{o}^N|)$, and Lemma 3.

The remainder part of the theorem is proved by analyzing the upper bound in (23) for different values of M and N. In particular, 2) is obtained using Lemma 15. ■

Proof of Lemma 5: We group the elements of $\mathcal{S}_\mathrm{i}$ into $\lceil |\mathcal{S}_\mathrm{i}|/2 \rceil$ pairs, denoted collectively as $\mathcal{S}_\mathrm{i}^{(2)}$, where each element of $\mathcal{S}_\mathrm{i}$ appears in exactly one pair. When $|\mathcal{S}_\mathrm{i}|$ is even, all pairs have distinct entries. When $|\mathcal{S}_\mathrm{i}|$ is odd, exactly one pair has the two entries same and the other pairs have distinct entries.

For each pair $(x, x') \in \mathcal{S}_\mathrm{i}^{(2)}$, fix $y_{x,x'}$ such that $Q(y_{x,x'}|x) \geq \varepsilon_Q$ and $Q(y_{x,x'}|x') \geq \varepsilon_Q$. Define $\mathcal{Z}$ as the collection of $z = (z_x, x \in \mathcal{Q}_\mathrm{i})$ such that $z_x = y_{x,x'}$ and $z_{x'} = y_{x,x'}$ for all pairs

$(x,x') \in \mathcal{S}_\mathrm{i}^{(2)}$. Let $\mathcal{S}_\mathrm{o} = \{y_{x,x'} : (x,x') \in \mathcal{S}_\mathrm{i}^{(2)}\}$. Therefore, $|\mathcal{S}_\mathrm{o}| \leq \lceil |\mathcal{S}_\mathrm{i}|/2 \rceil$. Hence for any $x \in \mathcal{S}_\mathrm{i}$ and $z \in \mathcal{Z}$, $\alpha(x,z) = z_x \in \mathcal{S}_\mathrm{o}$. When $\mathcal{A}$ is even,

$$P(Z \in \mathcal{Z}) = \prod_{(x,x') \in \mathcal{S}_\mathrm{i}^{(2)}} P(Z[x] = y_{x,x'}) P(Z_{x'} = y_{x,x'}) \tag{60}$$

$$= \prod_{(x,x') \in \mathcal{S}_\mathrm{i}^{(2)}} Q(y_{x,x'}|x) Q(y_{x,x'}|x') \geq \prod_{(x,x') \in \mathcal{S}_\mathrm{i}^{(2)}} \varepsilon_Q^2 = \varepsilon_Q^{|\mathcal{S}_\mathrm{i}|}. \tag{61}$$

When $\mathcal{A}$ is odd,

$$P(Z \in \mathcal{Z}) = \prod_{(x,x') \in \mathcal{S}_\mathrm{i}^{(2)} : x \neq x'} P(Z[x] = y_{x,x'}) P(Z_{x'} = y_{x,x'}) \prod_{(x,x) \in \mathcal{S}_\mathrm{i}^{(2)}} P(Z[x] = y_{x,x}) \tag{62}$$

$$= \prod_{(x,x') \in \mathcal{S}_\mathrm{i}^{(2)} : x \neq x'} Q(y_{x,x'}|x) Q(y_{x,x'}|x') \prod_{(x,x) \in \mathcal{S}_\mathrm{i}^{(2)}} Q(y_{x,x}|x) \geq \varepsilon_Q^{|\mathcal{S}_\mathrm{i}|}. \tag{63}$$

■

Proof of Theorem 6: Consider a line network of length L of general DMCs Q_ℓ with $\varepsilon_{Q_\ell} \geq \epsilon > 0$ and a GBNC as described in Sec. II. Without loss of optimality, we assume a deterministic recoding scheme, i.e., ϕ_ℓ are deterministic. Channel $Q_\ell^{\otimes N}$ can be modelled by the function α_ℓ^N with the channel status variable $Z_\ell = (Z_\ell[\mathbf{x}], \mathbf{x} \in \mathcal{Q}_\mathrm{i}^N)$ so that

$$\mathbf{Y}_\ell = \alpha_\ell^N(\mathbf{U}_\ell, Z_\ell). \tag{64}$$

As $\varepsilon_{Q_\ell^{\otimes N}} \geq \varepsilon_{Q_\ell}^N > 0$, the condition of applying Lemma 5 on $Q_\ell^{\otimes N}$ is satisfied.

Let $\mathcal{S}_\mathrm{i}^{(1)} = \mathcal{Q}_\mathrm{i}^N$. Applying Lemma 5 on $Q_1^{\otimes N}$ w.r.t. $\mathcal{S}_\mathrm{i}^{(1)}$, there exists subsets $\mathcal{Z}^{(1)}$ and $\mathcal{S}_\mathrm{o}^{(1)} \subseteq \mathcal{Q}_\mathrm{o}^N$ with $|\mathcal{S}_\mathrm{o}^{(1)}| \leq \lceil |\mathcal{S}_\mathrm{i}^{(1)}|/2 \rceil$ such that $\alpha_1^N(\mathbf{x}, z_1) \in \mathcal{S}_\mathrm{o}^{(1)}$ for any $\mathbf{x} \in \mathcal{S}_\mathrm{i}^{(1)}$ and $z_1 \in \mathcal{Z}^{(1)}$, and $P(Z_1 \in \mathcal{Z}^{(1)}) \geq \varepsilon^{N|\mathcal{Q}_\mathrm{i}|^N}$. Fix an integer $K = \lceil N \log |\mathcal{Q}_\mathrm{i}| \rceil$. For $i = 2, 3, \ldots, K$, define recursively $\mathcal{S}_\mathrm{i}^{(i)}$, $\mathcal{S}_\mathrm{o}^{(i)}$ and $\mathcal{Z}^{(i)}$ as follows: $\mathcal{S}_\mathrm{i}^{(i)} = \left\{\mathbf{x} \in \mathcal{Q}_\mathrm{i}^N : \mathbf{x} = \phi_{i-1}(\mathbf{y}) \text{ for certain } \mathbf{y} \in \mathcal{S}_\mathrm{o}^{(i-1)}\right\}$, and $\mathcal{S}_\mathrm{o}^{(i)}$ and $\mathcal{Z}^{(i)}$ are determined as in the proof of Lemma 5 w.r.t. $Q_i^{\otimes N}$ and $\mathcal{S}_\mathrm{i}^{(i)}$ so that $\alpha_i^{\otimes N}(\mathbf{x}, z) \in \mathcal{S}_\mathrm{o}^{(i)}$ for any $\mathbf{x} \in \mathcal{S}_\mathrm{i}^{(i)}$ and $z \in \mathcal{Z}^{(i)}$, and $P(Z_i \in \mathcal{Z}^{(i)}) \geq \varepsilon^{N|\mathcal{S}_\mathrm{i}^{(i)}|}$.

According to the construction, $|\mathcal{S}_\mathrm{i}^{(i)}| \leq |\mathcal{S}_\mathrm{o}^{(i-1)}|$ and $|\mathcal{S}_\mathrm{o}^{(i)}| \leq \lceil |\mathcal{S}_\mathrm{i}^{(i)}|/2 \rceil$. Hence $|\mathcal{S}_\mathrm{o}^{(K)}| \leq \lceil |\mathcal{S}_\mathrm{i}^{(1)}|/2^K \rceil = 1$. Since the set $\mathcal{S}_\mathrm{o}^{(K)}$ is non-empty, we have $|\mathcal{S}_\mathrm{o}^{(K)}| = 1$, i.e., there exists an output of $Q_K^{\otimes N}$ that occurs with a positive probability for all inputs of $Q_1^{\otimes N}$. Define the channel $G_1 = Q_1^{\otimes N} \phi_1 Q_2^{\otimes N} \cdots \phi_{K-1} Q_K^{\otimes N}$. Under the condition $Z_i \in \mathcal{Z}^{(i)}, i = 1, \ldots, K$, the output of G_1 must be unique for all possible channel inputs, i.e., G_1 is canonical. Note that

$$P(Z_i \in \mathcal{Z}^{(i)}, i = 1, \ldots, K) \geq \varepsilon^{N \sum_{i=1}^K |\mathcal{A}_i|} \geq \varepsilon^{N(2|\mathcal{Q}_\mathrm{i}|^N + K)}. \tag{65}$$

Let $L' = \lfloor L/K \rfloor$. For $i = 2, \ldots, L'$, define $G_i = Q_{K(i-1)+1}^{\otimes N} \phi_{K(i-1)+1} Q_{K(i-1)+2}^{\otimes N} \cdots \phi_{Ki-1} Q_{Ki}^{\otimes N}$. Similar as G_1, we know that G_i, $i = 2, \ldots, L'$ are all canonical. We see that G_i, $i = 1, \ldots, L'$ forms a length-L' network. Let $\tilde{W}_{L'} = \phi_0 G_1 \phi_K G_2 \phi_{2K} \cdots G_{L'}$, which is the end-to-end transition matrix of a GBNC with inner block length 1 for the length-L' network of canonical channels G_i. By the data processing inequality, $I(p_{\mathbf{X}}, W_L) \leq I(p_{\mathbf{X}}, \tilde{W}_{L'})$. Based on this relation, we are ready to prove the theorem, similar to that of Theorem 4. ∎

Appendix C
Proofs about Achievability

Proof of Lemma 8: Suppose that the node $\ell - 1$ transmits $u_\ell(x)$ for N times, where $x \in \mathcal{A}$. We know that the entries of $\mathbf{y}_\ell$ are i.i.d. random variables with distribution $Q_\ell(\cdot \mid u_\ell(x))$. The error probability for ML decoding at the node ℓ satisfies

$$\epsilon_\ell(x) \leq P\left(\vee_{\overline{x} \neq x} \mathcal{L}_\ell(\overline{x}; \mathbf{y}_\ell) \geq \mathcal{L}_\ell(x; \mathbf{y}_\ell)\right\} \tag{66}$$

$$\leq \sum_{\overline{x} \in \mathcal{A}:\ \overline{x} \neq x} P\left(\mathcal{L}_\ell(\overline{x}; \mathbf{y}_\ell) \geq \mathcal{L}_\ell(x; \mathbf{y}_\ell)\right), \tag{67}$$

where the second inequality follows from the union bound. For fixed $\overline{x} \in \mathcal{A}$ so that $\overline{x} \neq x$, we bound the probability $P\left(\mathcal{L}_\ell(\overline{x}; \mathbf{Y}_\ell) \geq \mathcal{L}_\ell(x; \mathbf{Y}_\ell)\right)$ by considering two cases.

If there exists a non-empty subset $\mathcal{Y}_0 \subseteq \mathcal{Q}_\mathrm{o}$ so that for any $y_0 \in \mathcal{Y}_0$, $Q_\ell(y_0 \mid u_\ell(x)) > 0$ but $Q_\ell(y_0 \mid u_\ell(\overline{x})) = 0$, as long as $\mathbf{y}_\ell[i] \in \mathcal{Y}_0$ for some i, we can assert that $\mathcal{L}_\ell(\overline{x}; \mathbf{y}_\ell) < \mathcal{L}_\ell(x; \mathbf{y}_\ell)$. Therefore,

$$P\left(\mathcal{L}_\ell(\overline{x}; \mathbf{y}_\ell) \geq \mathcal{L}_\ell(x; \mathbf{y}_\ell)\right) \leq P\left(\mathbf{Y}_\ell[i] \notin \mathcal{Y}_0, i = 1, \ldots, N\right) \tag{68}$$

$$= \left[\sum_{y \notin \mathcal{Y}_0} Q_\ell(y \mid u_\ell(x))\right]^N = \exp\left(-N \log \frac{1}{\sum_{y \notin \mathcal{Y}_0} Q_\ell(y \mid u_\ell(x))}\right), \tag{69}$$

where $\sum_{y \notin \mathcal{Y}_0} Q_\ell(y \mid u_\ell(x)) = 1 - \sum_{y \in \mathcal{Y}_0} Q_\ell(y \mid u_\ell(x)) < 1$.

Otherwise, consider that the support of $Q_\ell(\cdot \mid u_\ell(x))$ belongs to the support of $Q_\ell(\cdot \mid u_\ell(\overline{x}))$. For $i = 1, \ldots, N$, define the random variable $D_i = \log \frac{Q_\ell(\mathbf{Y}_\ell[i] | u_\ell(\overline{x}))}{Q_\ell(\mathbf{Y}_\ell[i] | u_\ell(x))}$. We see that D_i are i.i.d., and satisfy

$$\log \varrho_\ell \leq D_i \leq -\log \varrho_\ell, \tag{70}$$

where $\varrho_\ell = \min_{x \in \mathcal{Q}_\mathrm{i}, y \in \mathcal{Q}_\mathrm{o}: Q_\ell(y|x) > 0} Q_\ell(y|x)$, and

$$\mathbb{E}[D_i] = E'_\ell \triangleq -\mathcal{D}_{\mathrm{KL}}\left(Q_\ell(\cdot \mid u_\ell(x)) \| Q_\ell(\cdot \mid u_\ell(\overline{x}))\right), \tag{71}$$

where $\mathcal{D}_{\mathrm{KL}}$ denotes the Kullback-Leibler divergence. We see that $E'_\ell > -\infty$. Moreover, as $u_\ell(x) \neq u_\ell(\bar{x}) \in \mathcal{Q}_{\mathrm{i}}^\ell$, $Q_\ell(\cdot \mid u_\ell(x)) \neq Q_\ell(\cdot \mid u_\ell(\overline{x}))$ and hence $E'_\ell \neq 0$. Applying Hoeffding's inequality, we obtain

$$P\left(\mathcal{L}_\ell(\overline{x};\mathbf{y}_\ell) \geq \mathcal{L}_\ell(x;\mathbf{y}_\ell)\right) = P\left(\sum_{i=1}^N D_i \geq 0\right) \tag{72}$$

$$= P\left(\sum_{i=1}^N \left(D_i - E'_\ell\right) \geq -NE'_\ell\right) \tag{73}$$

$$\leq \exp\left(-\frac{NE'^2_\ell}{2\log^2 \varrho_\ell}\right). \tag{74}$$

The proof is completed by combining both cases. ■

Proof of Lemma 10: Suppose Q has size $m \times n$. As $C(Q) > \epsilon > 0$, $m \geq 2$. Let $\mathbf{a} = (a_1, \ldots, a_n)$ be a row of Q, and construct a new $m \times n$ stochastic matrix $\tilde{Q}$ with all the rows $\mathbf{a}$. We have $C(\tilde{Q}) = 0$ and hence $|C(Q) - C(\tilde{Q})| > \epsilon$. Since channel capacity as a function of stochastic matrices is uniformly continuous [13, Lemma I.1], there exists a constant $\delta > 0$ depending on ϵ such that $\|\tilde{Q} - Q\|_\infty > \delta$. As a consequence, there exists another row $\mathbf{a}' = (a'_1, \ldots, a'_n)$ of Q such that $\|\mathbf{a} - \mathbf{a}'\|_\infty > \delta$. Denote by j the index such that $|a_j - a'_j| > \delta$.

Using the example of uniform reduction with $s = 2$, we can choose R so that RQ is formed by $\mathbf{a}$ and $\mathbf{a}'$. Then we can find W so that $RQW = U_2(\rho_1)$, where

$$\rho_1 = \sum_{k: a_k + a'_k > 0} \frac{a_k^2}{a_k + a'_k} = 1 - \sum_{k: a_k + a'_k > 0} \frac{a_k a'_k}{a_k + a'_k}. \tag{75}$$

Based on the relation that

$$\frac{1}{2} - \sum_{k: a_k + a'_k > 0} \frac{a_k a'_k}{a_k + a'_k} = \frac{1}{4} \sum_{k: a_k + a'_k > 0} \frac{(a_k - a'_k)^2}{a_k + a'_k} \geq \frac{1}{4} \frac{(a_j - a'_j)^2}{a_j + a'_j} \geq \frac{\delta^2}{8}, \tag{76}$$

we have the lower bound $\rho_1 \geq B$ with $B = \frac{1}{2} + \frac{\delta^2}{8} > 1/2$. For any ϱ such that $1/2 < \varrho \leq B$, we have $U_2(\varrho) = U_2(\rho_1) U_2(\frac{\rho_1 + \varrho - 1}{2\rho_1 - 1})$, and hence $RQWU_2(\frac{\rho_1 + \varrho - 1}{2\rho_1 - 1}) = U_2(\varrho)$. ■

Proof of Lemma 12: As $\operatorname{rank}(Q) = r \geq s$, we can find stochastic matrices R and W such that $\min\operatorname{inv}(RQW) = \kappa_s(Q)$. Let $B = (RQW)^{-1}$, and $K = BU_s(\varrho)$. As $RQWK = U_s(\varrho)$, we only need to show that for $1/s < \varrho \leq \rho_s(Q)$, K is a stochastic matrix. Let $\mathbf{1}$ be the all-one vector of certain length. We see that $K\mathbf{1} = BU_s(\rho)\mathbf{1} = B\mathbf{1} = \mathbf{1}$, where the last equality follows because $RQW\mathbf{1} = \mathbf{1}$ and RQW is invertible.

It remains to show that all the entries of K are nonnegative. Let b_{ij} be the (i,j) entry of B. The (i,j) entry of K is $k_{ij} = \frac{1}{s-1}\left[(1-\varrho) + b_{ij}(s\varrho - 1)\right] \geq \frac{1}{s-1}\left[(1-\varrho) + \kappa_s(Q)(s\varrho - 1)\right].$

When $\kappa_s(Q) \geq 0$, we have $k_{ij} \geq 0$ for any $\varrho \in (1/s, 1]$. When $\kappa_s(Q) < 0$, we have $k_{ij} \geq 0$ for any $\varrho \in (1/s, \frac{\kappa_s(Q)-1}{s\kappa_s(Q)-1}]$. ■

Proof of Theorem 13: Recall the Markov chain relation in (41), where the transition matrix $\mathbf{P}$ is an $(M+1) \times (M+1)$ matrix with the (i, j) entry ($0 \leq i, j \leq M$):

$$p_{i,j} = \begin{cases} 0 & i < j, \\ \sum_{k=j}^{N} f(k; N, \epsilon)\zeta_j^{i,k} & i \geq j, \end{cases} \tag{77}$$

where $f(k; N, \epsilon) = \binom{N}{k}(1-\epsilon)^k \epsilon^{N-k}$ is the probability mass function (PMF) of the binomial distribution with parameters N and $1 - \epsilon$, and $\zeta_j^{i,k}$ is the probability that the $i \times k$ matrix with independent entries uniformly distributed over the field $\mathbb{F}_q$ has rank j. We know that (ref. [15, (2.4)]) $\zeta_j^{i,k} = \frac{\zeta_j^i \zeta_j^k}{\zeta_j^j q^{(i-j)(k-j)}}$, where

$$\zeta_r^m = \begin{cases} 1 & r = 0, \\ (1-q^{-m})(1-q^{-m+1})\cdots(1-q^{-m+r-1}) & 1 \leq r \leq m. \end{cases} \tag{78}$$

As shown in [32], the matrix $\mathbf{P}$ admits the eigendecomposition $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{V} = (v_{i,j})_{0 \leq i,j \leq M}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_M)$. Here $\lambda_j = \sum_{k=j}^{N} f(k; N, \epsilon)\zeta_j^k$, $v_{i,j} = \zeta_j^i$ for $i \geq j$ and otherwise $v_{i,j} = 0$. It can be checked that $\lambda_0 > \lambda_1 > \cdots > \lambda_M$. Denote the (i, j) entry $0 \leq i, j \leq M$ of V^{-1} by $u_{i,j}$. We know that $u_{i,j} = 0$ for $i < j$ and $u_{i,i} = 1/\zeta_i^i$. Based on the formulation above, we have

$$\begin{aligned} \mathbf{E}[\pi_L] &= \pi_0 \mathbf{V}\mathbf{\Lambda}^L\mathbf{V}^{-1} \begin{bmatrix} 0 & 1 & \cdots & M \end{bmatrix}^\top = \sum_{i=1}^{M} \lambda_i^L v_{M,i} \sum_{j=1}^{i} j u_{i,j} && (79) \\ &= \lambda_1^L v_{M,1} u_{1,1} \left(1 + \sum_{i=2}^{M} \frac{\lambda_i^L v_{M,i}}{\lambda_1^L v_{M,1} u_{1,1}} \sum_{j=1}^{i} j u_{i,j}\right) && (80) \\ &= \Theta(\lambda_1^L), && (81) \end{aligned}$$

where (80) follows from the fact that $v_{M,1}u_{1,1} > 0$, and (81) is obtained by noting that

$$\sum_{i=2}^{M} \frac{\lambda_i^L v_{M,i}}{\lambda_1^L v_{M,1} u_{1,1}} \sum_{j=1}^{i} j u_{i,j} = o(1) \tag{82}$$

as $\lambda_i \leq \lambda_1$ for $i \geq 2$. By (78), we further have

$$\begin{aligned} \lambda_1 &= \sum_{k=1}^{N} f(k; N, \epsilon)(1 - q^{-k}) = \sum_{k=1}^{N} f(k; N, \epsilon) - \sum_{k=1}^{N} f(k; N, \epsilon) q^{-k} && (83) \\ &= 1 - f(0; N, \epsilon) - \sum_{k=1}^{N} \binom{N}{k} (1-\epsilon)^k \epsilon^{N-k} q^{-k} = 1 - (\epsilon + (1-\epsilon)/q)^N. && (84) \end{aligned}$$

The proof is completed. ■

Proof of Lemma 14: For a DMC Q, two channel inputs x_1 and x_2 are said to be *adjacent* if there exists an output y such that $Q(y|x_1)Q(y|x_2) > 0$. Denote by $M_0(Q)$ the largest number of inputs in which adjacent pairs do not exist. For a DMC with $C_0 > 0$, we have that $M_0(Q) \geq 2$ and then $C_0(Q) \geq 1$, since otherwise it is easy to verify $M_0(Q^{\otimes n}) \leq 1$ for any n which leads to $C_0(Q) = 0$.

When the channel Q satisfies $C_0(Q) > 0$, we have $M_0(Q) \geq 2$. Define R as a two-row deterministic stochastic matrix that selects two rows of Q that correspond to two non-adjacent inputs. Denote by a_{ij} the (i, j) entry of RQ. We have $a_{1j}a_{2j} = 0$ for all $j = 1, \ldots, n$. Let S be defined same as the matrix W in defined in (37). ■