On the Capacity of an Infinite Cascade of Channels

Urs Niesen Christina Fragouli Daniela Tuninetti

École Polytechnique Fédérale de Lausanne
Computer & Communication Sciences, LCM
CH-1015 Lausanne, Switzerland
urs.niesen@ieee.org, christina.fragouli@epfl.ch

University of Illinois at Chicago
Electrical & Computer Engineering
60607 Chicago, IL, USA
daniela@ece.uic.edu

Abstract

We consider communication through an infinite cascade of identical discrete memoryless channels. We allow the source and destination nodes to use coding schemes of arbitrary complexity, but restrict the intermediate (relay) nodes to process blocks of a fixed blocklength. We calculate the optimal end-to-end rate, maximized over all possible processings at the relays, and show that it coincides with the end-to-end zero-error capacity. The optimal processing is shown to be identical at each relay and to correspond to a zero-error code. We also show that the rate of convergence to the asymptotic value is exponential in the length of the cascade.

1 Introduction

Consider a communication network where a source node transmits information to a destination node along a path that comprises L consecutive links of the network. We assume that each link corresponds to an identical Discrete Memoryless Channel (DMC). Thus, we can model the communication path between the source and the destination using a line network that consists of L DMCs. We assume that intermediate nodes (relays) are allowed to process blocks of N symbols, while the source and destination node can code and decode across an infinite number of such blocks. We are interested in characterizing the optimal information-theoretic rate that the source can convey to the destination node as a function of N and L.

It is well-known that as $N \to \infty$, we can use a capacity achieving code over each of the cascaded channels to achieve the min-cut bound [1]. That is, the capacity of the overall channel, optimized over all possible processing at the intermediate nodes, equals the capacity of the single DMC the cascade is composed of. To the other extreme, consider the case where N is finite but $L \to \infty$. This is the case where intermediate nodes have complexity and delay constraints, and messages have to traverse a large network to reach the destination. Using finite N limits both the complexity of the processing performed at the intermediate nodes as well as the total delay incurred during transmission through the network, and thus is well suited to complexity and delay constrained networks.

In this paper, we characterize the optimal processing at the relays and the resulting capacity. We will show that as $L \to \infty$ the optimal processing is identical at each relay and corresponds to a zero-error code [2]. The resulting capacity coincides with the end-to-end zero-error capacity. An intuitive interpretation of our results is that, as $L \to \infty$, the zero-error capacity is the only part of the transmitted information rate that we may hope to preserve.

In the paper we also study the rate, in terms of the number of cascaded channels L, at which this limiting capacity is achieved. We show that the rate of convergence is exponential and give tight upper and lower bounds for the exponent. This implies that even for long, but not infinite, cascades, the limiting results found before are still meaningful.

Capacity of line networks with finite complexity intermediate processing have also been investigated in [3], which contains lower bounds for large N. Cascades of channels without processing at the relays have been considered for example in [4] (which investigates the cascade of channels with invertible channel transition matrix) and in [5] (which looks at cascaded binary channels). Optimal orderings of several different binary channels such that the capacity of the overall cascade is maximized have been considered in [6].

The paper is organized as follows. Section 2 formally introduces the network model under consideration, proves properties of the optimal processing and reviews mathematical background on stochastic matrices. Section 3 presents the main results of this paper. In particular, we give exact expressions for the optimal rate and the speed at which convergence to the limiting expression of capacity for large L takes place. We also discuss the connections with zero-error capacity. Section 4 concludes the paper.

2 Problem Statement and Mathematical Background

2.1 Network Model

We consider a cascade of L identical DMCs, as depicted in Figure 1. The source A_0 sends information to the destination A_L via the relay nodes $\{A_i\}_{i=1}^{L-1}$. Each channel has a finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and transition probability matrix $\boldsymbol{V} \in \mathbb{R}_+^{|\mathcal{X}|\times|\mathcal{Y}|}$. While the source and the destination can perform coding of arbitrary complexity, the relays $\{A_i\}_{i=1}^{L-1}$ can only perform operations on *blocks* of N symbols. At a block-level, the channel between consecutive relays is a DMC with input alphabet $\mathcal{X}^N$, output alphabet $\mathcal{Y}^N$, and channel transition matrix

$$\boldsymbol{W} \triangleq \boldsymbol{V}^{\otimes N}, \tag{1}$$

where $\otimes$ denotes the matrix Kronecker product.

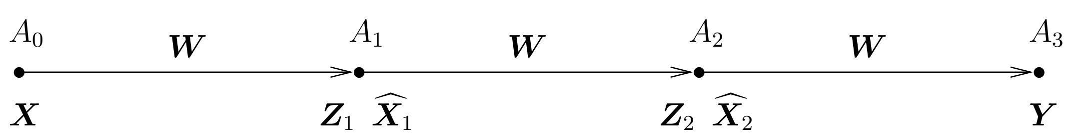

Figure 1: Three cascaded channels with two relays ($L = 3$).

Let $\widehat{\boldsymbol{X}}_i \in \mathcal{X}^N$ denote the input of channel i, and $\boldsymbol{Z}_i \in \mathcal{Y}^N$ the output of channel $i-1$. In general, $\widehat{\boldsymbol{X}}_i$ is a (not necessarily deterministic) function of $\boldsymbol{Z}_i$, which can be represented as a transition probability matrix $\boldsymbol{M}_i \in \mathbb{R}_+^{|\mathcal{Y}|^N\times|\mathcal{X}|^N}$, i.e., $\boldsymbol{M}_i$ specifies, for each realization $\boldsymbol{z}$ of $\boldsymbol{Z}_i$ and each possible value $\hat{\boldsymbol{x}}$ of $\widehat{\boldsymbol{X}}_i$, the probability that $\boldsymbol{z}$ is mapped into $\hat{\boldsymbol{x}}$. We are interested in finding the set of processings $\{\boldsymbol{M}_i\}_{i=1}^{L-1}$ that maximizes the end-to-end achievable rate between the source A_0 and the destination A_L, when $L \to \infty$. We assume that A_0 can encode over an unconstrained number of length-N blocks and A_L can perform any decoding function. The optimal end-to-end rate is hence the capacity of the equivalent channel

$$\boldsymbol{W}_{\text{eq}} \triangleq \boldsymbol{W} \prod_{i=1}^{L-1} \left(\boldsymbol{M}_i \boldsymbol{W}\right), \tag{2}$$

normalized by the number of channel uses N, i.e.,

$$C_{N,L}(\boldsymbol{V}) \triangleq \max_{\{\boldsymbol{M}_i\}_{i=1}^{L-1}} \frac{1}{N} C\big(\boldsymbol{W}_{\text{eq}}\big) \triangleq \max_{\{\boldsymbol{M}_i\}_{i=1}^{L-1}} \max_{\boldsymbol{p}} \frac{1}{N} I\big(\boldsymbol{p}, \boldsymbol{W}_{\text{eq}}\big). \tag{3}$$

The notation $I(\boldsymbol{p}, \boldsymbol{W})$ indicates the mutual information between the input and the output of channel $\boldsymbol{W}$ when the input is distributed according to $\boldsymbol{p}$.

2.2 Properties of the Optimal Processing

We next show that the optimal intermediate processings are *deterministic*, i.e., $\widehat{\boldsymbol{x}}_i = f_i(\boldsymbol{z}_i)$ for some function $f_i(\cdot)$ at the i-th relay, and that $f_i(\cdot)$ can be interpreted as a *decoding and re-encoding* operation.

Proposition 1. *The optimal processings $\{\boldsymbol{M}_i\}_{i=1}^{L-1}$ define a deterministic mapping, that is, every $\boldsymbol{M}_i$ is a binary stochastic matrix.*

Proof. To simplify notation, we consider one relay only, and we drop the subscript 1 in $\boldsymbol{M}_1$. The proof extends straightforwardly to the general case.

For any fixed input distribution $\boldsymbol{p}$, the mutual information $I\big(\boldsymbol{p}, \boldsymbol{W}_{\text{eq}}\big)$ is a convex function of $\boldsymbol{W}_{\text{eq}} = \boldsymbol{W}\boldsymbol{M}\boldsymbol{W}$ [7]. Since $\boldsymbol{W}_{\text{eq}}$ is a linear function of $\boldsymbol{M}$, $I\big(\boldsymbol{p}, \boldsymbol{W}_{\text{eq}}\big)$ is also convex in $\boldsymbol{M}$. Moreover, the set of all transition probability matrices $\boldsymbol{M}$ is a convex set whose extreme points are the binary stochastic matrices. It is a well-known result that the maximum of a convex function over a convex domain is achieved at an extreme point [8]. Hence the result follows. □

Proposition 2. *A binary stochastic matrix $\boldsymbol{M}$ of dimension $|\mathcal{Y}|^N \times |\mathcal{X}|^N$ and of rank ρ can be written as $\boldsymbol{M} = \boldsymbol{M}_D \boldsymbol{M}_E$, where $\boldsymbol{M}_D$ and $\boldsymbol{M}_E$ are again binary stochastic matrices of dimension $|\mathcal{Y}|^N \times \rho$ and $\rho \times |\mathcal{X}|^N$ respectively.*

Proof. A binary stochastic matrix $\boldsymbol{M}$ has exactly one 1 in each row. Since $\boldsymbol{M}$ has rank ρ, it contains a set of ρ linearly independent non-zero columns. Denote by $\{i_1, \ldots, i_\rho\}$ the positions of those columns. Let $\boldsymbol{m}_i$ be the i-th column of $\boldsymbol{M}$ and $\boldsymbol{e}_i$ be the row vector containing all zeros except a 1 in position i. Then the desired decomposition is

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{m}_{i_1} & \cdots & \boldsymbol{m}_{i_\rho} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_{i_1} \\ \vdots \\ \boldsymbol{e}_{i_\rho} \end{pmatrix} \triangleq \boldsymbol{M}_D \boldsymbol{M}_E. \tag{4}$$

□

In the following, without loss of optimality, we restrict our attention to deterministic mappings and we think of $\boldsymbol{M}_D$ as a decoder (mapping the $|\mathcal{Y}|^N$ possible channel output symbols into one of ρ possible "source" symbols), and $\boldsymbol{M}_E$ as an encoder (mapping these ρ "source" symbols back into one of the $|\mathcal{X}|^N$ possible channel input symbols).

2.3 Canonical Decomposition of Stochastic Matrices

In this section, we briefly review the canonical decomposition of a non-negative stochastic matrix $\boldsymbol{Q}$. We will then compute the limit of $\boldsymbol{Q}^L$ as $L \to \infty$. This result will be used in the following sections to characterize the limiting capacity of an arbitrary channel cascaded L times with itself. Our exposition closely follows [9].

Let $\boldsymbol{Q}$ be a square non-negative matrix and denote by $\mathcal{J} \triangleq \{1, \dots, m\}$ the set of its (row and column) indices. Let $q_{ij}^{(k)}$ be the (i, j)-th entry of $\boldsymbol{Q}^k$. We say that the index i *leads* to index j, and write $i \to j$, if $q_{ij}^{(k)} > 0$ for some $k \geq 1$. If $i \to j$ and $j \to i$, we say that i and j *communicate*. An index i is called *essential* if $i \to j$ implies $j \to i$. If i is not essential, it is called *inessential*. This partitions the set of indices $\mathcal{J}$ into the set of essential indices $\mathcal{E}$ and inessential indices $\mathcal{I}$. The set of essential indices $\mathcal{E}$ can furthermore be partitioned into communicating classes $\mathcal{C}$, such that all indices communicating with each other are in the same class.

The canonical form of a matrix $\boldsymbol{Q}$ is obtained by relabeling its indices in such a way that all indices of the same essential communicating class are consecutive, and every inessential index is greater than any essential index. Formally, this corresponds to pre- and post-multiplying $\boldsymbol{Q}$ by some permutation matrix $\boldsymbol{\Pi}$. This results in a matrix of the canonical form

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{\Pi}\boldsymbol{Q}\boldsymbol{\Pi}^T = \begin{pmatrix} \boldsymbol{P}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_2 & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{P}_{|\mathcal{C}|} & \boldsymbol{0} \\ \boldsymbol{R}_1 & \boldsymbol{R}_2 & \cdots & \boldsymbol{R}_{|\mathcal{C}|} & \boldsymbol{S} \end{pmatrix}. \tag{5}$$

The square matrix $\boldsymbol{P}_i$ in (5) contains the transition probabilities within the i-th essential communicating class, $\boldsymbol{S}$ the transition probabilities between the inessential indices $\mathcal{I}$, and $\boldsymbol{R}_i$ the transition probabilities from the inessential indices to the i-th essential communicating class. The submatrices $\boldsymbol{P}_i$ are by definition *irreducible*.

The *period* of an index i is defined as the greatest common divisor of those k for which $q_{ii}^{(k)} > 0$. All indices in the same communicating class have the same period, which is referred to as the period of the class. Denote by d_i the period of the submatrix $\boldsymbol{P}_i$. If $d_i = 1$, then $\boldsymbol{P}_i$ is called *primitive*, i.e., it is irreducible and aperiodic. If $d_i > 1$, then $\boldsymbol{P}_i$ can be written in a canonical form (again by permuting indices) such that, for any integer L, $\boldsymbol{P}_i^{d_i L}$ is of the form

$$\boldsymbol{P}_i^{d_i L} = \begin{pmatrix} \boldsymbol{P}_{i,1}^L & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{i,2}^L & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{P}_{i,d_i}^L \end{pmatrix}, \tag{6}$$

where again the square matrices $\{\boldsymbol{P}_{i,j}\}_{j=1}^{d_i}$ on the main diagonal are primitive.

The following theorem gives the limiting expression of $\boldsymbol{Q}^L$ when $L \to \infty$ for certain $\boldsymbol{Q}$. As we shall see in the next section, the class of $\boldsymbol{Q}$ covered by the theorem is general enough for our purposes.

Theorem 3. *Let $\boldsymbol{Q}$ be a square stochastic matrix in canonical form as in* (5) *and such that all its diagonal irreducible submatrices $\{\boldsymbol{P}_i\}_{i=1}^{|\mathcal{C}|}$ are primitive, i.e., have period* 1*,* *then*

$$\lim_{L\to\infty} \boldsymbol{Q}^L = \begin{pmatrix} \mathbf{1}\boldsymbol{\pi}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}\boldsymbol{\pi}_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}\boldsymbol{\pi}_{|\mathcal{C}|} & \mathbf{0} \\ \boldsymbol{a}_1\boldsymbol{\pi}_1 & \boldsymbol{a}_2\boldsymbol{\pi}_2 & \cdots & \boldsymbol{a}_{|\mathcal{C}|}\boldsymbol{\pi}_{|\mathcal{C}|} & \mathbf{0} \end{pmatrix}, \tag{7}$$

where the row vector $\boldsymbol{\pi}_i$ *is the unique stationary distribution of* $\boldsymbol{P}_i$*, and column vector* $\boldsymbol{a}_i \triangleq (\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{R}_i\mathbf{1}$ *(*$\boldsymbol{I}$ *indicates the identity matrix and* $\mathbf{1}$ *the column vector of all ones). Moreover, the convergence to the limit in* (7) *is exponentially fast in* L*, and depends on the Second Largest Eigenvalue Modulus (SLEM), that is, the eigenvalue of* $\boldsymbol{Q}$ *with largest modulus strictly less than the spectral radius.*

Proof. The proof follows from [9, Th. 4.1, Th. 4.2, Th. 4.3, Th. 4.7]. □

The following example illustrates these definitions.

Example 1. Let $p \in (0,1)$. Consider

$$\boldsymbol{Q} = \begin{pmatrix} 1-p & 0 & 0 & p \\ 0 & 1 & 0 & 0 \\ 0 & p & 1-p & 0 \\ p & 0 & 0 & 1-p \end{pmatrix}, \ \widetilde{\boldsymbol{Q}} = \begin{pmatrix} 1-p & p & 0 & 0 \\ p & 1-p & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & p & 1-p \end{pmatrix}, \ \widetilde{\boldsymbol{Q}}^\infty = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix}.$$

For the stochastic matrix $\boldsymbol{Q}$, we have $\mathcal{E} = \{1,2,4\}$, $\mathcal{I} = \{3\}$, and two essential communicating classes $\mathcal{C} = \big\{\{1,4\},\{2\}\big\}$, both aperiodic. The canonical form $\widetilde{\boldsymbol{Q}}$ has

$$\boldsymbol{P}_1 = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix} 1 \end{pmatrix}, \quad \boldsymbol{S} = \begin{pmatrix} 1-p \end{pmatrix}, \quad \boldsymbol{R}_1 = \begin{pmatrix} 0 & 0 \end{pmatrix}, \quad \boldsymbol{R}_2 = \begin{pmatrix} p \end{pmatrix}.$$

The limit of $\widetilde{\boldsymbol{Q}}^L$ for $L \to \infty$ is $\widetilde{\boldsymbol{Q}}^\infty$. ◊

3 Infinite Cascade of Channels

Determining $C_{N,L}(\boldsymbol{V})$ as defined in (3) for any pair (N,L) is not an easy problem. From the min-cut-set bound [1], we know that $\lim_{N\to\infty} C_{N,L}(\boldsymbol{V}) = C(\boldsymbol{V})$, where $C(\boldsymbol{V})$ is the capacity of the underlying channel $\boldsymbol{V}$. In this section, we establish a connection between $\lim_{L\to\infty} C_{N,L}(\boldsymbol{V})$ and the zero-error capacity of the underlying channel $\boldsymbol{V}$.

3.1 Capacity of an Infinite Cascade of Channels

In this section, we will use the results of Section 2.3 to find the capacity of an arbitrary channel cascaded with itself an infinite number of times.

The next theorem shows that, without loss of optimality, we can assume that all intermediate processings are identical, that is, $\boldsymbol{M} \triangleq \boldsymbol{M}_1 = \ldots = \boldsymbol{M}_{L-1} = \boldsymbol{M}_D\boldsymbol{M}_E$.

Theorem 4. *For a cascade of* L *identical DMCs, identical processing at the relays is optimal as* $L \to \infty$.

Proof. Let $\boldsymbol{W}$ be the channel transition probability matrix of the DMC and, as before, $\{\boldsymbol{M}_i\}_{i=1}^{L}$ the processing at the relays. Call $\boldsymbol{Q}_i \triangleq \boldsymbol{W}\boldsymbol{M}_i \in \mathcal{S}_{n,n}$. An *interval chain* σ of length ℓ is defined to be a sequence of intervals $\{\sigma_i\}_{i=1}^{\ell}$, where the σ_i are each integer intervals $\{l_i, \dots, r_i\}$ and have the property that $l_i = r_{i-1} + 1$ for all $i \in \{2, \dots, \ell\}$. Consider the product $\prod_{i=1}^{L} \boldsymbol{Q}_i$ and define $\boldsymbol{Q}_{\sigma_j} \triangleq \prod_{i=l_j}^{r_j} \boldsymbol{Q}_i$ for any integer interval $\sigma_j \subseteq \{1, \dots, L\}$. We will adapt an idea from [10] to show that, as $L \to \infty$, there exists an interval chain σ of arbitrary length ℓ such that all $\boldsymbol{Q}_{\sigma_j}$ are almost identical. More precisely, for every L there exists an ℓ with the property that $\ell \to \infty$ as $L \to \infty$ such that

$$C\Big(\prod_{i=1}^{L} \boldsymbol{Q}_i\Big) = C\big(\boldsymbol{P}\boldsymbol{Q}_{\sigma_1}^{\ell}\widetilde{\boldsymbol{P}}\big) + \varepsilon(L) \tag{8}$$

for some stochastic matrices $\boldsymbol{P}$ and $\widetilde{\boldsymbol{P}}$ and with $\lim_{L\to\infty} |\varepsilon(L)| = 0$. By the data processing inequality, we have

$$C\Big(\prod_{i=1}^{L} \boldsymbol{Q}_i\Big) \le C\big(\boldsymbol{Q}_{\sigma_1}^{\ell}\big) + \varepsilon(L). \tag{9}$$

But any stochastic matrix $\boldsymbol{Q}_{\sigma_1}$ resulting from this procedure can be written as $\boldsymbol{M}\boldsymbol{W}\widetilde{\boldsymbol{M}}$ for some stochastic matrices $\boldsymbol{M}$ and $\widetilde{\boldsymbol{M}}$ and hence $\boldsymbol{Q}_{\sigma_1}^{\ell}$ can be constructed from a cascade of ℓ channels $\boldsymbol{W}$ by using the same processing at each relay. Hence as $L \to \infty$ (and therefore also $\ell \to \infty$) we can restrict our attention to identical processing at the relays.

For a fixed $k \in \mathbb{N}$ construct $\widehat{\boldsymbol{Q}}$ from $\boldsymbol{Q}$ by quantizing every component of $\boldsymbol{Q}$ to the closest of the points $\{j/k\}_{j=0}^{k}$. The set of all possible quantized matrices (which are, in general, not stochastic) has cardinality $K \triangleq (k+1)^{n^2}$. By a lemma, originally due to Erdös and Szekeres (see [10] for a proof), we have that if $L \ge \ell^K$ then there exists an interval chain σ of length ℓ such that $\widehat{\boldsymbol{Q}}_{\sigma_1} = \widehat{\boldsymbol{Q}}_{\sigma_j}$ for all $j \in \{1, \dots, \ell\}$. Note that $\widehat{\boldsymbol{Q}}_{\sigma_j}$ is defined as the quantized version of $\boldsymbol{Q}_{\sigma_j}$ and hence $\widehat{\boldsymbol{Q}}_{\sigma_j}$ and $\boldsymbol{Q}_{\sigma_j}$ differ componentwise by at most $1/k$. By the above argument the product $\prod_{i=1}^{\ell} \boldsymbol{Q}_{\sigma_i}$ and $\boldsymbol{Q}_{\sigma_1}^{\ell}$ differ componentwise at most by $\frac{1}{k}(an)^{\ell}$ for some constant a independent of k. By choosing k large enough we can make this difference as small as desired. As mutual information is continuous in the channel transition probability matrix we can, for any input distribution $\boldsymbol{p}$, make the difference $|I(\boldsymbol{p}, \boldsymbol{Q}_{\sigma_1}^{\ell}) - I(\boldsymbol{p}, \prod_{i=1}^{\ell} \boldsymbol{Q}_{\sigma_i})|$ also as small as desired. Since ℓ is arbitrary, the result follows. □

Let $\boldsymbol{Q} \triangleq \boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D \in \mathbb{R}_{+}^{\rho\times\rho}$, where ρ is the rank of $\boldsymbol{M}$. Even though using an inner encoder $\boldsymbol{M}_E$ at the source A_0 and an inner decoder $\boldsymbol{M}_D$ at the destination A_L is in general suboptimal, in the limit for large L, this is not the case. In fact, by the data processing inequality

$$I(\boldsymbol{p}, \boldsymbol{Q}^{L-2}) \ge I(\boldsymbol{p}, \boldsymbol{W}\boldsymbol{M}_D\boldsymbol{Q}^{L-2}\boldsymbol{M}_E\boldsymbol{W}) \ge I(\boldsymbol{p}, \boldsymbol{Q}^{L}), \tag{10}$$

and hence

$$\lim_{L\to\infty} C(\boldsymbol{W}\boldsymbol{M}_D\boldsymbol{Q}^{L-2}\boldsymbol{M}_E\boldsymbol{W}) = \lim_{L\to\infty} C(\boldsymbol{Q}^{L}). \tag{11}$$

Without loss of generality we can assume that matrix $\boldsymbol{Q}$ is in canonical form. Indeed, we get from (5)

$$C(\widetilde{\boldsymbol{Q}}^{L}) = C(\boldsymbol{\Pi}\boldsymbol{Q}\boldsymbol{\Pi}^{T}\boldsymbol{\Pi}\boldsymbol{Q}\dots\boldsymbol{Q}\boldsymbol{\Pi}^{T}) = C(\boldsymbol{\Pi}\boldsymbol{Q}^{L}\boldsymbol{\Pi}^{T}) = C(\boldsymbol{Q}^{L}). \tag{12}$$

Theorem 5. *Consider a square stochastic matrix $\boldsymbol{Q}$, and let $\mathcal{C}_d$ be the set of irreducible classes of $\boldsymbol{Q}$ with period d. Then*

$$\lim_{L\to\infty} C(\boldsymbol{Q}^L) = \log\Big(\sum_d d|\mathcal{C}_d|\Big). \tag{13}$$

Proof. The notation is the same as in (5) and in Theorem 3. Let d_i be the period of $\boldsymbol{P}_i$ and denote by d the least common multiple of the $\{d_i\}$. By the data processing inequality $I(\boldsymbol{p}, \boldsymbol{Q}^L)$ is decreasing in L for any $\boldsymbol{p}$. As $C(\boldsymbol{Q}^L) \geq 0$ for any L, this implies that $\lim_{L\to\infty} C(\boldsymbol{Q}^L)$ exists. Hence

$$\lim_{L\to\infty} C(\boldsymbol{Q}^L) = \lim_{\ell\to\infty} C(\boldsymbol{Q}^{\ell d}). \tag{14}$$

From (6), we know that the part of $\boldsymbol{Q}^{\ell d}$ corresponding to essential indices is block diagonal for any ℓ. Moreover, there are exactly $D \triangleq \sum_{d=1}^{\infty} d|\mathcal{C}_d|$ such blocks and each block is a primitive matrix. Hence, from Theorem 3, we know the limit $\boldsymbol{Q}^\infty \triangleq \lim_{\ell\to\infty} \boldsymbol{Q}^{\ell d}$ in closed-form, i.e., $\boldsymbol{Q}^\infty$ is a block diagonal matrix where each block is a rank-one matrix.

As capacity is upper bounded by the logarithm of the rank of the channel transition probability matrix [11],

$$C(\boldsymbol{Q}^\infty) \leq \log \operatorname{rank}(\boldsymbol{Q}^\infty) = \log D. \tag{15}$$

Moreover, $\log D$ is easily seen to be an achievable rate. Recalling that $D \triangleq \sum_{k=1}^{\infty} k|\mathcal{C}_k|$ yields the desired result. □

It is interesting to notice that the optimal coding and decoding strategy for the channel $\boldsymbol{Q}^\infty$ are very simple. The encoder uses only one input from each essential communicating class with uniform probability. The decoder declares as transmitted symbol the index of the essential communicating class to which the received output symbol belongs to. Since essential communicating classes are disjoint, and inessential indices are never used, each relay recovers the transmitted information with zero probability of error. We shall return to this point in Section 3.3, where we will show that the optimal intermediate processing corresponds to a zero-error code and that the limiting capacity $\log(D)$ is closely related to the zero-error capacity of the underlying channel $\boldsymbol{V}$.

To conclude this section, we give a simple characterization of D that does not rely on the canonical decomposition of $\boldsymbol{Q}$.

Corollary 6. *Call D the number of eigenvalues of modulus 1 of the stochastic matrix $\boldsymbol{Q}$. Then*

$$\lim_{L\to\infty} C(\boldsymbol{Q}^L) = \log D. \tag{16}$$

Proof. The result is a consequence of the fact that the set of eigenvalues of a block triangular matrix is the union of the eigenvalues of its diagonal blocks [12], and that primitive stochastic matrices have a single eigenvalue of maximum modulus [9]. □

3.2 Asymptotic Rate of Decay

Theorem 3 states that the rate of convergence of $\boldsymbol{Q}^L$ to $\boldsymbol{Q}^\infty$ is exponential in L and depends on the SLEM of $\boldsymbol{Q}$. In this section, we determine the speed of convergence of $C(\boldsymbol{Q}^L)$ to $C(\boldsymbol{Q}^\infty)$. More precisely we are interested in

$$E_L(\boldsymbol{Q}) \triangleq \liminf_{L\to\infty} -\frac{1}{L}\log\Big(C(\boldsymbol{Q}^L) - \log(D)\Big). \tag{17}$$

Theorem 7. *Let $\boldsymbol{Q}$ be a stochastic matrix and call $\widetilde{\boldsymbol{Q}}$ the stochastic matrix obtained by deleting all inessential indices from $\boldsymbol{Q}$. Then*

$$-\log|\lambda_2(\boldsymbol{Q})| \leq E_L(\boldsymbol{Q}) \leq -2\log|\lambda_2(\widetilde{\boldsymbol{Q}})|. \tag{18}$$

where $|\lambda_2(\boldsymbol{M})|$ denotes the SLEM of the matrix $\boldsymbol{M}$.

Proof. The proof can be found in [13] where it is also shown that the right-hand-side of (18) is tight if $\boldsymbol{Q} = \widetilde{\boldsymbol{Q}}$, i.e., if $\boldsymbol{Q}$ contains no inessential indices. □

Example 2. As an example, consider for $p \in (0,1)$ and $t \in (0,1]$

$$\boldsymbol{Q} = \begin{pmatrix} 1-p & p & 0 \\ p & 1-p & 0 \\ t/2 & t/2 & 1-t \end{pmatrix}. \tag{19}$$

The eigenvalues of $\boldsymbol{Q}$ are $\{1, 1-2p, 1-t\}$. The asymptotic rate of decay of capacity can be computed analytically in this case and is given by

$$E_L(\boldsymbol{Q}) = -\log\big(\max\{1-t, (1-2p)^2\}\big). \tag{20}$$

With the right choice of the parameters p and t both the upper and the lower bound in Theorem 7 can be achieved. Hence both bounds in Theorem 7 are tight. ◊

3.3 Connections to Zero Error Capacity

In this section, we explore the connection between the code that achieves the capacity $C_{N,L}(\boldsymbol{V})$ in the limit of large L and the zero-error capacity of the underlying channel.

The zero-error capacity of a DMC, specified by its transition probability matrix $\boldsymbol{V}$, is the maximum rate at which communication is possible over this channel with *zero-error*. The concept of zero-error capacity was first introduced in [2].

Two input letters $x_1, x_2 \in \mathcal{X}$ of $\boldsymbol{V}$ are said to be *adjacent* if there exists an output letter $y \in \mathcal{Y}$ such that $v(y|x_1) > 0$ and $v(y|x_2) > 0$. Let $G(\boldsymbol{V})$ be the graph associated with $\boldsymbol{V}$ that has as vertex set the possible inputs of $\boldsymbol{V}$ and in which two vertices are connected if the corresponding input letters are adjacent. Denote by $M_0\big(G(\boldsymbol{V})\big)$ the largest number of vertices in $G(\boldsymbol{V})$ no two of which are connected by an edge (or, equivalently, the largest number of input letters of $\boldsymbol{V}$ no two of which are adjacent). It is shown in [2] that the zero-error capacity of $\boldsymbol{V}$ is given by

$$C_0\big(G(\boldsymbol{V})\big) \triangleq \sup_n \frac{1}{n}\log M_0\big(G(\boldsymbol{V}^{\otimes n})\big). \tag{21}$$

We shall now prove that for any finite N

$$\lim_{L\to\infty} C_{N,L}(\boldsymbol{V}) = \lim_{L\to\infty} \max_{\boldsymbol{M}_E,\boldsymbol{M}_D} C\Big((\boldsymbol{M}_E\boldsymbol{V}^{\otimes N}\boldsymbol{M}_D)^L\Big) = \frac{1}{N}\log M_0\big(G(\boldsymbol{V}^{\otimes N})\big). \tag{22}$$

The proof of this result uses the following theorem, which asserts that the zero-error capacity obeys a sort of data processing inequality.

Theorem 8. *Consider a cascade of L channels $\{\boldsymbol{Q}_i\}_{i=1}^L$. Then for any finite N and any $j = 1, \ldots, L$*

$$M_0\Big(G\big((\prod_{i=1}^L \boldsymbol{Q}_i)^{\otimes N}\big)\Big) \leq M_0\big(G(\boldsymbol{Q}_j^{\otimes N})\big). \tag{23}$$

Proof. By definition $M_0\big(G(\boldsymbol{Q})\big) = D$ if and only if there exists an encoder $\boldsymbol{M}_E$ and decoder $\boldsymbol{M}_D$ such that $\boldsymbol{M}_E\boldsymbol{Q}\boldsymbol{M}_D$ is an identity matrix of dimension D.

Call $(\boldsymbol{M}_E, \boldsymbol{M}_D)$ the optimal encoder and decoder for the matrix $(\prod_{i=1}^{L} \boldsymbol{Q}_i)^{\otimes N}$. By the properties of the Kronecker product [14], we have

$$\boldsymbol{I} = \boldsymbol{M}_E(\prod_{i=1}^{L} \boldsymbol{Q}_i)^{\otimes N}\boldsymbol{M}_D = \boldsymbol{M}_E\prod_{i=1}^{L}(\boldsymbol{Q}_i^{\otimes N})\boldsymbol{M}_D = \widetilde{\boldsymbol{M}}_E\boldsymbol{Q}_j^{\otimes N}\widetilde{\boldsymbol{M}}_D, \tag{24}$$

where $\widetilde{\boldsymbol{M}}_E \triangleq \boldsymbol{M}_E(\prod_{i=1}^{j-1} \boldsymbol{Q}_i^{\otimes N})$ and $\widetilde{\boldsymbol{M}}_D \triangleq (\prod_{i=j+1}^{L} \boldsymbol{Q}_i^{\otimes N})\boldsymbol{M}_D$. Hence there exists at least one zero-error encoder and decoder for $\boldsymbol{Q}_j^{\otimes N}$ yielding the same rate, which shows the result. □

From the definition of $C_0\big(G(\boldsymbol{Q})\big)$ in (21) we see that Theorem 8 implies a "min-cut" condition on the zero-error capacity:

$$C_0\big(G(\prod_{i=1}^{L} \boldsymbol{Q}_i)\big) \leq C_0\big(G(\boldsymbol{Q}_j)\big). \tag{25}$$

With the result of Theorem 8 we can now prove (22), which states that, for large L, the optimal $(\boldsymbol{M}_E, \boldsymbol{M}_D)$ pair is the best (in the sense of highest rate) possible zero-error code for the channel $\boldsymbol{V}$ of blocklength N.

Theorem 9. *Let $\boldsymbol{V}$ be the transition probability matrix of an arbitrary DMC and $\boldsymbol{W} \triangleq \boldsymbol{V}^{\otimes N}$. Then*

$$\begin{aligned}\lim_{L\to\infty} C\big((\boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D)^L\big) &= C_0\Big(G\big(\lim_{L\to\infty}(\boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D)^L\big)\Big) &(26)\\ &= \log M_0\Big(G\big(\lim_{L\to\infty}(\boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D)^L\big)\Big) &(27)\\ &\leq \log M_0\big(G(\boldsymbol{W})\big), &(28)\end{aligned}$$

and we have equality in the last line if the $(\boldsymbol{M}_E, \boldsymbol{M}_D)$ pair defines an optimal zero-error code for the channel $\boldsymbol{V}$ for the given blocklength N.

Proof. Equality in (26) follows since in the limit of large L, the usual capacity and the zero-error capacity of the channel $(\boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D)^L$ coincide. Equality in (27) follows since, again in the limit of large L, it is possible to construct an optimal zero-error code for the channel $(\boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D)^L$ with blocklength one. Finally (28) is just an application of Theorem 8.

To see that we have equality in (28) if $(\boldsymbol{M}_E, \boldsymbol{M}_D)$ is an optimal zero-error code for $\boldsymbol{W}$, observe that in this case $(\boldsymbol{M}_E\boldsymbol{W}\boldsymbol{M}_D)^L = (\boldsymbol{I})^L = \boldsymbol{I}$. □

Equation (22) follows now from interchanging the limit and the maximization operation (which can be shown to be possible here) and applying Theorem 9.

4 Conclusion

In this work, we have shown that for a cascade of L identical channels the limiting capacity when $L \to \infty$ can be easily computed as the logarithm of the number of eigenvalues of modulus one of the channel transition probability matrix. In this case, the optimal finite complexity processing performed at the relays is identical at each relay and corresponds to using an optimal zero-error code. We have also shown that with identical processing at the relays the limiting capacity for $L \to \infty$ is approached exponentially in L with the rate of decay being related to the second largest eigenvalue of the channel transition probability matrix.

References

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[2] C. E. Shannon. The zero error capacity of a noisy channel. *IEEE Transactions on Information Theory*, 2(3):8–19, September 1956.

[3] D. Tuninetti and C. Fragouli. Processing along the way: Forwarding vs. coding. In *Proceedings of the International Symposium on Information Theory and its Applications*, Parma, Italy, October 2004.

[4] M. K. Simon. On the capacity of a cascade of identical discrete memoryless nonsingular channels. *IEEE Transactions on Information Theory*, 16(1):100–102, January 1970.

[5] R. A. Silverman. On binary channels and their cascade. *IEEE Transactions on Information Theory*, 1(3):19–27, December 1955.

[6] A. B. Kiely and J. T. Coffey. On the capacity of a cascade of channels. *IEEE Transactions on Information Theory*, 39(4):1310–1321, July 1993.

[7] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[8] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1996.

[9] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer Verlag, 1981.

[10] M. Saks and A. Condon. A limit theorem for sets of stochastic matrices. *Linear Algebra and Applications*, 381:61–76, 2004.

[11] C. E. Shannon. Some geometrical results in channel capacity. *Nachrichtentechnische Zeitschrift*, 10, 1957.

[12] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[13] U. Niesen. Cascaded channels with finite complexity intermediate processing. Master's thesis, EPFL, August 2005.

[14] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.