On Capacity of Line Networks

Daniela Tuninetti (daniela@ece.uic.edu)
in collaboration with Urs Niesen (uniesen@mit.edu) and Christina Fragouli (christina.fragouli@epfl.ch)

***Abstract*— We consider communication through a cascade of L identical Discrete Memoryless Channels (DMCs). The source and destination node are allowed to use coding schemes of arbitrary complexity, but the intermediate relay nodes are restricted to process only blocks of N symbols. We are interested in the set of length-N processing performed at the relays that maximize the achievable rate between the source and the destination.**

It is well known [1] that for any finite L and $N \to \infty$ the relays can use a capacity achieving code and communicate reliably as long as the rate of this code is below the capacity of the underlying DMC. The capacity of the cascade is hence equal to the network *min-cut capacity*. In this work we summarize several results [2], [3], [4], [5] on the capacity of infinite long line networks with finite processing length N, and on scaling laws of N with L necessary and sufficient to achieve a certain end-to-end rate.

We show [3] that for finite N and $L \to \infty$ the capacity of the cascade coincides with the highest rate zero-error code of length N of the underlying DMC, and that it can be achieved by using the same processing at each relay. In other words, the *zero-error capacity* is the maximum achievable rate when the processing length is restricted to be finite and the cascade is infinitely long.

We also characterize how N must scale with L in order to achieve rates in between the zero-error and the min-cut capacity. By considering a specific communication scheme with identical processing at each relay, we show [4] that $N = \Theta(\log L)$ [1] is *sufficient* to achieve any rate below the min-cut capacity. By developing a novel upper bound on the capacity of cascades with optimal intermediate processing that applies for any (N, L) pairs, we then show [5] that $N = \Theta(\log L)$ is *necessary* to achieve certain rates above the zero-error capacity. We finally outline a promising way to extend our logarithmic scaling law to *any* rate above the zero-error capacity.

I. Introduction

Communication systems today are organized in large scale networks, with Internet the most conspicuous example, where information needs to traverse multiple hops in order to reach its destination. Another such example are wireless ad-hoc networks where the average number of hops between a source-destination pair scales as the square root of the number of nodes in the network [6]. Each of the hops may introduce errors, that become more and more pronounced as the size of the network grows.

Motivated by these observations, in [3] we investigated what benefits finite complexity processing at intermediate nodes may offer. We modeled the communication path between the source and the destination as a line network that consists of L cascaded identical DMCs, and allowed each intermediate node to process blocks of N symbols. This is a reasonable definition of complexity as it allows to bound not only processing complexity, but also the delay and memory requirements at intermediate nodes. Moreover, it is well suited to an environment where information is transmitted in packets.

In [3] we showed that if the network length increases ($L \to \infty$) but the blocklength N is fixed, the optimal processing is identical at each relay and corresponds to an optimal (highest rate) zero-error code of length N. Thus, the capacity of the cascade coincides with the rate of this code and it is always below the *zero-error capacity* of the underlying channel. The zero-error capacity is the maximum rate at which information can be communicated over a channel with zero probability of error [8]. An intuitive interpretation of this result 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. On the other hand, when $N \to \infty$ the relays can use a capacity achieving code and communicate reliably as long as the rate of this code is below the capacity of the DMC. That is, for $N \to \infty$ we can achieve the *min-cut capacity* [1].

Since the zero-error capacity and the min-cut capacity might differ quite substantially, a natural question to ask is what happens if we allow N to grow with L. In [4] we showed that $N = \Theta(\log L)$ is *sufficient to achieve any rate below the min-cut capacity*. In order to do so, we assume that all nodes use the same error correcting code of length N. Random coding arguments [10] show that there exist good codes with average probability of error decaying to zero exponentially fast. We use such a good code as our inner code. We then find a lower bound on the achievable end-to-end rate with this scheme. With this, we show that $N = \Theta(\log L)$ is sufficient for the lower bound to tend to the min-cut capacity when $L \to \infty$.

In [5] we showed that $N = \Theta(\log L)$ is *necessary to achieve certain rates above the zero-error capacity*. We start by deriving an upper bound on the capacity of cascades with optimal intermediate processing that are valid for any L and N. The main idea is to decompose the channel transition matrix into a linear combination of two stochastic matrices and develop an upper bound that depends on the smallest rank of these two matrices. With this, we show that $N = \Theta(\log L)$ is necessary for the upper bound to tend to the min-cut capacity when $L \to \infty$, for certain rates above the zero-error capacity.

To extend out logarithmic scaling law to *all* rates above the zero-error capacity, we develop here another upper bound that deals directly with the mutual information between the source and the destination. The idea is to provide some feedback to the destination and show that the advantage provided by feedback vanishes as N increases.

The paper is organized as follows. Section II introduces the network model; Section III derives the capacity for finite N

[1] We use Knuth's notation: $f(n) = O(g(n))$ means that there exists a constant c and integer n_0 such that $f(n) \leq c\, g(n)$ for $n > n_0$. $f(n) = \Theta(g(n))$ denotes that $f(n) = O(g(n))$ as well as $g(n) = O(f(n))$.

and $L \to \infty$ for identical intermediate processing; Section IV derives the capacity for finite N and $L \to \infty$ and optimal intermediate processing; Section V presents upper and lower bounds on the capacity with optimal intermediate processing valid for any (N, L) pairs, which are then used in Section VI to derive the scaling law. Section VII gives an example of application of the derived theory. Section VIII concludes the paper.

II. Network Model

We consider line networks with $L-1$ relays as depicted in Figure 1. The source A_0 sends information to the destination A_L via relays $\{A_i\}_{i=1}^{L-1}$. Each link corresponds to the same DMC with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and arbitrary transition probability matrix $\boldsymbol{V}$. We assume that all the DMCs in the cascade are the same.

We restrict the relays $\{A_i\}_{i=1}^{L-1}$ to perform operations from blocks of N symbols in $\mathcal{Y}$ to blocks of N symbols in $\mathcal{X}$ in a memoryless fashion across blocks. Using N times the channel $\boldsymbol{V}$ between A_i and A_{i+1}, amounts to connecting A_i and A_{i+1} through an equivalent DMC with input alphabet $\mathcal{X}^N$, output alphabet $\mathcal{Y}^N$, and transition probability matrix $\boldsymbol{W} \triangleq \boldsymbol{V}^{\otimes N}$ where $\otimes$ denotes the Kronecker product. For the node A_i, we denote by $\boldsymbol{X}_i \in \mathcal{X}^N$ what the relay sends and with $\boldsymbol{Y}_i \in \mathcal{Y}^N$ what the relay receives. The output $\boldsymbol{X}_i$ is then a (not necessarily deterministic) function of $\boldsymbol{Y}_i$. This function can be described by a transition probability matrix $\boldsymbol{M}_i$ specifying for each realization $\boldsymbol{x}$ of $\boldsymbol{X}_i$ and $\boldsymbol{y}$ of $\boldsymbol{Y}_i$ the probability $\Pr[\boldsymbol{X}_i = \boldsymbol{x} | \boldsymbol{Y}_i = \boldsymbol{y}]$.

We allow the source A_0 and the destination A_L to perform coding and decoding of arbitrary complexity, across a possibly infinite number of symbols in $\mathcal{X}^N$ and $\mathcal{Y}^N$.

We are interested in identifying the set of processings $\{\boldsymbol{M}_i\}_{i=1}^{L-1}$ that achieve

$$C_{N,L}(\boldsymbol{V}) \triangleq \max_{\{\boldsymbol{M}_i\}_{i=1}^{L-1}} \frac{1}{N} C\left(\boldsymbol{W} \prod_{i=1}^{L-1} (\boldsymbol{M}_i \boldsymbol{W})\right). \quad (1)$$

Here, $C(\boldsymbol{Q}) = \max_{\boldsymbol{p}} I(\boldsymbol{p}, \boldsymbol{Q})$ where $\boldsymbol{p}$ is the input distribution and $\boldsymbol{Q}$ the channel transition matrix.

In this paper, we also use the notion of the zero-error capacity of the underlying channel $\boldsymbol{V}$. The zero-error capacity is defined in [8] as the maximum rate at which communication is possible with zero probability of error and can be computed as follows. For a channel with transition matrix $\boldsymbol{V}$, we call two input letters k and ℓ adjacent if there exists an output letter j such that $[\boldsymbol{V}]_{k,j} > 0$ and $[\boldsymbol{V}]_{\ell,j} > 0$. We then construct a graph $G(\boldsymbol{V})$ corresponding to the stochastic matrix $\boldsymbol{V}$ having as vertex set the possible inputs of $\boldsymbol{V}$ and in which two vertices are connected by an edge if the corresponding input letters are adjacent. Denote by $M_0(\boldsymbol{V})$ the largest number of input letters of $\boldsymbol{V}$ no two of which are adjacent. The zero-error capacity of $\boldsymbol{V}$ is

$$C_0(\boldsymbol{V}) \triangleq \sup_n \frac{1}{n} \log M_0(\boldsymbol{V}^{\otimes n}). \quad (2)$$

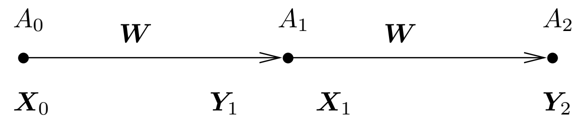

Fig. 1. A line network with two channels and one relay ($L = 2$).

Clearly, for any DMC $\boldsymbol{V}$, any intermediate processing of length N and any network length L, we have

$$\frac{1}{N} M_0(\boldsymbol{V}^{\otimes N}) \le C_{N,L}(\boldsymbol{V}) \le C(\boldsymbol{V}). \quad (3)$$

The upper bound in (3) is achievable by using a capacity achieving code at each node in the network. That no other coding strategy can do better than $C(\boldsymbol{V})$ is clear from the min-cut bound. Hence, the upper bound is tight for $N \to \infty$ [1]. The lower bound in (3) is achievable by using the same zero-error code of length N at each node in the network. We shall prove that the lower bound is tight for $L \to \infty$ and finite N.

It can be easily seen that:

Proposition II.1 ([3]). *The optimal intermediate processings $\{\boldsymbol{M}_i\}_{i=1}^{L-1}$ can be taken to be* deterministic *functions, i.e., $\boldsymbol{x}_i = f_i(\boldsymbol{y}_i)$ for some function $f_i(\cdot)$ at the i-th relay.*

Proposition II.2 ([3]). *Any deterministic processing of length N and rank ρ can be interpreted as a* decoding and re-encoding *operation of a rate $\log(\rho)/N$ code.*

Hence, we shall restrict our attention to deterministic mappings of the form $\boldsymbol{M}_i = \boldsymbol{M}_{i,D} \boldsymbol{M}_{i,E}$. We think of $\boldsymbol{M}_{i,D}$ as a decoder (mapping the $|\mathcal{Y}|^N$ possible channel output symbols into one of $\rho_i = \operatorname{rank}(\boldsymbol{M}_i)$ possible "source" symbols), and $\boldsymbol{M}_{i,E}$ as an encoder (mapping these ρ_i "source" symbols back into one of the $|\mathcal{X}|^N$ possible channel input symbols).

III. Capacity of Infinite Cascade of Identical Channels without intermediate processing

In this section, we derive the capacity of an infinite cascade of DMCs with square transition probability matrix $\boldsymbol{Q}$, that is, we compute $\lim_{L\to\infty} C(\boldsymbol{Q}^L)$. This case arises when either the channel $\boldsymbol{Q}$ is cascaded with itself *without any intermediate processing*, or it is the result of applying the *same intermediate processing* at each relay, that is $\boldsymbol{Q} = \boldsymbol{M}_E \boldsymbol{W} \boldsymbol{M}_D$.

The limit $\lim_{L\to\infty} C(\boldsymbol{Q}^L)$ can be easily evaluated when considering the canonical decomposition of the non-negative matrix $\boldsymbol{Q}$ and the limiting behavior of its powers. Namely,

Lemma III.1 ([7]). *For any non-negative matrix $\boldsymbol{Q}$, there exist a permutation matrix $\boldsymbol{\Pi}$ and an integer B such that*

$$\boldsymbol{\Pi} \boldsymbol{Q} \boldsymbol{\Pi}^T = \begin{pmatrix} \boldsymbol{P}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_2 & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{P}_B & \boldsymbol{0} \\ \boldsymbol{R}_1 & \boldsymbol{R}_2 & \cdots & \boldsymbol{R}_B & \boldsymbol{S} \end{pmatrix}, \quad (4)$$

where the submatrices $\boldsymbol{P}_i$ and the matrix $\boldsymbol{S}$ are irreducible.

Furthermore, for each diagonal block $\boldsymbol{P}_i$, there exists an integer $d_i \geq 1$ (called period*) such that, for any integer $\ell \geq 1$,*

$$\boldsymbol{P}_i^{d_i\,\ell} = \begin{pmatrix} \boldsymbol{P}_{i,1}^{\ell} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{i,2}^{\ell} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{P}_{i,d_i}^{\ell} \end{pmatrix}, \tag{5}$$

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

Lemma III.2 ([7]). *Let $\boldsymbol{Q}$ be a square* stochastic *matrix in canonical form as in* (4)*. If all the diagonal irreducible matrices $\{\boldsymbol{P}_i\}_{i=1}^{B}$ have period $d_i = 1$, then $\lim_{L\to\infty} \boldsymbol{Q}^L$ exists and it is a matrix with rank equal to B.*

Moreover, the convergence to the limit $\lim_{L\to\infty} \boldsymbol{Q}^L$ is exponentially fast in L. The exponent is a function on the Second Largest Eigenvalue Modulus (SLEM), that is, the eigenvalue of $\boldsymbol{Q}$ with largest modulus strictly less than one.

Lemma III.1 and Lemma III.2 imply that:

Corollary III.3 ([3]). *For a square* stochastic *matrix $\boldsymbol{Q}$*

$$\sum_{i=1}^{B} d_i = \operatorname{rank}\big(\lim_{L\to\infty} \boldsymbol{Q}^{d\,L}\big) = D(\boldsymbol{Q}), \tag{6}$$

where $D(\boldsymbol{Q})$ is the multiplicity, algebraic and geometric, of $\lambda_1(\boldsymbol{Q}) = 1$ and d is the least common multiple of $d_1, \cdots, d_B$.

From Corollary III.3 we can prove that:

Theorem III.4 ([3]). *For a square stochastic matrix $\boldsymbol{Q}$ let $D(\boldsymbol{Q})$ denote the number of eigenvalues on modulus 1. Then*

$$\lim_{L\to\infty} C(\boldsymbol{Q}^L) = \log D(\boldsymbol{Q}). \tag{7}$$

The convergence to the limiting capacity in (7) is exponentially fast in L. By defining the exponential asymptotic rate of decay as

$$E_L(\boldsymbol{Q}) \triangleq \liminf_{L\to\infty} -\frac{1}{L} \log\big(C(\boldsymbol{Q}^L) - \log D(\boldsymbol{Q})\big)\big), \tag{8}$$

we have:

Theorem III.5 ([3]). *Let $\boldsymbol{Q}$ be a square stochastic matrix in canonical form as in* (4)*. Let $\widehat{\boldsymbol{Q}}$ be the stochastic matrix obtained by deleting from $\boldsymbol{Q}$ all indices corresponding to $\boldsymbol{S}$. Then*

$$-\log|\lambda_2(\boldsymbol{Q})| \leq E_L(\boldsymbol{Q}) \leq -2\log|\lambda_2(\widehat{\boldsymbol{Q}})|, \tag{9}$$

where $\lambda_2(\boldsymbol{Q})$ denotes the second largest eigenvalue modulus of the channel matrix $\boldsymbol{Q}$.

Both bounds are tight. In particular, the upper bound in (9) *is tight if $\boldsymbol{Q} = \widehat{\boldsymbol{Q}}$.*

Theorem III.5 implies that for long, but still finite, cascades of identical channels, the limiting result of Theorem III.4 is meaningful. It also highlights the necessity of a well chosen intermediate processing.

IV. Capacity of Infinite Cascade of Identical Channels with optimal intermediate processing

In this section, we characterize the optimal finite length intermediate processing for an infinite cascade and establish connections with the zero-error capacity.

The next theorem shows that for an infinite cascade of identical DMCs, identical processing at the relays is optimal. This theorem is crucial as it allows us to optimize over only one intermediate processing instead of having to optimize over an infinite sequence of processing $\{\boldsymbol{M}_i\}_{i=1}^{\infty}$. With this, we show that the optimal identical intermediate processing at the relays corresponds to using the *best zero-error code of blocklength N for the channel $\boldsymbol{V}$*. The resulting capacity of the cascade equals the rate of this zero-error code. The proof uses a *coloring* argument originally due to Erdös and Szekeres.

Theorem IV.1 ([3]). *For a cascade of L identical DMCs there exits an optimal processing $\boldsymbol{M}^\star$ such that*

$$\begin{aligned}\lim_{L\to\infty} C_{N,L}(\boldsymbol{V}) &= \lim_{L\to\infty} \frac{1}{N} C\big((\boldsymbol{M}^\star \boldsymbol{W})^L\big) \\ &= \frac{1}{N} \log M_0\big(\boldsymbol{V}^{\otimes N}\big) \leq C_0(\boldsymbol{V}).\end{aligned} \tag{10}$$

It is interesting to notice that, when $L \to \infty$, although the source could use encoding schemes of arbitrary complexity (infinitely many length-N blocks) the optimal coding strategy at the source is also to employ the zero-error code used at the intermediate relays.

Finally, notice that *any rate* strictly below the zero-error capacity can be achieved with *finite* blocklength processing N. This is so because the limit superior in the definition of zero-error capacity in (2) is a true limit [9].

V. Bounds on $C_{N,L}(\boldsymbol{V})$

Since $C_0(\boldsymbol{V})$ (the ultimate limit for finite N and $L \to \infty$) and $C(\boldsymbol{V})$ (the ultimate limit for finite L and $N \to \infty$) might differ quite substantially, a natural question to ask is what happens if both N and L are allowed to grow.

In this section, we derive a lower bound for $C_{N,L}(\boldsymbol{V})$ to show that $N = \Theta\big(\log L\big)$ is sufficient to achieve

$$C_{N,L}(\boldsymbol{V}) \geq R(\alpha) \triangleq (1-\alpha)\, C_0(\boldsymbol{V}) + \alpha\, C(\boldsymbol{V}), \tag{11}$$

for all $\alpha \in [0,1]$. We also derive an upper bound for $C_{N,L}(\boldsymbol{V})$ to show that of $N = \Theta\big(\log L\big)$ is necessary to achieve a rate $R(\alpha)$ for $\alpha \in [\beta, 1]$ for some $\beta \leq 1$.

A. Lower Bound

In order to do derive a lower bound on $C_{N,L}(\boldsymbol{V})$, we choose a possibly suboptimal communication scheme and find a lower bound on the rate achievable with this scheme. In particular, we assume that all nodes in the network use the same inner encoder $\boldsymbol{M}_E$ and the corresponding maximum likelihood decoder $\boldsymbol{M}_D$. This corresponds to using as intermediate processing $\boldsymbol{M}_i = \boldsymbol{M}_D \boldsymbol{M}_E$ for all i, that is, this scheme constructs an overall channel between source and destination equal to $\boldsymbol{Q}^L = (\boldsymbol{M}_E \boldsymbol{W} \boldsymbol{M}_D)^L$. The source A_0 uses an outer code over the channel $\boldsymbol{Q}^L$ and the destination A_L uses the corresponding maximum likelihood decoder.

Using random coding arguments [10], we know that there exist good codes (defined by the pair $(\boldsymbol{M}_E, \boldsymbol{M}_D)$) in the sense that the average probability of error is bounded by $\mathrm{P_e}(r) \le \exp\big(-NE(r)\big)$, where $E(r)$ is the random coding error exponent for the channel $\boldsymbol{V}$ as a function of the rate r. We use such a good code as our inner code. Roughly speaking, with probability $\big(1-\exp(-N\,E(r))\big)^L$ the sequence on N bits sent by the source traverses the L channels and arrives at the destination without errors. Hence, the average number of bits delivered error-free to the destination is approximately $r\big(1-\exp(-N\,E(r))\big)^L$. With this scheme we can show:

Theorem V.1 ([4]).

$$C_{N,L}(\boldsymbol{V}) \ge \max_{r\in[0,C(\boldsymbol{V})]} r\big(1-\exp\big(-NE(r)\big)\big)^L - \frac{1}{N}. \quad (12)$$

Furthermore, the RHS of (12) *tends to the min-cut capacity* $C(\boldsymbol{V})$ *as* $N \to \infty$.

B. Upper Bound

We derive here an upper bound for $C_{N,L}(\boldsymbol{V})$ expressed as a linear combination of the min-cut capacity $C(\boldsymbol{V})$ and of a term reminiscent of the limiting capacity $\frac{1}{N}\log M_0(\boldsymbol{V}^{\otimes N})$ derived in Theorem IV.1. The basic idea is to decompose $\boldsymbol{V}^{\otimes N}$, the equivalent channel transition matrix between every pairs of relays, into a linear combination of two stochastic matrices, one of which has rank as close as possible to $M_0(\boldsymbol{V}^{\otimes N})$.

Theorem V.2 ([5]). *For any stochastic matrix* $\boldsymbol{V}$ *and any integer* N*, if there exist two stochastic matrices* $\boldsymbol{A}_N$ *and* $\boldsymbol{B}_N$*, and* $\delta_N \in (0,1)$ *such that*

$$\boldsymbol{V}^{\otimes N} = \delta_N\,\boldsymbol{A}_N + (1-\delta_N)\,\boldsymbol{B}_N \quad (13)$$

then

$$C_{N,L}(\boldsymbol{V}) \le \Big(1-(1-\delta_N)^{L-1}\Big)\frac{\log \operatorname{rank}(\boldsymbol{A}_N)}{N} + (1-\delta_N)^{L-1} C(\boldsymbol{V}). \quad (14)$$

In order to obtain the tightest bound for any given N, $\boldsymbol{A}_N$ should be chosen to have the *smallest rank possible*. The choice $\boldsymbol{A}_N = \boldsymbol{A}_1^{\otimes N}$ does not give the best possible bound in general [2]. It can be shown [2] that, for any matrix $\boldsymbol{A}_N$ such that (13) holds, we have $C_0(\boldsymbol{V}) \le \frac{1}{N}\log\operatorname{rank}(\boldsymbol{A}_N)$. It is however not clear whether this inequality becomes an equality in the limit for $N \to \infty$.

It can also be shown [2] that, if for some N, we find a decomposition such that $\operatorname{rank}(\boldsymbol{A}_N) = M_0(\boldsymbol{V}^{\otimes N})$, then $C_0(\boldsymbol{V}) = \frac{1}{N}\log\operatorname{rank}(\boldsymbol{A}_N)$. In this case, the bound in (14) is tight in the limit as $L\to\infty$ and the decay of $C_{N,L}(\boldsymbol{V})$ to the limiting capacity is exponentially fast in L. In Theorem III.5 we showed that if we impose identical processing at the relays, then the limiting capacity is equal to the logarithm on the number of eigenvalues of modulus one of $\boldsymbol{Q} = \boldsymbol{W}\boldsymbol{M}$ and the limit is approached exponentially fast in L. If $\operatorname{rank}(\boldsymbol{A}_N) = M_0(\boldsymbol{V}^{\otimes N})$, then the exponential decay also applies to non identical processing. This implies that even for long, but not infinite, cascades the derived limiting result in Theorem IV.1 is meaningful. It also highlights the importance of well chosen intermediate processing.

There exists an interesting connection between the computation of the minimum rank $\boldsymbol{A}_N$, an instance of the *Set Cover Problem*, and the computation of $M_0(\boldsymbol{V}^{\otimes N})$, an instance of the *Maximum Independent Set Problem*. The relaxations of those two integer programs are in fact (strong) dual.

VI. Scaling Laws

In this section we characterize

$$N^\star(L,\alpha) \triangleq \min\ \{N : C_{N,L}(\boldsymbol{V}) \ge R(\alpha)\},$$

where $R(\alpha)$ is defined in (11) for $\alpha\in[0,1]$.

The next theorem uses Theorem V.1 to establish that logarithmic growth of N with L is sufficient to achieve $R(\alpha)$ for all for $\alpha \in [0,1]$.

Theorem VI.1 ([5]). *For every* $\varepsilon > 0$ *there is a* N_0 *such that for all* $N \ge N_0$

$$N^\star(L,\alpha) \le \min_{r\in[R(\alpha)+\varepsilon, C(\boldsymbol{V})]} \frac{1}{E(r)} \log\left(\frac{L}{1-\frac{R(\alpha)+\varepsilon}{r}}\right) \quad (15)$$

is sufficient to achieve $R(\alpha)$.

The next theorem uses Theorem V.2 to establish that logarithmic growth of N with L is sufficient to achieve $R(\alpha)$ for all $\alpha \ge \beta_m$ where β_m is a non-negative constant.

Theorem VI.2 ([5]).

$$N^\star(L,\alpha) \ge \frac{\log(L-1) - \log\log\frac{1}{\alpha-\beta_m}}{\frac{1}{m}\log\frac{1}{\delta_m}} \quad (16)$$

is necessary to achieve $R(\alpha)$ *for all*

$$\alpha \ge \beta_m \triangleq \frac{\frac{1}{m}\log\operatorname{rank}(\boldsymbol{A}_m) - C_0(\boldsymbol{V})}{C(\boldsymbol{V}) - C_0(\boldsymbol{V})}, \quad (17)$$

where m *is any integer such that the stochastic matrices* $\boldsymbol{A}_m$*,* $\boldsymbol{B}_m$*, and the real-valued constant* $\delta_m \in (0,1)$ *in the decomposition* $\boldsymbol{V}^{\otimes m} = \delta_m\boldsymbol{A}_m + (1-\delta_m)\boldsymbol{B}_m$ *satisfies* $\frac{1}{m}\log\operatorname{rank}(\boldsymbol{A}_m) \le C(\boldsymbol{V})$.

The bound in (16) would prove that logarithmic growth is necessary to achieve *all* rates above the zero-error capacity if we could show that $\lim_{m\to\infty}\beta_m = 0$, that is, $\lim_{N\to\infty}\log\operatorname{rank}(\boldsymbol{A}_N)/N = C_0(\boldsymbol{V})$.

Toward showing that logarithmic growth is indeed necessary to achieve all rates above the zero-error capacity we derive an upper bound for $C_{N,L}(\boldsymbol{V})$ that deals directly with the mutual information between the source and the destination. The idea is to partition the input alphabet into two sets. An upper bound on the mutual information is obtained by revealing to the destination to which of the two subsets the input that generated the observed output belongs to. Providing the destination with some feedback from the source clearly increases the achievable end-to-end rate.

Theorem VI.3 (NEW). *For every N and L, there exist a $\delta_{N,L} \in [0,1]$ and an $\eta_{N,L} > 0$ such that*

$$C_{N,L}(\boldsymbol{V}) \leq \delta_{N,L}\, C(\boldsymbol{V}) + (1-\delta_{N,L})\, \frac{1}{N} \log M_0(\boldsymbol{V}^{\otimes N}) + \eta_{N,L}$$

where

$$\delta_{N,L} \triangleq \prod_{\ell=2}^{L} p_{N,\ell}, \quad \text{and} \quad \eta_{N,L} \triangleq \frac{1}{N} \sum_{\ell=2}^{L} \Big(\prod_{j=\ell+1}^{L} p_{N,j} \Big)$$

for

$$p_{N,L} = \min_{\boldsymbol{P} \in \mathcal{P}_{N,L}, \mathcal{Z}_N} \boldsymbol{P}[\mathcal{Z}_N]$$

where $\mathcal{Z}_N$ denotes the set of codewords of a rate-optimal length-N zero-error code and $\mathcal{P}_{N,L}$ denotes the set of input distributions that achieve $C_{N,L}(\boldsymbol{V})$.

Theorem VI.3 would prove that logarithmic growth of N with L is necessary to achieve *all* rates above the zero-error capacity if we could show that $\delta \triangleq \sup_{\ell \geq 2} \{p_{N,\ell}\} < 1$. In this case in fact, the bound in Theorem VI.3 reduces to

$$\delta^{L-1}\, C(\boldsymbol{V}) + (1-\delta^{L-1})\, \frac{1}{N} \log M_0(\boldsymbol{V}^{\otimes \boldsymbol{N}}) + \frac{1}{N} \frac{1-\delta^L}{1-\delta}$$

and $\frac{1}{N}\frac{1-\delta^L}{1-\delta} \to 0$ as $N \to \infty$. Verifying that $\delta < 1$ seems quite hard, as the evaluation of $\{p_{N,L}\}_{N,L}$ requires the knowledge of $\mathcal{P}_{N,L}$ and $\mathcal{Z}_N$.

Proving either $\lim_{N\to\infty} \log \operatorname{rank}(\boldsymbol{A}_N)/N = C_0(\boldsymbol{V})$ or $\delta < 1$ is on-going work.

VII. The Pentagon Channel Example

Consider the "pentagon" channel whose transition matrix $\boldsymbol{V}$ and the corresponding graph $G(\boldsymbol{V})$ are depicted in Figure 2.

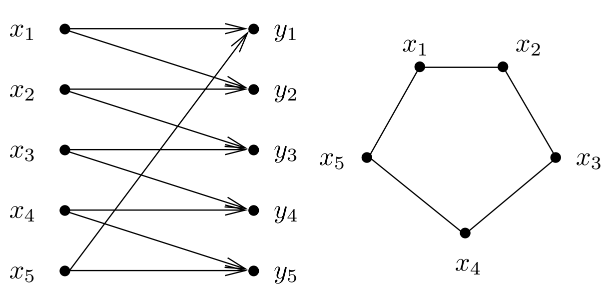

Fig. 2. The "pentagon" channel $\boldsymbol{V}$ and its graph $G(\boldsymbol{V})$.

For this channel, we have $M_0\big(\boldsymbol{V}\big) = 2$ (for example x_1 and x_3 are non adjacent), $M_0\big(\boldsymbol{V}^{\otimes 2}\big) = 5$ (for example (x_1, x_1), (x_2, x_3), (x_3, x_5), (x_4, x_2), (x_5, x_4) are non adjacent). It can be shown [9] that

$$C_0(\boldsymbol{V}) = \frac{1}{2} \log 5$$

that is, a zero-error code of blocklength 2 is optimal.

Theorem IV.1 states that for an infinite cascade of "pentagon" channels

$$\lim_{L\to\infty} C_{1,L}(\boldsymbol{V}) = \log 2, \quad \text{and} \quad \lim_{L\to\infty} C_{2,L}(\boldsymbol{V}) = \frac{1}{2} \log 5.$$

Moreover, for any other finite N, we can never achieve more than $\frac{1}{2}\log 5$. Hence, for an infinite cascade of "pentagon" channels an intermediate processing of length $N = 2$ is optimal if N is restricted to be finite. The optimal limiting capacity can be achieved by using at all intermediate nodes and at the source a length-2 zero-error encoder, and at all intermediate nodes and at the destination a length-2 zero-error decoder. Notice that in this example, if the intermediate nodes simply forward the incoming data, then the limiting capacity is (by Theorem III.4)

$$\lim_{L\to\infty} C(\boldsymbol{V}^L) = \log D(\boldsymbol{V}) = \log 1 = 0,$$

and this limit is approached exponentially fast with exponent (by Theorem III.5 in the case $\boldsymbol{Q} = \widehat{\boldsymbol{Q}}$)

$$-2\log|\lambda_2(\boldsymbol{V})| = -\log(1 - 2p(1-p)\sin^2(2\pi/5)),$$

where $p = \Pr[y_{i+1}|x_i]$, $i = 1, ..., 5$. In other words, intermediate processing is *necessary* if a non vanishing throughput is to be achieved in a long line network. Even a non-trivial one-symbol processing suffices to achieve a strictly positive end-to-end rate.

About the matrix decomposition in Theorem V.2, we can find $\operatorname{rank}(\boldsymbol{A}_1) = 3$ and $\operatorname{rank}(\boldsymbol{A}_2) = 8 < \operatorname{rank}(\boldsymbol{A}_1)^2 = 9$. This shows that the choice $\boldsymbol{A}_N = \boldsymbol{A}_1^{\otimes N}$ does not give the best possible bound for $C_{N,L}(\boldsymbol{V})$ in (14) in general. However $\beta_2 > 0$, hence we cannot state that logarithmic growth is necessary in the range $\alpha < \beta_2$.

VIII. Conclusions

In this paper, we investigated communication through a cascade of L channels, where intermediate nodes can perform processing on blocks of N symbols. When N is fixed and L goes to infinity, we showed that the network capacity can never exceed the zero-error capacity of the underlying channel. On the other hand, when $N \to \infty$ the network capacity coincides with the min-cut capacity. By deriving bounds on the capacity of finite length cascades, we showed that $N = \Theta(\log L)$ is sufficient to achieve any rate below the min-cut capacity and necessary to achieve certain rates above the zero-error capacity. We conjecture that logarithmic growth is actually necessary to achieve *any* rate above the zero-error capacity.

References

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

[2] U. Niesen, C. Fragouli, D. Tuninetti, "On Capacity of Line Networks", *submitted to IEEE Trans. on Info. Theory on Dec. 2005*.

[3] U. Niesen, C. Fragouli, D. Tuninetti, "On the Capacity of an Infinite Cascade of Channels", *Allerton 2005*, Monticello, IL (USA) Sept. 2005.

[4] D. Tuninetti, C. Fragouli, "Processing along the way: forwarding vs. coding", *ISITA 2004*, Parma (Italy) Oct. 2004.

[5] U. Niesen, C. Fragouli, D. Tuninetti, "Scaling Laws for Line Networks: From Zero-Error to Min-Cut Capacity", *submitted to ISIT 2006*.

[6] P. Gupta, P. R. Kumar, "The Capacity of Wireless Networks", *IEEE Trans. on Info. Theory*, Vol. 46(2), March 2000.

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

[8] C. E. Shannon, "The Zero Error Capacity of a Noisy Channel", *IEEE Trans. on Info. Theory*, Vol. 2(3), Sept. 1956.

[9] J. Körner, A. Orlitsky, "Zero-Error Information Theory", *IEEE Trans. on Info. Theory*, Vol. 44(6), Oct. 1998.

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