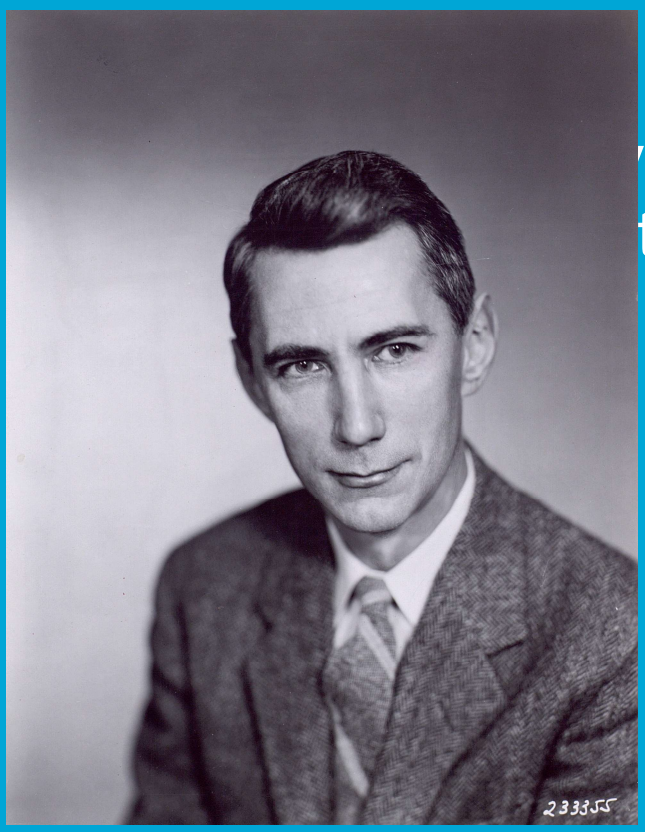
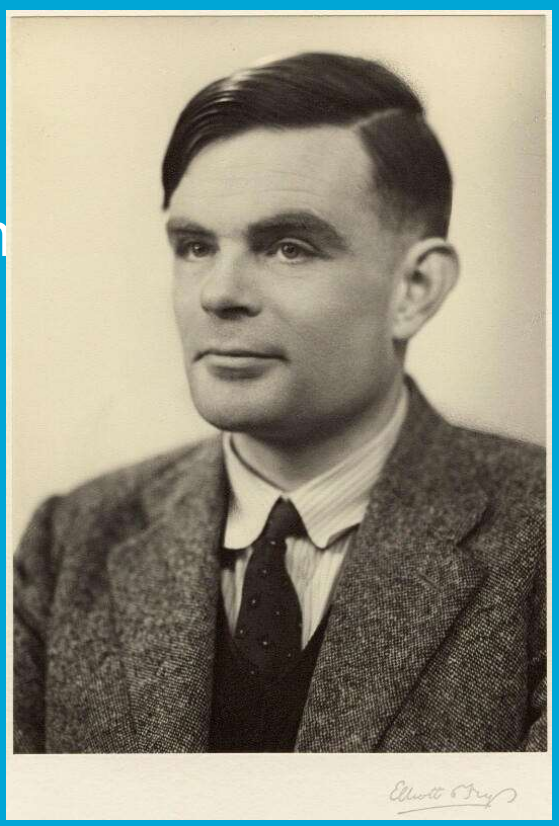
UNCOMPUTABILITY IN INFORMATION THEORY

Bassem Safieldeen
2019

Some info theoretic stuff is undecidable, dude.

Cool...
Show me.





TU Delft

Uncomputability in Information Theory

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

by

Bassem Safieldeen

August 2019

The work in this thesis was made in

David Elkouss's group
QuTech
Delft University of Technology

Supervisors:	Prof.dr. David Elkouss
Thesis Committee:	Prof. Niel Yorke-Smith (Algorithmics group, EWI, TU Delft)
	Prof. David Elkouss (QuTech, TU Delft)
	Prof. Martijn Caspers (Analysis group, EWI, TU Delft)

Abstract

We present a powerful approach for learning about uncomputability and undecidability in information theory. Our approach is to use automata from automata theory that have undecidable properties to construct channels for which an information-theoretic quantity is uncomputable or undecidable. We demonstrate this approach by showing that, for channels with memory, capacity is uncomputable and information-stability is undecidable.

Acknowledgements

I wanted to work on an idea that I came up with, and to have a senior scientist help me so that I think and learn quickly.
I wanted to study (quantum) information theory.
I wanted to learn to be precise and prove stuff properly.
I needed to learn how to write well.
David gave me an opportunity to do all that.
Thank you, David. I've learned a lot.

I wanted to have cool science friends who studied and thought about big stuff with me.
The students at QuTech, especially Kenneth Goodenough, gave me that.
Thank you, students of QuTech, especially Kenneth Goodenough.

There a lot of other people who have substantially helped me achieve whatever good qualities and achievements I have thus far achieved. I'm not going to mention them in detail here because it is difficult to properly do so without being dramatic. And because I like to think my journey has not yet reached its climax, I am going to save the dramatic acknowledgements till that time. However, just in case I do not reach such climax – perhaps because of an untimely demise due to [insert funny way to die here] – a full dramatic acknowledgement can be found in a paper in my wallet.

Anyway, I sincerely hope you live long and prosper, all of you.
...

Note on style

A lot of researchers, especially young ones who are not familiar with the jargon of the field, spend too much time trying to understand an badly written paper. Sometimes, when reading papers, I myself feel as if the authors and their colleagues are gathered in a small huddle whispering ideas to each other and they are making it harder than it should be for me to get in. Therefore, I have tried in this thesis — although I do not consider my attempt a definitive success — to be as inviting as possible to the readers.

One of the problems I have specifically tried to address is the lack of good visualization aids in scientific literature. Sometimes visualization is essential for quick and satisfying understanding. The authors of many papers are aware of this, as they needed to visualize the concepts themselves to be able to say something new about them. The new theorems they create often build upon their visual understanding of the concepts, and require visual aids themselves to be understood by other scientists. Instead of supplementing their papers with proper visual aids, the authors describe the picture in their minds in text, in the constrained style allowed by scientific journals, and throw the burden on the reader to decipher the text in order to reconstruct the picture.

As visual aids, figures are good; but often not good enough. Some concepts, like ones that involve assembling components in some order (see for example the definition of Post's Correspondence Problem in section 4.2.1), are best understood via animations or hand motions or some other time-dependent visual aid like that. To address this deficiency, through out the thesis I have put videos that explain such concepts. The links to the videos can be found next to curly brackets on the left margins of the pages.

Contents

1 Introduction

Information theory encompasses the study of communication channels. A communication channel is any medium through which information can pass; for example, a telephone line, or an optical fiber. In general, communication channels can be put into two categories: channels that change every time they are used, and channels that do not. The former are known as channels with memory, because they way they change depends in general on the input they received in past uses; the latter are known as memoryless channels.

Suppose we are given a communication channel $\mathbf{W}$ that we are told we can use a limited number of times. A natural question we might ask is the following: What is the maximmum amount of information we can transmit over $\mathbf{W}$ if we use it as efficiently as possible — if we do not send any unnecessarily redundant information? In the limit in which the number of times we are allowed to use the channel approaches ∞, the maximum amount of information that we can transmit per channel use is known as the channel's capacity. The capacity of memoryless channels is given by the formula [23]

$$C(\mathbf{W}) = \max_{X} I(X;Y), \tag{1}$$

where $I(X;Y)$ (see Section 2.2) is called the mutual information, and represents the amount of correlation between the random-variable input to the channel, X, and the random-variable induced at the output of the channel, Y. As given by (1), the quantity $C(\mathbf{W})$ is known to be efficiently computable: given any memoryless channel, we can use the so-called Blahut-Arimoto algorithm [3, 1] to efficiently approximate $C(\mathbf{W})$ to within any desired precision.

In terms of computability of capacity, the situation for channels with memory is different: It has long been unknown whether an algorithm like the Blahut-Arimoto algorithm exists, and, hence, whether capacity is computable. Although algorithms have been found that approximate the capacity of channels with memory ([19, 8, 22, 2, 24, 16, 14, 21, 27, 10] and more recently [28]), these algorithms either only work for special cases or cannot guarantee the precision of the approximation.

The aim of this paper is to present a powerful approach to studying the decidability and computability of information-theoretic properties of communication channels with memory. And as a demonstration of this approach, we will resolve the aforementioned open question: we will show that capacity is, in fact, uncomputable for channels with memory. A second demonstration of our approach will be to show that another information-theoretic property of channels, called information-stability, is undecidable.

The channels we use to prove our theorems are instances of what are known as finite-state machine channels (FSMC), which are channels whose behavior is dictated by a finite-state automaton. More precisely, the memory of an FSMC is modeled by the state of the finite-state automaton, and the way the memory of the channel changes after each input follows how the underlying automaton transitions between its different states in response to different inputs. This is the essence of our approach: by choosing the underlying automaton to be one that has a certain undecidable property, we can construct a channel that has an uncomputable or undecidable property.

2 Communication channels

Mathematically, a communication channel $\mathbf{W}$ is a sequence of conditional probability distributions:

$$\mathbf{W} = (W_1(y_1|x_1), W_2(y_2|x_2), W_3(y_3|x_3), \dots)\,. \tag{2}$$

The ith conditional probability distribution in the sequence determines the probability that the ith output of the channel will be y_i if the ith input was x_i. For memoryless channels, because the channel does not change from one input to another, all the conditional probability distributions in the sequence are equal: $W_1(y|x) = W_2(y|x) = \cdots = W(y|x)$. On the other hand, for channels with memory the distributions are not in general equal, because the ith conditional probability distribution depends on the state of the memory in the ith use.

The set of input symbols that can be input to a channel is its input alphabet; the set of symbols that it can output is its output alphabet. For example, a channel with input alphabet $\{0, 1\}$ and output alphabet $\{0, 1, 2\}$ can receive 0 or 1 at the input and produce 0, 1, or 2 at the output. In addition to input and output alphabets, channels with memory have a set of memory states.

2.1 Finite-state machine channels (FSMC)

Finite-state machine channels (FSMC) are a type of channels with memory whose memory is modeled by a finite-state automaton. Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}$ denote finite sets that represent the FSMC's input alphabet, output alphabet, and the set of states, respectively. An FSMC can be fully characterized by a set of conditional probabilities $\{p(y, s|x, s')\}_{s,y,x,s'}$, where each $p(y, s|x, s')$ is the probability that the FSMC is going to output y and transition to the state s given that the input was x and it was in the state s', where $s, s' \in \mathcal{S}$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$. The sequence that determines the action of the channel (Eq. (2)) then satisfies $W_i(y_i|x_i) \in \{p(y, s|x, s')\}_{s,y,x,s'}$.

2.2 Mutual information and information-stability

2.2.1 Mutual information

For any communication channel, the information transmission capability of the channel can be quantified in terms of the amount of correlation between what we input to the channel and what it outputs. This correlation is captured by the concept of the mutual information $I(X;Y)$ between a random variable X that we input to the channel and the random variable Y that the channel outputs. The mutual information appears in formulas for the capacity, like Eq. (1) and Eq. (5).

Mutual information is defined as follows. For any two random variables X and Y with joint probability distribution $p_{X,Y}(x, y)$, define the random variable $i_{X,Y}(x; y)$ as

$$i_{X,Y}(x; y) \equiv \log \frac{p(y|x)}{p(y)}.$$

The random variable $i_{X,Y}(x; y)$ is called the information-density, and its expected value is the mutual information $I(X;Y)$ between X and Y:

$$I(X;Y) \equiv \langle i_{X,Y}(x; y)\rangle_{XY} = \sum_{x,y} p(x, y) \log \frac{p(y|x)}{p(y)}.$$

The mutual information is also given by the following expression in terms of the Shannon entropy:

$$I(X;Y) = H(Y) - H(Y|X),$$

where, for any sequence of random variables $\mathbf{X^n}$, $H(\mathbf{X^n})$ – the Shannon entropy of $\mathbf{X^n}$ – is the expected value of $h_{\mathbf{X^n}}(x^n) \equiv -\log(p_{\mathbf{X^n}}(x^n))$.

2.2.2 Information-stability

There are several formulas for computing the capacity of communication channels. Although all of them involve mutual information, the formula to be used to compute the capacity of a given channel depends on whether the channel is *information-stable*. Here is a rough intuition for when a channel is information-stable. Suppose there is a store that sells channels. In the store there are boxes arranged on the shelves, and all of them contain the same channel $\mathbf{W}$. Someone goes into the store, buys one instance of $\mathbf{W}$, takes it home and starts using it. After a while of using the channel, the owner gets a feeling for the "usefulness" of the channel for transmitting information – how often the channel successfully passes the input to the output. If this "usefulness" is equal to the average "usefulness" of all the other $\mathbf{W}$ channels in the store, then $\mathbf{W}$ is information-stable.

To illustrate this intuition, here is an example of a channel that is NOT information-stable. Consider a channel $\mathbf{W}$ that, when the "on" button is pressed, becomes an noiseless channel with probability 1/2 or a completely noisy channel with probability 1/2, and stays like that forever. Suppose a channel store is stocked with this channel $\mathbf{W}$. A customer comes in and buys one instance of $\mathbf{W}$. On the first use, the channel, unbeknownst to the lucky customer, switches to being a forever noiseless channel. After using the channel for a while and seeing how the channel is perfectly useful, this person might guess that all the store only sells perfect channels. However, this guess would be misplaced, because approximately half of channels in the store, when used, are completely noisy and useless. Since the usefulness of one instance of $\mathbf{W}$ leads to a wrong guess on how useful, on average, a store-bought channel $\mathbf{W}$ is going to be, $\mathbf{W}$ is not an information-stable channel.

This intuition is captured precisely by the following mathematical definition of information-stability.

Definition 2.1. (Information-stability [5, 25]) A channel $\mathbf{W}$ is said to be information-stable if and only if there exists a sequence of inputs $\{X_i\}_{i=1}^{\infty}$ that satisfies the condition that, for any $\lambda > 0$,

$$\lim_{n\to\infty} \Pr\left[\left|\frac{i_{\mathbf{X^n},\mathbf{W^n}}(\mathbf{X^n};\mathbf{Y^n})}{\sup_{\mathbf{X^n}} I(\mathbf{X^n};\mathbf{Y^n})} - 1\right| > \lambda\right] = 0. \tag{3}$$

The random variable $i_{\mathbf{X^n},\mathbf{W^n}}(\mathbf{X^n};\mathbf{Y^n})$ represents the "usefulness" of a single instance of $\mathbf{W}$ that a customer takes home from the store in the example above; the term $\sup_{\mathbf{X^n}} I(\mathbf{X^n};\mathbf{Y^n})$ represents the average usefulness of all the channels in the store. The information-stability condition (3) then just means the following: in the limit of infinitely many uses, the probability that the usefulness of a single instance of the channel is different from the average usefulness of all the channels in the store is 0.

Another way of saying this is that a channel is information-stable if and only if the input process that maximizes the mutual information and the process it induces at the output of the channel both behave ergodically. It is because of this ergodicity that the random variable $i_{\mathbf{X^n},\mathbf{W^n}}(\mathbf{X^n};\mathbf{Y^n})$ converges (for the right input process) by the law of large numbers to its average $\sup_{\mathbf{X^n}} I(\mathbf{X^n};\mathbf{Y^n})$, thereby satisfying condition (3).

3 Statement of the results and sketch of their proofs

Now that we have formally defined what communication channels are — basically a sequence of conditional probability distributions $(W_i(x_i|y_i))$ —, we are ready to state our results.

In a nutshell, our results are the following: The first result is that there exist channels with memory whose capacity is uncomputable; the second result is that there exist channels with memory whose information-stability is undecidable.

Formally, our results are the following.

Theorem 3.1. (Capacity is uncomputable) Any function that takes as input the probability assignments $\{p(y,s|x,s')\}_{s,y,x,s'} = \{p(y|x,s')p(s|x,s')\}_{s,y,x,s'}$ of an information-stable finite-state machine channel $\mathbf{W}$ with input, output, and number of states point-wise equal or larger than $(7,2,30)$ and outputs a rational number c such that the capacity $C(\mathbf{W})$ satisfies

$$|C(\mathbf{W}) - c| < 1/4$$

must be uncomputable.

Theorem 3.2. (Information-stability is undecidable) Given the probability assignments $\{p(y,s|x,s')\}_{s,y,x,s'} = \{p(y|x,s')p(s|x,s')\}_{s,y,x,s'}$ of an finite-state machine channel $\mathbf{W}$ with input, output, and number of states point-wise equal to or larger than $(7,2,30)$, it is undecidable whether $\mathbf{W}$ is information-stable or not.

To arrive at the above results, we use FSMCs whose memory is based on finite automata that have been shown in the automata theory literature to have an undecidable property. Let us be more specific: The automata we use are called probabilistic finite automata (PFA). These are finite automata for which the transitions between the different states are probabilistic. Given a PFA, its so-called value is the supremum – over all possible strings of input symbols – of the probability that the automaton ends up in a certain state.

In Ref. [7], a family of PFAs $\mathcal{B}_\alpha$ was given for which it is undecidable whether the value is 1 or α, with $\alpha \in [1/2, 1)$. In Section 4.3 we update the construction of $\mathcal{B}_\alpha$ to reduce its number of states; we call the new construction $\mathcal{B}'_\alpha$. With $\mathcal{B}'_\alpha$ in hand, we can construct an FSMC as follows (full details in Section 2.1). In each use, we can input a symbol from $\{a, b, c, c^*, 0, 1, rt\}$ into the channel input. A PFA $\mathcal{B}'_\alpha$ has two types of states, accepting states and non-accepting states. Right after a symbol is input, two things happen. The first thing that happens is that the channel checks if $\mathcal{B}'_\alpha$ is in an accepting state or a non-accepting state; if non-accepting, the channel outputs 0 or 1 uniformly at random; if accepting, the channel checks if the input was 0 or 1, or if it was one of the other symbols; if 0 or 1, the channel passes the input to the output as is; if one of the other symbols, the channel outputs 0 or 1 uniformly at random. The second thing that happens is, after the channel produces an output, $\mathcal{B}'_\alpha$ transitions to a new state depending on the input.

This particular channel construction allows us to relate the expected information transmission capability of any of these channels to the value of the underlying PFA (Lemma 6.7):

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_\alpha}. \tag{4}$$

The expression on left-hand side of this equation, as it turns out [26, 5], gives the capacity of information-stable channels; i.e, for information-stable channels

$$C(\mathbf{W}) = \lim_{n\to\infty} \sup_{\mathbf{X^n}} \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}). \tag{5}$$

Using this fact, we connect the undecidability in $val_{\mathcal{B}'_\alpha}$ to the capacity of our channels by showing that those of our channels whose memory is modeled by PFAs $\mathcal{B}'_\alpha$ with $\alpha = 1/2$ are information-stable (Lemmas 7.1 and 7.2). Upon doing so, we will be able to write

$$C(\mathbf{W}) = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_{\alpha=1/2}}. \tag{6}$$

Then, since it is undecidable whether $val_{\mathcal{B}'_\alpha} = \alpha = 1/2$ or $val_{\mathcal{B}'_\alpha} = 1$, it is undecidable whether $C(\mathbf{W}) = \alpha$ or $C(\mathbf{W}) = 1$; this implies that $C(\mathbf{W})$ is uncomputable to any precision $< \frac{1-\alpha}{2} = 1/4$, proving our first result.

As for the second result (Theorem 3.2), we prove it as follows. In Section 7 (Lemmas 7.1 and 7.2) we prove that the channels built on top of some PFAs $\mathcal{B}'_\alpha$ are information-stable when $\mathcal{B}'_\alpha$ has value 1 and is not information-stable when $\mathcal{B}'_\alpha$ has value α. Then, since it is undecidable whether $val_{\mathcal{B}'_\alpha} = \alpha$ or $val_{\mathcal{B}'_\alpha} = 1$, it is undecidable whether these channels are information-stable or not.

4 Probabilistic finite automata (PFA)

It is time to lay down the technical foundations we will need to build our results. This section is devoted to reviewing PFAs in general (Section 4.1), reviewing the PFA family $\mathcal{B}_\alpha$ from Ref. [7] in particular (Section 4.2), and finally presenting a PFA family $\mathcal{B}'_\alpha$ with a smaller number of states than $\mathcal{B}_\alpha$ (Section 4.3); recall that $\mathcal{B}'_\alpha$ is the family of PFAs with an undecidable property on which our channel construction (next section) is based. This section can be skipped on first reading. However, the reader is advised to read the statement of Lemma 4.8 and to look at the design of the PFA $\mathcal{B}_\alpha$ in Fig. 2, as they will be referred to throughout the rest of the text.

4.1 Definition

A PFA $\mathcal{P}$ is given by a tuple $\mathcal{P} = (\mathcal{Q}, \mathcal{N}, \mathcal{K}, \mathbf{v}, \mathcal{F})$, where $\mathcal{Q}$ is a finite set containing the states of $\mathcal{P}$, $\mathcal{N}$ is a finite set containing the input alphabet of $\mathcal{P}$, $\mathcal{K}$ is a finite set containing stochastic matrices – one matrix for every input symbol in $\mathcal{N}$–, $\mathbf{v}$ is a vector containing the initial probability distribution over the states $\mathcal{Q}$, and $\mathcal{F} \subset \mathcal{Q}$ is the set of accepting states of $\mathcal{P}$.

We denote by $p[q_a \xrightarrow{w} q_b]$ the probability that the PFA transitions from the state q_a to the state q_b upon reading the input symbol w; this probability is given by

$$p[q_a \xrightarrow{w} q_b] = \pi_{q_b}^T K_w \pi_{q_a},$$

where π_q is a column vector with one in the position of the state q and zeroes elsewhere, and $K_w \in \mathcal{K}$ is the stochastic transition matrix corresponding to the symbol w. Using this notation, the probability that a PFA transitions from q_a to q_b upon reading the word $\mathbf{w} \equiv w_1 \dots w_{|\mathbf{w}|} \in \mathcal{N}^{|\mathbf{w}|}$ is given by

$$p[q_a \xrightarrow{\mathbf{w}} q_b] = \pi_{q_b}^T K_{w_{|\mathbf{w}|}} \cdot \ldots \cdot K_{w_1} \pi_{q_a}.$$

We denote by $val_{\mathcal{P}}(\mathbf{w})$ the probability that the PFA $\mathcal{P}$ accepts upon reading the word $\mathbf{w}$: the probability that $\mathcal{P}$ starts from the initial probability distribution $\mathbf{v}$ and, upon reading $\mathbf{w}$, ends up in any of the accepting states $\mathcal{F}$ of $\mathcal{P}$. $val_{\mathcal{P}}(\mathbf{w})$ is given by

$$val_{\mathcal{P}}(\mathbf{w}) = \sum_{F \in \mathcal{F}} \pi_F^T K_{w_{|\mathbf{w}|}} \cdot \ldots \cdot K_{w_1} \mathbf{v}.$$

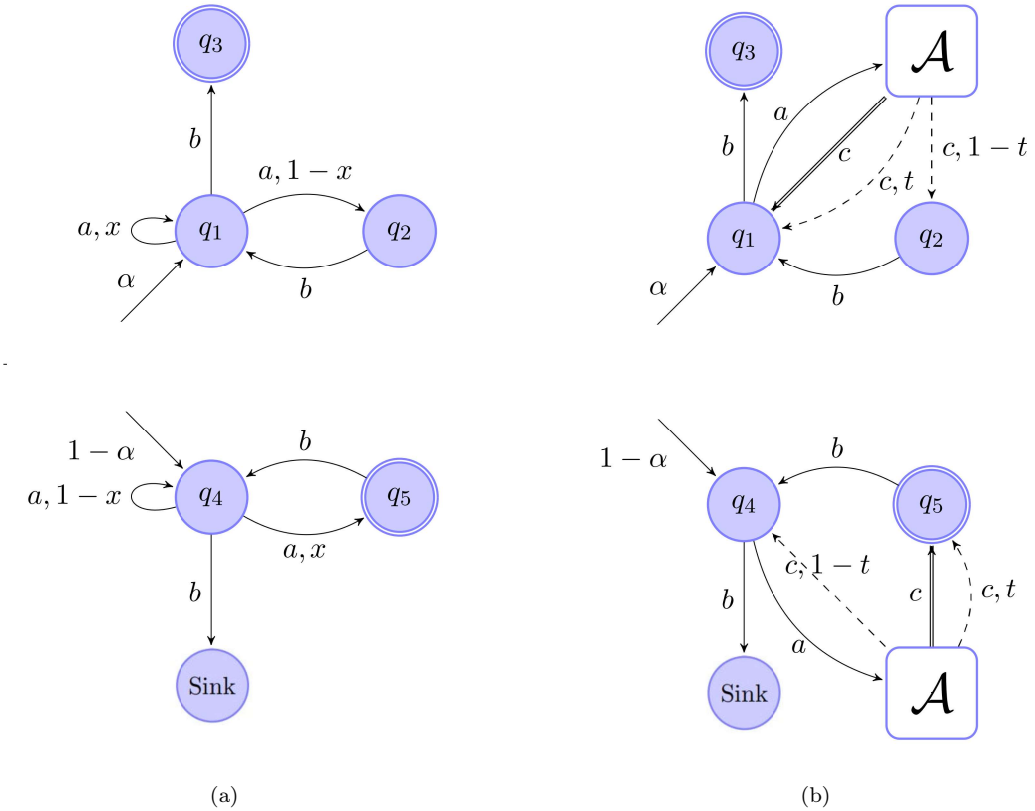
The supremum of acceptance probabilities over all input words to a PFA $\mathcal{P}$ is called the value of $\mathcal{P}$, and is denoted by $val_{\mathcal{P}}$:

$$val_{\mathcal{P}} \equiv \sup_{\mathbf{w} \in \mathcal{N}^*} val_{\mathcal{P}}(\mathbf{w}),$$

where $\mathcal{N}^*$ is the set of finite length words in $\mathcal{N}$.

In our diagramatic representation of the different automaton constructions (Fig. 1, Fig. 2, and Fig. 4) we will adopt the following conventions. A state is represented by a circle. If the state is accepting, the circle has a double line around it. Initial states are represented by circles to which origin-less arrows point. If the automaton transitions from a state q_a to a state q_b with probability p upon reading the letter w, this transition is depicted on the diagram by an arrow $\xrightarrow{w,p}$ that points from q_a to q_b. To avoid clutter, we do not show self-loops that occur with probability 1, and if a transition occurs with probability 1 but is not a self-loop we drop the probability and simply write $\xrightarrow{w}$. Additionally, if all input symbols trigger the same transition with the same probability p we simply depict that transition with $\xrightarrow{p}$.

Moreover, if after the first use one of the PFAs in this paper started in the upper(lower) branch, i.e. started in the state $q_1(q_4)$ shown on the figures, we will use $val_{up(low)}(\mathbf{w})$ to denote the probability that the PFA ends up in an accepting state when the word $\mathbf{w}$ is read. It will be clear from context which PFA we are talking about.



(a) (b)

Figure 1: The PFA construction on the left (with α fixed to $1/2$) is originally due to [7]. Note that this construction has 6 states. The PFA $\mathcal{A}$ from Lemma 4.1 is embedded in the construction on the left to give the construction on the right, which we refer to as $\mathcal{B}_\alpha$. Note that this construction has $6 + 2n_\mathcal{A}$, where $n_\mathcal{A}$ is the number of states of $\mathcal{A}$ and in this paper will take values $20, 25, 30, \ldots$, according to Lemma 4.3. The symbol c in $\mathcal{B}_\alpha$ takes $\mathcal{B}_\alpha$ from $\mathcal{A}$ to a state outside $\mathcal{A}$; the double line denotes that the transition was from an accepting state in $\mathcal{A}$ and the dashed line denotes that the transition was from a non-accepting state. Besides the symbols shown on the figure, $\mathcal{B}_\alpha$ has a reset symbol which takes it to q_1 if it started in upper branch and to q_4 if it started in the lower branch. Transitions inside $\mathcal{A}$ are triggered by the symbols 0 and 1, which are the same symbols that constitute the channels' output alphabet. t is a rational number that is chosen such that the symbol c causes $\mathcal{B}_\alpha$ to transition from $\mathcal{A}$ to $q_1(q_5)$ with maximum probability $val_\mathcal{A} + t(1 - val_\mathcal{A})$. In particular, t is chosen to be a fixed number such that the symbol c causes $\mathcal{B}_\alpha$ to transition from $\mathcal{A}$ to $q_1(q_5)$ with maximum probability $> 1/2$ when $val_\mathcal{A} > \delta$ and with maximum probability $\leq 1/2$ when $val_\mathcal{A} \leq \delta$, where δ is a rational number such that it is undecidable whether $val_\mathcal{A} > \delta$ or $val_\mathcal{A} \leq \delta$. This is needed for the proof of Lemma 4.7. The PFA $\mathcal{B}_\alpha$ has input alphabet $\{a, b, c, 0, 1, rt\}$: 6 symbols.

4.2 The PFA family $\mathcal{B}_\alpha$

In Ref. [7], the authors presented a family of PFAs for which it is undecidable whether the value is 1 or α, where α is some number such that $0 < \alpha < 1$. To construct $\mathcal{B}_\alpha$, the authors of [7] start with the PFA shown in Fig. 1(a), which they show (see Lemma 4.4) has value 1 when $x > 1/2$ and value α when $x \leq 1/2$. To get $\mathcal{B}_\alpha$, the PFA in Fig. 1(a) is then modified such that it is effectively undecidable whether $x > 1/2$ or $x \leq 1/2$. This is accomplished by embedding another PFA due to Hirvensalo [13] – which we refer to as the PFA $\mathcal{A}$ – in the PFA in Fig. 1(a), in a way such that, for some rational number δ, $val_{\mathcal{A}} > \delta$ has the same effect as $x > 1/2$ and $val_{\mathcal{A}} \leq \delta$ has the same effect as $x \leq 1/2$. This produces the PFA $\mathcal{B}_\alpha$ shown in Fig. 1(b). Because $\mathcal{A}$ is such that it is undecidable whether $val_{\mathcal{A}} > \delta$ or $val_{\mathcal{A}} \leq \delta$, it is then undecidable whether $val_{\mathcal{B}_\alpha} = 1$ or $val_{\mathcal{B}_\alpha} = \alpha$.

4.2.1 The PFA $\mathcal{A}$

The undecidability in $val_{\mathcal{A}}$ obeys the following lemma:

Lemma 4.1. [13] Let $\mathcal{A}$ be a PFA with $(5k-10)$ states and an alphabet of size 2, where k is an integer such that $k \geq 7$. And let δ be a rational number such that $\delta > 1/(5k-10)$. It is undecidable whether $val_{\mathcal{A}} > \delta$ or $val_{\mathcal{A}} \leq \delta$.

The PFA $\mathcal{A}$ inherits this undecidability in its value from Post's Correspondence Problem (PCP), which is defined as follows:

Definition 4.1. (PCP) Given k pairs of words (u_i, v_i), where $u_i, v_i \in \Sigma^*$ and Σ is a finite alphabet, decide whether there exists $i_1, \ldots, i_n \in \{1, \ldots, k\}^+$ such that $u_{i_1} u_{i_2} \ldots u_{i_{n-1}} u_{i_n} = v_{i_1} v_{i_2} \ldots v_{i_{n-1}} v_{i_n}$.

At the time [13] was published, PCP was known to be undecidable for $k \geq 7$, and to prove Lemma 4.1 Hirvensalo [13] basically mapped an instance of PCP to a PFA. Recently, Neary [20] improved the PCP undecidability result by showing that it is undecidable for five pairs of words ($k \geq 5$). More precisely, Neary proved the following:

Lemma 4.2. ([20]) Given the five pairs of words

$$\{(p_1, q_1), (p_2, q_2), (p_3, q_3), (p_4, q_4), (p_5, q_5)\}$$
$$= \{(1, 1u), (1\underbrace{0\ldots0}_{\beta}1, 110), (1\underbrace{0\ldots0}_{\beta}1, 0), (1, 0), (1\underbrace{0\ldots0}_{\beta}1111, 1111)\},$$

where $u \in \{0,1\}^*$, $\beta \in \mathbb{N}$, and $\beta \geq 1$, it is undecidable whether there exists $i_1, \ldots, i_n \in \{1,2,3,4\}^+$ such that $p_1 p_{i_1} p_{i_2} \ldots p_{i_n} = q_1 q_{i_1} q_{i_2} \ldots q_{i_n}$.

This recent result can be used to modify Lemma 4.1 so that it applies to PFAs with smaller memory sizes. This is straightforward but has to be done with care, since Hirvensalo [13] uses what are known as Claus instances of PCP to prove Lemma 4.1, while the the five pairs of words in Lemma 4.2 do not constitute such an instance.

The Claus formulation of PCP can be defined as follows.

Definition 4.2. (PCP: Claus) [4, 9] Given k pairs of words (u_i, v_i), where $u_i, v_i \in \Sigma^*$ and Σ is a finite alphabet, decide whether there exists $i_1, \ldots, i_n \in \{2, \ldots, k-1\}^+$ such that $u_1 u_{i_1} \ldots u_{i_n} u_k = v_1 v_{i_1} \ldots v_{i_n} v_k$.

In other words, the standard PCP (Definition 4.1) and the Claus formulation of PCP (Definition 4.2) differ in that the latter requires that the first and last word pairs, (u_1, v_1) and (u_k, v_k), be used only at the beginning and end of the string.

The five pairs of words in Lemma 4.2 are almost a Claus PCP instance in that the first pair is fixed. To adapt Lemma 4.2 such that it is usable in Lemma 4.1, we can simply add a dummy pair of words to the set of five words, resulting in a Claus PCP instance of six pairs of words where the first pair is fixed to the beginning and the dummy pair is fixed to the end. This gives us the following new lemma for the PFA $\mathcal{A}$:

Lemma 4.3. Let $\mathcal{A}$ be a PFA with $(5k-10)$ states and an alphabet of size 2, where k is an integer such that $k \geq 6$. And let δ be a rational number such that $\delta > 1/(5k-10)$. It is undecidable whether $val_{\mathcal{A}} > \delta$ or $val_{\mathcal{A}} \leq \delta$.

4.2.2 The value of $\mathcal{B}_\alpha$ is undecidable

To derive the undecidability in the value of $\mathcal{B}_\alpha$, we begin by showing that the PFA in Fig. 1(a) has value 1 when $x > 1/2$ and value α otherwise (Lemma 4.4), then we embed $\mathcal{A}$ in the PFA in Fig. 1(a) – producing $\mathcal{B}_\alpha$ (Fig. 1(b)) – such that $val_{\mathcal{A}} > \delta$ has the same effect as $x > 1/2$ and $val_{\mathcal{A}} \leq \delta$ has the same effect as $x \leq 1/2$, from which it will follow that $val_{\mathcal{B}_\alpha} = 1$ when $val_{\mathcal{A}} > \delta$ and $val_{\mathcal{B}_\alpha} = \alpha$ when $val_{\mathcal{A}} \leq \delta$ (Lemma 4.6). It will then immediately follow that it is undecidable whether $val_{\mathcal{B}_\alpha} = 1$ or $val_{\mathcal{B}_\alpha} = \alpha$ (Lemma 4.7).

Lemma 4.4. ([7]) The value of the PFA in Fig. 1(a) is 1 when $x > 1/2$ and α when $x \leq 1/2$.

Proof. First we are going to show that for the PFA in Fig. 1(a) $val = \alpha$ when $x \leq 1/2$. Note that if we enter $'b'$, the upper branch ends up in an accepting state and the lower branch ends up in q_6 with certainty, and so for this word the acceptance probability is α. We can check that this is the maximum acceptance probability from the following chain of inequalities:

$$\begin{aligned} val(\mathbf{w}_k) &= \alpha\, val_{up}(\mathbf{w}_k) + (1-\alpha)\, val_{low}(\mathbf{w}_k) \\ &\leq \alpha\,(1-\epsilon) + (1-\alpha)\,\epsilon \\ &\leq \alpha\,(1-\epsilon) + \alpha\,\epsilon \\ &= \alpha, \end{aligned}$$

where the first inequality follows from Lemma 4.5 below. Thus, since α is the maximum probability that any word could give when $x \leq 1/2$, we have that the value is α when $x \leq 1/2$.

Now we are going to prove that the value is 1 when $x > 1/2$, and we do that by showing that for any $\epsilon \in (0, x)$ there exists a word $\mathbf{w}_k$ of the form given in (13) such that

$$p[q_4 \to q_6] \leq \epsilon \tag{7}$$

$$p[q_1 \to q_3] \geq 1 - \epsilon. \tag{8}$$

Let the lengths $n_2, \ldots, n_k$ in (13) be given by

$$n_k = \left\lceil \log_x \frac{1}{k} + C_\epsilon \right\rceil, \tag{9}$$

where $C_\epsilon = \frac{1}{b} \log_x \left(\frac{b-1}{b}\epsilon\right)$, and let $b > 1$ be a number such that $x^b = 1 - x$. The following chain of inequalities holds:

$$\begin{aligned} p[q_4 \to q_6] &= (1-x)^{n_2} + (1-(1-x)^{n_2})(1-x)^{n_3} + \cdots \\ &\leq \sum_{i=2}^{k} (1-x)^{n_i} \\ &= \sum x^{b n_i} \\ &= \sum x^{b\left\lceil \log_x \frac{1}{i} + C_\epsilon \right\rceil} \\ &\leq x^{bC_\epsilon} \sum x^{b \log_x \frac{1}{i}} \\ &= x^{bC_\epsilon} \sum_{i=2}^{k} \frac{1}{i^b}. \end{aligned} \tag{10}$$

Note that the sum in (10) when k goes to ∞ is very similar to the Riemann zeta function evaluated at a real argument strictly larger than one, which can be upper bounded by [15] $\zeta(b) = \sum_{n=1}^{\infty} \frac{1}{n^b} \leq \frac{b}{b-1}$. If we apply this bound to Eq. (10) we obtain

$$\begin{aligned}\lim_{k\to\infty} p[q_4 \to q_6] &\leq \lim_{k\to\infty} x^{bC_\epsilon} \sum_{i=2}^{k} \frac{1}{i^b} \\ &\leq x^{bC_\epsilon} \frac{b}{b-1} \\ &= \epsilon. \end{aligned} \tag{11}$$

Eq. (11) remains an upper bound for any finite k since we are only dropping positive contributions. Hence, ineq. (7) holds for all k.

Now let us check that (8) also holds. Consider the following sum

$$\begin{aligned}\sum_{i=2}^{k} x^{n_i} &\geq \sum x^{\log_x \frac{1}{i} + C_\epsilon + 1} \\ &= x^{C_\epsilon+1} \sum x^{\log_x \frac{1}{i}} \\ &= x^{C_\epsilon+1} \sum_{i=2}^{k} \frac{1}{i}. \end{aligned} \tag{12}$$

This sum diverges for any non-zero x and finite C_ϵ, which implies that $\lim_{k\to\infty} \prod_{i=2}^{k}(1-x^{n_i}) = 0$ and that there exists finite k such that $\prod_{i=2}^{k}(1-x^{n_i}) \leq \epsilon$. Then, $p[q_1 \to q_3] \geq 1-\epsilon$.

□

Lemma 4.5. When $x \leq 1/2$, any input word that causes the PFA in Fig. 1(a) to accept with probability $1-\epsilon$ if it started in the upper branch must cause it to accept with probability $\leq \epsilon$ if it started in the lower branch; i.e., if $val_{up}(\mathbf{w}) = 1-\epsilon$ then $val_{low}(\mathbf{w}) \leq \epsilon$.

Proof. Note that any word that ends with a $'b'$ cannot cause the lower branch to end up in the accepting state q_5, and for these words too the statement of the lemma holds because $val_{low}(\mathbf{w}) = 0 \leq \epsilon$ no matter what ϵ is. The same holds for any word that contains two consecutive $'b'$ and any word that starts with a $'b'$.

Next, note that any word that does not contain any $'b'$ cannot cause the upper branch to accept ($val_{up}(\mathbf{w}) = 0$), and for these words the statement of the lemma holds trivially because $val_{low}(\mathbf{w}) \leq \epsilon = 1$.

Besides those words, any remaining word that we could enter must be of the form

$$\mathbf{w}_k = a^{n_2} b a^{n_3} b \dots b a^{n_k}. \tag{13}$$

We know that for any word $\mathbf{w}$, $val_{up}(\mathbf{w}) = p[q_1 \to q_3]$ and

$$\begin{aligned} val_{low}(\mathbf{w}) &= p[q_4 \to q_5] \\ &= 1 - p[q_4 \to q_6] - p[q_4 \to q_4] \\ &\leq 1 - p[q_4 \to q_6]. \end{aligned}$$

We can also easily see that, for words of the form given in (13), we have that $p[q_1 \to q_3] = 1 - \Pi_{i=1}^{k}(1-x^{n_i})$ and $p[q_4 \to q_6] = 1 - \Pi_{i=1}^{k}(1-(1-x)^{n_i})$. Since $x \leq 1-x$ whenever $x \leq 1/2$, it follows that

$$1 - \Pi_{i=1}^{k}(1-x^{n_i}) \leq 1 - \Pi_{i=1}^{k}(1-(1-x)^{n_i}). \tag{14}$$

Thus, if $val_{up}(\mathbf{w}_k) = p[q_1 \to q_3] = 1-\epsilon$, ineq. (14) says that $p[q_4 \to q_6] \geq 1-\epsilon$, which implies that $val_{low}(\mathbf{w}_k) \leq 1 - p[q_4 \to q_6] \leq \epsilon$. □

Now we are going to embed the PFA $\mathcal{A}$ from Lemma 4.1 into the PFA in Fig. 1(a) to get the PFA $\mathcal{B}_\alpha$ shown in Fig. 1(b). The main idea is that x is replaced by the probability that $\mathcal{A}$ accepts a word $\mathbf{w}_\mathcal{A}$.

Lemma 4.6. The PFA $\mathcal{B}_\alpha$ shown in Fig. 1(b) has value 1 when $val_\mathcal{A} > \delta$ and value α when $val_\mathcal{A} \leq \delta$ (recall from Lemma 4.1 that δ is a rational number such that it is undecidable whether $val_\mathcal{A} > \delta$ or $val_\mathcal{A} \leq \delta$).

Proof. First, note that $\mathcal{B}_\alpha$ transitions to $\mathcal{A}$ it continues inside $\mathcal{A}$ until the symbol $'c'$, which is a symbol outside the alphabet of $\mathcal{A}$, is entered. Let $\mathbf{w}_\mathcal{A}$ be an arbitrary input word into $\mathcal{A}$. Suppose we enter the symbol $'a'$, followed by $w_\mathcal{A}$, followed by the symbol $'c'$. Then

$$p[q_1 \xrightarrow{a\mathbf{w}_\mathcal{A}c} q_1] = p[q_4 \xrightarrow{a\mathbf{w}_\mathcal{A}c} q_5] = val_\mathcal{A}(\mathbf{w}_\mathcal{A}) + t(1 - val_\mathcal{A}(\mathbf{w}_\mathcal{A})),$$

where t is a rational number chosen such that $\delta + t(1-\delta) = 1/2$, which has the effect that the probability that the symbol $'c'$ takes $\mathcal{B}_\alpha$ to the state $q_1(q_5)$ with probability $> 1/2$ when $val_\mathcal{A} > \delta$ and with probability $\leq 1/2$ when $val_\mathcal{A} \leq \delta$.

Let us assume first that $val_\mathcal{A} > \delta$. Then there exists some word $\mathbf{w}_\mathcal{A}$ such that $val_\mathcal{A}(\mathbf{w}_\mathcal{A}) > \delta$. Hence we can construct the input word

$$\mathbf{w}_k = (a\mathbf{w}_\mathcal{A}c)^{n_2} \dots (a\mathbf{w}_\mathcal{A}c)^{n_k} \tag{15}$$

with the lengths $n_2 \dots n_k$ given by Eq. (9). As with the automaton in Fig. 1(a), we have that for $\epsilon > 0$, there exists a k such that $\mathbf{w}_k$ satisfies conditions (7) and (8).

Now let us assume that $val_\mathcal{A} \leq \delta$. Since the word "$bb$" still gets us value y, we can restrict our attention to words of the form $(a\mathbf{w}^1_\mathcal{A}c)^{n_1}b \dots b(a\mathbf{w}^k_\mathcal{A}c)^{n_k}$. Furthermore, letting $g(\mathbf{w}) \equiv val_\mathcal{A}(\mathbf{w}) + t(1 - val_\mathcal{A}(\mathbf{w}))$, for any word $\mathbf{w}_\mathcal{A}$ we have that $g(\mathbf{w}_\mathcal{A}) \leq 1 - g(\mathbf{w}_\mathcal{A})$ and in consequence

$$1 - \prod_{i=1}^{k}(1 - g(\mathbf{w}^i_\mathcal{A})^{n_i}) \leq 1 - \prod_{i=1}^{k}(1 - (1 - g(\mathbf{w}^i_\mathcal{A})^{n_i})),$$

just like in ineq. (14).

For $\epsilon > 0$, the above inequality implies that for any word $\mathbf{w}_k$ such that $p[q_1 \to q_4] = 1 - \epsilon$ we have that $p[q_4 \to q_6] \geq 1 - \epsilon$ and $val_{\mathcal{B}_\alpha}(\mathbf{w}_k) \leq \alpha$, as is the case with the PFA in Fig. 1(a).

So far, we have not considered any word that contains the symbol rt. However, note that any word $\mathbf{w} = \mathbf{w}_1\, rt\, \mathbf{w}_2\, rt \dots rt\, \mathbf{w}_t$, where $\mathbf{w}_i$ are words that do not contain rt, has acceptance probability $val(\mathbf{w}_t)$, and any word that ends with a rt has acceptance probability 0.

Thus, $val_{\mathcal{B}_\alpha} = 1$ when $val_\mathcal{A} > \delta$ and $val_{\mathcal{B}_\alpha} = \alpha$ otherwise.

□

Note that it follows from this proof that Lemma 4.5 also holds for $\mathcal{B}_\alpha$, with $x \leq 1/2$ replaced by $val_\mathcal{A} \leq \delta$; i.e. the following corollary holds

Corollary 4.6.1. When $val_\mathcal{A} \leq \delta$ (or equivalently when $val_{\mathcal{B}_\alpha} = \alpha$), any input word that causes the upper branch of $\mathcal{B}_\alpha$ to accept with probability $1-\epsilon$ must cause the lower branch to accept with probability $\leq \epsilon$; i.e., if $val_{up}(\mathbf{w}) = 1 - \epsilon$ then $val_{low}(\mathbf{w}) \leq \epsilon$.

Now we are finally prove the undecidability in the value of $\mathcal{B}_\alpha$:

Lemma 4.7. It is undecidable whether the PFA $\mathcal{B}_\alpha$ has value 1 or α.

Proof. From Lemma 4.6, $\mathcal{B}_\alpha$ has value 1 when $val_\mathcal{A} > \delta$ and value α when $val_\mathcal{A} \leq \delta$. Since it is undecidable whether $val_\mathcal{A} > \delta$ or $val_\mathcal{A} \leq \delta$ (Lemma 4.1), it is undecidable whether $val_{\mathcal{B}_\alpha} = 1$ or $val_{\mathcal{B}_\alpha} = \alpha$. □

4.3 The PFA family $\mathcal{B}'_\alpha$ (like $\mathcal{B}_\alpha$ but smaller)

One of the main contributions of this work is to present a new PFA family whose value is undecidable but with fewer states than $\mathcal{B}_\alpha$ from [7]. Whereas the smallest PFA $\mathcal{B}_\alpha$ has 46 states, the smallest PFA $\mathcal{B}'_\alpha$ has 30. We achieve this reduction by modifying the construction of $\mathcal{B}_\alpha$ so that, instead of using two $\mathcal{A}$ automata in parallel, the new PFA construction $\mathcal{B}'_\alpha$ (shown in Fig. 2) uses only one $\mathcal{A}$ automaton sequentially. Recall from Section 4.2.1 that $\mathcal{A}$ is the source of undecidability in the value of $\mathcal{B}_\alpha$. The PFA $\mathcal{B}'_\alpha$ is designed to emulate $\mathcal{B}_\alpha$, in the sense that when $val_\mathcal{A} > \delta$ the automaton $\mathcal{B}'_\alpha$ has value 1, and when $val_\mathcal{A} \leq \delta$ it has value α.

The new PFA family $\mathcal{B}'_\alpha$ allows us to update Lemma 4.7 into the following:

Lemma 4.8. It is undecidable whether the PFA $\mathcal{B}'_\alpha$ has value 1 or α.

Proof. Let us begin with Lemma 4.6. Lemma 4.6 says that the automaton $\mathcal{B}_\alpha$ (see Fig. 1(b)) has value 1 when $val_\mathcal{A} > \delta$, and has value α when $val_\mathcal{A} \leq \delta$, where δ is some rational number. We would like to show that this lemma holds also for the automaton $\mathcal{B}'_\alpha$ (Fig. 2). First, it is easy to verify by looking at Fig. 1(b) and Fig. 2 that words of the form

$$\mathbf{w}_k = (a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_k} \tag{16}$$

cause $\mathcal{B}'_\alpha$ to transition to the same states (with the same transition probabilities) as $\mathcal{B}_\alpha$ when the latter reads $(a\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c)^{n_k}$. More precisely, if we examine what happens to the automaton $\mathcal{B}_\alpha$ and the automaton $\mathcal{B}'_\alpha$ when the former reads $(a\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c)^{n_k}$ and the latter reads $(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_k}$, we will find that both automata have identical probability distributions over the states q_1, q_2, q_3, q_4, q_5, and $Sink$. This observation implies that, since a word of the form $(a\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c)^{n_k}$ causes $\mathcal{B}_\alpha$ to have value 1 when $val_\mathcal{A} > \delta$ (Lemma 4.6), a word of the form $(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_k}$ causes $\mathcal{B}'_\alpha$ to have value 1 when $val_\mathcal{A} > \delta$. Thus, the first half of Lemma 4.6 holds for $\mathcal{B}'_\alpha$. Let us see how the second half also holds.

When $val_\mathcal{A} \leq \delta$, we can easily verify by looking at Fig. 2 that the word $'b'$ causes $\mathcal{B}'_\alpha$ to accept with probability α. We need to check that this is the supremum of achievable probabilities when $val_\mathcal{A} \leq \delta$. From the same observation used above, since (by Lemma 4.6) any word of the form $(a\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c)^{n_k}$ cannot cause $\mathcal{B}_\alpha$ to accept with probability greater than α, any word of the form $(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_k}$ cannot cause $\mathcal{B}'_\alpha$ to accept with probability greater than α. Finally, lemma 4.9 below shows that no other word can perform better than words of the form $(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_k}$. Thus, Lemma 4.6 also holds for $\mathcal{B}'_\alpha$.

Since it is undecidable whether $val_\mathcal{A} > \delta$ or $val_\mathcal{A} \leq \delta$, it is undecidable whether $val_{\mathcal{B}'_\alpha} = 1$ or $val_{\mathcal{B}'_\alpha} = \alpha$.

□

Lemma 4.9 (helper lemma to Lemma 4.8)**.** Any word not of the form $(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_2}b\ldots b(a\mathbf{w}_\mathcal{A}c^*\mathbf{w}_\mathcal{A}c)^{n_k}$ causes $\mathcal{B}'_\alpha$ to accept with probability at most α.

Proof. Note that any word that causes either branch of $\mathcal{B}'_\alpha$ to end up in the sink with certainty leads to an acceptance probability of at most α. The question now is, which words do not cause $\mathcal{B}'_\alpha$ to end up in the sink with certainty? We will use a set of rules based on the automaton in Fig. 3 to eliminate all words that cause $\mathcal{B}'_\alpha$ to end up in the sink with certainty, eventually showing that the only words that remain are words of the form $(ac^*c)^{n_2}b\ldots b(ac^*c)^{n_k}$, where in between the symbols we may insert $\mathbf{w}_\mathcal{A}$ – recall that $\mathbf{w}_\mathcal{A} \in \{0,1\}^*$.

Video

The rules, which can be straightforwardly deduced (and verified) by looking at Fig. 3, are the following:

1. any word that contains "aa", "bb", "cc", "c^*c^*", "ab", "ac", "bc^*", "c^*a", "c^*b" causes $\mathcal{B}'_\alpha$ to end up in the Sink with certainty.
2. If the automaton reads "bc" it will always end up in the same state as it had read a "b"; i.e., a "c" does nothing when it comes after a "b". Similarly, if the automaton reads "cc^*c" it will always end up in the same state as it had read a "c".
3. any word that does not take the automaton to Sink with certainty must begin with an "a".

To verify the first rule, it is sufficient to check that it holds for any state the automaton can be in. This can be done using Fig. 3. For example, to verify that any word containing "aa" takes either the upper or lower branch to Sink with certainty, we start from the state q_1/q_4 in Fig. 3 and check that it goes to Sink when "aa" is read; then we do the same check but starting from the state $c_1/\mathcal{A}$; then starting from all the other states. That is it. To verify that words containing "bb" takes either the upper or lower branch to Sink with certainty we go through the same procedure. And similarly for the rest of the sequences listed in the first rule above. These checks do not take long to perform, and are left as an exercise for the reader. Thus, the first rule is verified.

To verify the second rule we similarly start from every state in Fig. 3 and check that entering "bc" has the same effect as just entering a "b"; and similarly for "cc^*c" and "c".

Verifying the third rule is easy. The automaton begins in the state q_1/q_4, as indicated by the free arrow near the top of Fig. 3. From this starting position, we can easily see that the automaton can proceed to the next state (that is not a Sink) if and only if it reads an "a".

With the rules verified, let us now try to construct a word that does not cause $\mathcal{B}'_\alpha$ to end up in the sink with certainty. From the third rule, we know that we have to begin with an "a". Then, from the first rule we can see that to avoid falling in the Sink "a" must be followed by nothing but a "c^*", and "c^*" must be followed by nothing but a "c". Therefore, we have that our word can only begin with "ac^*c". Now, from the first rule above, "c" cannot be followed by another "c", therefore we are left with three possibilities: 1) "$ac^*c\ a$", 2) "$ac^*c\ b$", and 3) "$ac^*c\ c^*$". The third possibility can be discarded: Since "c^*" can only be followed by a "c", "$ac^*c\ c^*$" can only continue on to be "$ac^*c\ c^*c$", which by the second rule above is equivalent to "ac^*c", taking us back to where we were. This leaves us with the two possibilities 1) "$ac^*c\ a$" and 2) "ac^*cb". Let us think about the first possibility. Like before, we must follow the "a" by "c^*c". In fact, any "a" must be followed by "c^*c" from now on, and so "ac^*c" can be seen as a block of symbols that must occur together. Moreover, a "ac^*c" block can be followed by another "ac^*c" block, so we can have "$\underbrace{ac^*c}\underbrace{ac^*c}\underbrace{ac^*c}\underbrace{ac^*c}\ldots$". Now let us think about the second possibility. By the first rule, "b" must be followed by an "a". Since "ac^*c" must come together as a block, the second possibility continues on to be "$\underbrace{ac^*c}\ \overbrace{b}\ \underbrace{ac^*c}$". So we can see that "$b$" acts as a separator between blocks of "ac^*c". To conclude the above discussion, we have that after the initial "ac^*c", we can build our word only using "ac^*c" blocks and the separator "b". The only words that we can build using these two are words of the form "$(ac^*c)^{n_2}b\ldots b(ac^*c)^{n_k}$".

Thus, we can see that any word that does not cause $\mathcal{B}'_\alpha$ to end up in Sink with certainty (and hence have its acceptance probability limited to α) must be of the form "$(ac^*c)^{n_2}b\ldots b(ac^*c)^{n_k}$". Plugging back the $\mathbf{w}_\mathcal{A}$ words we get words of the form shown in (16). □

The new $\mathcal{B}'_\alpha$ construction also allows us to update Corollary 4.6.1 to the following:

Corollary 4.9.1. When $val_\mathcal{A} \leq \delta$ (or equivalently when $val_{\mathcal{B}'_\alpha} = \alpha$), any input word that causes the upper branch of $\mathcal{B}'_\alpha$ to accept with probability $1-\epsilon$ must cause the lower branch to accept with probability $\leq \epsilon$; i.e., if $val_{up}(\mathbf{w}) = 1-\epsilon$ then $val_{low}(\mathbf{w}) \leq \epsilon$.

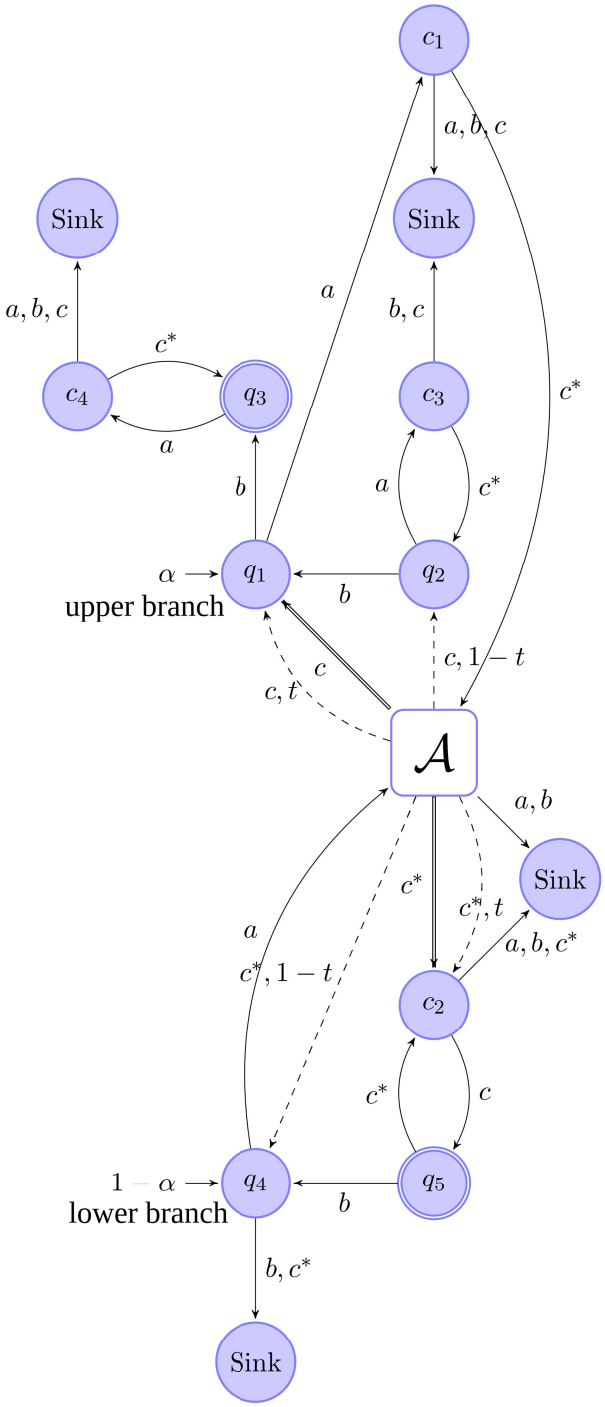
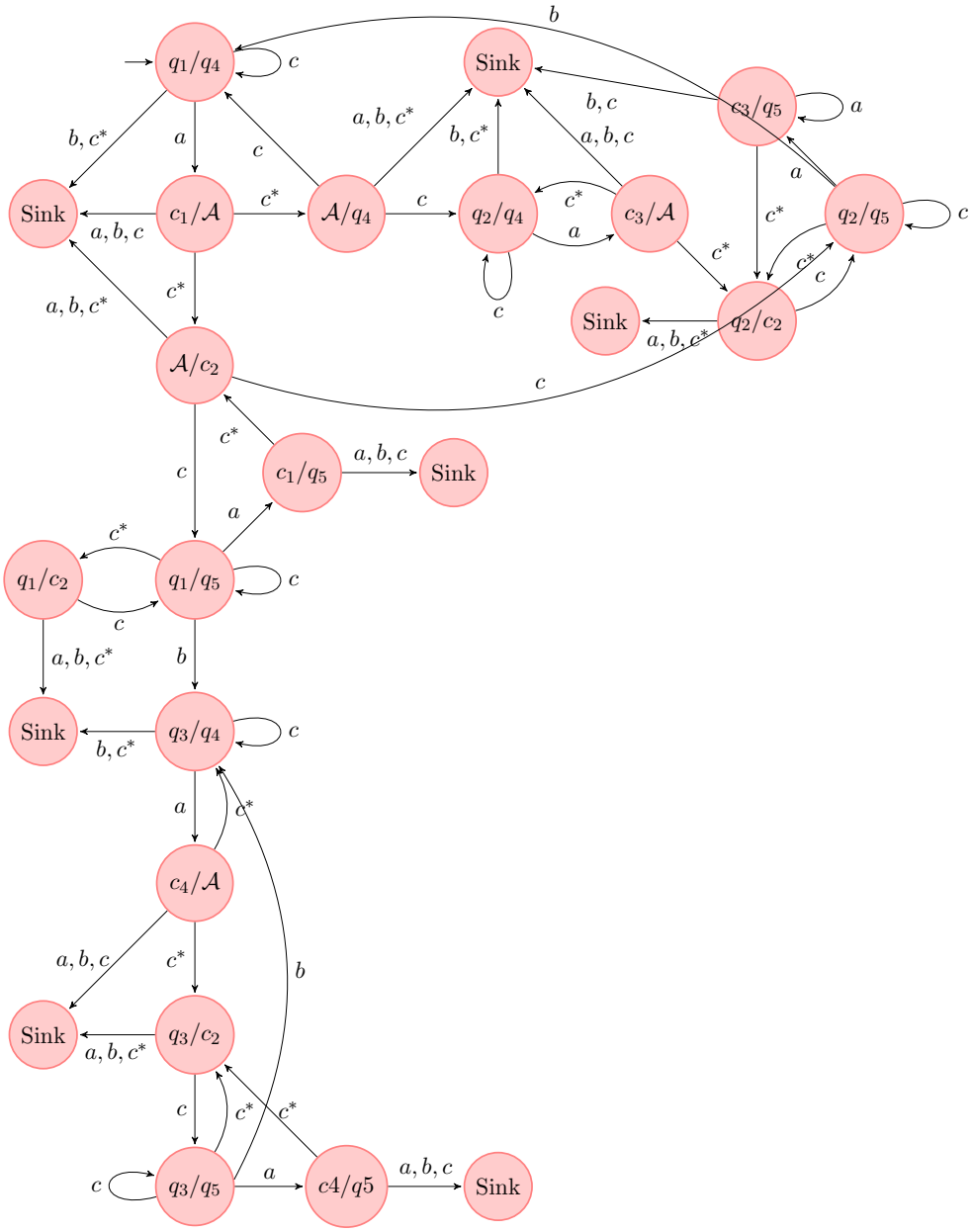


Figure 2: The PFA construction $\mathcal{B}'_\alpha$ shown in this figure is designed to emulate $\mathcal{B}_\alpha$ in Fig. 1(b) for input words with a specific form. It requires only one $\mathcal{A}$ and so has fewer states than $\mathcal{B}_\alpha$. More precisely, it has $10 + n_\mathcal{A}$, where $n_\mathcal{A}$ is the number of states of $\mathcal{A}$ and can take values $20, 25, 30, \ldots$. Note that all *Sink* states are one state. $\mathcal{B}'_\alpha$ has input alphabet $\{0, 1, a, b, c, c^*, rt\}$: 7 symbols. The reset symbol "rt" takes the automaton to the state q_1 if it started from q_1 (upper branch) in the first use and takes it to the state q_4 if it started from q_4 (lower branch).



Video

Figure 3: The automaton shown in this figure represents the behavior of the automaton $\mathcal{B}'_\alpha$ in Fig. 2: each state of this automaton contains the state of the PFA $\mathcal{B}'_\alpha$ if it started in the upper branch is written on the left and if it started in the lower branch (written on the right). For example, the situation in which a word causes $\mathcal{B}'_\alpha$ to transition from the state q_1 to the state c_1 if it started in the upper branch and to transition from q_4 to q_5 if it started in the lower branch corresponds to this automaton starting from the state q_1/q_4 then transitioning after each symbol so that it ends up in the state c_1/q_5. The Sink nodes on the graph correspond to either of the upper or lower branches of $\mathcal{B}'_\alpha$ ending up in the state Sink.

5 Our channel construction

Now that we have a family of PFAs $\mathcal{B}'_\alpha$ with an undecidable property let us construct channels on top of them. We will see in the next sections how this channel construction causes the undecidability in $\mathcal{B}'_\alpha$ to be inherited by the information-theoretic properties of the channel.

5.1 The construction

Recall from our definition of FSMC in Section 2.1 that we use $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}$ to denote finite sets that represent the FSMC's input alphabet, output alphabet, and the set of states, respectively. Recall also that an FSMC is fully characterized by a set of conditional probabilities $\{p(y, s|x, s')\}_{s,y,x,s'}$, where each $p(y, s|x, s')$ is the probability that the FSMC is going to output y and transition to the state s given that the input was x and it was in the state s'; where $s, s' \in \mathcal{S}$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$. Moreover, to eliminate any ambiguity regarding whether our uncomputability result is due to problems with approximating $p(y, s|x, s')$, in this paper we only consider (w.l.o.g) FSMCs for which the probabilities $p(y, s|x, s')$ are rational. For simplicity, we also only consider (w.l.o.g) FSMCs for which $p(y, s|x, s')$ factorizes to $p(y|x, s')p(s|x, s')$; that is, FSMCs for which $\{p(y, s|x, s')\}_{s,y,x,s'} = \{p(y|x, s')p(s|x, s')\}_{s,y,x,s'}$.

All channels in this paper will be built on top of $\mathcal{B}'_\alpha$, and these channels have the input alphabet $\mathcal{X} = \{0, 1, a, b, c, c^*, rt\}$ and output alphabet $\mathcal{Y} = \{0, 1\}$

For all FSMCs in this paper, $p(y|x, s')$ can be expressed as follows:

$$p(y|x, s') = \begin{cases} \delta_{yx} & \text{if } s' \text{ is accepting and } x \in \{0, 1\} \\ 1/2 & \text{otherwise,} \end{cases} \tag{17}$$

where δ_{yx} is 1 when $x = y$ and 0 otherwise. After the channel produces an output, the underlying PFA transitions to a new state with probability $p(s|x, s')$, the value of which is shown on the figures; for example, from Fig. 2 we can see that $p(c_3|'a', q_2) = 1$.

5.2 Upper channel and lower channel

For channels in this paper, it will be useful to think of the channel as a mixture of an "upper channel", the one based on the upper branch of the PFA $\mathcal{B}'_\alpha$ shown in Fig. 2, and a "lower channel", the one based on the lower branch of $\mathcal{B}'_\alpha$. Mathematically, we can express the channel as $W^n(y^n|x^n) = \alpha\, W^n_{up}(y^n|x^n) + (1-\alpha)\, W^n_{low}(y^n|x^n)$, where $W^n(y^n|x^n) = W_1(y_1|x_1)W_2(y_2|x_2)\cdots W_n(y_n|x_n)$ is the probability that the channel is going to produce the output sequence $(y_1, y_2, \ldots, y_n)$ when it gets as input the sequence $(x_1, x_2, \ldots, x_n)$, and $W_i(y_i|x_i) \in \{p(y, s|x, s')\}_{s,y,x,s'}$. In other words, the channel can be thought of as a switched channel: After the first use, the channel either switches to being the upper channel (in which case the underlying PFA can only be in one of the states of the upper branch) with probability α, or switches to being the lower channel (in which case the underlying PFA can only be in one of the states of the lower branch) with probability $1-\alpha$. And then the switch position does not change for the duration of the transmission. In the following sections, we will employ this interpretation of the channel as a switched channel via the following lemmas.

Lemma 5.1. Given the mixed channel $W^n(y^n|x^n) = \alpha_1 W^n_{up}(y^n|x^n) + \alpha_2 W^n_{low}(y^n|x^n)$, where $\alpha_1 + \alpha_2 = 1$, it holds that

$$I(\mathbf{X^n}; \mathbf{Y^n}) \geq \alpha\, I(\mathbf{X^n}; \mathbf{Y^n_{up}}) + (1-\alpha)\, I(\mathbf{X^n}; \mathbf{Y^n_{low}}) - 1. \tag{18}$$

Proof. Let Z be a random variable that takes value 1 when the PFA goes to the upper branch and value 0 when it goes to the lower branch, and so $p(Z = 1) = \alpha$ and $p(Z = 0) = 1 - \alpha$. Note that Z is independent of the input $\mathbf{X^n}$, so $I(\mathbf{X^n}; Z) = 0$. We can write

$$\begin{aligned} I(\mathbf{X^n}; \mathbf{Y^n}) &= I(\mathbf{X^n}; \mathbf{Y^n}|Z) + I(\mathbf{X^n}; Z) - I(\mathbf{X^n}; Z|\mathbf{Y^n}) \\ &\geq I(\mathbf{X^n}; \mathbf{Y^n}|Z) - 1 \\ &= \alpha\, I(\mathbf{X^n}; \mathbf{Y^n_{up}}) + (1-\alpha)\, I(\mathbf{X^n}; \mathbf{Y^n_{low}}) - 1, \end{aligned}$$

where the inequality follows because $I(\mathbf{X^n}; Z|\mathbf{Y^n}) \leq 1$. □

Lemma 5.2. Given the mixed channel $W^n(y^n|x^n) = \alpha_1 W^n_{up}(y^n|x^n)+\alpha_2 W^n_{low}(y^n|x^n)$, where $\alpha_1+\alpha_2 = 1$, it holds that

$$I(\mathbf{X^n};\mathbf{Y^n}) \leq \alpha I(\mathbf{X^n};\mathbf{Y^n_{up}}) + (1-\alpha) I(\mathbf{X^n};\mathbf{Y^n_{low}}) \tag{19}$$

Proof. Let Z be a random variable that takes value 1 when the PFA goes to the upper branch and value 0 when it goes to the lower branch, and so $p(Z=1)=\alpha$ and $p(Z=0)=1-\alpha$. Note that Z is independent of the input $\mathbf{X^n}$, so $I(\mathbf{X^n};Z)=0$. We can write

$$\begin{aligned} I(\mathbf{X^n};\mathbf{Y^n}) &= I(\mathbf{X^n};\mathbf{Y^n}|Z) + I(\mathbf{X^n};Z) - I(\mathbf{X^n};Z|\mathbf{Y^n}) \\ &\leq I(\mathbf{X^n};\mathbf{Y^n}|Z) \\ &= \alpha I(\mathbf{X^n};\mathbf{Y^n_{up}}) + (1-\alpha) I(\mathbf{X^n};\mathbf{Y^n_{low}}). \end{aligned}$$

□

The two lemmas above can be seen as special cases of the following more general lemma.

Lemma 5.3 (adaptation of Lemma 1.4.2 and Remark 1.4.1 in [11])**.** Given the mixed channel $W^n(y^n|x^n) = \alpha_1 W^n_{up}(y^n|x^n) + \alpha_2 W^n_{low}(y^n|x^n)$, where $\alpha_1+\alpha_2=1$, and an arbitrary sequence of functions $f^n(\cdot)$, the following holds for any λ:

$$\begin{aligned} &\alpha_1 \Pr\left[f^n\left(\frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n_{up}}}(x^n;y^n_{up}) - \beta_2 + \beta_1\right) > \lambda\right] + \alpha_2 \Pr\left[f^n\left(\frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n_{low}}}(x^n;y^n_{low}) - \beta_3 + \beta_1\right) > \lambda\right] \\ \geq &\Pr\left[f^n\left(i_{\mathbf{X^n},\mathbf{W^n}}(x^n;y^n)\right) > \lambda\right] \geq \\ &\alpha_1(1-e^{-n\gamma_n}) \Pr\left[f^n\left(\frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n_{up}}}(x^n;y^n_{up}) - \beta_2 + \beta_1\right) > \lambda\right] \\ &+ \alpha_2(1-e^{-n\gamma_n}) \Pr\left[f^n\left(\frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n_{low}}}(x^n;y^n_{low}) - \beta_3 + \beta_1\right) > \lambda\right], \end{aligned} \tag{20}$$

where $\{\gamma^n\}$ is an arbitrary sequence such that $\gamma_1 > \gamma_2 > \ldots > \gamma_n > 0$, $\gamma_n \to 0$ and $n\gamma_n \to \infty$, $0 \leq \beta_1 \leq c_0/n$, $c_0 \equiv -\log\min(\alpha_1,\alpha_2)$, and $0 \leq \beta_2, \beta_3 \leq \gamma_n$.

Proof. We use the following simple inequality from [12]:

$$\frac{1}{n}\log\min(\alpha_1,\alpha_2) + \max(u,v) \leq \frac{1}{n}\log[\alpha_1 e^{nu} + \alpha_2 e^{nv}] \leq \max(u,v). \tag{21}$$

By setting $u = \frac{1}{n}\log W^n_{up}(y^n)$ and $v = \frac{1}{n}\log W^n_{low}(y^n)$, where $W^n(y^n) = \sum_{x^n} p_{\mathbf{X^n}}(x^n) W^n(y^n|x^n)$, we have

$$-\frac{c_0}{n} + \Delta_n(y^n) \leq \frac{1}{n}\log W^n(y^n) \leq \Delta_n(y^n), \tag{22}$$

where $c_0 \equiv -\log\min(\alpha_1,\alpha_2)$ and $\Delta_n(y) \equiv \max\left(\frac{1}{n}\log W^n_{up}(y), \frac{1}{n}\log W^n_{low}(y)\right)$.

Now, given the sequence $\{\gamma^n\}$, it holds that

$$\begin{aligned} &\Pr\left[\frac{1}{n}\log W^n_{up}(y^n_{up}) - \frac{1}{n}\log W^n_{low}(y^n_{up}) \leq -\gamma_n\right] \\ &= \sum_{y\in B_n} W^n_{up}(y) \\ &\leq \sum_{y\in B_n} W^n_{low}(y) e^{-n\gamma_n} \\ &\leq e^{-n\gamma_n} \to 0 \,(\text{as } n\to\infty), \end{aligned}$$

where $B_n \equiv \{y \in \mathcal{Y}^n | \frac{1}{n}\log W^n_{up}(y) - \frac{1}{n}\log W^n_{low}(y) \leq -\gamma_n\}$. This implies that with probability $\geq 1-e^{-n\gamma_n}$

$$\frac{1}{n}\log W^n_{up}(y^n_{up}) \geq \frac{1}{n}\log W^n_{low}(y^n_{up}) - \gamma_n.$$

Therefore, with probability $\geq 1 - e^{-n\gamma_n}$, we have

$$\frac{1}{n}\log W^n_{up}(y^n_{up}) \leq \max\left(\frac{1}{n}\log W^n_{up}(y^n_{up}), \frac{1}{n}\log W^n_{low}(y^n_{up})\right) \leq \frac{1}{n}\log W^n_{up}(y^n_{up}) + \gamma_n,$$

i.e.,

$$\frac{1}{n}\log W^n_{up}(y^n_{up}) \leq \Delta_n(y^n_{up}) \leq \frac{1}{n}\log W^n_{up}(y^n_{up}) + \gamma_n. \tag{23}$$

Similarly, we have that with probability $\geq 1 - e^{-n\gamma_n}$

$$\frac{1}{n}\log W^n_{low}(y^n_{low}) \leq \Delta_n(y^n_{low}) \leq \frac{1}{n}\log W^n_{low}(y^n_{low}) + \gamma_n. \tag{24}$$

Equation (22) implies that

$$\frac{1}{n}\log W^n(y^n) = \Delta_n(y^n) - \beta_1, \tag{25}$$

where $0 \leq \beta_1 \leq \frac{c_0}{n}$. Equations (23) and (24) respectively imply that with probability $\geq 1 - e^{-n\gamma_n}$

$$\frac{1}{n}\log W^n_{up}(y^n_{up}) = \Delta_n(y^n_{up}) - \beta_2 \tag{26}$$

and that with probability $\geq 1 - e^{-n\gamma_n}$

$$\frac{1}{n}\log W^n_{low}(y^n_{low}) = \Delta_n(y^n_{low}) - \beta_3, \tag{27}$$

where $0 \leq \beta_2, \beta_3 \leq \gamma_n$. Note that β_1, β_2, β_3 are all random variables.
We can now finally write the following chain of inequalities. Let $L(y) \equiv \frac{1}{n}\log\frac{p(x^n,y^n)}{p(x^n)}$:

$$\begin{aligned}
&\Pr\left[f^n\left(\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n}}(x^n;y^n)\right) > \lambda\right] = \Pr\left[f^n\left(L(y^n) - \frac{1}{n}\log W^n(y^n)\right) > \lambda\right] \\
&= \Pr\left[f^n\left(L(y^n) - \Delta_n(y^n) + \beta_1\right) > \lambda\right] \\
&= \alpha_1 \Pr\left[f^n\left(L(y^n_{up}) - \Delta_n(y^n_{up}) + \beta_1\right) > \lambda\right] + \alpha_2 \Pr\left[f^n\left(L(y^n_{low}) - \Delta_n(y^n_{low}) + \beta_1\right) > \lambda\right] \\
&\geq \alpha_1(1 - e^{-n\gamma_n})\Pr\left[f^n\left(L(y^n_{up}) - \frac{1}{n}\log W^n_{up}(y^n_{up}) + \beta_1 - \beta_2\right) > \lambda\right] \\
&+ \alpha_2(1 - e^{-n\gamma_n})\Pr\left[f^n\left(L(y^n_{low}) - \frac{1}{n}\log W^n_{low}(y^n_{low}) + \beta_1 - \beta_3\right) > \lambda\right] \\
&= \alpha_1(1 - e^{-n\gamma_n})\Pr\left[f^n\left(\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n}_{up}}(x^n;y^n_{up}) + \beta_1 - \beta_2\right) > \lambda\right] \\
&+ \alpha_2(1 - e^{-n\gamma_n})\Pr\left[f^n\left(\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{low}}}(x^n;y^n_{low}) + \beta_1 - \beta_3\right) > \lambda\right]
\end{aligned}$$

where the third equality follows from Lemma 5.4. This establishes the lower bound in (20). Noting that, trivially, (26) and (27) also hold with probability ≤ 1, so we can also write

$$
\begin{aligned}
&\Pr\left[f^n\left(\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n}}(x^n;y^n)\right)>\lambda\right]=\Pr\left[f^n\left(L(y^n)-\frac{1}{n}\log W^n(y^n)\right)>\lambda\right]\\
&=\Pr\left[f^n\left(L(y^n)-\Delta_n(y^n)+\beta_1\right)>\lambda\right]\\
&=\alpha_1\Pr\left[f^n\left(L(y^n_{up})-\Delta_n(y^n_{up})+\beta_1\right)>\lambda\right]+\alpha_2\Pr\left[f^n\left(L(y^n_{low})-\Delta_n(y^n_{low})+\beta_1\right)>\lambda\right]\\
&\leq\alpha_1\Pr\left[f^n\left(L(y^n_{up})-\frac{1}{n}\log W^n_{up}(y^n_{up})+\beta_1-\beta_2\right)>\lambda\right]\\
&+\alpha_2\Pr\left[f^n\left(L(y^n_{low})-\frac{1}{n}\log W^n_{low}(y^n_{low})+\beta_1-\beta_3\right)>\lambda\right]\\
&=\alpha_1\Pr\left[f^n\left(\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{up}}}(x^n;y^n_{up})+\beta_1-\beta_2\right)>\lambda\right]\\
&+\alpha_2\Pr\left[f^n\left(\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{low}}}(x^n;y^n_{low})+\beta_1-\beta_3\right)>\lambda\right],
\end{aligned}
$$

which establishes the upper bound in Eq. (20). □

Lemma 5.4 (helper lemma to Lemma 5.3. [17])**.** Let $\left(Z_n^{(1)}\right)_{n=1}^{\infty}$ and $\left(Z_n^{(2)}\right)_{n=1}^{\infty}$ be two sequences of random variables taking values in $\mathcal{Z}_n$, and define $(Z_n)_{n=1}^{\infty}$ by

$$\mathrm{P}_{Z_n}(z)=\alpha_1\mathrm{P}_{Z_n^{(1)}}(z)+\alpha_2\mathrm{P}_{Z_n^{(2)}}(z)$$

for all $z\in\mathcal{Z}_n$. Then for an arbitrary sequence of functions $(f_n)_{n=1}^{\infty}$ defined over $\mathcal{Z}_n$, we have that for any λ

$$\Pr\left[f_n(Z_n)>\lambda\right]=\alpha_1\Pr\left[f_n(Z_n^{(1)})>\lambda\right]+\alpha_2\Pr\left[f_n(Z_n^{(2)})>\lambda\right].$$

6 Relating mutual information of the channel to the value of the underlying PFA

Here is an intuitive argument for why the way we have defined the channels in Eq. (17) might allow us to relate the information transmission capabilities of the channel to the value of the underlying automaton: Eq. (17) just means that if the underlying automaton is in an accepting state and the input is 0 or 1 the channel is going to act like an ideal channel; and if the underlying automaton is in a non-accepting state the channel is going to be completely noisy. Suppose for the sake of argument that the value of the underlying automaton is $val_{\mathcal{B}'_\alpha} = 0$. This implies that the automaton can never be in an accepting state, and hence the channel is always going to be completely noisy. Since you cannot transmit any information over a channel that is always completely noisy, the mutual information between the input and the output is always going to be $I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_\alpha} = 0$. On the other hand, if the underlying automaton had value $val_{\mathcal{B}'_\alpha} = 1$ then the automaton is always in an accepting state, and the channel is always ideal. This corresponds to $I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_\alpha} = 0$. In this section we are going to show that this intuition holds not just for $val_{\mathcal{B}'_\alpha} = 0$ and $val_{\mathcal{B}'_\alpha} = 1$ but for all intermediate values as well.

More precisely, the main aim of this section is to prove that the supremum of all values of mutual information that the channel can create between the input and the output is equal to $val_{\mathcal{B}'_\alpha}$;i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_\alpha}.$$

To prove this relation we need to prove two things: 1) we need to prove that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \leq val_{\mathcal{B}'_\alpha}$ and 2) we need to prove that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \geq val_{\mathcal{B}'_\alpha}$.

Let us begin with the latter. There are only two possible values for $val_{\mathcal{B}'_\alpha}$: 1 and α. If we can prove that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \geq val_{\mathcal{B}'_\alpha}$ for both of these cases, we would be done.

Lemma 6.1. For channels built on top of $\mathcal{B}'_\alpha$: when $val_{\mathcal{B}'_\alpha} = 1$, for any $\epsilon > 0$ there exists an arbitrarily long input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{up}) \geq 1 - \epsilon$ and $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{low}) \geq 1 - \epsilon$.

Proof. We have from the proof of Lemma 4.8 that for any $\eta > 0$ there exists a word $\mathbf{w}$ with length m (i.e., $\mathbf{w} = w_1 \dots w_m$) such that $val_{up}(\mathbf{w}) \geq 1 - \eta$ and $val_{low}(\mathbf{w}) \geq 1 - \eta$. Consider the sequence

$$\mathbf{X} = (\mathbf{w}, \mathbf{data}, rt), \tag{28}$$

where the data symbols $\mathbf{data} \in \{0,1\}^k$ are chosen uniformly at random. (From this point on, when a word $\mathbf{w}$ is part of a sequence that gets input to a channel, it is to be understood that what is actually input to the channel is a random variable that takes value $\mathbf{w}$ with certainty. This is because, formally, an information channel takes random variables as input). If we enter the sequence (28) into the channel input the following chain of inequalities holds simultaneously for the upper and lower channels:

$$\begin{aligned} H(\mathbf{Y_{up(low)}}|\mathbf{X}) &= \sum_{i=1}^{m+k+1} H(Y_{up(low)_i}|\mathbf{Y_{up(low)}}_{[1,i-1]}\mathbf{X}) \\ &= m + 1 + H(\mathbf{Y_{up(low)}}_{[m+1,m+k]}|\mathbf{Y_{up(low)}}_{[1,m]}\mathbf{X}_{[1,m+k]}) \\ &\leq m + 1 + H(\mathbf{Y_{up(low)}}_{[m+1,m+k]}|\mathbf{X}_{[1,m+k]}) \\ &\leq m + 2 + k\eta, \end{aligned}$$

where the first equality follows from the chain rule, the second equality from the fact that the channel output is uniformly random and independent of the input during the first m uses and the last use (see Section 5), the first inequality by removing the conditioning on $\mathbf{Y_{up(low)}}_{[1,m]}$, and the final inequality by bounding the entropy by the following bound:

after the first m uses, the channel behaves like a noiseless channel with probability at least $1 - \eta$ and like a completely noisy channel with the complementary probability, so we can bound the conditional entropy of the output of the uses $m + 1$ to $m + k$ as follows:

$$\begin{aligned} H(\mathbf{Y_{up(low)}}_{[m+1,m+k]}|\mathbf{X}_{[1,m+k]}) &\leq H((1-\eta+\eta 2^{-k})\phi + \eta(1-2^{-k})\rho) \\ &= h(1-\eta+\eta 2^{-k}) + \eta(1-2^{-k})\log(2^k-1) \\ &\leq 1+k\eta, \end{aligned}$$

where $\phi = (1, 0, \cdots, 0)$ is a vector with length 2^k, $\rho = (0, \frac{1}{2^k-1}, \ldots, \frac{1}{2^k-1})$ is a vector of length 2^k-1, and $h(p) \equiv -p\log(p) - (1-p)\log(1-p)$ is the binary entropy function. Using the above and the fact that $H(\mathbf{Y_{up(low)}}) = m+k+1$ because the output of the channel is completely uncertain if one does not know the input, we can now write

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y_{up(low)}}) &= H(\mathbf{Y_{up(low)}}) - H(\mathbf{Y_{up(low)}}|\mathbf{X}) \\ &= m+k+1 - H(\mathbf{Y_{up(low)}}|\mathbf{X}) \\ &\geq k(1-\eta) - 1. \end{aligned}$$

For any given $\eta > 0$, by choosing any k larger than $(1+(1-2\eta)(m+1))/\eta$ we get

$$\frac{1}{m+k+1} I(\mathbf{X}; \mathbf{Y_{up(low)}}) \geq 1-2\eta,$$

which implies that, for any given $\epsilon > 0$, there always exists an arbitrarily long input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{up(low)}) \geq 1-\epsilon$. □

Knowing how the upper and lower branches of the channel behave, we can use the lemmas we developed in Section 5.2 to say the following about the whole channel.

Lemma 6.2. When $val_{\mathcal{B}'_\alpha} = 1$, it holds that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \geq val_{\mathcal{B}'_\alpha}$.

Proof. When $val_{\mathcal{B}'_\alpha} = 1$, it follows from Lemma 6.1 that, for any $\epsilon > 0$, there exists an arbitrarily long (arbitrarily large n) input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{up}) > 1-\epsilon$ and $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{low}) > 1-\epsilon$. Inputting that sequence into the channel, Eq. (18) gives us $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}) \geq 1-\epsilon-\frac{1}{n}$. Since $1/n$ can be made arbitrarily small, we have that there is always an input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}) \geq 1-\eta = val_{\mathcal{B}'_\alpha} - \eta$ for any $\eta > 0$. □

The proof for the $val_{\mathcal{B}'_\alpha} = \alpha$ case proceeds similarly.

Lemma 6.3. For channels built on top of $\mathcal{B}_\alpha$: when $val_{\mathcal{B}'_\alpha} = \alpha$, for any $\epsilon > 0$ there exists an arbitrarily long input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up}}) \geq 1-\epsilon$ and $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{low}}) = 0$.

Proof. Recall from Section 4.3 that a word that achieves value α is $\mathbf{w} = \ 'b'$. Using this word $\mathbf{w}$ in the sequence in (28), we have from Lemma 6.5 below that $\frac{1}{m+k+1} I(\mathbf{X}; \mathbf{Y_{low}}) \leq val_{low}((b, \mathbf{data})) = val_{low}(b) = 0$. To get the $\frac{1}{m+k+1} I(\mathbf{X}; \mathbf{Y_{up}})$ term, we can write

$$\begin{aligned} H(\mathbf{Y_{up}}|\mathbf{X}) &= \sum_{i=1}^{m+k+1} H(Y_{up_i}|\mathbf{Y_{up}}_{[1,i-1]}\mathbf{X}) \\ &= m+1 + H(\mathbf{Y_{up}}_{[m+1,m+k]}|\mathbf{Y_{up}}_{[1,m]}\mathbf{X}_{[1,m+k]}) \\ &= m+1, \end{aligned}$$

where the last equality follows because after entering $'b'$ we know with certainty that the PFA is in the accepting state if started in the upper branch, and so the channel output is completely known if we are given the input. Then, since $H(\mathbf{Y_{up}}) = m+k+1$ (because without knowing the input the output induced by sequence (28) is completely uncertain), we have

$$\begin{aligned} \frac{1}{m+k+1} I(\mathbf{X}; \mathbf{Y_{up}}) &= \frac{1}{m+k+1} \left(H(\mathbf{Y_{up}}) - H(\mathbf{Y_{up}}|\mathbf{X})\right) \\ &\geq \frac{k}{m+k+1}. \end{aligned}$$

This implies that for any given $\epsilon > 0$, by choosing any k larger than $(m+1)(1-\epsilon)/\epsilon$ we get

$$\frac{1}{m+k+1} I(\mathbf{X}; \mathbf{Y_{up}}) \geq 1 - \epsilon.$$

□

Lemma 6.4. When $val_{\mathcal{B}'_\alpha} = \alpha$, it holds that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \geq val_{\mathcal{B}'_\alpha}$.

Proof. When $val_{\mathcal{B}'_\alpha} = \alpha$, it follows from Lemma 6.3 that, for any $\epsilon > 0$, there exists an arbitrarily long input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{up}) > 1 - \epsilon$ and $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}_{low}) = 0$. For this sequence Eq. (18) becomes $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}) \geq \alpha\,(1-\epsilon) - \frac{1}{n}$. Since $1/n$ can be made arbitrarily small, we have that there is always an input sequence $\mathbf{X^n}$ such that $\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}) \geq \alpha - \eta = val_{\mathcal{B}'_\alpha} - \eta$ for any $\eta > 0$. □

Now let us prove that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \leq val_{\mathcal{B}'_\alpha}$.

Lemma 6.5. Let $\mathbf{X}$ be an arbitrary sequence of the form

$$\mathbf{X} = (X_1, \ldots, X_n, rt)\,.$$

Let $\mathbf{X}$ be input to a channel built on top of some PFA $\mathcal{P}$ such that the channel acts as a noiseless channel when $\mathcal{P}$ is in an accepting state and as a completely noisy channel otherwise, and the reset symbol erases all memory of previous uses of the channel. Then we have that

$$\frac{1}{n+1} I(\mathbf{X}; \mathbf{Y}) \leq val_{\mathcal{P}}(\mathbf{X}_{[1,n]}). \tag{29}$$

Proof. Consider the following chain of inequalities:

$$\begin{aligned}
H(\mathbf{Y}|\mathbf{X}) &= \sum_{j=1}^{n+1} H(Y_j|\mathbf{Y}_{[1,j-1]}\mathbf{X}) \\
&\geq \sum_{j=1}^{n+1} H(Y_j|\mathbf{X}S_{j-1}) \\
&= \sum_{j=1}^{n+1} p(S_{j-1} \in F) H(Y_j|\mathbf{X}, S_{j-1} \in F) + p(S_{j-1} \notin F) H(Y_j|\mathbf{X}, S_{j-1} \notin F) \\
&\geq \sum_{j=1}^{n+1} p(S_{j-1} \notin F) H(Y_j|\mathbf{X}, S_{j-1} \notin F) \\
&= \sum_{j=1}^{n+1} p(S_{j-1} \notin F) \\
&\geq (n+1)(1 - val_{\mathcal{P}}(\mathbf{X}_{[1,n]})).
\end{aligned}$$

The first equality follows from the chain rule. The first inequality follows from the fact that knowing the state of the PFA tells us at least as much about what the next output could be as knowing the previous outputs. The second equality is clear. The second inequality is clear. The third equality follows from the fact that if the automaton is in a non-accepting state, the output is random regardless of the input. The final inequality follows from the fact that the probability of not being at an accepting state is always at least as big as $1 - val_{\mathcal{P}}(\mathbf{X}_{[1,n]})$.

Knowing that, trivially, $H(\mathbf{Y}) \leq n+1$, we can finally write

$$\begin{aligned}
\frac{1}{n+1} I(\mathbf{X}; \mathbf{Y}) &= \frac{1}{n+1} \left(H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X})\right) \\
&\leq \frac{1}{n+1} \left(n + 1 - (n+1)(1 - val_{\mathcal{P}}(\mathbf{X}_{[1,n]}))\right) \\
&= val_{\mathcal{P}}(\mathbf{X}_{[1,n]}).
\end{aligned}$$

□

The following corollary follows from Lemma 6.5:

Corollary 6.5.1. Let $\mathbf{X}$ be a length n sequence of the form

$$\mathbf{X} = \left(\mathbf{X^{n_1+1}}, \mathbf{X^{n_2+1}}, \ldots, \mathbf{X^{n_i+1}}, \ldots\right), \tag{30}$$

where $\mathbf{X^{n_i+1}} \equiv (X_1, \ldots, X_{n_i}, rt)$, and note that $\sum_i (n_i + 1) = n$. Let $\mathbf{X}$ be input to a channel built on top of some PFA $\mathcal{P}$ such that the channel acts as a noiseless channel when $\mathcal{P}$ is in an accepting state and as a completely noisy channel otherwise, and the reset symbol erases all memory of previous uses of the channel. Then we have that

$$\frac{1}{n} I(\mathbf{X}; \mathbf{Y}) \leq \frac{1}{n} \sum_i (n_i + 1)\, val_{\mathcal{P}}(\mathbf{X^{n_i+1}}_{[1,n_i]}). \tag{31}$$

Proof. The presence of the reset button at the end of each $\mathbf{X^{n_i+1}}$ makes the channel outputs corresponding to all $\mathbf{X^{n_i+1}}$ independent of each other, which allows us to write $I(\mathbf{X}; \mathbf{Y}) = \sum_i I(\mathbf{X^{n_i+1}}; \mathbf{Y^{n_i+1}})$. By Lemma 6.5 it then follows that $\frac{1}{n} I(\mathbf{X}; \mathbf{Y}) = \frac{1}{n} \sum_i I(\mathbf{X^{n_i+1}}; \mathbf{Y^{n_i+1}}) \leq \frac{1}{n} \sum_i (n_i+1)\, val_{\mathcal{P}}(\mathbf{X^{n_i+1}}_{[1,n_i]})$. □

Lemma 6.6. For channels built on top of $\mathcal{B}'_\alpha$, it holds that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \leq val_{\mathcal{B}'_\alpha}$.

Proof. We can write

$$\begin{aligned}
\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n}) &= \alpha\, \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up}}) + (1-\alpha)\, \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{low}}) \\
&\leq \alpha\, \frac{1}{n} \sum_i (n_i + 1)\, val_{up}(\mathbf{X^{n_i+1}}_{[1,n_i]}) + (1-\alpha)\, \frac{1}{n} \sum_i (n_i + 1)\, val_{low}(\mathbf{X^{n_i+1}}_{[1,n_i]}) \\
&= \frac{1}{n} \sum_i (n_i + 1)\, val_{\mathcal{B}'_\alpha}(\mathbf{X^{n_i+1}}_{[1,n_i]}) \\
&\leq val_{\mathcal{B}'_\alpha} \frac{1}{n} \sum_i (n_i + 1) \\
&= val_{\mathcal{B}'_\alpha},
\end{aligned} \tag{32}$$

where the first equality follows from Lemma 5.2, the first inequality follows from Corollary 6.5.1 and we used that any sequence $\mathbf{X^n}$ can be written in the form (30). □

Now the relation we set out to prove at beginning of this section follows immediately.

Lemma 6.7. For channels built on top of $\mathcal{B}'_\alpha$, it holds that

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_\alpha}.$$

Proof. Since $val_{\mathcal{B}'_\alpha} = \alpha$ and $val_{\mathcal{B}'_\alpha} = 1$ are the only two possible values for $val_{\mathcal{B}'_\alpha}$, we have from Lemmas 6.2 and 6.4 that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \geq val_{\mathcal{B}'_\alpha}$. Additionally. from Lemma 6.6 we have that $\lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) \leq val_{\mathcal{B}'_\alpha}$. This proves the lemma. □

Discussion

Lemma 6.7 constitutes the first of two steps towards relating the capacity of the channels we constructed in Section 5 to the value of the underlying PFA $\mathcal{B}'_\alpha$ via Eq. (6). The second and final step will be taken in Section 7. The ultimate goal is to prove Theorem 3.1.

7 The information-stability of our channels

Our results depend on the information-stability of the channels we have just constructed (see definition of information-stability in Section 2.2.2). The PFAs $\mathcal{B}'_\alpha$ on which the channels are built are parametrized by a number α, which is the probability that the PFA enters the upper branch in the first use (see Fig. 2). As we will see in this section, the information-stability of our channels depends on the parameter α of the underlying PFA.

The proofs in this section will require that we treat the upper and lower branches of the PFA as supporting channels of their own, i.e., and upper channel and a lower channel, and prove the results for these channels individually. We introduced Lemma 5.3 in Section 5.2 in preparation for this.

7.1 When $val_{\mathcal{B}'_\alpha} = 1$, the channels are information-stable

Lemma 7.1. FSMCs based on PFAs $\mathcal{B}'_\alpha$ are information-stable when $val_{\mathcal{B}'_\alpha} = 1$.

Proof. From Lemma 6.1 we have that when $val_{\mathcal{B}'_\alpha} = 1$ there always exists an input sequence $\mathbf{X^n}$ such that

$$\frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up(low)}}) \geq 1 - \epsilon = val_{\mathcal{B}'_\alpha} - \epsilon \tag{33}$$

for any $\epsilon > 0$. Assuming without loss of generality that $n_{t+1} \geq n_t$, this implies that there exists an input sequence $\mathbf{X^t} = (X_1, \ldots, X_{n_t - 1}, rt)$ such that $\frac{I(\mathbf{X^t}; \mathbf{Y^t_{up(low)}})}{n_t} \geq val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^t}$ for any $\mu > 0$. Suppose we concatenate these sequences to get the following sequence:

$$\mathbf{V^n} = \left(\underbrace{\mathbf{X^1}, \ldots, \mathbf{X^1}}_{m_1 \text{ times}}, \ldots, \underbrace{\mathbf{X^t}, \ldots, \mathbf{X^t}}_{m_t \text{ times}}, \underbrace{\mathbf{X^{t+1}}, \ldots, \mathbf{X^{t+1}}}_{\kappa \text{ times}}, \mathbf{X^{t+1}}_{[1,\tau]} \right). \tag{34}$$

For this sequence, each use of the channel can be uniquely identified by the triple (t, κ, τ), with $t \in \mathbb{N}$, $\kappa \in [0, m_{t+1} - 1]$, and $\tau \in [0, n_{t+1}]$, and for the triple that corresponds to the n-th use of the channel the following holds:

$$n = \sum_{i=1}^{t} m_i n_i + \kappa n_{t+1} + \tau. \tag{35}$$

The numbers m_i are chosen such that

$$\frac{I(\mathbf{V^n}; \mathbf{W^n_{up(low)}})}{n} \geq val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^{t-1}}, \tag{36}$$

where $\mathbf{W^n_{up(low)}}$ is the random variable induced by $\mathbf{V^n}$ at the output of the upper(lower) channel, and n is related to t by (35). Note that the input sequence $\mathbf{V^n}$ is composed of independent random variables because each $\mathbf{X^t}$ ends in a reset, and hence

$$i_{\mathbf{V^n}, \mathbf{W^n_{up(low)}}} = \sum_{i=1}^{t} m_i \, i_{\mathbf{X^i}, \mathbf{Y^i_{up(low)}}} + \kappa \, i_{\mathbf{X^{t+1}}, \mathbf{Y^{t+1}_{up(low)}}} + i_{\mathbf{X^{t+1}}_{[1,\tau]}, \mathbf{Y^{t+1}_{up(low)}}_{[1,\tau]}}. \tag{37}$$

Also, note that for all t and $\tau \in [0, n_t]$, we have $0 \leq i_{\mathbf{X^t}_{[1,\tau]}, \mathbf{Y^t_{up(low)}}_{[1,\tau]}} \leq \tau$.

Eq. (37) implies the following:

$$\begin{aligned} I(\mathbf{V^n}, \mathbf{W^n_{up(low)}}) &= \sum_{i=1}^{t} m_i I(\mathbf{X^i}; \mathbf{Y^i_{up(low)}}) + \kappa I(\mathbf{X^{t+1}}; \mathbf{Y^{t+1}_{up(low)}}) + I(\mathbf{X^{t+1}}_{[1,\tau]}; \mathbf{Y^{t+1}_{up(low)}}_{[1,\tau]}) \\ &\geq \sum_{i=1}^{t} m_i n_i (val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^i}) + \kappa n_{t+1} (val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^{t+1}}). \end{aligned}$$

From that, we can see that (36) is verified if m_t is chosen to be larger than

$$\left\lceil \frac{2^t}{n_t \mu} \left(\sum_{i=1}^{t-1} m_i n_i \mu \left(\frac{1}{2^i} - \frac{1}{2^{t-1}} \right) + n_{t+1} \left(val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^{t-1}} \right) \right) \right\rceil, \tag{38}$$

so that $\sum_{i=1}^{t} m_i n_i \left(val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^i} \right) \geq \left(\sum_{i=1}^{t} m_i n_i + n_{t+1} \right) \left(val_{\mathcal{B}'_\alpha} - \frac{\mu}{2^{t-1}} \right)$.

However, for technical reasons pertinent to the bounds that follow we choose

$$m_t = \max\{Eq.(38), (n_{t+1})^2\}. \tag{39}$$

Recall the definition of information-stability in (3), and let $C_n \equiv 1/n \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n})$. From Lemma 5.3 we can write

$$\begin{aligned} &\Pr\left[\left| \frac{i_{\mathbf{V^n}, \mathbf{W^n}}}{nC_n} - 1 \right| \geq \eta\mu \right] \\ &\leq \alpha \Pr\left[\left| \frac{i_{\mathbf{V^n}, \mathbf{W^n_{up}}} - (\beta_2 - \beta_1)}{nC_n} - 1 \right| \geq \eta\mu \right] + (1-\alpha) \Pr\left[\left| \frac{i_{\mathbf{V^n}, \mathbf{W^n_{low}}} - (\beta_3 - \beta_1)}{nC_n} - 1 \right| \geq \eta\mu \right]. \end{aligned} \tag{40}$$

For the first term we can write

$$\begin{aligned} &\Pr\left[\left| \frac{i_{\mathbf{V^n}, \mathbf{W^n_{up}}} - (\beta_2 - \beta_1)}{nC_n} - 1 \right| \geq \eta\mu \right] \\ &= \Pr\left[\frac{i_{\mathbf{V^n}, \mathbf{W^n_{up}}} - (\beta_2 - \beta_1)}{nC_n} - 1 \geq \eta\mu \right] + \Pr\left[1 - \frac{i_{\mathbf{V^n}, \mathbf{W^n_{up}}} - (\beta_2 - \beta_1)}{nC_n} \geq \eta\mu \right] \\ &\leq \Pr\left[\frac{i_{\mathbf{V^n}, \mathbf{W^n_{up}}} + c_0/n}{nC_n} - 1 \geq \eta\mu \right] + \Pr\left[1 - \frac{i_{\mathbf{V^n}, \mathbf{W^n_{up}}} - \gamma_n}{nC_n} \geq \eta\mu \right]. \end{aligned} \tag{41}$$

For the input sequence described above the following bounds hold:

$$
\begin{aligned}
&\leq 1+\zeta && (42)\\
&\leq \frac{1}{1-\epsilon-1/n}\\
&= \frac{I(\mathbf{V^n};\mathbf{W^n_{up(low)}})/n}{1-\epsilon-1/n}\\
&\leq \frac{I(\mathbf{V^n};\mathbf{W^n_{up(low)}})/n}{\alpha(1-\epsilon)+(1-\alpha)(1-\epsilon)-1/n} && (43)\\
&\leq \frac{I(\mathbf{V^n};\mathbf{W^n_{up(low)}})/n}{\alpha\, I(\mathbf{V^n};\mathbf{W^n_{up}})/n+(1-\alpha)\, I(\mathbf{V^n};\mathbf{W^n_{low}})/n-1/n}\\
&\leq \frac{I(\mathbf{V^n};\mathbf{W^n_{up(low)}})/n}{\sup_{\mathbf{V^n}}\left(\alpha\, I(\mathbf{V^n};\mathbf{W^n_{up}})/n+(1-\alpha)\, I(\mathbf{V^n};\mathbf{W^n_{low}})/n-1/n\right)} && (44)\\
&= \frac{I(\mathbf{V^n};\mathbf{W^n_{up(low)}})/n}{\sup_{\mathbf{X^n}} I/n}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left[\frac{i_{\mathbf{V^n},\mathbf{W^n_{up(low)}}}}{nC_n}\right]\\
&\geq \mathbb{E}\left[\frac{i_{\mathbf{V^n},\mathbf{W^n_{up(low)}}}}{n\, val_{\mathcal{B}'_\alpha}}\right]\\
&\geq \frac{1}{val_{\mathcal{B}'_\alpha}}\left(val_{\mathcal{B}'_\alpha}-\frac{\mu}{2^{t-1}}\right)\\
&= 1-\frac{\mu}{val_{\mathcal{B}'_\alpha}2^{t-1}}, && (45)
\end{aligned}
$$

for any number $\zeta>0$, and where (44) follows from (18) and (43) follows from (33). Note that ζ can be made to go to 0 as n goes to ∞. Substituting by the upper bound in (42) and the lower bound in (45) into the first and second terms in (41), respectively, we get

$$
\begin{aligned}
&\Pr\left[\left|\frac{i_{\mathbf{V^n},\mathbf{W^n_{up}}}-(\beta_2-\beta_1)}{nC_n}-1\right|\geq\eta\mu\right]\\
&\leq \Pr\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}-\mathbb{E}\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}\right]\geq\eta\mu nC_n-c_0/n-nC_n\zeta\right]\\
&+\Pr\left[\mathbb{E}\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}\right]-i_{\mathbf{V^n},\mathbf{W^n_{up}}}\geq nC_n\left(\eta\mu-\frac{\mu}{val_{\mathcal{B}'_\alpha}2^{t-1}}\right)-\gamma_n\right]\\
&\leq \Pr\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}-\mathbb{E}\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}\right]\geq nC_n\left(\eta\mu-\zeta-\frac{\mu}{val_{\mathcal{B}'_\alpha}2^{t-1}}\right)-c_0/n-\gamma_n\right]\\
&+\Pr\left[\mathbb{E}\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}\right]-i_{\mathbf{V^n},\mathbf{W^n_{up}}}\geq nC_n\left(\eta\mu-\zeta-\frac{\mu}{val_{\mathcal{B}'_\alpha}2^{t-1}}\right)-c_0/n-\gamma_n\right]\\
&=\Pr\left[\left|i_{\mathbf{V^n},\mathbf{W^n_{up}}}-\mathbb{E}\left[i_{\mathbf{V^n},\mathbf{W^n_{up}}}\right]\right|\geq nC_n\left(\eta\mu-\zeta-\frac{\mu}{val_{\mathcal{B}'_\alpha}2^{t-1}}\right)-c_0/n-\gamma_n\right]
\end{aligned}
$$

Now we use the fact that $i_{\mathbf{V^n},\mathbf{W^n_{up}}}$ can be expressed as sum of $l=\sum_{i=1}^{t}m_i+\kappa+1$ independent random variables. For these sums we can bound the two-tailed probability via Hoeffding's inequality. More concretely, let $\{X_i\}_{i=1}^{l}$ be a sequence of l independent random variables, let $d\geq 0$, and let $a_i\leq X_i\leq b_i$, then

$$
\Pr\left[\left|\sum_{i=1}^{l}X_i-\mathbb{E}\left[\sum_{i=1}^{l}X_i\right]\right|\geq d\right]\leq 2\exp\left(\frac{-2d^2}{\sum_{i=1}^{l}|b_i-a_i|^2}\right).
$$

We can bound the denominator in the exponential as follows:

$$\begin{aligned}\sum_{i=1}^{l}|b_i-a_i|^2 &= \sum_{i=1}^{t} m_i(n_i)^2+\kappa(n_{t+1})^2+\tau^2\\ &\leq \sum_{i=1}^{t} m_i n_i n_{t+1}+\kappa n_{t+1}n_{t+1}+\tau n_{t+1}\\ &\leq n^{3/2},\end{aligned}$$

where the last inequality follows because, from Eq. (39), we have $(n_{t+1})^2 \leq m_t \leq n$. Now the exponential becomes

$$2\exp\left(-\frac{2\left(nC_n\left(\eta\mu-\zeta-\frac{\mu}{val_{\mathcal{B}'_\alpha}2^{t-1}}\right)-c_0/n-\gamma_n\right)^2}{n^{3/2}}\right), \tag{46}$$

which goes to 0 for all $\eta>0$ as $n\to\infty$.

The exact same can be done for the second term in (40), and hence we have that, when $val_{\mathcal{B}'_\alpha}=1$, there exists a sequence for which $\lim_{n\to\infty}\Pr\left[\left|\frac{i_{\mathbf{V^n};\mathbf{W^n}}}{nC_n}-1\right|>\eta\mu\right]=0$, and so the channel is information-stable.

□

7.2 When $val_{\mathcal{B}'_\alpha}=\alpha$, channels parametrised by $\alpha=1/2$ are information-stable and channels parametrised by $1/2<\alpha<1$ are not information-stable.

Lemma 7.2. FSMCs based on PFAs $\mathcal{B}'_\alpha$ with $\alpha=1/2$, are information-stable when $val_{\mathcal{B}'_\alpha}=\alpha$.

Proof. We are going to show that the following input sequence causes the channel to be information-stable when $\alpha=1/2$:

$$\mathbf{X}=\left(\mathbf{w^0},\mathbf{data},rt,\mathbf{w^1},\mathbf{data},rt\right), \tag{47}$$

where $\mathbf{w^0}\equiv w_1^0\ldots w_{m_0}^0$ is a word that causes the PFA to accept with probability 1 if it started in the upper branch and with probability 0 if it started in the lower branch, $\mathbf{w^1}\equiv w_1^1\ldots w_{m_1}^1$ is a word that causes the PFA to accept with probability 0 if it started in the upper branch and with probability 1 if it started in the lower branch, $m\equiv m_0+m_1$ and $n\equiv n_0+n_1$, and the data symbols $\mathbf{data}\in\{0,1\}^n$ are chosen uniformly at random. For the above sequence we have that $H(\mathbf{Y_{up}})=H(\mathbf{Y_{low}})=m+n+2$, because the output of the channel is completely uncertain if one does not know the input. Moreover, for the upper channel we can write

$$\begin{aligned}H(\mathbf{Y_{up}}|\mathbf{X}) &= \sum_{i=1}^{m+n+2} H(Y_{up_i}|\mathbf{Y_{up}}_{[1,i-1]}\mathbf{X})\\ &= m+2+\sum_{i=1}^{n} H(Y_{up_i}|\mathbf{Y_{up}}_{[1,i-1]}\mathbf{X})\\ &\leq m+2+\sum_{i=1}^{n_1} H(Y_{up_i}|\mathbf{Y_{up}}_{[1,i-1]}\mathbf{X})\\ &\leq m+2+n_1,\end{aligned}$$

where the first inequality follows from bounding the entropy of the uses in which w_0 and w_1 and the two resets are entered by $m+2$, the second inequality follows from the fact that when the upper branch accepts with probability 1 there is no uncertainty in the output given the input, and the final inequality follows by bounding the entropy of the n_1 uses by n_1; and so $\frac{1}{m+n+2}I(\mathbf{X};\mathbf{Y_{up}})\geq\frac{n_0}{m+n_0+n_1}$.

Similarly, for the lower channel we have $\frac{1}{m+n+2}I(\mathbf{X};\mathbf{Y_{low}})\geq\frac{n_1}{m+n_0+n_1}$.

We can equate both of these lower bounds to get the n_0 and n_1 at which the lower bounds are simultaneously $\geq 1/2 - \epsilon$ for any $\epsilon > 0$. If we let $n_x \equiv n_0 = n_1$, we can see that there always exists an n_x such that $\frac{n_x}{m+2n_x} \geq \frac{1}{2} - \epsilon = val_{\mathcal{B}'_\alpha} - \epsilon$.

The rest of the proof is exactly like the proof in the previous section starting from Eq. (34).

□

Lemma 7.3. FSMCs based on PFAs $\mathcal{B}'_\alpha$ with $1/2 < \alpha < 1$ are not information-stable when $val_{\mathcal{B}'_\alpha} = \alpha$.

Proof. Letting $C_n \equiv \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n})$, recall that a channel $\mathbf{W}$ is said to be information-stable if and only if there exists a sequence of inputs $\{X_i\}_{i=1}^{\infty}$ that satisfies the condition

$$\lim_{n\to\infty} \Pr\left[\left|\frac{i_{\mathbf{X^n};\mathbf{W^n}}}{nC_n} - 1\right| > \lambda\right] = 0, \tag{48}$$

for any $\lambda > 0$. For $1/2 < \alpha < 1$, we are going to show that condition can never be satisfied when $val_{\mathcal{B}'_\alpha} = \alpha$, or, equivalently, when $val_{\mathcal{A}} \leq \delta$.

Consider the following chain of inequalities:

$$\begin{aligned}
&\Pr\left[\left|\frac{i_{\mathbf{X^n},\mathbf{W^n}}}{nC_n} - 1\right| > \lambda\right] \\
&= \Pr\left[\left|\frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} - C_n\right| > \lambda C_n\right] \\
&= \Pr\left[\frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} - C_n > \lambda C_n\right] + \Pr\left[C_n - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} > \lambda C_n\right] \\
&\geq \Pr\left[C_n - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} > \lambda C_n\right] \\
&\geq \Pr\left[C_n - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} > \lambda\right],
\end{aligned} \tag{49}$$

where the last inequality follows because $C_n \leq 1$.

Since we have from Lemma 6.7 that when $val_{\mathcal{B}'_\alpha} = \alpha$ it holds that $\lim_{n\to\infty} C_n = val_{\mathcal{B}'_\alpha} = \alpha$, we know that for a sufficiently large n we have that $C_n > \alpha - \epsilon$ for any $\epsilon > 0$. At one such n, the chain of inequalities above continues on to be

$$\begin{aligned}
&\Pr\left[C_n - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} > \lambda\right] \\
&\geq \Pr\left[\alpha - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n}} > \lambda + \epsilon\right] && (50) \\
&\geq \alpha(1 - e^{-n\gamma_n}) \Pr\left[\alpha - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n_{up}}} + \beta_2 - \beta_1 > \lambda + \epsilon\right] && (51) \\
&+ (1-\alpha)(1 - e^{-n\gamma_n}) \Pr\left[\alpha - \frac{1}{n} i_{\mathbf{X^n},\mathbf{W^n_{low}}} + \beta_3 - \beta_1 > \lambda + \epsilon\right], && (52)
\end{aligned}$$

where the second inequality follows from Lemma 5.3.

Lemma 7.4 below implies that $\max\left(\alpha - \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up}}), \alpha - \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{low}})\right) \geq \alpha - 1/2$. Let us assume that we are in the case

$$\alpha - \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up}}) \geq \alpha - \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{low}}) \geq \alpha - 1/2, \tag{53}$$

and let us choose $\lambda = \left(\alpha - \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up}})\right)/4$ and $\epsilon = \left(\alpha - \frac{1}{n} I(\mathbf{X^n}; \mathbf{Y^n_{up}})\right)/4 + (\beta_2 - \beta_1)$. Before proceeding let us check that with this choice of ϵ the chain of inequalities starting from (50) still holds. From (53) we have that $\epsilon \geq (\alpha - 1/2)/4 + (\beta_2 - \beta_1) \geq (\alpha - 1/2)/4 - c_0/n$. As n increases, the c_0/n term vanishes, and we have that ϵ is bounded away from 0 by a constant, which guarantees that at a sufficiently large n we can have that $C_n > \alpha - \epsilon$, and hence that the chain of inequalities holds. Now, with that choice of λ and ϵ, (52) becomes

$$\alpha(1-e^{-n\gamma_n})\Pr\left[\alpha-\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{up}}}+\beta_2-\beta_1>\lambda+\epsilon\right]$$
$$+(1-\alpha)(1-e^{-n\gamma_n})\Pr\left[\alpha-\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{low}}}+\beta_3-\beta_1>\lambda+\epsilon\right]$$
$$\geq \alpha(1-e^{-n\gamma_n})\Pr\left[\alpha-\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{up}}}>\frac{\alpha-\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})}{2}\right]. \tag{54}$$

Now, note that $\alpha-\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})=\mathbb{E}\left[\alpha-\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{up}}}\right]$. Denoting the random variable $\alpha-\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{up}}}$ by V, (54) can be rewritten as $\alpha(1-e^{-n\gamma_n})\Pr\left[V>\frac{\mathbb{E}[V]}{2}\right]$. Since $V\equiv\alpha-\frac{1}{n}i_{\mathbf{X^n},\mathbf{W^n_{up}}}(\mathbf{X^n};\mathbf{Y^n_{up}})\leq 1$, we have from Lemma 7.5 below that

$$\alpha(1-e^{-n\gamma_n})\Pr\left[V>\frac{\mathbb{E}[V]}{2}\right]\geq\alpha(1-e^{-n\gamma_n})\frac{\mathbb{E}[V]}{2-\mathbb{E}[V]}. \tag{55}$$

From (53) we have that $1\geq\mathbb{E}[V]\geq\alpha-1/2>0$, and so the right hand side of (55) is always strictly positive, and all that finally means that there exists a λ at which condition (48) does not hold for any sequence of inputs when n is sufficiently large. The same conclusion holds in case

$$\alpha-\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\geq\alpha-\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})\geq\alpha-1/2,$$

and we chose $\lambda=\left(\alpha-\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\right)/4$ and $\epsilon=\left(\alpha-\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\right)/4+(\beta_3-\beta_1)$. Note that both $(\beta_2-\beta_1)$ and $(\beta_3-\beta_1)$ can both be made arbitrarily small for a sufficiently large n.

□

Lemma 7.4 (helper lemma to Lemma 7.3)**.** For the PFA $\mathcal{B}'_\alpha$, when $val_{\mathcal{B}'_\alpha}=\alpha$ it holds that for any input process $\mathbf{X^n}$

$$\min\left(\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}}),\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\right)\leq 1/2.$$

Proof. Recall that we can think of the upper branch and lower branch of $\mathcal{B}'_\alpha$ as two channels whose reset erases all memory of previous uses. We have from Corollary 6.5.1 that $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})\leq\frac{1}{n}\sum_i(n_i+1)\,val_{up}(\mathbf{X^{n_i+1}}_{[1,n_i]})$ and $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\leq\frac{1}{n}\sum_i(n_i+1)\,val_{low}(\mathbf{X^{n_i+1}}_{[1,n_i]})$. From Corollary 4.9.1 we then have $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})\leq\frac{1}{n}\sum_i(n_i+1)\,(1-\epsilon_i)$ and $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\leq\frac{1}{n}\sum_i(n_i+1)\,\epsilon_i$.

Note that $\frac{1}{n}\sum_i(n_i+1)(1-\epsilon_i)=\frac{1}{n}\sum_i(n_i+1)-\frac{1}{n}\sum_i(n_i+1)\epsilon_i\geq 1/2$ implies that $\frac{1}{n}\sum_i(n_i+1)\epsilon_i\leq 1/2$, and conversely $\frac{1}{n}\sum_i(n_i+1)\,\epsilon_i\geq 1/2$ implies that $\frac{1}{n}\sum_i(n_i+1)\,(1-\epsilon_i)\leq 1/2$. From this it follows that if $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})\geq 1/2$ it must be that $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\leq 1/2$, and that if $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{low}})\geq 1/2$ it must be that $\frac{1}{n}I(\mathbf{X^n};\mathbf{Y^n_{up}})\leq 1/2$. Hence the statement of the lemma. □

Lemma 7.5 (helper lemma to Lemma 7.3. [18])**.** Let V be a random variable such that $V\leq 1$. Then $\Pr\left[V>\frac{\mathbb{E}[V]}{2}\right]\geq\frac{\mathbb{E}[V]}{2-\mathbb{E}[V]}$.

Proof. If we introduce the random variable Z which takes value 1 when $V>\frac{\mathbb{E}[V]}{2}$ and value 0 otherwise, we can write

$$\begin{aligned}\mathbb{E}[V]&=p(Z=1)\,\mathbb{E}[V|Z=1]+(1-p(Z=1))\,\mathbb{E}[V|Z=0]\\&\leq p(Z=1)+(1-p(Z=1))\,\mathbb{E}[V|Z=0]\\&\leq p(Z=1)+(1-p(Z=1))\,\frac{\mathbb{E}[V]}{2},\end{aligned}$$

From this chain of inequalities we have that $p(Z=1)=\Pr\left[V>\frac{\mathbb{E}[V]}{2}\right]\geq\frac{\mathbb{E}[V]}{2-\mathbb{E}[V]}$.

□

Discussion

The take-away from this section is the following: The channels we constructed in Section 5 are information-stable when the value of the underlying PFA is 1 or 1/2 (Lemmas 7.1 and 7.2, respectively), and not information-stable otherwise (Lemma 7.3).

8 Capacity is uncomputable

proof of Theorem 3.1. We saw in Section 5 that we can build FSMCs on top of the PFAs $\mathcal{B}'_\alpha$ such that the channels have input alphabet, output alphabet, and memory size pointwise equal to or greater than $(7, 2, 30)$. For PFAs $\mathcal{B}'_\alpha$ with $\alpha = 1/2$, Lemmas 7.1 and 7.2 tell us that the channels are always information-stable. Combined with Lemma 6.7, this allows us to write

$$C(\mathbf{W}) = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathbf{X^n}} I(\mathbf{X^n}; \mathbf{Y^n}) = val_{\mathcal{B}'_\alpha}. \tag{56}$$

Then, since it is undecidable whether $val_{\mathcal{B}'_\alpha} = \alpha = 1/2$ or $val_{\mathcal{B}'_\alpha} = 1$, it is undecidable whether $C(\mathbf{W}) = \alpha$ or $C(\mathbf{W}) = 1$; this implies that $C(\mathbf{W})$ is uncomputable to any precision $< \frac{1-\alpha}{2} = 1/4$. □

9 Information-stability is undecidable

proof of Theorem 3.2. We saw in Section 5 that we can build FSMCs on top of the PFAs $\mathcal{B}'_\alpha$ such that the channels have input alphabet, output alphabet, and memory size pointwise equal to or greater than $(7, 2, 30)$. For PFAs $\mathcal{B}'_\alpha$ with $1/2 < \alpha < 1$, Lemma 7.1 tells us that the channel is information-stable when $val_{\mathcal{B}'_\alpha} = 1$ and Lemma 7.3 tells us that the channel is not information-stable when $val_{\mathcal{B}'_\alpha} = \alpha$. Since it is undecidable whether $val_{\mathcal{B}'_\alpha} = 1$ or $val_{\mathcal{B}'_\alpha} = \alpha$ (Lemma 4.8), it is undecidable whether the channel is information-stable or not. □

10 Conclusion

We have demonstrated a new approach for exploring uncomputability in information theory. In this approach, we use tools from automata theory to show, for channels with memory, that capacity is in general uncomputable, information-stability is in general undecidable. Both of these results have been shown to hold for channels with input alphabet, output alphabet, and number of states point-wise equal or larger than $(7, 2, 30)$. We expect that our approach can be used to show other problems in information theory and quantum information theory undecidable or uncomputable.

In the appendix, we present two corollaries that follow from our results. The first corollary, which follows from Theorem 3.1, is the following. Suppose you are given an information source $\mathbf{X}$ and a channel $\mathbf{W}$. It is undecidable whether it is possible transmit $\mathbf{X}$ over $\mathbf{W}$ such that the receiving party receives $\mathbf{X}$ with arbitrarily small error. The second corollary, which follows from our new PFA construction shown in Fig. 2, is the following. For a PFA with input alphabet of size 5 and 36 states, it is undecidable whether the value of the PFA is 1. This is known as the value-1 problem, and it is a well known problem in automata theory. In fact, it has already been shown before that the value-1 problem is undecidable. Our corollary only improves the undecidability result so that it applies for PFAs with input alphabet with size 5 and 36 memory states, instead of the previous best of 5 and 61 [7].

11 For future research

11.1 Reducing the memory size further

The capacity of channels with 1 memory state (memoryless channels) is computable [3, 1]. The capacity of channels with 30 memory states is uncomputable (Theorem 3.1). At what number of memory states does capacity transition from being computable to being uncomputable? This is an interesting question to answer in future research.

The channels we use to prove Theorem 3.1 inherit the uncomputability in their capacity from Post's Correspondence problem (see section 4.2.1), which we know is undecidable for $k > 5$ pairs of words. At the time of writing this paper, it is unknown whether Post's Correspondence Problem is decidable for $3 \leq k \leq 5$ pairs of words. If it turns out to be undecidable for $k \geq 3$, it would immediately follow from our analysis that the results hold for channels with input alphabet, output alphabet, and number of states point-wise equal or larger than $(7, 2, 15)$. This is one way through the memory size of channels with uncomputable capacity could be reduced.

Another way is find a cleverer PFA construction than ours (shown in Fig. 2). Such a construction would presumably still have be based on an undecidable problem, although it might not need to be Post's Correspondence Problem..

In section C of the appendix we detail some of our failed attempts to reduce the memory size of the channels. These failed attempts eventually led to the automaton shown in Fig. 2.

11.2 Proving other problems in (quantum) information theory undecidable

11.2.1 Quantum channels with uncomputable capacity

Since classical channels are special cases of quantum channels – they are quantum channels that can only transmit classical information and no quantum information –, the results of this work trivially apply to quantum channels as well.

Non-trivially, however, the results might be extensible to quantum channels that can transmit quantum information. That is, it might be possible – and we suspect that it is – to show, for instance, that quantum capacity for quantum channels with memory is uncomputable. This could be done using the same technique we used in the work: simply by defining the channels built on top of the automata to take qudits as input and produce qudits at the output instead of classical symbols.

References

[1] Suguru Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20, 1972.

[2] Dieter-Michael Arnold, H-A Loeliger, Pascal O Vontobel, Aleksandar Kavcic, and Wei Zeng. Simulation-based computation of information rates for channels with memory. *IEEE Transactions on information theory*, 52(8):3498–3508, 2006.

[3] Richard Blahut. Computation of channel capacity and rate-distortion functions. *IEEE transactions on Information Theory*, 18(4):460–473, 1972.

[4] V Claus. Some remarks on pcp(k) and related problems. *Bulletin of The European Association for Theoretical Computer Science - EATCS*, 12, 01 1980.

[5] RL Dobrushin. General formulation of shannon's main theorem in information theory. *American mathematical society translations*, 33:323–438, 1963.

[6] David Elkouss and David Pérez-García. Memory effects can make the transmission capability of a communication channel uncomputable. *Nature Communications*, 9(1):1149, 2018.

[7] Hugo Gimbert and Youssouf Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In *International Colloquium on Automata, Languages, and Programming*, pages 527–538. Springer, 2010.

[8] Andrea J Goldsmith and Pravin P Varaiya. Capacity, mutual information, and coding for finite-state markov channels. *IEEE Transactions on Information Theory*, 42(3):868–886, 1996.

[9] Vesa Halava, Tero Harju, and Mika Hirvensalo. Undecidability bounds for integer matrices using claus instances. *International Journal of Foundations of Computer Science*, 18(05):931–948, 2007.

[10] Guangyue Han. A randomized algorithm for the capacity of finite-state channels. *IEEE Transactions on Information Theory*, 61(7):3651–3669, 2015.

[11] Te Sun Han. Information-spectrum methods in information theory. 2014.

[12] Te Sun Han and Sergio Verdú. Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772, 1993.

[13] Mika Hirvensalo. Improved undecidability results on the emptiness problem of probabilistic and quantum cut-point languages. In *International Conference on Current Trends in Theory and Practice of Computer Science*, pages 309–319. Springer, 2007.

[14] Tim Holliday, Andrea Goldsmith, and Peter Glynn. Capacity of finite state channels based on lyapunov exponents of random matrices. *IEEE Transactions on Information Theory*, 52(8):3509–3532, 2006.

[15] G. J. O. Jameson. *The Prime Number Theorem*. Cambridge University Press, 2013.

[16] Aleksandar Kavcic. On the capacity of markov sources over noisy channels. In *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, volume 5, pages 2997–3001. IEEE, 2001.

[17] H Koga et al. *Information-spectrum methods in information theory*, volume 50. Springer Science & Business Media, 2013.

[18] leonbloy (https://math.stackexchange.com/users/312/leonbloy). Lower bound on the probability that a random variable is greater than half of its mean. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/3168259 (version: 2019-03-30).

[19] Mordechai Mushkin and Israel Bar-David. Capacity and coding for the gilbert-elliott channels. *IEEE Transactions on Information Theory*, 35(6):1277–1290, 1989.

[20] Turlough Neary. Undecidability in binary tag systems and the post correspondence problem for five pairs of words. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 30. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

[21] Henry D. Pfister. *The capacity of finite-state channels in the high-noise regime*, pages 179–222. London Mathematical Society Lecture Note Series. Cambridge University Press, 2011.

[22] Henry D Pfister, Joseph B Soriaga, and Paul H Siegel. On the achievable information rates of finite state isi channels. In *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, volume 5, pages 2992–2996. IEEE, 2001.

[23] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[24] Vinod Sharma and SK Singh. Entropy and channel capacity in the regenerative setup with applications to markov channels. In *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No. 01CH37252)*, page 283. IEEE, 2001.

[25] Sridhar Vembu, Sergio Verdu, and Yossef Steinberg. The source-channel separation theorem revisited. *IEEE Transactions on Information Theory*, 41(1):44–54, 1995.

[26] Sergio Verdú et al. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, 1994.

[27] Pascal O Vontobel, Aleksandar Kavcic, Dieter M Arnold, and Hans-Andrea Loeliger. A generalization of the blahut–arimoto algorithm to finite-state channels. *IEEE Transactions on Information Theory*, 54(5):1887–1918, 2008.

[28] Chengyu Wu, Guangyue Han, and Brian Marcus. A deterministic algorithm for the capacity of finite-state channels. *arXiv preprint arXiv:1901.02678*, 2019.

Appendix

A Corollary of Theorem 3.1: Reliable transmission is undecidable

As a corollary from Theorem 3.1, it follows that it is undecidable whether an information source can be reliably transmitted over a channel.

A source is said to be reliably transmissible over a channel if it can be decoded at the output of the channel with arbitrarily small error in the limit of infinitely many uses of the channel. A necessary and sufficient condition for a (stationary) source to be reliably transmitted over a channel was given in [25]. We show that for the family of states in Theorem 3.1, it is undecidable whether that condition is satisfied. The corollary can be stated precisely as follows:

Corollary A.0.1. Given a stationary information source $\mathbf{X}$ with minimum source-coding rate $1/2 < T(\mathbf{X}) < 1$, and an information-stable FSMC $\mathbf{W}$ with rational product probabilities and with input, output, and number of states point-wise equal or larger than $(7, 2, 30)$, it is undecidable whether $\mathbf{X}$ can be reliably transmitted over $\mathbf{W}$ or not.

Before proving Corollary A.0.1, let us first make some definitions.

Definition A.1 (limsup in probability)**.** Define the limsup in probability of a sequence of random variables $A^n \equiv \{A_n\}$ to be the smallest extended real number β such that, for all $\epsilon > 0$, $\lim_{n\to\infty} \Pr[A_n \geq \beta + \epsilon] = 0$.

Definition A.2 (source coding rate)**.** An (n, M, ϵ) fixed-length source code for $\mathbf{X^n}$ is a collection of M n-tuples $\{a_1^n, \cdots, a_M^n\}$ such that $\Pr[\mathbf{X^n} \notin \{a_1^n, \cdots, a_M^n\}] \leq \epsilon$. R is an ϵ-achievable source coding rate for the source $\mathbf{X}$ if for every $\gamma > 0$ there exist, for all sufficiently large n, (n, M, ϵ) codes with $\frac{1}{n} \log M < R + \gamma$. R is an achievable (fixed-length) source coding rate for $\mathbf{X}$ if it is ϵ-achievable for all $\epsilon > 0$. We denote by $T(\mathbf{X})$ the minimal achievable source coding rate for $\mathbf{X}$.

If we replace "for all sufficiently large" by "for infinitely many" in the definition above, R becomes the "optimistic" achievable source coding rate. We denote by $\underline{\mathrm{T}}(\mathbf{X})$ the minimal optimistic achievable source coding rate for $\mathbf{X}$.

For an expression for the minimal achievable source coding rate, we have the following lemma.

Lemma A.1 (Theorem 4 in [25])**.** For any finite alphabet source $\mathbf{X}$, $T(\mathbf{X}) = \overline{H}(\mathbf{X})$, where $\overline{H}(\mathbf{X})$ denotes the sup-entropy rate, which is given by the limsup in probability of the sequence of random variables $\{(1/n)h_{\mathbf{X^n}}(\mathbf{X^n})\}_{n=1}^{\infty}$.

One of the main components of our proof is a source-channel coding theorem. The classical source-channel coding theorem proved by Shannon in [23] was only proved for memoryless channels, so we cannot use it. Instead, we are going to use a more general source-channel coding theorem from [25], which applies to any channel provided that the source is stationary. We re-construct this theorem below.

Lemma A.2 (Theorem 6 in [25])**.** For any source $\mathbf{X}$ and channel $\mathbf{W}$, if the condition $T(\mathbf{X}) < C(\mathbf{W})$ is satisfied then $\mathbf{X}$ can be reliably transmitted over $\mathbf{W}$.

This lemma constitutes the direct part the general source-channel coding theorem. We are now going to re-construct the converse part.

Lemma A.3 (Theorem 12 in [25])**.** Let $T_C(\mathbf{X}) \equiv \inf_{\mathbf{V}\in\mathcal{V}} C(\mathbf{W})$, where $\mathcal{V}$ is the set of all channels over which the source $\mathbf{X}$ can be reliably transmitted. For any source $\mathbf{X}$ and channel $\mathbf{W}$, the following two conditions are equivalent:

- $T_C(\mathbf{X}) = T(\mathbf{X})$.
- If $\mathbf{X}$ is reliably transmissible over $\mathbf{W}$, then it must hold that $T(\mathbf{X}) \leq C(\mathbf{W})$.

Note that the second condition is the converse part of the general source-channel coding theorem. For a source $\mathbf{X}$, if this second condition holds, then (with Lemma A.2) we have a source-channel coding theorem. It turns out that this converse part does not hold in general, but it does hold for stationary sources (because the first condition holds for stationary sources) [25], and so (with Lemma A.2) we finally have that a stationary source $\mathbf{X}$ can be reliably transmitted over any channel $\mathbf{W}$ if $T(\mathbf{X}) < C(\mathbf{W})$ and only if $T(\mathbf{X}) \leq C(\mathbf{W})$

We are now ready to prove corollary A.0.1.

proof of corollary A.0.1. From Theorem 3.1, it is undecidable whether the capacity $C(\mathbf{W})$ is $1/2$ or 1. It then follows trivially from the above discussion that for any stationary source for which $1/2 < T(\mathbf{X}) < 1$, it is undecidable whether the condition $T(\mathbf{X}) < C(\mathbf{W})$ is satisfied, and hence it is undecidable whether reliable transmission is possible.

Such source can be very simple: for example, an i.i.d source that outputs 0 with probability p and 1 with probability $1-p$, for which it holds that $1/2 < -p\log(p) - (1-p)\log(1-p) < 1$. This is because for such a source we can write

$$\begin{aligned} T(\mathbf{X}) &= \overline{H}(\mathbf{X}) \\ &= H(X) \\ &= -p\log(p) - (1-p)\log(1-p), \end{aligned}$$

where the first equality follows from Lemma A.1, the second equality follows from the fact that for an i.i.d source $\overline{H}(\mathbf{X}) = H(X)$ [11].

□

B New automaton for the value 1 problem

In this section we show that the new automaton $\mathcal{B}'_\alpha$ – after a few minor modifications – improves on the undecidability result of the value 1 problem [7]. The value 1 problem is a well-known problem in automata theory. It can be defined as follows:

Definition B.1. (The value 1 problem [7]) Given a PFA $\mathcal{P}$, determine whether it has value 1.

The new PFA construction above allows us to prove the following corollary about the value 1 problem.

Corollary B.0.1. The value 1 problem is undecidable for a PFA with 36 states and 5 input symbols.

Proof. We modify the automaton in Fig. 2 so that it has 5 input symbols instead of 7. This results in the PFA $\mathcal{B}''$ shown in Fig. 4, which has 5 input symbols and 36 states.

Since we are now concerned with building a PFA for the value 1 problem and not for information transmission purposes, the reset symbol is not needed. After removing the reset symbol, we have 6 symbols. Next, we are going to remove the symbol $'a'$, and we do this as follows. Taking $\mathcal{B}'_\alpha$ as a starting point, we are going to replace every $'a'$ transition (e.g. $q_1 \xrightarrow{a} c_1$) by a $'b'$ transition to an intermediate state then a $'1'$ transition to the final state. And we replace every $'b'$ transition by first a $'b'$ transition to an intermediate state then a $'0'$ transition to the final state. For example, the transition $q_1 \xrightarrow{a} c_1$ in Fig. 2 is replaced by the transition $q_1 \xrightarrow{b} d_5 \xrightarrow{1} c_1$ in Fig. 4, and the transition $q_1 \xrightarrow{b} q_3$ is replaced by the transition $q_1 \xrightarrow{b} d_5 \xrightarrow{0} q_3$. Note that for transitions in $\mathcal{B}'_\alpha$ in which $'a'$ and $'b'$ lead to the same state, no intermediate state is required; e.g. the transition $\mathcal{A} \xrightarrow{a,b}$ Sink.

Recalling that $'0'$ and $'1'$ are the control symbols for $\mathcal{A}$, one might think that entering a control word with the symbols $'0'$ and $'1'$ intended to trigger transitions outside of $\mathcal{A}$ might cause unintended transitions if $\mathcal{B}''$ happens to be in $\mathcal{A}$. However, note that for all words that achieve the value in $\mathcal{B}'_\alpha$, namely words of the form (16) and the word $'b'$, the symbols $'a'$ and $'b'$ are never entered when the automaton is inside $\mathcal{A}$, and so replacing $'a'$ and $'b'$ by $'b1'$ and $'b0'$, respectively, does not lead to unintended transitions inside $\mathcal{A}$. Moreover, any word in which $'b0'$ or $'b1'$ is entered when the automaton is inside $\mathcal{A}$ causes the automaton to fall into the sink, and these words can never lead to a higher acceptance probability than 1 and y.

Thus, $\mathcal{B}''$ emulates $\mathcal{B}'_\alpha$, and therefore has value 1 when $val_{\mathcal{A}} > \delta$ and value α when $val_{\mathcal{A}} \leq \delta$. So, the value 1 problem is undecidable for a PFA with 36 memory states and 5 input symbols. □

Discussion

We have improved on the undecidability result of the value 1 problem [7] by making it apply to smaller automata: the value 1 problem is now undecidable for automata with input alphabet, and number of states pointwise equal to or greater than $(5, 36)$ instead of the previous $(5, 61)$ in [7].

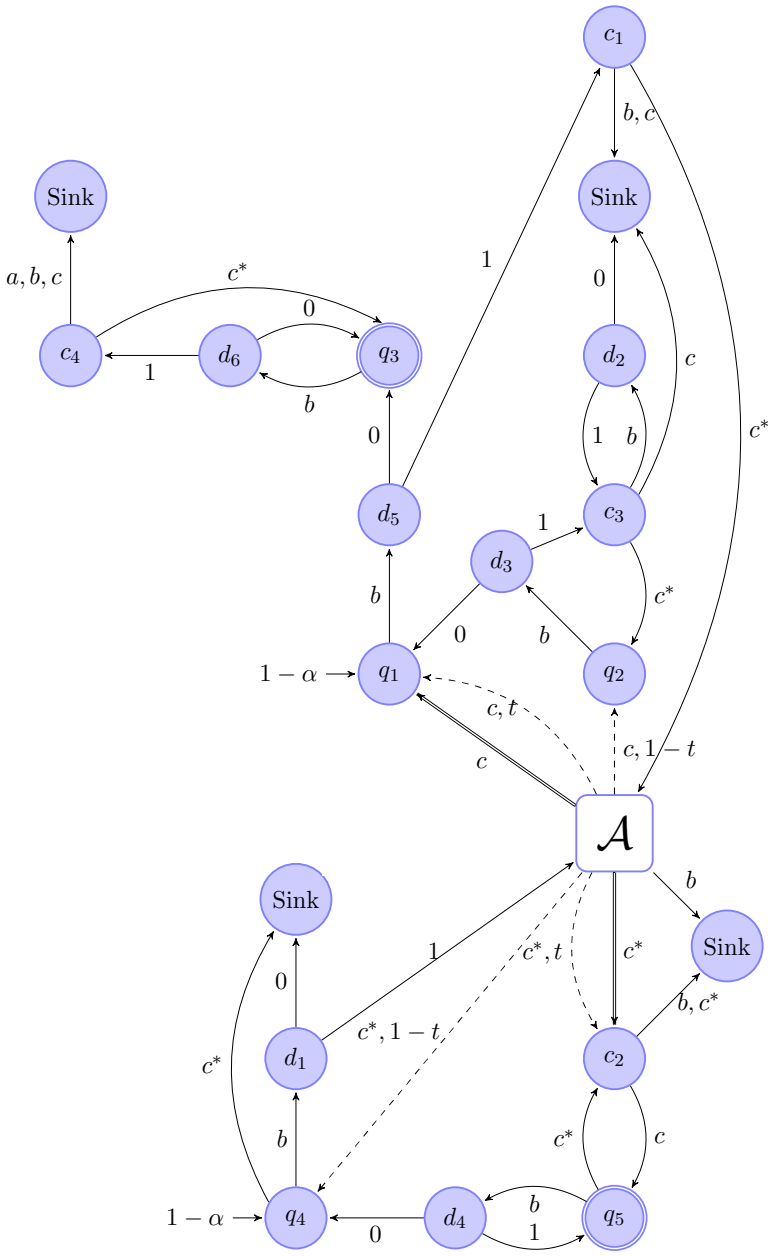


Figure 4: The PFA construction $\mathcal{B}''_\alpha$: designed to emulate $\mathcal{B}'_\alpha$ in Fig. 2 but with 5 input symbols instead of 7 (the symbols rt and a are removed). If any symbol except 0 and 1 is entered while the PFA is in one of the states d_i it goes to the $Sink$. This is not shown on the figure to avoid clutter. $\mathcal{B}''_\alpha$ has $16 + states_{\mathcal{A}}$ states.

C Reducing the memory size of the channels: failed designs

The automaton shown in fig. 2 is smaller than the original automaton used in the paper [6], and so allows for smaller channel constructions with uncomputable capacity. In this section we detail our analysis of some of prototypes of the automaton design shown in fig. 2.

At first, I did not have an idea how to begin making a smaller automaton, but I was intrigued by the fact that the when $x \leq 1/2$ (see fig. 1) the automaton has value $1/2$ and when $x > 1/2$ it had value 1. This jump in the value is essential for eventually showing capacity uncomputable. I started to wonder if we could still get a gap if the x in the upper branch and lower branch of the automaton were different (see fig. 5 left). Not having to have the same x in the upper and lower branch would mean that we might be able to make do with only a single $\mathcal{A}$ automaton instead of two as in the original desgin. This thinking led to the following analysis.

C.1 Failed design 1

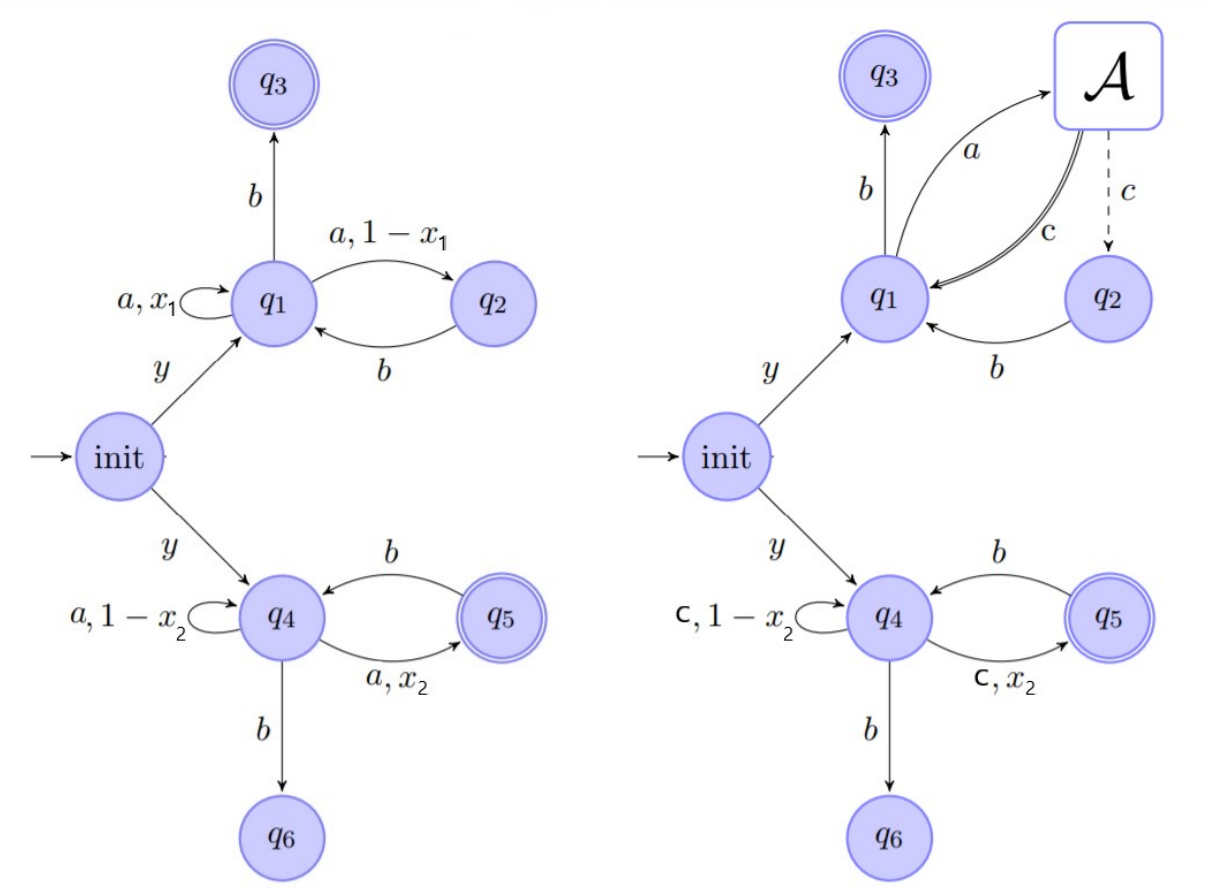


Figure 5: Here, $y = x_2 = 1/2$

Consider the automaton in fig. 5, with $y = x_2 = 1/2$. We aim to show that the value of the automaton jumps from one value to another as x_1 transitions from $\leq 1/2$ to $> 1/2$; i.e. that there is a gap in the value. We begin with the following lemma.

Lemma C.1. The value of the automaton in fig. 5 is $1/2$ when $x_1 \leq 1/2$.

Proof. We know that

$$\text{val}(A, w) \leq \frac{1}{2} p[q_1 \to q_3] + \frac{1}{2}(1 - p[q_4 \to q_6]),$$

and that the upper bound is reachable.
Note that

$$p[q_1 \to q_3] = 1 - \Pi_{i=1}^{t}(1 - x_1^{n_i}),$$

and

$$p[q_4 \to q_6] = 1 - \Pi_{i=1}^{t}(1 - (1 - x_2)^{n_i}).$$

Consider the case $x_1 \leq 1/2$. In that case, $x_1 \leq 1 - x_2 = 1/2$, and therefore

$$\underbrace{1 - \Pi_{i=1}^{t}(1 - x_1^{n_i})}_{p[q_1 \to q_3]} \leq \underbrace{1 - \Pi_{i=1}^{t}(1 - (1 - x_2)^{n_i})}_{p[q_4 \to q_6]}.$$

We can then use that to write

$$\begin{aligned}
\text{val}(A, w) &\leq \frac{1}{2}\, p[q_1 \to q_3] + \frac{1}{2}\,(1 - p[q_4 \to q_6]) \\
&\leq \frac{1}{2}\, p[q_1 \to q_3] + \frac{1}{2}\,(1 - p[q_1 \to q_3]) \\
&= \frac{1}{2}.
\end{aligned}$$

Since the upper bound is reachable – simply by entering $'b'$ into the automaton–, we have that the value is $1/2$. □

If we can now show that when x_1 transitions to $x_1 > 1/2$ the value jumps from $1/2$ to some other value, we would be done. For this we have the following lemma.

Lemma C.2. The value of the automaton in fig. 5 is $\geq 1/2 + \epsilon$ (for some finite $0 < \epsilon \leq 1/2$) when $x_1 > 1/2$.

Proof. We are going to prove the lemma by giving a word w for which

$$\text{val} \geq \text{val}(A, w) > 1/2$$

for any $x_1 > 1/2$.
This word is $w = aabaaa \cdots$ until ∞. For this specific word we can easily see that

$$\text{val}(A, w) = \frac{1}{2}(1 - (1 - x_1)) + \frac{1}{2}x_2^2 \underbrace{\sum_{i=0}^{\infty}(1 - x_2)^i}_{1/x_2} = \frac{x_1}{2} + \frac{1}{4}.$$

Thus, for $x_1 = 0.5 + \epsilon$ and any $0.5 \geq \epsilon > 0$,

$$\text{val}(A, w) = 0.5 + \frac{\epsilon}{2} > 0.5.$$

The term $(1 - (1 - x_1))$ is the complement of the probability that the upper branch of the automaton will go to the state q_2 when the second "a" in w is entered. The "a"s after the "b" do not add to the acceptance probability of the upper branch.

As for the acceptance probability of the lower branch: if we label the event that automaton loops back to q_4 when it is in q_4 by "lp", the event that the automaton goes forward from q_4 to q_5 by "fd", and the event that the automaton goes back from q_5 to q_4 by "bk", the term $x_2^2 \sum_{i=0}^{\infty}(1 - x_2)^i$ is the addition of the following probabilities:

$$
\begin{aligned}
x_2^2 \sum_{i=0}^{\infty} (1-x)^i = & p(fd \to bk \to fd) \\
& + p(fd \to bk \to lp \to fd) \\
& + p(fd \to bk \to lp \to lp \to fd) \\
& + p(fd \to bk \to lp \to lp \to lp \to fd) \\
& + p(fd \to bk \to lp \to lp \to lp \to lp \to fd) \\
& + \cdots .
\end{aligned}
$$

The x_2^2 factor is there because in all the terms there are two "fd"s.

□

With Lemma C.1, we have that the value of the automaton is $= 1/2$ when $x_1 \leq 1/2$. With Lemma C.2, we have that the value of the automaton $> 1/2 + \epsilon/2$ when $x_1 = 1/2 + \epsilon > 1/2$. We have indeed established a gap in the value. However, the usefulness of this gap for the purpose of proving uncomputability of the capacity is questionable for the following reason. As x_1 gets arbitrarily close to $1/2$, the gap becomes arbitrarily small, and this leads to an ambiguity: is capacity really uncomputable or does it just seem so because we can never in principle compute the value of any function to within infinitely small precision in a finite amount of time?

Therefore, this is a failed approach.

However, the idea of using a single $\mathcal{A}$ automaton instead of two inspired the approach detailed in the next section.

C.2 Failed design 2

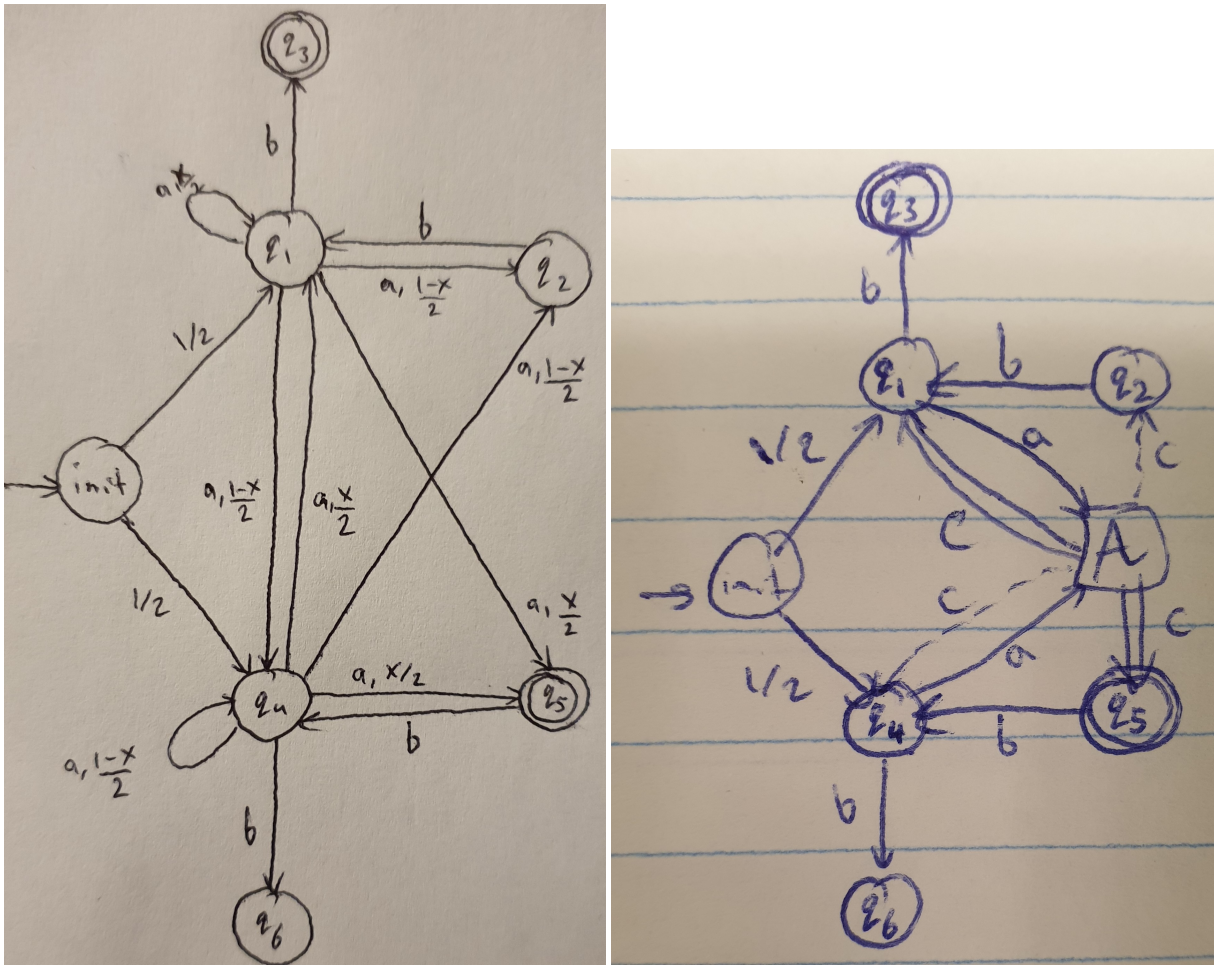


Figure 6: The automaton on the left is the operational equivalent to the one of the right.

Consider the automaton in fig. 6 (left). This corresponds to the design in fig. 6 (right). Like before, we would like to know if we can get a gap in the value using this design. Below, we will investigate what happens when we enter the general word $a^{n_2}b\cdots ba^{n_k}$, where n_i can assume any non-negative integer value. This word form includes all possible words that we could enter into the automaton.

First, let's think about what happens when the part a^{n_2} is entered. The probability that the automaton is going to end up at q_2 is that same regardless of whether the automaton went from *init* to q_1 or q_4, because a has exactly the same effect regardless of whether the automaton is in q_1 or q_4. Now, the probability of ending up at q_2 is equal to the probability that the first a took us to q_2 ($\frac{1-x}{2}$), plus the probability that the first a took us to q_1 or q_4 and the second a took us to q_2 ($(\frac{x}{2}+\frac{1-x}{2})(\frac{1-x}{2})$), and so on... This all adds up to

$$\begin{aligned} p_{a^{n_2}}(q_2) &= (\frac{1-x}{2}) + \frac{1}{2}(\frac{1-x}{2}) + \cdots + \frac{1}{2^{n_2-1}}(\frac{1-x}{2}) \\ &= \sum_{i=0}^{n_2-1} \frac{1}{2^i}(\frac{1-x}{2}) \\ &= \frac{(1-x)(2^{n_2}-1)}{2^{n_2}}. \end{aligned}$$

Similarly,

$$\begin{aligned} p_{a^{n_2}}(q_5) &= (\frac{x}{2}) + \frac{1}{2}(\frac{x}{2}) + \cdots + \frac{1}{2^{n_2-1}}(\frac{x}{2}) \\ &= \sum_{i=0}^{n_2-1} \frac{1}{2^i}(\frac{x}{2}) \\ &= \frac{x(2^{n_2}-1)}{2^{n_2}}. \end{aligned}$$

Now, when the first b is entered, if the automaton was in q_1 or q_4 it will end up in the final states q_3 or q_6, and if that happens the probability of ending up at q_2 drops to 0 and nothing that comes afterwards can change that. However, if the automaton was in q_2 or q_5 (which happens with probability $p_{a^{n_2}}(q_2) + p_{a^{n_2}}(q_5) = 1 - \frac{1}{2^{n_2}}$), the b takes the automaton to q_1 or q_4, respectively. Then, the same reasoning as above gives us that the probability of ending up at q_2 and q_5 for the word $a^{n_2}ba^{n_3}$ is

$$p_{a^{n_2}ba^{n_3}}(q_2) = (1 - \frac{1}{2^{n_2}})\frac{(1-x)(2^{n_3}-1)}{2^{n_3}},$$

and

$$p_{a^{n_2}ba^{n_3}}(q_2) = (1 - \frac{1}{2^{n_2}})\frac{x(2^{n_3}-1)}{2^{n_3}},$$

respectively.

So, after the entire word $a^{n_2}b\cdots ba^{n_k}$ is entered, the probabilities of ending up at q_2 and q_5 are

$$\boxed{p(q_2) = (1-x)\prod_{i=2}^{k}(1 - \frac{1}{2^{n_i}}),}$$

and

$$\boxed{p(q_5) = x\prod_{i=2}^{k}(1 - \frac{1}{2^{n_i}}).}$$

Now let's calculate $p(q_1)$ and $p(q_4)$. Let's think about what happens when the part a^{n_2} is entered. Note that to get to q_1, it must be that none of the a's took us to q_2 or q_5, because once we enter these two states there is no getting out until a b comes. The probability of ending up at q_1 after the a^{n_2} is equal to the probability of going to q_1 then going to q_1 then going to q_1 $\cdots$ then finally going to q_1 $((\frac{x}{2})^{n_2})$, plus the probability of going to q_4 then going to q_1 then going to q_1 $\cdots$ then finally going to q_1 $((\frac{1-x}{2})(\frac{x}{2})^{n_2-1})$, plus the probability of going to q_1 then going to q_4 then going to q_1 $\cdots$ then finally going to q_1 $((\frac{x}{2})(\frac{1-x}{2})(\frac{x}{2})^{n_2-2})$, and so on, adding up to

$$\begin{aligned} p_{a^{n_2}}(q_1) &= \sum_{i=0}^{n_2-1} \binom{n_2-1}{i} (\frac{x}{2})^{n_2-i} (\frac{1-x}{2})^i \\ &= (\frac{x}{2})^{n_2} \sum_{i=0}^{n_2-1} \binom{n_2-1}{i} (\frac{1-x}{x})^i \\ &= (\frac{x}{2})^{n_2} (\frac{1-x}{x} + 1)^{n_2-1} \\ &= \frac{x}{2^{n_2}}. \end{aligned}$$

Similarly,

$$
\begin{aligned}
p_{a^{n_2}}(q_4) &= \sum_{i=0}^{n_2-1} \binom{n_2-1}{i} \left(\frac{1-x}{2}\right)^{n_2-i} \left(\frac{x}{2}\right)^{i} \\
&= \left(\frac{1-x}{2}\right)^{n_2} \sum_{i=0}^{n_2-1} \binom{n_2-1}{i} \left(\frac{x}{1-x}\right)^{i} \\
&= \left(\frac{1-x}{2}\right)^{n_2} \left(\frac{x}{1-x}+1\right)^{n_2-1} \\
&= \frac{1-x}{2^{n_2}}.
\end{aligned}
$$

Now, when the b comes, the automaton will go to q_3 if it was in q_1, so $p_{a^{n_2}b}(q_3) = p_{a^{n_2}}(q_1) = x/2^{n_2}$. So, $p_{a^{n_2}ba^{n_3}b}(q_3)$ is equal to the probability that the automaton went to q_3 when the first b came, plus the probability that the automaton was in q_2 or q_5 before the first b and that it went to q_3 when the second b came, which adds up to

$$
p_{a^{n_2}ba^{n_3}b}(q_3) = \frac{x}{2^{n_2}} + (1 - \frac{1}{2^{n_2}})(\frac{x}{2^{n_3}}).
$$

Following this reasoning, we get that after the whole word is entered (note that a^{n_k} does not contribute to the probability of ending up in q_3 since no b comes after it), the probability of being at q_3 is

$$
\boxed{p(q_3) = x \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}).}
$$

Also,

$$
\boxed{p(q_1) = x \sum_{i=2}^{k} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}).}
$$

Similarly,

$$
\boxed{p(q_6) = (1-x) \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}),}
$$

and

$$
\boxed{p(q_4) = (1-x) \sum_{i=2}^{k} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}).}
$$

From all of the above, we can write that the value of the automaton, given a word w of the form $a^{n_2}b \cdots ba^{n_k}$, is

$$
\boxed{val(w) = p(q_3) + p(q_5) = x \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) + x \prod_{i=2}^{k} (1 - \frac{1}{2^{n_i}}).} \tag{57}
$$

This expression can be simplified as follows. Note that

$$
\begin{aligned}
1 &= p(q_1) + p(q_2) + p(q_3) + p(q_4) + p(q_5) + p(q_6) \\
&= x \sum_{i=2}^{k} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) + (1-x) \prod_{i=2}^{k} (1 - \frac{1}{2^{n_i}}) \\
&+ x \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) + (1-x) \sum_{i=2}^{k} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) \\
&+ x \prod_{i=2}^{k} (1 - \frac{1}{2^{n_i}}) + (1-x) \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) \\
&= \sum_{i=2}^{k} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) + \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) + \prod_{i=2}^{k} (1 - \frac{1}{2^{n_i}}) \\
&= \frac{1}{2^{n_k}} \prod_{j=2}^{k-1} (1 - \frac{1}{2^{n_j}}) + 2 \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) + \prod_{i=2}^{k} (1 - \frac{1}{2^{n_i}}) \\
&= \prod_{i=2}^{k-1} (1 - \frac{1}{2^{n_i}}) + 2 \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}).
\end{aligned}
$$

So,

$$
2 \sum_{i=2}^{k-1} \frac{1}{2^{n_i}} \prod_{j=2}^{i-1} (1 - \frac{1}{2^{n_j}}) = 1 - \prod_{i=2}^{k-1} (1 - \frac{1}{2^{n_i}}).
$$

Substituting that last line into equation (57), we get

$$
val(w) = \frac{x}{2} + \frac{x}{2} (1 - \frac{1}{2^{n_k}}) \prod_{j=2}^{k} (1 - \frac{1}{2^{n_i}}),
$$

which implies that

$$
val(w) \leq x
$$

regardless of x. The conclusion is that the automaton shown in fig. 6 cannot get us a gap.

The failure of this design to get us a gap in the value might be rooted in the fact that a word could cause the automaton to begin in the upper branch and then "leak" into the lower branch or vice versa. The design analysed in the next section was made to address this particular problem.

C.3 Failed design 3

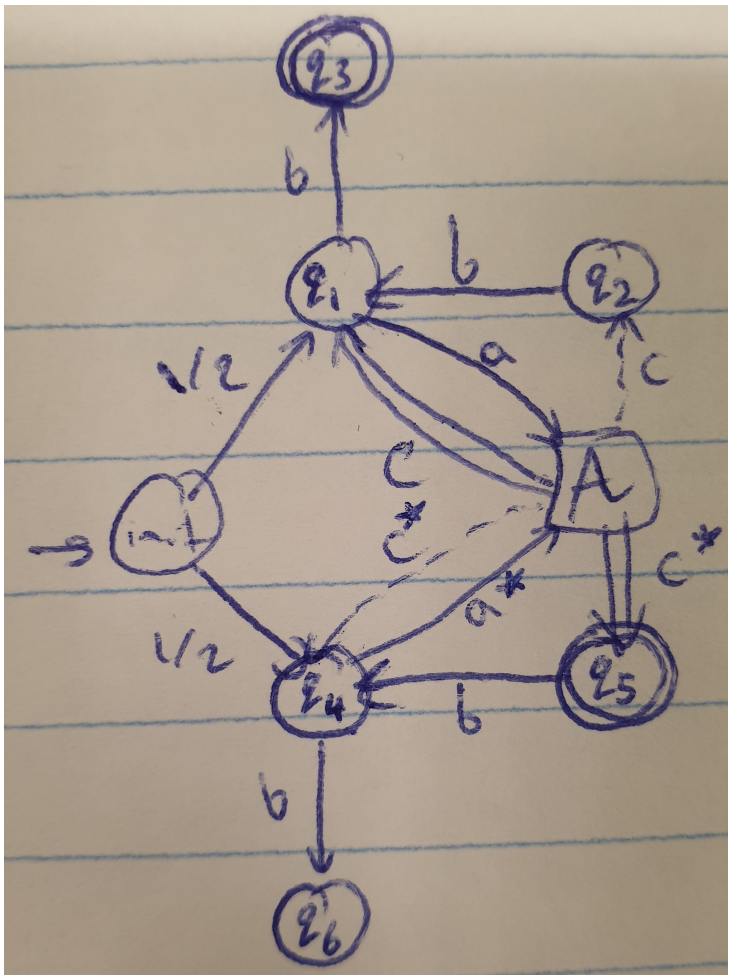


Figure 7

The failure in the design in the previous section – the absence of a gap in the value – was in part due to the fact that the branches could move between each other; there was a connection between them. The design in this section takes the first steps towards patching this problem.

We begin by modifying the automaton so that we use different entry and exit (into $\mathcal{A}$) symbols "a, c" and "a^*, c^*" for the upper and lower branches, respectively, as shown in fig. 7. The leakage problem is not fixed yet because the word "aw_Ac^*" could take the automaton from the upper branch to the lower branch.

Before proceeding further, let us consider exactly how the leakage problem harms our ability to prove uncomputability. Note first that, using the automaton in fig. 7, we can still get value 1 when $L_{\lambda>1/2}$ is not empty. This can be achieved using a word of the form

$$"[(aw_Ac)(a^*w_Ac^*)]^{n_2}b[(aw_Ac)(a^*w_Ac^*)]^{n_3}b\cdots b[(aw_Ac)(a^*w_Ac^*)]^{n_k}",$$

because this word causes the automaton to behave exactly like the automaton in fig. 1, which we know can achieve value 1. We do not get a gap, however, because, if $L_{\lambda>1/2}$ is empty we can use the word "$a^*w_Abbc^*$" to get value 1. So we have no gap from which we can get uncomputability. The value is always 1.

Let's think about what the word "$a^*w_Abbc^*$" makes the automaton do. Suppose the automaton was in init, then slid up to q_1. When a^*w_A is entered the automaton remains in q_1. Then when bb is entered, the automaton accepts. Now suppose the automaton was in init, then slid down to q_4. The automaton just acts normal, and repeating the same word for infinity gets the automaton to accept. The problem

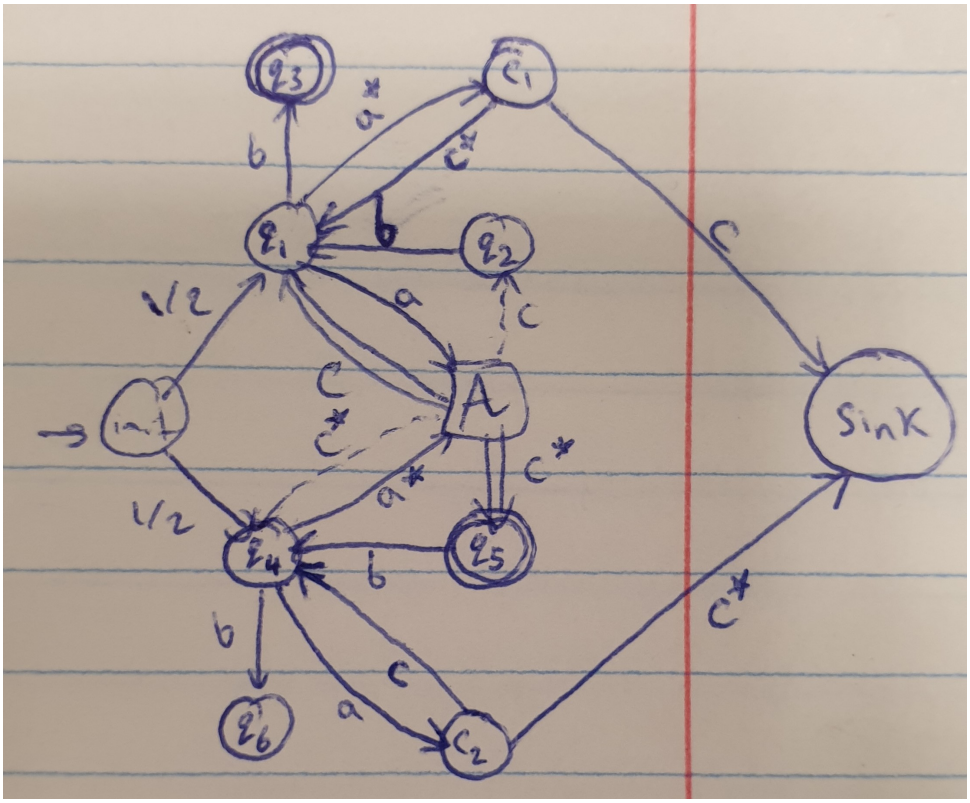


Figure 8

can be stated as follows: when one of the branches is operating, the other branch can just do whatever it wants without affecting the other branch. The branches are decoupled, which is bad because it gives the automaton the ability to always accept. This particular hole can be fixed by adding the states c_1 and c_2, and the Sink as shown in figure 8.

By making sure that the upper branch is debilitated when the lower branch is operating, the decoupling problem is fixed. And by making sure that leakages from one branch to another are punished, other problems are fixed. Generally speaking, by patching the automaton such that the presence of a connection between the upper and lower branches cannot be exploited to get value more than $1/2$ in the case $L_{\lambda>1/2}$ empty, we eventually arrive at the automaton shown in fig. 2, which we prove to have value $1/2$ when $L_{\lambda>1/2}$ is empty.

Colophon

This document was typeset using LaTeX. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from AndrÃľ Miede.