Variable-Length Coding for Zero-Error Channel Capacity

Nicolas Charpenay, Mael Le Treust

HAL Id: hal-03334091
https://hal.science/hal-03334091

Variable-Length Coding for Zero-Error Channel Capacity

Nicolas Charpenay[*] and Maël Le Treust[†]

ETIS UMR 8051, Université Paris Seine, Université Cergy-Pontoise, ENSEA, CNRS,

6, avenue du Ponceau, 95014 Cergy-Pontoise CEDEX, FRANCE

Email: {nicolas.charpenay ; mael.le-treust}@ensea.fr

Abstract

The zero-error channel capacity is the maximum asymptotic rate that can be reached with error probability exactly zero, instead of a vanishing error probability. The nature of this problem, essentially combinatorial rather than probabilistic, has led to various researches both in Information Theory and Combinatorics. However, the zero-error capacity is still an open problem, for example the capacity of the noisy-typewriter channel with 7 letters is unknown. In this article, we propose a new approach to construct optimal zero-error codes, based on the concatenation of words of variable-length, taken from a generator set. Three zero-error variable-length coding schemes, referred to as "variable-length coding", "intermingled coding" and "automata-based coding", are under study. We characterize their asymptotic performances via linear difference equations, in terms of simple properties of the generator set, e.g. the roots of the characteristic polynomial, the spectral radius of an adjacency matrix, the inverse of the convergence radius of a generator series. For a specific example, we construct an "intermingled" coding scheme that achieves asymptotically the zero-error capacity.

Index Terms

Zero-Error Information Theory, Channel Coding, Analytic Combinatorics, Graph Theory, Combinatorics on Words, Linear Difference Equation, Automata

I. INTRODUCTION

In [1], Shannon investigates the zero-error information transmission by considering codes that must allow for a correct decoding with probability one, instead of an asymptotic probability one. This subtle difference radically changes the nature of the problem, as the exact values of the non-null transition probabilities of the channel do not appear anymore. Shannon defined the *zero-error capacity* of a channel as the maximum asymptotic rate that can be reached with error probability exactly zero. The characterization of the zero-error capacity of an arbitrary channel is a wide open problem, that shares deep connections with Graph Theory. Equivalently, the zero-error capacity

is the asymptotic limit of the independence number of iterated strong products of channel graphs. This problem inspired Berge's notion of perfect graphs [2, Chap 16], for which the zero-error capacity is given by the one-shot independence number [3, Theorem 4.18]. Over time, this open problem has attracted a lot of attention both from the Information Theory and Combinatorics communities, see [4, Chap. 27].

A. Motivations

The zero-error capacity problem has several applications. In data center storage systems, the automatic treatment a large amount of data imposes reliability constraints which do not support any positive probability of error, even if arbitrarily small. In [5], Kovačević investigated the zero-error capacity, for the duplication channels in *DNA-based data storage systems*, and in [6], [7], for the timing channels in *molecular communication*.

The analysis of the zero-error capacity problem provides an important insight into the exponential decrease of the error probability for the classical channel coding problem, see the discussion in [8, pp. 203]. In particular, it shares deep connections with the channel coding problem in the *finite-blocklength regime* [9], of interests for the transmission of short packet in *IoT networks*. For example, the error-exponent goes to infinity when the coding rate approaches the largest rate of a zero-error code [10], [11].

When considering channel with memory [12], the zero-error capacity problem covers the problems of *coding for error-correction* [13]. In [14], [15], Dalai uses the zero-error capacity tools in order to bound on the minimum distance of codes. In [16], Bose et al. describe codes which achieve zero error capacities in limited magnitude error channels.

B. Tools and bounds for zero-error capacity

In [1], Shannon gives a sufficient condition to determine the zero-error capacity, based on the existence of an adjacency-reducing mapping for the channel graph. This condition works well to almost all channels with 5 symbols or less, and boils down to specific instances of tesselation covers, studied by Abreu et al. in [17]. In [18], Lovasz gives bounds on the zero-error capacity with the well-known θ function, and determines the capacity for the specific class of auto-complementary and vertex-transitive channel graphs. The θ number possesses the property of being multiplicative with respect to the strong product. Another graph invariant, called the Rosenfeld number [19], also presents this property. In [20], Hales studies this number and shows that Shannon's adjacency-reducing mappings condition in [1] is not necessary. In [21], Haemers proves another upper bound on the zero-error capacity of a channel, based on the rank of the adjacency matrix of the channel graph. In [22], Alon builds a counterexample for the zero-error capacity of a union of channels, disproving the conjecture formulated by Shannon in [1], which states that the zero-error capacity, defined without the logarithm, is additive with respect to disjoint union of channels. In [23], the authors study a variant of the zero-error capacity problem in which the strong product of graphs is replaced by the direct product. Instead of considering the independence number of product graphs, Hahn et al. study in [24], the ratio between the independence number and the number of vertices, and they derive bounds based on the chromatic and the fractional chromatic numbers. In [25], Körner and Orlitsky give a review of the literature on zero-error capacity and its variants.
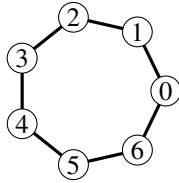
Fig. 1: The zero-error capacity of the graph C_7, corresponding to the noisy-typewriter channel with 7 letters, is unknown. Some recent lower and upper bounds are stated in [26].

C. About cycles

In [27], Gallai extends Shannon's condition to bipartite channel graphs, that is a graph without odd cycles, leading to further interests in determining the zero-error capacity for odd cycles. The product of odd cycles of different sizes is investigated by Sonnemann and Krafft, in [28], and by Vesel, in [29], while Bohman and Holzman study the complementary graphs of odd cycles, in [30]. The channel graph with odd cycles of 7 vertices, denoted by C_7, is depicted in Fig. 1. Its zero-error capacity is still unknown, despite several attempts to build zero-error codes on these channels, see [31], [32], and [26]. In Fig. 2, we depict the best known bounds on the zero-error rate for small number of channel uses, for the channel graphs C_7 and C_9. An other interpretation of the zero-error capacity problem for cycles is the tiling problem, studied in [33], whose solutions also provide upper bounds for every finite number of channel uses. In [34], Badalyan and Markosyan determine the maximum independent sets of products of cycles-powers, that are cycles with edges added towards the vertices of distance at most k. In [35] and [36], Bohman characterizes the asymptotic zero-error capacity of odd cycles, when the size of the cycle goes to infinity. In [37], Mathew and Östergård improve several lower bounds on the capacities of odd cycles using stochastic search methods.

The graphs with odd cycles are also related to Berge's conjecture [38], later proved in [39] by Chudnovsky et al., namely "a graph G is perfect if and only if either G or its complementary graph $\bar{G}$, have odd cycles of length 5 or more". Since the zero-error capacity of the cycle graph C_5 is known, as well as the zero-error capacity for perfect graph, see [3, Theorem 4.18], C_7 is the minimal connected graph for which the zero-error capacity is still an open problem.

D. Variable-length coding

In this paper, we investigate the zero-error capacity by constructing variable-length coding schemes. In [40], Shannon determines the asymptotic performances of variable-length coding via the characteristic polynomial of linear difference equations. In [41], Weidmann et al. study variable-length arithmetic coding scheme for a joint source-channel coding scenario. In [42], Flajolet and Sedgewick investigate rational expressions and variable-length coding schemes through their respective generating series. In [43], Devroye propose a zero-error variable-length communication scheme, assuming that the transmitter has a perfect, but rate limited channel feedback. In [44], Guo and Watanabe present a family of graph where no finite-length code can achieve the zero-error capacity.
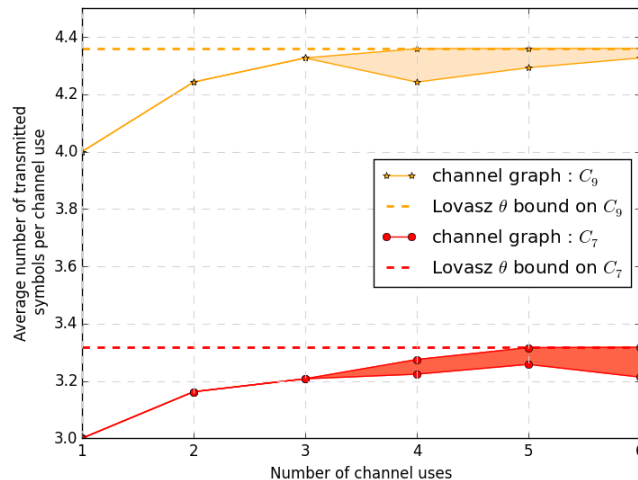
Fig. 2: Best known lower and upper bounds on the maximum zero-error rate achievable for small number of channel uses, see [18], [26], [30]–[33], [37]

E. Recent information-theoretic literature

In [45]–[47], Devroye et al. investigate the zero-error capacity of the primitive relay channel, by proposing a one-shot relaying scheme, termed color-and-forward. They highlight the connection to the zero-error source coding problem with receiver's side information, studied by Witsenhausen in [48]. In [49], the authors define a new notion of product of graph that allows to compute recursively the optimal relaying scheme. Several multi-user channels, such as the relay, the multiple-access, the broadcast, and the interference channels, are investigated in [50], where necessary and sufficient conditions regarding the positivity of the zero-error capacity are provided. The zero-error capacity with noisy channel feedback is studied in [51] and [52], where dynamic programming provide lower and upper bounds.

In [53], [54], Wang and Shayevitz investigate the combination of zero-error source and channel coding schemes, by introducing the notion of "graph information ratio", also related to the relative Shannon capacity of two graphs, introduced by Körner and Marton in [55]. In [56], Shayevitz investigates the zero-error broadcasting problem by introducing the ρ-capacity function, for which upper and lower bounds are derived. In [57], Ordentlich and Shayevitz investigate the zero-error capacity region of the multiple access channel, called the binary adder. They provide a new outer bound that strictly improves upon the bound obtained by Urbanke and Li, in [58].

In [59]–[61], Wiese et al. define the zero-error wiretap codes by requiring that every output at the eavesdropper can be generated by at least two inputs. They define the zero-error secrecy capacity as the supremum of rates for which there exists a zero-error wiretap code, and they show it either equals zero or the zero-error capacity of the channel between the encoder and the legitimate receiver.

In [62], Ruiz and Pérez-Cruz construct linear codes over rings, and provide a lower bound on the zero-error capacity for the noisy-typewriter channel with odd letters of the form $2^n + 1$, that outperforms Bohman's bound in

[30]. In [63], Cullina et al. introduced a different notion of product of graph by removing edges between sequences which differ in more than d positions. They provide upper and lower bound on the asymptotic independence number of such an iterated product of graph. In [64], Dalai improves the bound on the zero-error list-decoding capacity, introduced by Elias in [65]. In [66], Xu and Radziszowski study the construction of lower bounds for multicolor Ramsey numbers of product graphs, and their relation to the zero-error capacity. In particular, the authors prove that the supremum of the zero-error capacity over all graphs with independence number equal to 2, cannot be achieved by any finite graph power.

F. Scenarios and contributions

In this paper, we design three coding algorithms that are based on a generator set of zero-error words, refereed to as variable-length, intermingled, and automata-based coding schemes. We characterize their respective asymptotic performances via the root of the characteristic polynomial, in Theorem III.4, the spectral radius of the adjacency matrix, in Theorem IV.5, the inverse of the convergence radius of the generator series, in Theorem VI.10.

- *Variable-length* coding scheme, in Sec. III, produces channel inputs sequences by concatenation of zero-error words from a generator set.
- *Intermingled* coding scheme, in Sec. IV, allows to stop the transmission and switch to another word from the generator set. Thus, additional information is embedded over the positions of such a stops and switches.
- *An example* with the channel graph $C_5 \boxplus \mathbf{1}$, is stated in Sec. V. We construct explicitly the generator set of zero-error words, for which the asymptotic rate of the intermingled coding achieves the zero-error capacity.
- *Automata-based* coding scheme, in Sec. VI, generalizes the two previous algorithms by allowing multiple interleaving of the same word from the generator set.

The paper is organized as follows. The definitions of the zero-error channel capacity and of the maximum independence number of the product graph of the channel are stated in Sec. II. The variable-length coding and the intermingled coding are studied in Sec. III and IV. In Sec V, we provide an example based on the channel graph $C_5 \boxplus \mathbf{1}$. The automata-based coding scheme is investigated in Sec. VI. The proofs are stated in the Appendices.

II. PRESENTATION OF THE MODEL

A. Notations

- ◆ Given a finite set $\mathcal{A}$, we denote by $\mathcal{P}(\mathcal{A})$ its power set, and $\#A$ its cardinality.
- ◆ We use the following notations for matrices slicing. Given a matrix $M \in \mathcal{M}_{n,p}(\mathbb{R})$ we define

$$M_{:,j} \doteq \begin{pmatrix} M_{1,j} \\ \vdots \\ M_{n,j} \end{pmatrix} \text{ and } M_{i,:} \doteq \begin{pmatrix} M_{i,1} & \cdots & M_{i,p} \end{pmatrix}. \tag{1}$$

We use the same notation for tuples or words : let $w = x_1 \cdots x_{|w|}$ be a word over the alphabet $\mathcal{X}$, we note $|w|$ the length of w and we define

$$w_{i:} = x_i x_{i+1} ... x_{|w|} \quad \text{and} \quad w_{:j} = x_1 ... x_j \tag{2}$$

◆ We note the support of a vector supp, that is the set of the indexes of its non-null components.

◆ Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $B \in \mathcal{M}_{p,q}(\mathbb{R})$ be two matrices, we note $A \otimes B \in \mathcal{M}_{mp,nq}(\mathbb{R})$ their Kronecker product, that is :

$$A \otimes B \doteq \left(A_{i_1,j_1} B_{i_2,j_2} \right)_{i \in [\![1,m]\!] \times [\![1,p]\!], j \in [\![1,n]\!] \times [\![1,q]\!]} \tag{3}$$

and we also note $A^{\otimes L} \doteq A \otimes ... \otimes A$ (L times).

B. Zero-error capacity

We consider a Discrete Memoryless Channel (DMC) where $\mathcal{X}$ denotes the input alphabet, $\mathcal{Y}$ denotes the output alphabet, and $W = (W_{x,y})_{x \in \mathcal{X}, y \in \mathcal{Y}}$ denotes the transition probabilities.

Definition II.1 (Channel graph) *The channel graph $G_W \doteq \big(\mathcal{V}(G_W), \mathcal{E}(G_W) \big)$ is defined by the input alphabet as set of vertices $\mathcal{V}(G_W) = \mathcal{X}$, and $xx' \in \mathcal{E}(G_W)$ if $W_{x,y} > 0$ and $W_{x',y} > 0$ for some output $y \in \mathcal{Y}$.*

Two inputs x and x' are distinguishable if they satisfy $W_{x,y} = 0$ or $W_{x',y} = 0$ for all input y, equivalently

$$x \text{ and } x' \text{ are distinguishable} \iff \max_{y \in \mathcal{Y}} \min(W_{x,y}, W_{x',y}) = 0 \iff xx' \notin \mathcal{E}(G_W). \tag{4}$$

A family of inputs is distinguishable if its elements are pairwise distinguishable.

The channel graph is the main tool for the characterization of the zero-error capacity, that is the maximum asymptotic number of bits that can be transmitted with zero error. Before stating the definition, we introduce Fekete's Lemma, see [67, Lemma 11.6, pp. 103].

Lemma 1 (Fekete) *For all superadditive sequence $(u_l)_{l \in \mathbb{N}}$, i.e. $u_{l+l'} \geq u_l + u_{l'}$ for all (l, l'), $\lim\limits_{l \to \infty} \frac{u_l}{l}$ exists and is equal to $\sup\limits_{l} \frac{u_l}{l}$ (it can be $+\infty$).*

Definition II.2 (Zero-error capacity) *Let $N(W, L)$ be the maximum number of distinguishable inputs for L uses of the channel W. Equivalently, $N(W, L)$ is the value of the optimization problem*

$$\max_{\mathcal{N} \subseteq \mathcal{X}^L} \quad \#\mathcal{N} \tag{5}$$

$$s.t. \quad \max_{y \in \mathcal{Y}^L} \# \mathrm{supp} \left(W_{x,y}^{\otimes L} \right)_{x \in \mathcal{N}} \leq 1. \tag{6}$$

Then the zero-error capacity of W is defined by

$$C_0(W) \doteq \limsup_{L \to \infty} \frac{1}{L} \log \big(N(W, L) \big) \tag{7}$$

$$\underset{[\textit{Fekete}]}{=} \sup_{L} \frac{1}{L} \log \big(N(W, L) \big), \tag{8}$$

where the second equality comes from Fekete's lemma as the sequence $\log N(W, \cdot)$ is superadditive for all W. Indeed, for $L+L'$ channel uses, there exists at least $N(W, L) \cdot N(W, L')$ distinguishable inputs, by using successively the codebook for L' channel uses and the codebook for L channel uses.

C. Maximum independent set of the strong product channel graph

The zero-error capacity relates to the graph theoretic notions of *strong product* and *maximum independent set*.
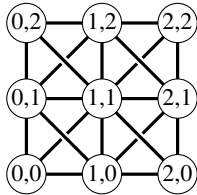
Definition II.3 (Strong product ⊠) *Let* $G \doteq (\mathcal{V}(G), \mathcal{E}(G))$ *and* $G' \doteq (\mathcal{V}(G'), \mathcal{E}(G'))$, *we define the strong product, i.e. the AND product,* $G \boxtimes G' \doteq (\mathcal{V}(G \boxtimes G'), \mathcal{E}(G \boxtimes G'))$ *by*

$$\mathcal{V}(G \boxtimes G') \doteq \mathcal{V}(G) \times \mathcal{V}(G'), \tag{9}$$

$$\forall (v_1, v_1') \neq (v_2, v_2'), \ (v_1, v_1')(v_2, v_2') \in \mathcal{E}(G \boxtimes G') \ if \tag{10}$$

$$\begin{pmatrix} v_1 v_2 \in \mathcal{E}(G) \\ \\ or \ v_1 = v_2 \end{pmatrix} AND \begin{pmatrix} v_1' v_2' \in \mathcal{E}(G') \\ \\ or \ v_1' = v_2' \end{pmatrix}.$$

As an example, we consider two channel graphs $G = G' = $ ⓪——①——②, then the product graph $G \boxtimes G'$ is the king's graph, corresponding to two channel uses.

We denote by $G_W^{\boxtimes L} = G_W \boxtimes ... \boxtimes G_W$, the L-times iterated strong product, and we give several equivalent interpretations.

- ◆ $(x_l)_{l \leq L}$ and $(x_l')_{l \leq L}$ are not distinguishable.
- ◆ For all $l \leq L$, there exists y_l such that $W_{x_l, y_l} > 0$ and $W_{x_l', y_l} > 0$.
- ◆ $xx' \in \mathcal{E}(G_W^{\boxtimes L})$.
- ◆ $\langle W_{x,:}^{\otimes L}, W_{x',:}^{\otimes L} \rangle > 0$.

Definition II.4 (Disjoint union ⊞) *Given two graphs* $G = (\mathcal{V}(G), \mathcal{E}(G))$ *and* $G' = (\mathcal{V}(G'), \mathcal{E}(G'))$, *we define* $G \boxplus G' = (\mathcal{V}(G \boxplus G'), \mathcal{E}(G \boxplus G'))$ *to be the disjoint union between* G *and* G', *that is :*

$$\mathcal{V}(G \boxplus G') = \mathcal{V}(G) \cup \mathcal{V}(G'), \tag{11}$$

$$vv' \in \mathcal{E}(G \boxplus G') \quad if \quad \begin{pmatrix} v, v' \in \mathcal{V}(G) \\ \\ and \ vv' \in \mathcal{E}(G) \end{pmatrix} OR \begin{pmatrix} v, v' \in \mathcal{V}(G') \\ \\ and \ vv' \in \mathcal{E}(G') \end{pmatrix}. \tag{12}$$

Remark II.5 *Since the set of finite graphs with the laws* ⊞, ⊠ *have a semiring structure, we denote* **1** *the graph with one vertex and* **0** *the graph with zero vertex.*

Definition II.6 (Independent set) *An independent set* $\mathcal{S}$ *is a subset of* $\mathcal{V}(G)$ *such that* $\forall s, s' \in \mathcal{S}, \ ss' \notin \mathcal{E}(G)$.

Definition II.7 (Independence number α**)** *The independence number of a graph* G *is the size of the largest independent set of* G. *It is denoted by* $\alpha(G)$.

Proposition II.8 *The maximum number of distinguishable inputs $N(W, L)$ is the independence number $\alpha(G_W^{\boxtimes L})$ of the product graph $G_W^{\boxtimes L}$.*

Thus it makes sense to work directly on channel graphs, independently of the channel transition probabilities $W = (W_{x,y})_{x \in \mathcal{X}, y \in \mathcal{Y}}$, that generated the channel graph.

III. VARIABLE-LENGTH CODING

In this section, we introduce a variable-length coding scheme based on a generator set of words, tailored for the zero-error transmission. We characterize the asymptotic coding rate via the unique positive root of the characteristic polynomial of the generator set.

Definition III.1 (Set of words) *For a given finite set $\mathcal{S}$, we define the set of words over $\mathcal{S}$ by*

$$\mathcal{S}^* \doteq \bigcup_{l \in \mathbb{N}} \mathcal{S}^l, \tag{13}$$

with the usual concatenation law, i.e. the neutral element for that law is the empty word, denoted by ϵ. The length of a word $w \in \mathcal{S}^*$ is the integer l such that $w \in \mathcal{S}^l$ and is denoted by $|w|$. For a given subset $\mathcal{S}' \subseteq \mathcal{S}^*$, and integers l, l' such that $l \le l'$, we define

$$\mathcal{S}'_{[l]} \doteq \{w \in \mathcal{S}' \mid |w| = l\}, \tag{14}$$

$$\mathcal{S}'_{[l:l']} \doteq \{w \in \mathcal{S}' \mid l \le |w| \le l'\}. \tag{15}$$

The variable-length coding is based on the following idea, instead of determining the maximum number of distinguishable inputs over L channel uses when L goes to infinity, we consider codes in $\mathcal{X}^*$ and we determine the asymptotic number of transmitted symbols per channel use.

Definition III.2 (Generator set) *The* generator set *is a finite subset $\mathcal{C}$ of $\mathcal{X}^*$ composed of words of variable length. The generator set $\mathcal{C}$ is zero-error for the channel W if*

$$\forall c, c' \in \mathcal{C} \text{ such that } c \ne c', \ |c| \le |c'|, \ \text{and } cc'_{:|c|} \notin \mathcal{E}(G_W^{\boxtimes|c|}), \tag{16}$$

with the convention that vertices in $G_W^{\boxtimes|c|}$ are auto-adjacent. At the end of its transmission, the word $c \in \mathcal{C}$ is distinguishable from any other word $c' \in \mathcal{C}$.

The variable-length coding scheme produces the channel input sequences by concatenating the words from the generator set $\mathcal{C}$. The set of all possible channel input sequences of length L, constructed with such a procedure, is denoted by $\mathcal{C}^*_{[L]}$. The zero-error property extends naturally from the generator set $\mathcal{C}$ to the set of channel input sequences $\mathcal{C}^*_{[L]}$.

Definition III.3 (Asymptotic rate of variable-length codes) *We consider the generator set $\mathcal{C}$ that is zero-error for the channel W. The asymptotic rate of $\mathcal{C}$ is defined by*

$$r(\mathcal{C}) \doteq \lim_{L \to \infty} \frac{1}{L} \log \#\mathcal{C}^*_{[L]} \underset{[Fekete]}{=} \sup_{L \in \mathbb{N}} \frac{1}{L} \log \#\mathcal{C}^*_{[L]}. \tag{17}$$

The average number of transmitted symbols per channel use is defined by

$$\nu(\mathcal{C}) \doteq \lim_{L\to\infty} \sqrt[L]{\#\mathcal{C}^*_{[L]}} \underset{[Fekete]}{=} \sup_{L\in\mathbb{N}} \sqrt[L]{\#\mathcal{C}^*_{[L]}} = 2^{r(\mathcal{C})}. \tag{18}$$

We can apply Fekete's lemma only if the values of $\log \#\mathcal{C}^*_{[l]}$ are finite for all l large enough, i.e. if and only if $\gcd(|c|, c \in \mathcal{C}) = 1$. *When* $\gcd(|c|, c \in \mathcal{C}) = d \neq 1$, *we define the rate as*

$$r(\mathcal{C}) \doteq \lim_{L\to\infty} \frac{1}{dL} \log \#\mathcal{C}^*_{[dL]}, \tag{19}$$

and we take again $\nu(\mathcal{C}) \doteq 2^{r(\mathcal{C})}$ *with this new definition.*

The asymptotic rate $r(\mathcal{C})$ corresponds to the asymptotic number of bits transmitted per channel use by concatenating the variable-length words from the generator set $\mathcal{C}$. For each generator set $\mathcal{C}$, we have $r(\mathcal{C}) \le C_0(W)$ if $\mathcal{C}$ is zero-error.

Theorem III.4 (Rate computation of variable-length codes) *Let W be a DMC and $\mathcal{C} \subseteq \mathcal{X}^*$ a generator set that is zero-error for the channel W. We denote by $\bar{l} > 0$, (resp. $\underline{l} > 0$), the maximal length, (resp. the minimal length), of the words in $\mathcal{C}$. Then $\nu(\mathcal{C})$ is the unique positive solution of the characteristic polynomial*

$$X^{\bar{l}} - \sum_{l=\underline{l}}^{\bar{l}} \#\mathcal{C}_{[l]} X^{\bar{l}-l} = 0, \tag{20}$$

where $\mathcal{C}_{[l]} = \{c \in \mathcal{C} \mid |c| = l\}$.

The proof of Theorem III.4 is stated in App. B and relies on standard properties of linear difference equations, pointed out by Shannon in [40, Part I] for discrete noiseless systems, see also [68, pp. 13]. In Sec. V, we provide an example based on a channel graph for which we compute explicitly the asymptotic rate $r(\mathcal{C})$ of the variable-length code.

Remark III.5 (Linear difference formulation) *We provide an equivalent formulation for the average number of transmitted symbols per channel use, for a variable-length code with generator set $\mathcal{C}$. Given $\#\mathcal{C}^*_{[l]}$ for $l \in]\!] L-\bar{l}, L]\!]$, one can compute $\#\mathcal{C}^*_{[L+1]}$ as a linear combination of the $(\#\mathcal{C}^*_{[l]})_{l\in]\!]L-\bar{l},L]\!]}$, i.e.*

$$\#\mathcal{C}^*_{[L]} = \sum_{l=\underline{l}}^{\bar{l}} \#\mathcal{C}_{[l]} \#\mathcal{C}^*_{[L-l]}. \tag{21}$$

In particular, the companion matrix of the characteristic polynomial (20)

$$M = \begin{pmatrix} 0 & 1 & 0 & & (0) \\ & 0 & 1 & \ddots & \\ & & \ddots & \ddots & 0 \\ (0) & & & 0 & 1 \\ \#\mathcal{C}_{[\bar{l}]} & \#\mathcal{C}_{[\bar{l}-1]} & \cdots & \#\mathcal{C}_{[2]} & \#\mathcal{C}_{[1]} \end{pmatrix}, \tag{22}$$

satisfies

$$M(\#\mathcal{C}^*_{[l]})_{l\in]\!]L-\bar{l},L]\!]} = (\#\mathcal{C}^*_{[l+1]})_{l\in]\!]L-\bar{l},L]\!]}. \tag{23}$$

Then the asymptotic rate corresponds to the dominant term after computing the Jordan decomposition of M.

Note that the number of generated words for any bounded length window $[\![L - l, L]\!]$ behaves equivalently asymptotically, for all $l \in \mathbb{N}$ we have

$$\sqrt[L]{\#\mathcal{C}^*_{[L-l:L]}} \underset{L\to\infty}{\to} \nu(\mathcal{C}). \tag{24}$$

Proposition III.6 $C_0(W) = \sup_{\mathcal{C}} \log \nu(\mathcal{C})$.

Proof. [Proposition III.6] On one hand, variable length codes cannot achieve a better rate than $C_0(W)$. On the other hand, the fixed length codes, i.e. subsets of $\mathcal{X}^L$, are included in the set of variable length codes, so we have

$$C_0(W) \geq \sup_{\mathcal{C}} \log \nu(\mathcal{C}) \geq \sup_{L} \frac{\log \alpha(G_W^{\boxtimes L})}{L} = C_0(W). \tag{25}$$

□

Thus finding an optimal variable-length code is equivalent to finding a family of fixed-length codes with optimal supremum rate. For some scenario, a variable-length code might be far easier to describe than an infinite family of fixed-length codes.

IV. INTERMINGLED CODING

In this section, we generalize the previous variable-length coding scheme, by revisiting Shannon's intermingled coding scheme [1, proof of Th. 4]. Instead of allowing only the concatenation of words taken from the generator set $\mathcal{C} \subseteq \mathcal{X}^*$, the intermingled coding scheme allows to stop the transmission of one word $c \in \mathcal{C}$ and resume it later, in an "intermingling" pattern, so that additional information is embedded on positions of such stops.

Definition IV.1 (Intermingled codes) *An intermingled code is composed of a generator set $\mathcal{C} \subseteq \mathcal{X}^*$ that is zero-error for the channel W, and a succession rule*

$$\rho : \prod_{c \in \mathcal{C}} [\![0, |c| - 1]\!] \to \mathcal{P}(\mathcal{C}), \tag{26}$$

such that $\rho(z)$ is nonempty for all $z \in \prod_{c \in \mathcal{C}} [\![0, |c| - 1]\!]$, which maps the possible transmission states to the possible transmittable characters at the next time step. The encoder chooses a time horizon L, then maps the message to some sequence $(x_l)_{l \leq L} \in \mathcal{X}^L$ based on the following algorithm :

- ◆ *Initialize $z \leftarrow (0, ..., 0)$ vector of size $\#\mathcal{C}$ and $l \leftarrow 0$*
- ◆ *While $l \leq L$:*
  - ◇ *Choose a word $w \in \rho(z)$ $(\subseteq \mathcal{C})$*
  - ◇ *Transmit $x_l = w_{z_w+1}$ over the channel*
  - ◇ $z_w \leftarrow z_w + 1 \mod |w|$
  - ◇ $l \leftarrow l + 1$

An intermingled code $(\mathcal{C}, \rho)$ is zero-error for the channel W if all the possible sequences generated by this algorithm are distinguishable, i.e. for all generated codewords x and x' from $\mathcal{X}^L$, there exists a time step $l \leq L$ such that $x_l x'_l \notin \mathcal{E}(G_W)$.

Remark IV.2 *The choice :*

$$\rho : z \mapsto \begin{cases} \mathcal{C} & \text{if } z = (0, ..., 0), \\ \{c \in \mathcal{C} \mid z_c \neq 0\} & \text{otherwise,} \end{cases} \tag{27}$$

corresponds to the variable-length coding scheme of Sec. III, based on the simple concatenation of words. The variable-length codes are a subclass of intermingled codes.

Definition IV.3 (Asymptotic rate of intermingled codes) *The asymptotic rate of an intermingled code $(\mathcal{C}, \rho)$ is defined by*

$$r(\mathcal{C}, \rho) \doteq \lim_{L \to \infty} \frac{1}{L} \log \# \mathcal{S}_L, \tag{28}$$

where $\mathcal{S}_L$ denotes the set of channel input sequences $x \in \mathcal{X}^L$ that are generated by the algorithm described in Definition IV.1 with time horizon L, with $z_w = (0, ..., 0)$ at the last time step. We also define the average number of transmitted symbols per channel use :

$$\nu(\mathcal{C}, \rho) \doteq 2^{r(\mathcal{C}, \rho)} \tag{29}$$

The existence of the limit is given by Fekete's lemma, as $\log \# \mathcal{S}_{L+L'} \geq \log \# \mathcal{S}_L + \log \# \mathcal{S}_{L'}$ for all $L, L' \in \mathbb{N}$. Similarly to the Definition III.3, if there exists a nonempty maximal family of integers $(k_i)_{i \in \mathcal{I}}$ such that $d = \gcd(k_i, \ i \in \mathcal{I}) \neq 1$ with $\# \mathcal{S}_{k_i L} = 0$ for all $i \in \mathcal{I}$ and $L \in \mathbb{N}$, then Fekete's lemma cannot be applied directly and we redefine the rate by

$$r(\mathcal{C}, \rho) \doteq \lim_{L \to \infty} \frac{1}{dL} \log \# \mathcal{S}_{dL}, \tag{30}$$

and we take again $\nu(\mathcal{C}) \doteq 2^{r(\mathcal{C}, \rho)}$ with this new definition.

Definition IV.4 (Transition graph) *Let $(\mathcal{C}, \rho)$ be a zero-error intermingled code, we define its transition graph $G = \big(\mathcal{V}(G), \mathcal{E}(G)\big)$ by*

$$\mathcal{V} = \prod_{c \in \mathcal{C}} [\![0, |c| - 1]\!], \tag{31}$$

$$\forall v, v' \in \mathcal{V}, \ vv' \in \mathcal{E} \quad \text{if} \quad \exists i \in \rho(v), \ \forall j \leq \bar{l}, \ v'_j = v_j + \mathbb{1}_{i=j} \mod j, \tag{32}$$

i.e. v' can be obtained by adding 1 to only one of the components of v that is included in $\rho(v)$, modulo $|v|$.

Theorem IV.5 (Rate computation of intermingled codes) *We consider the intermingled code $(\mathcal{C}, \rho)$ that is zero-error for the channel W. The asymptotic rate satisfies*

$$r(\mathcal{C}, \rho) = \log \max_i |\lambda_i(M_G)|, \tag{33}$$

where $(\lambda_i)_{i \leq \# \mathcal{V}}$ are the elements of the spectrum of M_G, which is the adjacency matrix of G the transition graph.

The proof of Theorem IV.5 is stated in App. C. In Sec. V, we provide an example based on the channel graph $C_5 \boxplus \mathbf{1}$, for which we compute explicitly the asymptotic rate $r(\mathcal{C})$ of the intermingled code.

Remark IV.6 (Optimality of intermingled codes) *In [1, Theorems 3 and 4], Shannon proves the optimality of intermingled codes for two parallel channels, when an adjacency-reducing mapping can be built over one of the two channels.*

Given the generator set of an intermingled code $\mathcal{C}$, one can define the maximal succession rule as

$$\operatorname*{argmax}_{\rho} r(\mathcal{C}, \rho) \tag{34}$$

$$\text{s.t. } (\mathcal{C}, \rho) \text{ is zero-error.} \tag{35}$$

Thus finding the optimal rate over the set of intermingled codes boils down to an optimization problem over $\mathcal{P}(\mathcal{C})^{\prod_{c \in \mathcal{C}} [\![0, |c|-1]\!]}$: the succession rule is in general far too complex to be described, in order to make this approach tractable. The main interest lies instead in the possible existence of an optimal intermingled generator of the family of fixed-length codes that asymptotically reaches the capacity, even if no finite fixed-length code can reach it.

V. EXAMPLE

Let us consider the channel graph $C_5 \boxplus \mathbf{1}$, depicted in Fig. 3 with vertices $\{1, 2, 3, 4, 5\} \cup \{0\}$. We consider two generator sets $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$ and $\mathcal{C}' = \{11, 23, 35, 42, 54\} \cup \{001, 003\}$, that are zero-error for any channel W whose graph is $C_5 \boxplus \mathbf{1}$.
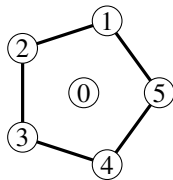

Fig. 3: The channel graph $C_5 \boxplus \mathbf{1}$

A. Variable-length coding scheme

The variable-length code obtain with the generator set $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$ is depicted in Fig. 4. The maximal length, (resp. the minimum length), is $\bar{l} = 2$, (resp. $\underline{l} = 1$). By Theorem III.4, the asymptotic rate of the variable-length code is $r(\mathcal{C}) = \log(\nu(\mathcal{C}))$ where $\nu(\mathcal{C})$ is solution of the characteristic polynomial $X^2 - X - 5 = 0$. This gives us $\nu(\mathcal{C}) = \frac{1+\sqrt{21}}{2} \simeq 2.791$ and a rate $r(\mathcal{C}) \simeq 1.422$, which is inferior to the zero-error capacity of this channel $C_0 = \log(1 + \sqrt{5}) \simeq 1.694$ and also $1 + \sqrt{5} \simeq 3.236$, obtained by combining Lovasz capacity result for the graph C_5 in [18], with the result of Shannon for adjacency reducing mapping, in [1].

Lemma 2 (Small number of channel uses) *For a small number of channel uses and* $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$, *the transmission rates are given by*

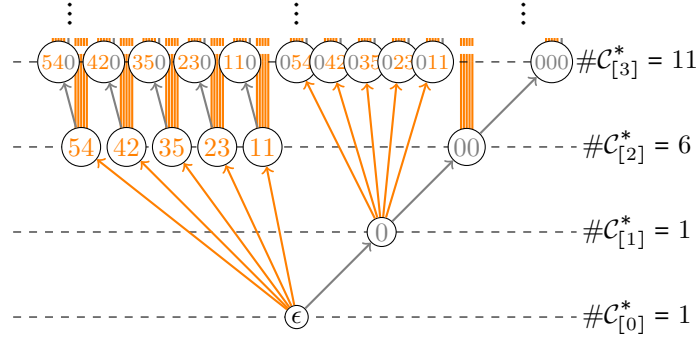| Channel uses: L	2	3	4	5
$\sqrt[L]{\#\mathcal{C}^*_{[L]}}$	$\sqrt{6} \simeq 2.449$	$\sqrt[3]{11} \simeq 2.224$	$\sqrt[4]{41} \simeq 2.530$	$\sqrt[5]{96} \simeq 2.491$

Fig. 4: Set of channel input sequences obtained by concatenation of words from the generator set

$$\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}.$$

Let us now consider the generator set $\mathcal{C}' = \{11, 23, 35, 42, 54\} \cup \{001, 003\}$ that is also zero-error for the graph is $C_5 \boxplus 1$. The maximal length, (resp. the minimum length), of words in $\mathcal{C}$ is $\bar{l} = 3$, (resp. $\underline{l} = 2$). The corresponding variable-length code achieves a rate $r(\mathcal{C}') = \log\big(\nu(\mathcal{C}')\big)$ where $\nu(\mathcal{C}')$ is solution of the characteristic polynomial $X^3 - 5X - 2 = 0$. This gives us $\nu(\mathcal{C}') = 1 + \sqrt{2} \simeq 2.414$ and a rate $r(\mathcal{C}') \simeq 1.272$, which is also inferior to the know zero-error capacity of this channel $C_0 = \log(1 + \sqrt{5}) \simeq 1.694$ and also $1 + \sqrt{5} \simeq 3.236$.

Lemma 3 *For a small number of channel uses and $\mathcal{C}' = \{11, 23, 35, 42, 54\} \cup \{001, 003\}$, the transmission rates are given by*

Channel uses: L	3	4	5
$\sqrt[L]{\#\mathcal{C}_{[L]}'^*}$	$\sqrt[3]{2} \simeq 1.260$	$\sqrt[4]{25} \simeq 2.236$	$\sqrt[5]{20} \simeq 1.821$

Remark V.1 (One-shot maximum independent set) *Coding on the one-shot maximum independent set with $\mathcal{C}'' \doteq \{0, 1, 3\}$ achieves a better rate than with the generator sets $\mathcal{C}$ and $\mathcal{C}'$, as $\nu(\mathcal{C}'') = 3$.*

B. Impact of non-dominant eigenvalues

As shown in the proof of Theorem III.4 in App. B, the asymptotic rate can be read in the dominant eigenvalue of the adjacency matrix of the transition graph. Studying the other eigenvalues gives the exact shape and decay-rate of the teeth patterns in the rate curves. A closed-form expression can be obtained with a Jordan decomposition of the adjacency matrix.

For example, the code $\mathcal{C}'$ generates the following transition matrix :

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 5 & 0 \end{pmatrix} = P \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 - \sqrt{2} & 0 \\ 0 & 0 & 1 + \sqrt{2} \end{pmatrix} P^{-1}. \tag{36}$$
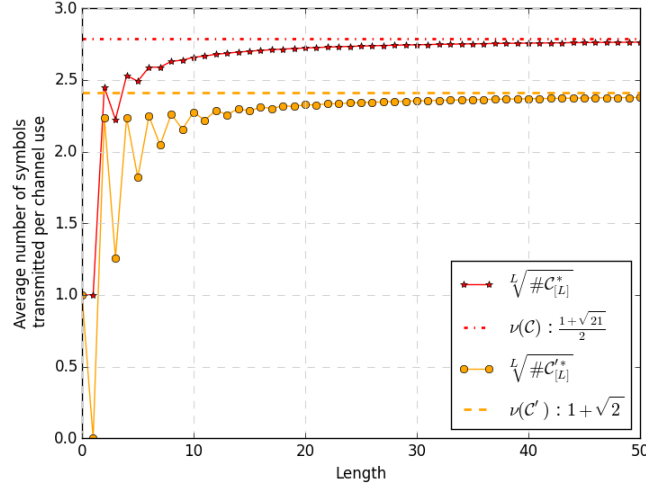
Fig. 5: Transmission rates of the generated fixed-length codes corresponding to the generator sets $\mathcal{C} = \{0, 11, 23, 35, 42, 54\}$ and $\mathcal{C}' = \{11, 23, 35, 42, 54, 001, 003\}$, with their respective limits.

We have for all $L \in \mathbb{N}$:

$$
\begin{pmatrix} \#\mathcal{C}'^{*}_{[L+1]} \\ \#\mathcal{C}'^{*}_{[L+2]} \\ \#\mathcal{C}'^{*}_{[L+3]} \end{pmatrix} = P \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1-\sqrt{2} & 0 \\ 0 & 0 & 1+\sqrt{2} \end{pmatrix}^{L} P^{-1} \begin{pmatrix} \#\mathcal{C}'^{*}_{[1]} \\ \#\mathcal{C}'^{*}_{[2]} \\ \#\mathcal{C}'^{*}_{[3]} \end{pmatrix}. \tag{37}
$$

The general solution for linear difference equations [68, pp. 13], writes

$$
\#\mathcal{C}'^{*}_{[L]} = h_1(1+\sqrt{2})^L + h_2(-2)^L + h_3(1-\sqrt{2})^L, \tag{38}
$$

where the parameters $(h_1, h_2, h_3) \in \mathbb{R}^3$ are determined by the initial values $(\#\mathcal{C}'^{*}_{[1]}, \#\mathcal{C}'^{*}_{[2]}, \#\mathcal{C}'^{*}_{[3]}) = (0, 5, 2)$. Straightforward computations lead to

$$
\#\mathcal{C}'^{*}_{[L]} = \frac{6 + 5\sqrt{2}}{28}(1+\sqrt{2})^L + \frac{4}{7}(-2)^L + \frac{6 - 5\sqrt{2}}{28}(1-\sqrt{2})^L. \tag{39}
$$

We observe two oscillating terms in this expression, which are asymptotically dominated by $(1+\sqrt{2})^L$, creating the vanishing teeth-shaped patterns on the rate-length curves. The oscillations' amplitude decrease at the rate $\left(\frac{|-2|}{|1+\sqrt{2}|}\right)^L = (2\sqrt{2} - 2)^L \simeq 0.828^L$. Since $\#\mathcal{C}'^{*}_{[L]} = h_1\nu(\mathcal{C}')^L + O((-2)^L)$ with $h_1 = \frac{6+5\sqrt{2}}{28} \neq 1$ and $\left|\frac{-2}{\nu(\mathcal{C}')}\right| < 1$ we have

$$
\sqrt[L]{\#\mathcal{C}'^{*}_{[L]}} - \nu(\mathcal{C}') = e^{\frac{1}{L}\ln(\#\mathcal{C}'^{*}_{[L]})} - e^{\ln(\nu(\mathcal{C}'))} \sim e^{\ln(\nu(\mathcal{C}'))}\left(\frac{1}{L}\ln(\#\mathcal{C}'^{*}_{[L]}) - \ln(\nu(\mathcal{C}'))\right) \tag{40}
$$

$$
\sim \frac{\nu(\mathcal{C}')^2}{L}\ln(h_1) + \nu(\mathcal{C}')\ln\left(1 + O\left(\frac{(-2)^L}{\nu(\mathcal{C}')^L}\right)\right) \sim \frac{\ln(h_1)\nu(\mathcal{C}')^2}{L}. \tag{41}
$$

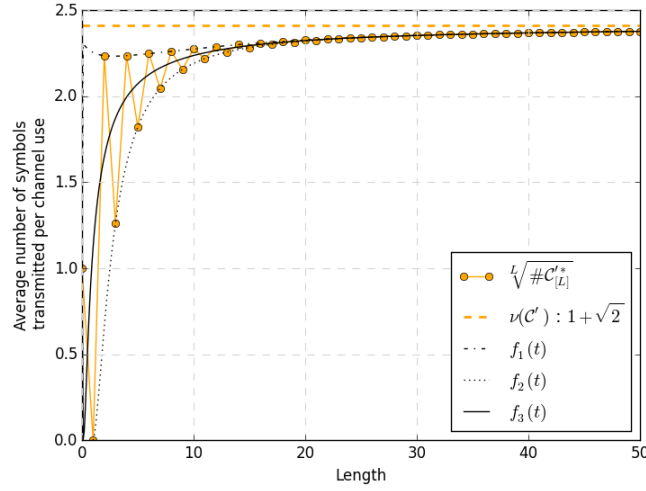Thus the speed of convergence towards the asymptotic rate is $O(\frac{1}{L})$.

Fig. 6: Value of $\sqrt[L]{\#\mathcal{C}_{[L]}^{\prime*}}$ with its oscillations

In Fig. 6, we represent $\sqrt[L]{\#\mathcal{C}_{[L]}^{\prime*}}$ along with the three following functions for $t \in (0, 50]$ that capture its oscillations.

$$f_1(t) = \left(\frac{6 + 5\sqrt{2}}{28}(1 + \sqrt{2})^t + \frac{4}{7}2^t + \frac{6 - 5\sqrt{2}}{28}(\sqrt{2} - 1)^t \right)^{\frac{1}{t}}, \tag{42}$$

$$f_2(t) = \left(\frac{6 + 5\sqrt{2}}{28}(1 + \sqrt{2})^t - \frac{4}{7}2^t - \frac{6 - 5\sqrt{2}}{28}(\sqrt{2} - 1)^t \right)^{\frac{1}{t}}, \tag{43}$$

$$f_3(t) = \left(\frac{6 + 5\sqrt{2}}{28}(1 + \sqrt{2})^t \right)^{\frac{1}{t}} = \left(\frac{6 + 5\sqrt{2}}{28} \right)^{\frac{1}{t}}(1 + \sqrt{2}). \tag{44}$$

C. Intermingled coding scheme

Now we consider the intermingled code $(\mathcal{C}, \rho)$, where $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$,

$$\rho : (z_0, z_{11}, z_{23}, z_{35}, z_{42}, z_{54}) \mapsto \begin{cases} \mathcal{C} \text{ if } (z_{11}, z_{23}, z_{35}, z_{42}, z_{54}) = (0, 0, 0, 0, 0), \\ \{0\} \cup \{c \in \mathcal{C} \mid z_c \neq 0\} \text{ otherwise.} \end{cases} \tag{45}$$

This code is zero-error as z has at most one positive component for all time step. Following the Definition IV.4, the corresponding transition graph is depicted in Fig. 7.

Proposition V.2 *The intermingled coding scheme $(\mathcal{C}, \rho)$ defined by $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$ and equation (45), achieves the zero-error capacity $C_0 = \log(1 + \sqrt{5})$ of the channel graph $C_5 \boxplus \mathbf{1}$.*

For a given generator set $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$, the intermingled coding scheme performs better than the variable-length coding since the asymptotic rate $r(\mathcal{C}) = \log\left(\frac{1 + \sqrt{21}}{2} \right) \simeq 1.422$ obtained in Sec. V-A, is strictly smaller than the channel capacity $C_0 = \log(1 + \sqrt{5}) \simeq 1.694$ obtained with the intermingled code $(\mathcal{C}, \rho)$. We also recall that since $(1 + \sqrt{5})^L \notin \mathbb{N}$ for all L, any fixed-length code is suboptimal, as noticed in [25].
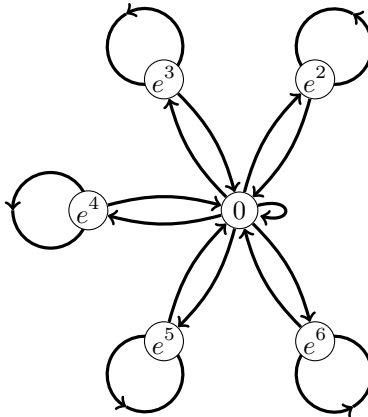
Fig. 7: Transition graph of the intermingled coding scheme $(\mathcal{C}, \rho)$ defined by (45).

Proof. [Proposition V.2] The adjacency matrix M_G of the transition graph depicted in Fig. 7, with the possible transmission states $(0, e^2, e^3, e^4, e^5, e^6)$, where (e^i) denotes the canonical basis of $\mathbb{R}^6$, is given by

$$M_G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{46}$$

The characteristic polynomial of this matrix is $(X - 1 - \sqrt{5})(X - 1 + \sqrt{5})(X - 1)^4$. The intermingled code has the rate equal to its maximal root, that is $1 + \sqrt{5}$, thus corresponding to the zero-error capacity $C_0 = \log(1 + \sqrt{5})$ of the channel graph $C_5 \boxplus \mathbf{1}$. □

D. Variable-length code for the channel graph C_7

We recall that the zero-error capacity corresponding to the channel graph C_7, depicted in Fig. 1, is still unknown, see [26]. Let us build a zero-error variable-length code on C_7. We consider the generator set $\widehat{\mathcal{C}} = \{0, 20, 22, 24, 40, 42, 44\}$, $\nu(\widehat{\mathcal{C}})$ is the positive solution of $X^2 = X + 6$, so $\nu(\widehat{\mathcal{C}}) = 3$. Note that it is equal to the cardinality of the one-shot maximum independent set, e.g. $\widehat{\mathcal{C}}' = \{0, 2, 4\}$.

VI. AUTOMATA-BASED CODING

Our previous Theorems III.4 and IV.5 share deep connections with the Automata Theory. In this section, we formulate the variable-length and intermingled coding schemes as particular automata. By using the concepts of *regular expressions* and *generator series*, we simplify the rate analysis of the optimal coding scheme for the channel graph $C_5 \boxplus \mathbf{1}$.

A. Main concepts

The definitions and characterization proofs in this subsection can be retrieved from [69].

Definition VI.1 (Deterministic finite automaton (DFA)) *A deterministic finite automaton (DFA) is a tuple $A = (\mathcal{X}, \mathcal{S}, \tau, s_{start}, \mathcal{S}_{accept})$, which consists in*

⋄ *a finite alphabet $\mathcal{X}$,*

⋄ *a finite set of states $\mathcal{S}$,*

⋄ *a transition function $\tau : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$,*

⋄ *an initial state $s_{start} \in \mathcal{S}$,*

⋄ *a subset of accept states $\mathcal{S}_{accept} \subseteq \mathcal{S}$.*

We extend τ to $\mathcal{S} \times \mathcal{X}^*$ by defining

$$\tau(s, x_1...x_n) \doteq \tau(\tau(...\tau(s, x_1), x_2), ...), x_n). \tag{47}$$

An example of automaton is depicted in Fig. 8. The automaton starts with the initial state s_{start}, receives a word $w \in \mathcal{X}^*$ as entry and at each time step $l \le |w|$ it applies the transition function to determine its new state s_{l+1}, that is $s_{l+1} = \tau(s_l, w_l)$. The automaton accepts the word if the final state belongs to $\mathcal{S}_{accept}$, that is $\tau(s_{start}, w) \in \mathcal{S}_{accept}$. Otherwise the word is rejected.

The set of words accepted by A is called language recognized by A and is noted $\mathcal{L}(A)$.

Definition VI.2 (Transition graph of a DFA) *We define the transition graph $G_A \doteq (\mathcal{V}(G_A), \mathcal{E}(G_A))$ of a DFA $A = (\mathcal{X}, \mathcal{S}, \tau, s_{start}, \mathcal{S}_{accept})$ by*

$$\mathcal{V}(G_A) = \mathcal{S}, \tag{48}$$

$$ss' \in \mathcal{E}(G_A) \quad if \quad \exists x \in \mathcal{X}, \tau(s, x) = \{s'\}. \tag{49}$$

Definition VI.3 (Regular expression) *The languages of DFA's are easily described by* regular expressions*, as in* (54)*. Given a regular expression E, the corresponding language is denoted by $\mathcal{L}(E)$. The regular expressions are built with the letters of $\mathcal{X}$, the symbols ϵ and $\varnothing$ and the operators $+$, $\cdot$ and * where*

$$\mathcal{L}(\varnothing) = \varnothing, \quad \mathcal{L}(\epsilon) = \{\epsilon\}, \quad and \quad \forall x \in \mathcal{X}, \mathcal{L}(x) = \{x\}. \tag{50}$$

Given two regular expressions E and E', we have

$$\mathcal{L}(E^*) = \mathcal{L}(E)^*, \tag{51}$$

$$\mathcal{L}(E + E') = \mathcal{L}(E) \cup \mathcal{L}(E'), \tag{52}$$

$$\mathcal{L}(E \cdot E') = \mathcal{L}(E) \cdot \mathcal{L}(E') = \{ww' \mid w \in \mathcal{L}(E), w' \in \mathcal{L}(E')\}. \tag{53}$$

Definition VI.4 (Regular language) *A regular language is a subset $\mathcal{A} \subseteq \mathcal{X}^*$ such that there exists a DFA A such that $\mathcal{A} = \mathcal{L}(A)$.*

There exists other equivalent characterizations for regular languages :

- ◆ Languages recognized by nondeterministic finite automata: the deterministic automata have the same recognition power as the nondeterministic ones.
- ◆ Languages generated by regular expressions.

Remark VI.5 (Connection with variable-length and intermingled coding in Sec. III and IV) *The variable-length coding scheme presented in Sec. III corresponds to a subclass of regular languages that can be expressed by using the generator set with at most 1 imbricated star, these are called star-height one languages. Furthermore, the automaton that recognises these languages can be found in App. B, with the graph G that can be straightforwardly completed as a transition graph between states, by adding an initial and final set $\mathcal{F}$.*

The intermingled coding scheme generates a particular class of regular languages as well, the transition function can be also derived from the transition graph of the code, with the adequate completion, as in Fig. 8.

B. Example of a deterministic finite automaton

We consider the deterministic finite automaton with

- ◆ alphabet $\mathcal{X} = [\![0, 5]\!]$,
- ◆ set of states $\mathcal{S} = \{s_0, ..., s_5\} \cup \{sink\}$,
- ◆ $s_{start} = s_0$ as the initial and only accepting state ($\mathcal{S}_{accept} = \{s_0\}$),
- ◆ a transition function described on Fig. 8. On each state s_i, an arrow starts towards each letter x of $\mathcal{X}$, and the state s_j at the extremity of this arrow corresponds to the result of the transition function $\{s_j\} = \tau(s_i, x)$.
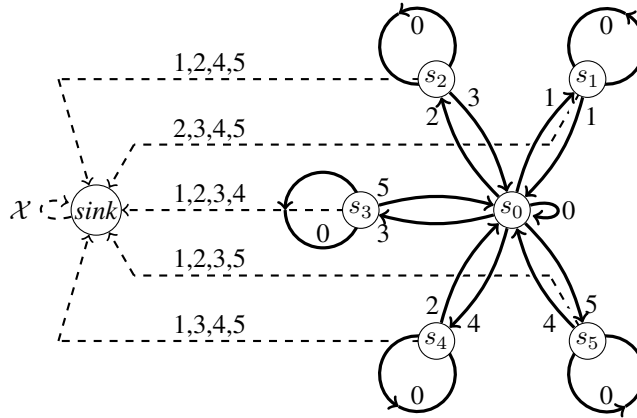
Fig. 8: Discrete finite automaton (DFA) corresponding to the intermingled coding scheme of Sec. V-C.

This automaton recognizes the language $\mathcal{L}(E)$ corresponding to the following regular expression

$$E = \left(0 + 1(0)^*1 + 2(0)^*3 + 3(0)^*5 + 4(0)^*2 + 5(0)^*4\right)^*. \tag{54}$$

This language is exactly the union over $L \in \mathbb{N}$ of the codewords of length L generated by the intermingled coding pattern over $C_5 \boxplus \mathbf{1}$, as in Sec. V-C.

C. Rational coding scheme

Definition VI.6 (Rational codes) *Let E be a regular expression such that $\mathcal{L}(E)$ is infinite and $E = (E')^*$, where E' is a regular expression over the alphabet $\mathcal{X}$. For all $L \in \mathbb{N}$, the codebook of the rational code is defined by $\mathcal{L}(E)_{[L]}$, that is the set of words of length L accepted by the automaton corresponding to the regular expression E.*

This code is zero-error if for all $L \in \mathbb{N}$, the codewords from $\mathcal{L}(E)_{[L]}$ are distinguishable, i.e. $\forall x, x' \in \mathcal{L}(E)_{[L]}$, $xx' \notin \mathcal{E}(G_W^{\boxtimes L})$.

Definition VI.7 (Rational code rate) *The asymptotic rate $r(E)$ of a rational code E is defined by*

$$r(E) \doteq \lim_{L \to \infty} \frac{1}{L} \log \big(\# \mathcal{L}(E)_{[L]} \big),$$

and its average number of transmitted symbols per channel use $\nu(E) \doteq 2^{r(E)}$.

Note that the existence of the limit is given by Fekete's lemma, as for all $L, L' \in \mathbb{N}$ we have

$$\log \# \mathcal{L}(E)_{[L+L']} \geq \log \# \mathcal{L}(E)_{[L]} + \log \# \mathcal{L}(E)_{[L']}. \tag{55}$$

Let E' be the regular expression such that $E = (E')^*$. If $\gcd(|w|, w \in \mathcal{L}(E')) \neq 1$, then Fekete's lemma cannot be applied directly and we redefine the rate by

$$r(E) \doteq \lim_{L \to \infty} \frac{1}{dL} \log \big(\# \mathcal{L}(E)_{[dL]} \big), \tag{56}$$

where $d = \gcd\big(|w|, w \in \mathcal{L}(E')\big)$ and $\nu(\mathcal{C}) \doteq 2^{r(\mathcal{C})}$, and we take again $\nu(E) \doteq 2^{r(E)}$ with this new definition.

The asymptotic rate of the rational code can be computed via the generator series of its associated regular expression.

Definition VI.8 (Generator series) *Given a regular expression E, we define its generator series by*

$$F_E : z \mapsto \sum_{l \in \mathbb{N}} \# \mathcal{L}(E)_{[l]} z^l. \tag{57}$$

The generator series F_E of a regular expression E is always a rational fraction, as the terms of the series follow a linear difference equation.

Proposition VI.9 (Recursive computation of the generator series) *Let us denote $E^l \doteq E \cdot \ldots \cdot E$ l times. If the regular expressions E and E' satisfy $\mathcal{L}(E) \cap \mathcal{L}(E') = \varnothing$, then we have*

$$F_{E+E'}(z) = F_E(z) + F_{E'}(z), \tag{58}$$

$$F_{EE'}(z) = F_E(z) F_{E'}(z). \tag{59}$$

If the sets $\big(\mathcal{L}(E^l)\big)_{l \in \mathbb{N}}$ are disjoint, then we have

$$F_{E^*}(z) = \frac{1}{1 - F_E(z)}. \tag{60}$$

The proof of Proposition VI.9 is stated in App. D. For regular languages, the sequence of the number of words of length $L \in \mathbb{N}$ satisfies a linear difference equation. By using the same arguments as in the proof of Theorem IV.5 in App. C, one can determine a closed-form expression for this sequence and easily derive its asymptotic rate.

Theorem VI.10 (Rate computation of rational codes) *Let E be a rational code, and let A be the automaton that recognize the language $\mathcal{L}(E)$. Then $\nu(E)$ equals the spectral radius of the adjacency matrix of the transition graph of A.*

Furthermore $\nu(E)$ is also the inverse of the convergence radius of the generator series F_E, or equivalently $\nu(E) = \frac{1}{|p|}$ where p is the pole of the generator series F_E with the smallest modulus.

Although this theorem is a well-known result in Automata Theory, see [42, Proposition 8.1, pp. 20], we provide the proof in App. E.

Theorem VI.11 (Channel generator series) *We define the channel generator series by*

$$\sum_{l \in \mathbb{N}} \alpha(G_W^{\boxtimes l}) z^l. \tag{61}$$

The zero-error capacity of the channel C_0 is the inverse of the convergence radius of the channel generator series.

Proof. [Theorem VI.11] This is a direct consequence of the Cauchy-Hadamard theorem for power series, see [70, Theorem 2.6, pp. 55], which states that the convergence radius of a power series $\sum_{l \in \mathbb{N}} a_l z^l$ is equal to $\frac{1}{\limsup_l \sqrt[n]{a_l}}$. □

D. Computation of the asymptotic rate based on the generator series

The Proposition V.2 shows that the intermingled code $(\mathcal{C}, \rho)$ defined by $\mathcal{C} = \{0\} \cup \{11, 23, 35, 42, 54\}$ and equation (45), is optimal for the optimal channel graph $C_5 \boxplus \mathbf{1}$. The corresponding regular expression writes

$$E \doteq \left(0 + 1(0)^* 1 + 2(0)^* 3 + 3(0)^* 5 + 4(0)^* 2 + 5(0)^* 4 \right)^*. \tag{62}$$

By using the properties (58)-(60) of the generator series F_E, we compute the asymptotic rate of the automaton defined by the regular expression E.

$$F_E(z) = \frac{1}{1 - F_{0+1(0)^*1+2(0)^*3+3(0)^*5+4(0)^*2+5(0)^*4}(z)} \tag{63}$$

$$= \frac{1}{1 - F_0(z) - F_{1(0)^*1}(z) - F_{2(0)^*3}(z) - F_{3(0)^*5}(z) - F_{4(0)^*2}(z) - F_{5(0)^*4}(z)} \tag{64}$$

$$= \frac{1}{1 - z - \frac{5z^2}{1-z}} \tag{65}$$

$$= \frac{z - 1}{4z^2 + 2z - 1}, \tag{66}$$

where equation (65) comes from (59) and

$$F_{1(0)^*1}(z) = F_1(z) F_{0^*}(z) F_1(z) = z \cdot F_{0^*}(z) \cdot z \tag{67}$$

$$= z^2 \sum_{l \in \mathbb{N}} z^l = \frac{z^2}{1-z}. \tag{68}$$

as it directly follows from the definition that $F_i(z) = z$ for all $i \in \{0, ..., 5\}$.

This rational fraction (66) has two poles $\frac{-1 \pm \sqrt{5}}{4}$. The pole with smallest modulus is $\frac{-1 + \sqrt{5}}{4}$, which has $\sqrt{5} + 1$ as inverse. We retrieved the optimal rate $\log(\sqrt{5} + 1)$ of Proposition V.2 with much simpler computations.

ACKNOWLEDGMENT

REFERENCES

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

[2] C. Berge, *Graphs and Hypergraphs*, ser. North-Holland mathematical library. Amsterdam, 1973.

[3] M. Grötschel, L. Lovász, and A. Schrijver, "Polynomial algorithms for perfect graphs," *Ann. Discrete Math*, vol. 21, pp. 325–356, 1984.

[4] S. Klavzar, R. Hammack, and W. Imrich, "Handbook of graph products," 2011.

[5] M. Kovačević, "Zero-error capacity of duplication channels," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6735–6742, Oct 2019.

[6] M. Kovačević and P. Popovski, "Zero-error capacity of a class of timing channels," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6796–6800, Nov 2014.

[7] M. Kovačević, M. Stojaković, and V. Y. F. Tan, "Zero-error capacity of p -ary shift channels and fifo queues," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7698–7707, Dec 2017.

[8] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[9] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[10] M. Dalai and Y. Polyanskiy, "Bounds on the reliability of a typewriter channel," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1715–1719.

[11] ——, "Bounds on the reliability function of typewriter channels," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6208–6222, Sep. 2018.

[12] G. Cohen, E. Fachini, and J. Körner, "Zero-error capacity of binary channels with memory," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 3–7, Jan 2016.

[13] R. Ahlswede, N. Cai, and Z. Zhang, "Zero-error capacity for models with memory and the enlightened dictator channel," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1250–1252, May 1998.

[14] M. Dalai, "An elias bound on the bhattacharyya distance of codes for channels with a zero-error capacity," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 1276–1280.

[15] ——, "Elias bound for general distances and stable sets in edge-weighted graphs," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2335–2350, May 2015.

[16] B. Bose, N. Elarief, and L. G. Tallini, "On codes achieving zero error capacities in limited magnitude error channels," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 257–273, Jan 2018.

[17] A. Abreu, L. Cunha, T. Fernandes, C. de Figueiredo, L. Kowada, F. Marquezino, D. Posner, and R. Portugal, "The graph tessellation cover number: extremal bounds, efficient algorithms and hardness," in *Latin American Symposium on Theoretical Informatics*. Springer, 2018, pp. 1–13.

[18] L. Lovász, "On the shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 1–7, 1979.

[19] M. Rosenfeld, "On a problem of ce shannon in graph theory," *Proceedings of the American Mathematical Society*, vol. 18, no. 2, pp. 315–319, 1967.

[20] R. Hales, "Numerical invariants and the strong product of graphs," *Journal of Combinatorial Theory, Series B*, vol. 15, no. 2, pp. 146–155, 1973.

[21] W. Haemers *et al.*, "An upper bound for the shannon capacity of a graph," in *Colloq. Math. Soc. János Bolyai*, vol. 25, 1978, pp. 267–272.

[22] N. Alon, "The shannon capacity of a union," *Combinatorica*, vol. 18, no. 3, pp. 301–310, 1998.

[23] P. K. Jha and S. Klavzar, "Independence in direct-product graphs," *Ars Combinatoria*, vol. 50, pp. 53–64, 1998.

[24] G. Hahn, P. Hell, and S. Poljak, "On the ultimate independence ratio of a graph," *European Journal of Combinatorics*, vol. 16, no. 3, pp. 253–261, 1995.

[25] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2207–2229, 1998.

[26] S. C. Polak and A. Schrijver, "New lower bound on the shannon capacity of C_7 from circular graphs," *Information Processing Letters*, vol. 143, pp. 37–40, 2019.

[27] T. Gallai, "Graphen mit triangulierbaren ungeraden vielecken," *Magyar Tud. Akad. Mat. Kutató Int. Közl*, vol. 7, pp. 3–36, 1962.

[28] E. Sonnemann and O. Krafft, "Independence numbers of product graphs," *Journal of Combinatorial Theory, Series B*, vol. 17, no. 2, pp. 133–142, 1974.

[29] A. Vesel, "The independence number of the strong product of cycles," *Computers & Mathematics with Applications*, vol. 36, no. 7, pp. 9–21, 1998.

[30] T. Bohman and R. Holzman, "A nontrivial lower bound on the shannon capacities of the complements of odd cycles," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 721–722, 2003.

[31] A. Vesel and J. Žerovnik, "Improved lower bound on the shannon capacity of C_7," *Information Processing Letters*, vol. 81, no. 5, pp. 277–282, 2002.

[32] B. Codenotti, I. Gerace, and G. Resta, "Some remarks on the shannon capacity of odd cycles," *Ars Combinatoria*, vol. 66, pp. 243–258, 2003.

[33] L. Baumert, R. McEliece, E. Rodemich, H. Rumsey, R. Stanley, and H. Taylor, "A combinatorial packing problem," *Computers in Algebra and Number Theory*, vol. 4, 1971.

[34] S. H. Badalyan and S. E. Markosyan, "On the independence number of the strong product of cycle-powers," *Discrete Mathematics*, vol. 313, no. 1, pp. 105–110, 2013.

[35] T. Bohman, "A limit theorem for the shannon capacities of odd cycles. I," *Proceedings of the American Mathematical Society*, vol. 131, no. 11, pp. 3559–3569, 2003.

[36] ——, "A limit theorem for the shannon capacities of odd cycles. II," *Proceedings of the American Mathematical Society*, vol. 133, no. 2, pp. 537–543, 2005.

[37] K. A. Mathew and P. R. Östergård, "New lower bounds for the shannon capacity of odd cycles," *Designs, Codes and Cryptography*, vol. 84, no. 1-2, pp. 13–22, 2017.

[38] C. Berge, "Farbung von graphen, deren samtliche bzw. deren ungerade kreise starr sind," *Wissenschaftliche Zeitschrift*, 1961.

[39] M. Chudnovsky, N. Robertson, P. D. Seymour, and R. Thomas, "Progress on perfect graphs," *Mathematical Programming*, vol. 97, no. 1-2, pp. 405–422, 2003.

[40] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[41] S. Ben-Jamaa, C. Weidmann, and M. Kieffer, "Analytical tools for optimizing the error correction performance of arithmetic codes," *IEEE Transactions on Communications*, vol. 56, no. 9, pp. 1458–1468, 2008.

[42] P. Flajolet and R. Sedgewick, "Analytic combinatorics: functional equations, rational and algebraic functions," *report available [on-line] at https://hal.inria.fr/inria-00072528*, 2001.

[43] M. Asadi and N. Devroye, "On the zero-error capacity of channels with rate limited noiseless feedback," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2018, pp. 1141–1146.

[44] F. Guo and Y. Watanabe, "On graphs in which the shannon capacity is unachievable by finite product," *IEEE Transactions on Information Theory*, vol. 36, no. 3, pp. 622–623, May 1990.

[45] Y. Chen and N. Devroye, "Zero-error relaying for primitive relay channels," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7708–7715, Dec 2017.

[46] Y. Chen, S. Shahi, and N. Devroye, "Colour-and-forward: Relaying "what the destination needs" in the zero-error primitive relay channel," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2014, pp. 987–995.

[47] Y. Chen and N. Devroye, "On the optimality of colour-and-forward relaying for a class of zero-error primitive relay channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1272–1276.

[48] H. Witsenhausen, "The zero-error side information problem and chromatic numbers (corresp.)," *IEEE Transactions on Information Theory*, vol. 22, no. 5, pp. 592–593, Sep. 1976.

[49] M. Asadi, K. Palacio-Baus, and N. Devroye, "A relaying graph and special strong product for zero-error problems in primitive relay channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 281–285.

[50] N. Devroye, "When is the zero-error capacity positive in the relay, multiple-access, broadcast and interference channels?" in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2016, pp. 672–678.

[51] M. Asadi and N. Devroye, "On the zero-error capacity of channels with noisy feedback," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 642–649.

[52] L. Zhao and H. H. Permuter, "Zero-error feedback capacity of channels with state information via dynamic programming," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2640–2650, June 2010.

[53] L. Wang and O. Shayevitz, "Graph information ratio," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 913–917.

[54] L. Wang and O. Shayevitz, "Graph information ratio," *SIAM Journal on Discrete Mathematics*, vol. 31, no. 4, pp. 2703–2734, 2017.

[55] J. Körner and K. Marton, "Relative capacity and dimension of graphs," *Discrete Mathematics*, vol. 235, no. 1, pp. 307 – 315, 2001.

[56] S. Hu and O. Shayevitz, "The ρ-capacity of a graph," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2241–2253, April 2017.

[57] O. Ordentlich and O. Shayevitz, "A vc-dimension-based outer bound on the zero-error capacity of the binary adder channel," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2366–2370.

[58] R. Urbanke and Quinn Li, "The zero-error capacity region of the 2-user synchronous bac is strictly smaller than its shannon capacity region," in *1998 Information Theory Workshop (Cat. No.98EX131)*, June 1998, pp. 61–.

[59] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Secure estimation and zero-error secrecy capacity," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1047–1062, March 2019.

[60] M. Wiese, K. H. Johansson, T. J. Oechtering, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Uncertain wiretap channels and secure estimation," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2004–2008.

[61] ——, "Secure estimation for unstable systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 5059–5064.

[62] F. J. R. Ruiz and F. Pãl'rez-Cruz, "Zero-error codes for the noisy-typewriter channel," in *2011 IEEE Information Theory Workshop*, Oct 2011, pp. 495–497.

[63] D. Cullina, M. Dalai, and Y. Polyanskiy, "Rate-distance tradeoff for codes above graph capacity," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1331–1335.

[64] M. Dalai, V. G. Carnegie, and J. Radhakrishnan, "An improved bound on the zero-error list-decoding capacity of the 4/3 channel," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1658–1662.

[65] P. Elias, "Zero error capacity under list decoding," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 1070–1074, Sep. 1988.

[66] X. Xu and S. P. Radziszowski, "Bounds on shannon capacity and ramsey numbers from product of graphs," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4767–4770, Aug 2013.

[67] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, 2nd ed. Cambridge University Press, 2001.

[68] D. H. Greene and D. E. Knuth, *Mathematics for the Analysis of Algorithms*. Springer Science & Business Media, 2007.

[69] P. Linz, *An introduction to formal languages and automata*. Jones & Bartlett Learning, 2006.

[70] S. Lang, *Complex analysis*. Springer Science & Business Media, 1999, vol. 103.

[71] P. Lax, *Functional analysis*, ser. Pure and applied mathematics. Wiley, 2002.

APPENDIX A

PROOF OF LEMMAS

A. Proof of Lemma 1 (Fekete)

Let $\epsilon > 0$, L' be an integer such that $\frac{u_{L'}}{L'} \geq \sup_l \frac{u_l}{l} + \epsilon$. Let $L \geq L'$ and (r, q) be the remainder and quotient of the euclidean division of L by L'. Then we have $L = L'q + r$ and

$$\frac{u_L}{L} \geq \frac{qu_{L'} + r}{L} \geq \frac{qL'}{L}\left(\sup_l \frac{u_l}{l} + \epsilon\right) + \frac{r}{L} \underset{L \to \infty}{\to} \sup_l \frac{u_l}{l} + \epsilon. \tag{69}$$

Thus the limit of $\left(\frac{u_L}{L}\right)_{L \in \mathbb{N}}$ equals its supremum.

B. Proof of Lemma 2

For $L = 2$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt{6} \simeq 2.449$:

$$\#\mathcal{C}^*_{[L]} = \#\mathcal{C}^*_{[2]} = \#\{00, 11, 23, 35, 42, 54\} = 6. \tag{70}$$

For $L = 3$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt[3]{11} \simeq 2.224$:

$$\#\mathcal{C}^*_{[L]} = \#\mathcal{C}^*_{[3]}$$
$$= \#\{000\} \cup \{011, 023, 035, 042, 054\} \tag{71}$$
$$\cup \{110, 230, 350, 420, 540\} = 11.$$

For $L = 4$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt[4]{41} \simeq 2.530$:

$$\#\mathcal{C}^*_{[L]} = \#\mathcal{C}^*_{[4]}$$
$$= \#\{0011, 0023, 0035, 0042, 0054\}$$
$$\cup \{0110, 0230, 0350, 0420, 0540\}$$
$$\cup \{1100, 2300, 3500, 4200, 5400\} \tag{72}$$
$$\cup \{11, 23, 35, 42, 54\}^2 \cup \{0000\}$$
$$= 41.$$

For $L = 5$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt[5]{96} \simeq 2.491$:

$$\#\mathcal{C}^*_{[L]} = \#\mathcal{C}^*_{[5]}$$
$$= \#\{0\}^5 + \#\{0\}^3 \times [5] + \#\{0\}^2 \times [5] \times \{0\}$$
$$+ \#\{0\} \times [5] \times \{0\}^2 + \#\{0\} \times [5]^2 + \#[5] \times \{0\}^3 \tag{73}$$
$$+ \#[5] \times \{0\} \times [5] + \#[5]^2 \times \{0\}$$
$$= 1 + 5 + 5 + 5 + 25 + 5 + 25 + 25 = 96.$$

C. Proof of Lemma 3

For $L = 3$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt[3]{2} \simeq 1.260$:

$$\#\mathcal{C}'^*_{[L]} = \#\mathcal{C}'^*_{[3]} = \#\mathcal{C}' = 2. \tag{74}$$

For $L = 4$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt[4]{25} \simeq 2.236$:

$$\#\mathcal{C}'^*_{[L]} = \#\mathcal{C}'^*_{[4]} = \#\{11, 23, 35, 42, 54\}^2 = 25. \tag{75}$$

For $L = 5$, the average number of transmitted symbols per channel use of the generated fixed length code is $\sqrt[5]{20} \simeq 1.821$:

$$\#\mathcal{C}_{[L]}^{\prime*} = \#\mathcal{C}_{[5]}^{\prime*}$$

$$= \{11, 23, 35, 42, 54\} \times \{001, 003\}$$

$$\cup \{001, 003\} \times \{11, 23, 35, 42, 54\} \tag{76}$$

$$= 10 + 10 = 20.$$

APPENDIX B

PROOF OF THEOREM III.4 FOR VARIABLE-LENGTH CODES

Let us index the generator set $\mathcal{C} = \{\kappa^1, ..., \kappa^{\#\mathcal{C}}\}$. Then we define the directed graph $G \doteq (\mathcal{V}, \mathcal{E})$ with

$$\mathcal{V} = \{(i, j) \mid \kappa^i \in \mathcal{C}, j \in [\![1, |\kappa^i|]\!]\}, \tag{77}$$

$$(i, j)(i', j') \in \mathcal{E} \quad \text{if} \quad (j = |\kappa^i| \text{ and } j' = 1) \text{ OR } (i = i' \text{ and } j' = j + 1). \tag{78}$$

We also define the set of *final* nodes by $\mathcal{F} \doteq \{(i, |\kappa_i|) \mid i \leq \#\mathcal{C}\}$.
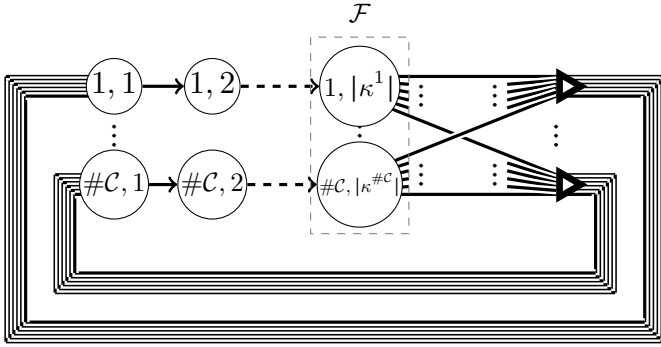

Fig. 9: The transition graph G defined by (77) and (78).

We recall that in a directed graph $G = (\mathcal{V}, \mathcal{E})$ with adjacency matrix M_G, the number of paths of length L from a vertex v to another vertex v' is given by $(M_G^L)_{vv'}$. Thus the number of paths of length L from the vertex $(1, |\kappa^1|)$ to the set $\mathcal{F}$ in G is given by

$$\sum_{v \in \mathcal{F}} (M_G^L)_{(1, |\kappa^1|), v} = \left\langle M_G^L, I_{(1, |\kappa^1|), \mathcal{F}} \right\rangle, \tag{79}$$

where $I_{(1, |\kappa^1|), \mathcal{F}} \doteq (\mathbb{1}_{v = (1, |\kappa^1|) \text{ and } v' \in \mathcal{F}})_{v, v' \in \mathcal{V}}$.

By construction, the number of distinct words of length L that can be achieved by concatenation of the elements from the generator set $\mathcal{C}$, is equal to the number of paths of length L from the vertex $(1, |\kappa^1|)$ to the set $\mathcal{F}$ in G. Indeed, each branch $(i, 1) \to (i, 2) \to ... \to (i, |\kappa^i|)$ corresponds to the transmission of the word κ^i, and each path from $(1, |\kappa^1|)$ to the set $\mathcal{F}$ is a succession of such branches. Thus the number of channel input sequences satisfies

$$\#\mathcal{C}_{[L]}^* = \sum_{v \in \mathcal{F}} (M_G^L)_{(1, |\kappa^1|), v} = \left\langle M_G^L, I_{(1, |\kappa^1|), \mathcal{F}} \right\rangle. \tag{80}$$

Lemma 4 $\nu(\mathcal{C}) = \max_i |\lambda_i(M_G)|$, where $(\lambda_i(M_G))_{i \leq \#\mathcal{V}}$ are the elements of the spectrum of M_G.

Proof. [Lemma 4] We denote by $\|\cdot\|$, the euclidean norm for matrices. If $\gcd(|c|, c \in \mathcal{C}) = 1$ we have

$$\sqrt[L]{\#\mathcal{C}^*_{[L]}} = \sqrt[L]{\left\langle M_G^L, I_{(1,|\kappa^1|),\mathcal{F}} \right\rangle} \tag{81}$$

$$= \sqrt[L]{\|M_G^L\|} \sqrt[L]{\left\langle \frac{M_G^L}{\|M_G^L\|}, I_{(1,|\kappa^1|),\mathcal{F}} \right\rangle}. \tag{82}$$

We use Gelfand formula

$$\sqrt[L]{\|M_G^L\|} \underset{L \to \infty}{\to} \max_i |\lambda_i(M_G)|, \tag{83}$$

where $(\lambda_i(M_G))_{i \leq \#\mathcal{V}}$ are the elements of the spectrum of M_G, see [71, Theorem 4, pp. 195].

Now let us show that there exists two positive constants $\overline{m}$ and $\underline{m}$ such that for all L large enough, $\underline{m} \leq \left\langle \frac{M_G^L}{\|M_G^L\|}, I_{(1,|\kappa^1|),\mathcal{F}} \right\rangle \leq \overline{m}$. The existence of $\overline{m}$ is given by Cauchy-Schwarz inequality :

$$\left\langle \frac{M_G^L}{\|M_G^L\|}, I_{(1,|\kappa^1|),\mathcal{F}} \right\rangle \leq \|I_{(1,|\kappa^1|),\mathcal{F}}\| = \overline{m}. \tag{84}$$

Since we assumed the code has a positive rate, $\left\langle \frac{M_G^L}{\|M_G^L\|}, I_{(1,|\kappa^1|),\mathcal{F}} \right\rangle$ cannot converge to 0, as $\sqrt[L]{\|M_G^L\|}$ is asymptotically bounded. Thus the value $\underline{m}$ exists. Therefore $\sqrt[L]{\left\langle \frac{M_G^L}{\|M_G^L\|}, I_{(1,|\kappa^1|),\mathcal{F}} \right\rangle}$ converges to 1 when L goes to infinity. Now we have

$$\sqrt[L]{\#\mathcal{C}^*_{[L]}} \underset{L \to \infty}{\to} \max_i |\lambda_i(M_G)|. \tag{85}$$

This proof can be straigtforwardly adapted to the case where $\gcd(|c|, c \in \mathcal{C}) = d \neq 1$ by taking dL instead of L. $\square$

Lemma 5 *There exists a unique positive number δ such that $\sum_{l=\underline{l}}^{\overline{l}} \#\mathcal{C}_{[l]} \delta^{-l} = 1$.*

Proof. [Lemma 5] The following function

$$f : \mathbb{R}^+_* \to \mathbb{R}^+_*, \tag{86}$$

$$x \mapsto \sum_{l=\underline{l}}^{\overline{l}} \#\mathcal{C}_{[l]} x^{-l}, \tag{87}$$

is strictly decreasing on $\mathbb{R}^+_*$, tends to $+\infty$ on 0^+ and goes to 0^+ on $+\infty$: there exists a unique δ such that $\sum_{l=\underline{l}}^{\overline{l}} \#\mathcal{C}_{[l]} \delta^{-l} = 1$. Note that $\delta \geq 1$ as $(\#\mathcal{C}_{[l]})_{l \leq \overline{l}}$ are integers with at least one nonzero term, i.e. $\overline{l} \neq 0$. $\square$

Lemma 6 *The value $\max_i |\lambda_i(M_G)|$ is the unique positive solution of $X^{\overline{l}} = \sum_{l=\underline{l}}^{\overline{l}} \#\mathcal{C}_{[l]} X^{\overline{l}-l}$.*

Proof. [Lemma 6] Let ν be an eigenvector for an eigenvalue $\lambda \neq 0$ of M_G. They must satisfy : $M_G\nu = \lambda\nu$. Thus for all $i \leq \#\mathcal{C}, j < |\kappa^i|$,

$$\lambda\nu_{(i,j)} = (M_G\nu)_{(i,j)} = \sum_{k \in \mathcal{V}} (M_G)_{(i,j),k}\nu_k \tag{88}$$

$$= \sum_{k \in \mathcal{V}} \mathbb{1}_{k=(i,j+1)}\nu_k = \nu_{(i,j+1)}, \tag{89}$$

and

$$\lambda \nu_{(i,|\kappa^i|)} = (M_G \nu)_{(i,|\kappa^i|)} = \sum_{k \in \mathcal{V}} (M_G)_{(i,|\kappa^i|),k} \nu_k \tag{90}$$

$$= \sum_{k \in \mathcal{V}} \mathbb{1}_{k=(\cdot,1)} \nu_k = \sum_{i' \le \#\mathcal{C}} \nu_{(i',1)}. \tag{91}$$

This gives us the equation

$$\sum_{i \le \#\mathcal{C}} \nu_{(i,1)} = \sum_{i \le \#\mathcal{C}} \lambda^{-|\kappa^i|+1} \nu_{(i,|\kappa^i|)} = \sum_{i \le \#\mathcal{C}} \lambda^{-|\kappa^i|} \sum_{i' \le \#\mathcal{C}} \nu_{(i',1)}. \tag{92}$$

Since $\sum_{i \le \#\mathcal{C}} \nu_{(i,1)} = 0$ would imply $\nu = 0$, λ must satisfy the polynomial equation

$$1 = \sum_{i \le \#\mathcal{C}} \lambda^{-|\kappa^i|} = \sum_{l \le \bar{l}} \#\mathcal{C}_{[l]} \lambda^{-l}. \tag{93}$$

Thus for all eigenvalue $\lambda \ne 0$ we have

$$1 \le \sum_{l \le \bar{l}} \#\mathcal{C}_{[l]} |\lambda|^{-l}. \tag{94}$$

There exists a unique positive real solution δ of $1 = \sum_{l \le \bar{l}} \#\mathcal{C}_{[l]} X^{-l}$, which is given by Lemma 5 ; if δ is an eigenvalue of M_G then it has the maximum modulus, since the equality is reached with the above modulus inequality.

Now we define $\nu_{\max} \doteq (\delta^{j-|\kappa^i|})_{(i,j) \in \mathcal{V}}$, let us show that $\nu_{\max}$ is an eigenvector for the eigenvalue δ. Let $i \le \#\mathcal{C}, j < |\kappa_i|$, then

$$(M_G \nu_{\max})_{(i,j)} \underset{[j<|\kappa_i|]}{=} (\nu_{\max})_{(i,j+1)} = \delta^{j+1-|\kappa^i|} = \delta(\nu_{\max})_{(i,j)}. \tag{95}$$

and

$$(M_G \nu_{\max})_{(i,|\kappa^i|)} = \sum_{i' \le \#\mathcal{C}} (\nu_{\max})_{(i',1)} = \delta \sum_{i' \le \#\mathcal{C}} \delta^{-|\kappa^{i'}|} \tag{96}$$

$$= \delta \sum_{i=\underline{l}}^{\bar{l}} \#\mathcal{C}_{[i]} \delta^{-i} \underset{[\text{Lemma 2}]}{=} \delta = \delta \cdot \delta^{|\kappa^i|-|\kappa^i|} = \delta(\nu_{\max})_{(i,|\kappa^i|)}. \tag{97}$$

The condition $M_G \nu_{\max} = \delta \nu_{\max}$ is verified on each vertex, i.e. $\nu_{\max}$ is an eigenvector for the eigenvalue δ. Thus $\delta = \max_i |\lambda_i(M_G)|$ and it is the unique positive solution of $X^{\bar{l}} = \sum_{l=\underline{l}}^{\bar{l}} \#\mathcal{C}_{[l]} X^{\bar{l}-l}$. $\square$

The desired result follows directly from lemmas 4 and 6.

APPENDIX C
PROOF OF THEOREM IV.5 FOR INTERMINGLED CODES

The following proof holds when Fekete's lemma can be applied directly to the sequence $\big(\log(\#\mathcal{S}_L)\big)_{L \in \mathbb{N}}$, that is when $\#\mathcal{S}_L \ne 0$ for all L large enough, but the proof can be straightforwardly adapted to the other cases by defining the rate as in (30).

Let $(\mathcal{C}, \rho)$ be an intermingled code over the channel W and $G = (\mathcal{V}, \mathcal{E})$ the transition graph of the transmission states. Let M_G be the adjacency matrix of G. Similarly to the previous section, we recall that for all L, the number

of paths from v to v' of length L is given by $(M_G^L)_{vv'}$. Now by construction, $\#\mathcal{S}_L$ is equal to the number of paths starting from $(0,...,0)$ of length L and finishing at $(0,...,0)$, that is

$$\#\mathcal{S}_L = (M_G^L)_{(0,...,0)(0,...,0)} = \langle M_G^L, I_{0,0}\rangle, \tag{98}$$

with $I_{0,0} = (\mathbb{1}_{v=(0,...,0) \text{ and } v'=(0,...,0)})_{v,v'\in\mathcal{V}}$.

Now we can proceed similarly as in the proof of Lemma 4 in App. B.

$$r(\mathcal{C},\rho) = \lim_{L\to\infty} \frac{1}{L}\log \#\mathcal{S}_L \tag{99}$$

$$= \lim_{L\to\infty} \frac{1}{L}\log\langle M_G^L, I_{0,0}\rangle \tag{100}$$

$$= \lim_{L\to\infty} \log\left(\sqrt[L]{\|M_G^L\|}\right) + \frac{1}{L}\log\left(\left\langle \frac{M_G^L}{\|M_G^L\|}, I_{0,0}\right\rangle\right). \tag{101}$$

The quantity $\langle \frac{M_G^L}{\|M_G^L\|}, I_{0,0}\rangle$ is positive, bounded as L goes to infinity by Cauchy-Schwartz inequality, and does not converge to zero because of the positive rate. Thus $\frac{1}{L}\log\left(\left\langle \frac{M_G^L}{\|M_G^L\|}, I_{0,0}\right\rangle\right) \to 0$ when L goes to infinity and by using Gelfand formula, $\log\left(\sqrt[L]{\|M_G^L\|}\right)$ converges to $\log\max_i |\lambda_i(M_G)|$ where $(\lambda_i)_{i\leq\#\mathcal{V}}$ are the elements of the spectrum of M_G.

APPENDIX D

PROOF OF PROPOSITION VI.9

For all regular expressions E and E' such that $\mathcal{L}(E) \cap \mathcal{L}(E') = \varnothing$, we have

$$F_{E+E'}(z) = \sum_{l\in\mathbb{N}} \#(\mathcal{L}(E) \cup \mathcal{L}(E'))_{[l]} z^l \underset{\text{[Hyp.]}}{=} \sum_{l\in\mathbb{N}} (\#\mathcal{L}(E)_{[l]} + \#\mathcal{L}(E')_{[l]})z^l = F_E(z) + F_{E'}(z). \tag{102}$$

For all regular expressions E and E' we have

$$F_{EE'}(z) = \sum_{l\in\mathbb{N}} \#(\mathcal{L}(E) \cdot \mathcal{L}(E'))_{[l]} z^l = \sum_{l'\in\mathbb{N}} \sum_{l''\in\mathbb{N}} \#\mathcal{L}(E)_{[l']}\#\mathcal{L}(E')_{[l'']} z^{l'+l''} = F_E(z)F_{E'}(z). \tag{103}$$

For all regular expression E, let us note $E^l \doteq E \cdot ... \cdot E$ l times. Assume that the sets $(\mathcal{L}(E^l))_{l\in\mathbb{N}}$ are disjoint, then we have

$$F_{E^*}(z) = \sum_{l\in\mathbb{N}} \#\mathcal{L}(E^*)_{[l]} z^l \underset{\text{[Hyp.]}}{=} \sum_{l\in\mathbb{N}} \left(\sum_{l'\in\mathbb{N}} \#\mathcal{L}(E^{l'})_{[l]}\right) z^l \tag{104}$$

$$= \sum_{l'\in\mathbb{N}} \left(\sum_{l\in\mathbb{N}} \#\mathcal{L}(E^{l'})_{[l]} z^l\right) = \sum_{l'\in\mathbb{N}} F_{E^{l'}}(z) = \sum_{l'\in\mathbb{N}} F_E(z)^{l'} = \frac{1}{1 - F_E(z)}. \tag{105}$$

APPENDIX E

PROOF OF THEOREM VI.10

The following proof holds when Fekete's lemma can be applied directly to $\left(\log(\#\mathcal{L}(E)_{[L]})\right)_{L\in\mathbb{N}}$, that is when $\#\mathcal{L}(E)_{[L]} \neq 0$ for all L large enough, but the proof can be straightforwardly adapted to the other cases by redefining the rate as in (56).

Let E be a rational code, $A = (\mathcal{X}, \mathcal{S}, \tau, s_{start}, \mathcal{S}_{accept})$ a DFA which recognizes $\mathcal{L}(E)$ and E' be the regular expression such that $E = (E')^*$. Let G_A be the transition graph of the DFA A and M_A its adjacency matrix. Similarly to the proofs in App. B and C, for all $L \in \mathbb{N}$ we have

$$\#\mathcal{L}(E)_{[L]} = \sum_{s \in \mathcal{S}_{accept}} (M_A^L)_{s_{start}s}, \tag{106}$$

$$\nu(E) = \lim_{L \to \infty} \sqrt[L]{\#\mathcal{L}(E)_{[L]}} \tag{107}$$

$$= \lim_{L \to \infty} \sqrt[L]{\|M_A^L\|} \sqrt[L]{\left\langle \frac{M_A^L}{\|M_A^L\|}, I_{s_{start}, \mathcal{S}_{accept}} \right\rangle}, \tag{108}$$

where $I_{s_{start}, \mathcal{S}_{accept}} = (\mathbb{1}_{s=s_{start}, s' \in \mathcal{S}_{accept}})_{s, s' \in \mathcal{S}}$. By Gelfand's formula $\sqrt[L]{\|M_A^L\|} \underset{L \to \infty}{\to} \max_s |\lambda_s(M_A)|$, where $(\lambda_s(M_A))_{s \in \mathcal{S}}$ is the spectrum of M_A. On the other hand, the quantity $\left\langle \frac{M_A^L}{\|M_A^L\|}, I_{s_{start}, \mathcal{S}_{accept}} \right\rangle$ is positive, bounded as L goes to infinity by Cauchy-Schwartz inequality, and does not converge to zero because of the positive rate. Thus $\sqrt[L]{\left\langle \frac{M_A^L}{\|M_A^L\|}, I_{s_{start}, \mathcal{S}_{accept}} \right\rangle} \to 1$ when L goes to infinity.