

A maximal entropy stochastic process for a timed automaton

Nicolas Basset

▶ **To cite this version:**

Nicolas Basset. A maximal entropy stochastic process for a timed automaton. ICALP 2013, Jul 2013, Riga, Latvia. pp.61-73, 10.1007/978-3-642-39212-2_9 . hal-00808909v2

HAL Id: hal-00808909
https://hal.science/hal-00808909v2

Submitted on 22 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A maximal entropy stochastic process for a timed automaton* **

Nicolas Basset[1,2]

[1] LIGM, University Paris-Est Marne-la-Vallée and CNRS, France.
[2] LIAFA, University Paris Diderot and CNRS, France
nbasset@liafa.univ-paris-diderot.fr

Abstract. Several ways of assigning probabilities to runs of timed automata (TA) have been proposed recently. When only the TA is given, a relevant question is to design a probability distribution which represents in the best possible way the runs of the TA. This question does not seem to have been studied yet. We give an answer to it using a maximal entropy approach. We introduce our variant of stochastic model, the stochastic process over runs which permits to simulate random runs of any given length with a linear number of atomic operations. We adapt the notion of Shannon (continuous) entropy to such processes. Our main contribution is an explicit formula defining a process Y^* which maximizes the entropy. This formula is an adaptation of the so-called Shannon-Parry measure to the timed automata setting. The process Y^* has the nice property to be ergodic. As a consequence it has the asymptotic equipartition property and thus the random sampling w.r.t. Y^* is quasi uniform.

1 Introduction

Timed automata (TA) were introduced in the early 90's by Alur and Dill [4] and then extensively studied, to model and verify the behaviours of real-time systems. In this context of verification, several probability settings have been added to TA (see references below). There are several reasons to add probabilities: this permits (i) to reflect in a better way physical systems which behave randomly, (ii) to reduce the size of the model by pruning the behaviors of null probability [9], (iii) to resolve undeterminism when dealing with parallel composition [18,19].

In most of previous works on the subject (see e.g. [13,2,14,18]), probability distributions on continuous and discrete transitions are given at the same time as the timed settings. In these works, the choice of the probability functions is left to the designer of the model. Whereas, she or he may want to provide only the TA and ask the following question: what is the "best" choice of the probability functions according to the TA given? Such a "best" choice must transform the

* An extended version of the present paper containing detailed proofs and examples is available on-line http://hal.archives-ouvertes.fr/hal-00808909.

** The support of Agence Nationale de la Recherche under the project EQINOCS (ANR-11-BS02-004) is gratefully acknowledged.

TA into a random generator of runs the least biased as possible, i.e it should generate the runs as uniformly as possible to cover with high probability the maximum of behaviours of the modeled system. More precisely the probability for a generated run to fall in a set should be proportional to the size (volume) of this set (see [19] for a same requirement in the context of job-shop scheduling). We formalize this question and propose an answer based on the notion of entropy of TA introduced in [7].

The theory developed by Shannon [25] and his followers permits to solve the analogous problem of quasi-uniform path generation in a finite graph. This problem can be formulated as follows: given a finite graph G, how can one find a stationary Markov-chain on G which allows one to generate the paths in the most uniform manner? The answer is in two steps (see Chapter 1.8 of [22] and also section 13.3 of [21]): (i) There exists a stationary Markov-chain on G with maximal entropy, the so called Shannon-Parry Markov-chain; (ii) This stationary Markov-chain allows to generate paths quasi uniformly.

In this article we lift this theory to the timed automata setting. We work with timed region graphs which are to timed automata what finite directed graphs are to finite state automata i.e. automata without labeling on edges and without initial and final states. We define stochastic processes over runs of timed region graphs (SPOR) and their (continuous) entropy. This generalization of Markov-chains for TA has its own interest, it is up to our knowledge the first one which provides a continuous probability distribution on starting states. Such a SPOR permits to generate step by step random runs. As a main result we describe a maximal entropy SPOR which is stationary and ergodic and which generalizes the Shannon-Parry Markov-chain to TA (Theorem 4). Concepts of maximal entropy, stationarity and ergodicity can be interesting by themselves, here we use them as the key hypotheses to ensure a quasi uniform sampling (Theorem 5). More precisely the result we prove is a variant of the so called Shannon-McMillan-Breiman theorem also known as asymptotic equipartition property (AEP).

Potential applications. There are two kind of probabilistic model checking: (i) the almost sure model checking aiming to decide if a model satisfies a formula with probability one (e.g. [16,3]); (ii) the quantitative (probabilistic) model checking (e.g. [14,18]) aiming to compare the probability of a formula to be satisfied with some given threshold or to estimate directly this probability.

A first expected application of our results would be a "proportional" model checking. The inputs of the problem are: a timed region graph $\mathcal{G}$, a formula φ, a threshold $\theta \in [0,1]$. The question is whether the proportion of runs of $\mathcal{G}$ which satisfy φ is greater than θ or not. A recipe to address this problem would be as follows: (i) take as a probabilistic model $\mathcal{M}$ the timed region graph $\mathcal{G}$ together with the maximum entropy SPOR Y^* defined in our main theorem; (ii) run a quantitative (probabilistic) model checking algorithm on the inputs $\mathcal{M}$, φ, θ (the output of the algorithm is yes or no whether $\mathcal{M}$ satisfies φ with a probability greater than θ or not) (iii) use the same output for the proportional model checking problem.

A random simulation with a linear number of operations wrt. the length of the run can be achieved with our probabilistic setting. It would be interesting to incorporate the simulation of our maximal entropy process in a statistical model checking algorithms. Indeed random simulation is at the heart of such kind of quantitative model checking (see [18] and reference therein).

The concepts handled in this article such as stationary stochastic processes and their entropy, AEP, etc. come from information and coding theory (see [17]). Our work can be a basis for the probabilistic counterpart of the timed channel coding theory we have proposed in [5]. Another application in the same flavour would be a compression method of timed words accepted by a given deterministic TA.

Related work. As mentioned above, this work generalizes the Shannon-Parry theory to the TA setting. Up to our knowledge this is the first time that a maximal entropy approach is used in the context of quantitative analysis of real-time systems.

Our models of stochastic real-time system can be related to numerous previous works. Almost-sure model checking for probabilistic real-time systems based on generalized semi Markov processes GSMPs was presented in [3] at the same time as the timed automata theory and by the same authors. This work was followed by [2,13] which address the problem of quantitative model checking for GSMPs under restricted hypotheses. The GSMPs have several differences with TA; roughly they behave as follows: in each location, clocks decrease until a clock is null, at this moment an action corresponding to this clock is fired, the other clocks are either reset, unchanged or purely canceled. Our probability setting is more inspired by [9,16,18] where probability densities are added directly on the TA. Here we add the new feature of an initial probability density function on states.

In [18], a probability distribution on the runs of a network of priced timed automaton is implicitly defined by a race between the components, each of them having its own probability. This allows a simulation of random runs in a non deterministic structure without state space explosion. There is no reason that the probability obtained approximates uniformness and thus it is quite incomparable to our objective.

Our techniques are based on the pioneering articles [7,8] on entropy of regular timed languages. In the latter article and in [5], an interpretation of the entropy of a timed language as information measure of the language was given.

Paper structure In section 2 we recall the theory of maximal entropy Markov-chain on finite graph. The aim of the rest of the paper is to lift results of this section to the timed setting. In section 3 we introduce stochastic processes over runs (SPOR) of a timed region graph is the timed analogous of Markov-chains on a finite graphs. We also gives definition of entropies of these dense objects inspired by [25] for the processes and by [7] for the timed region graph. In the main section (section 4), after giving the technical assumptions, we state the two main theorems: the existence of the maximal entropy SPOR which is ergodic and the resulted asymptotic equipartition properties. In section 5 we treat two

examples. We briefly discuss the computability issues and the perspectives in the conclusion (section 7).

Extended version The present paper is an extended version of the conference paper [10].

2 Maximal entropy Markov-chain on a graph

2.1 Markov-chain on a graph

A graph is defined by a finite set of states Q and a set of transitions Δ. Any transition $\delta \in \Delta$ has a starting state $\delta^- \in Q$ and an ending state $\delta^+ \in Q$ (there can be several transitions between the same two states).

A *path* $\delta_1 \cdots \delta_n \in \Delta^*$ is a word of consecutive transitions (${\delta_{i+1}}^- = {\delta_i}^+$, $i = 1, \ldots, n-1$). We denote by $\texttt{Path}_n(G)$ the set of paths of length n.

A Markov-chain on a graph G is given by

- initial state probabilities $p_0(q)$ for $q \in Q$ such that $\sum_{q \in Q} p_0(q) = 1$;
- conditional probabilities on transitions $p(\delta|\delta^-)$ such that for all $q \in Q$, $\sum_{\delta|\delta^- = q} p(\delta|q) = 1$ (and such that $p(\delta|q) = 0$ if $q \neq \delta^-$).

The following chain rule (1) defines a probability distribution p_n on $\texttt{Path}_n$ i.e. $\sum_{\pi \in \texttt{Path}_n} p_n(\pi) = 1$

$$p_n(\delta_1 \cdots \delta_n) = p_0({\delta_1}^-) p(\delta_1|{\delta_1}^-) \ldots p(\delta_n|{\delta_n}^-). \tag{1}$$

We also denote by p_n the induced probability measure on $\texttt{Path}_n$ i.e. for $A \subseteq \texttt{Path}_n$, $p_n(A) = \sum_{\pi \in A} p_n(A)$.

The conditional probabilities and the initial probabilities are respectively represented by a $Q \times Q$ stochastic matrix P and a row vector $\boldsymbol{p}_0$ such that:

$$P_{i,j} = \sum_{\delta|\delta^- = i, \delta^+ = j} p(\delta|i).$$

The stochastic process associated to a Markov-chain is *stationary* iff $\boldsymbol{p}_0 P = \boldsymbol{p}_0$. Moreover if P is irreducible (i.e. for all $i, j \in Q$, there exists $k \in \mathbb{N}$ such that $P^k_{i,j} > 0$) then the Markov-chain satisfies the ergodic property defines as follows:

Ergodicity. Given a set of infinite path $R \subseteq \Delta^\omega$ and $i, j \in \mathbb{N}$, we denote by $R_i^{i+j} \subseteq \Delta^{j+1}$ the set of finite path $\delta_i \cdots \delta_{i+j}$ that can occur between indices i and $i+j$ in an infinite path $(\delta_k)_{k \in \mathbb{N}}$ of R. Let p be a *stationary* Markoc chain then the sequence $p_n(R_0^{n-1})$ decreases and converges to a value called the probability of R and denoted by $p_\infty(R) = \lim_{n \to \infty} p_n(R_0^{n-1})$. The set R is *shift invariant* if for every $i, n \in \mathbb{N}$: $R_i^{i+n} = R_0^n$. The Markov-chain satisfies the *ergodic* property whenever it is stationary and every shift invariant set has probability 0 or 1. Definition of ergodicity for general probability measures can be found in [15].

2.2 Entropies

There are different notions of entropies. Their mutual connexion and their meanings are discussed in the rest of this section. Here we only summarize the definitions and propositions we lift to the timed setting. We refer the reader to [22,21,17] for more explanations about notions of Markov-chain, entropies, almost equipartition properties...

Proposition-definition 1 *Given a finite graph G, the following limit exists and defined the entropy of G:*

$$h(G) = \lim_{n\to\infty} \frac{1}{n} \log_2(|\mathtt{Path}_n(G)|).$$

Proposition-definition 2 *Let p be a stationary Markov-chain on a finite graph G, then*

$$-\frac{1}{n} \sum_{\pi\in\mathtt{Path}_n(G)} p_n(\pi) \log_2 p_n(\pi) \to_{n\to\infty} -\sum_{q\in Q} p_0(q) \sum_{\delta\in\Delta} p(\delta|q) \log_2 p(\delta|q).$$

This limit is called the entropy (rate) of the markov chain, denoted by $h(p)$.

2.3 The Shannon-McMillan-Breiman theorem

This classical theorem also known as asymptotic equipartition property (AEP) states that for Markov-chain satisfying the ergodic property almost every paths of a length n (for n large enough) have approximately the same probability which is roughly $2^{-nh(p)}$ where $h(p)$ is the entropy of the Markov-chain considered.

Theorem 1 (AEP for Markov chain). *If p is a Markov-chain satisfying the ergodic property then*

$$p_\infty[\{\delta_1\delta_2\cdots \mid -(1/n)\log_2 p_n(\delta_1\cdots\delta_n) \to_{n\to+\infty} h(p)\}] = 1$$

This theorem applied to a chain p^* such that $h(p^*) = h(G)$ means that long paths have a high probability to have a quasi uniform probability:

$$p^*_n(\delta_1\cdots\delta_n) \approx 2^{-nh(p^*)} \approx 1/|\mathtt{Path}_n(G)| \ (H(p^*) = h(g)).$$

2.4 The Shannon-Parry Markov-chain

In fact there exists a unique Markov-chain p^* such that $h(p^*) = h(G)$, the Shannon-Parry Markov-chain [25,23]:

Theorem 2 (Shannon-Parry). *If G is strongly connected then*

- *every stationary Markov-chain p on G satisfies $h(p) \leq h(G)$;*
- *there exists a unique stationary Markov-chain p^* such that $h(p^*) = h(G)$;*
- *p^* is ergodic.*

To define the Markov-chain p^* we need to remind the classical Perron-Frobenius theorem (see for instance [22]).

The *spectral radius* of a matrix is the maximal modulus of its eigenvalues.

Theorem 3 (Perron-Frobenius). *If M is the adjacency matrix of a strongly connected graph G (i.e. M has positive entries and is irreducible) then*

- *The spectral radius ρ of M is a simple eigenvalue of M;*
- *there exists a right (resp. left) positive eigenvector v (resp. w) of M for the eigenvalue ρ;*
- *any non-negative right (resp. left) eigenvector of M is collinear to v (resp. w).*

Proposition 1. *The entropy of G and the spectral radius ρ of its adjacency matrix are linked by the following equality $h(G) = log_2(\rho)$.*

Given a strongly connected graph G, let ρ, v, w be given by the Perron-Frobenius theorem above. Eigenvectors v and w are chosen such that $\langle v, w\rangle = \sum_{q\in Q} v_q w_q = 1$ (eigenvectors are defined up to a scalar constant). The Shannon-Parry Markov-chain p^* on G is given by: for every $q \in Q$, $\delta \in \Delta$,

$$p_0^*(q) = v_q w_q; \quad p^*(\delta|\delta^-) = \frac{v_{\delta^+}}{\rho v_{\delta^-}}. \tag{2}$$

The corresponding transition probability matrix is defined by: for every $i, j \in Q$, $P_{i,j} = M_{i,j} v_j / \rho v_i$.

3 Stochastic processes on timed region graphs

3.1 Timed graphs and their runs

In this section we define a timed region graph which is the underlying structure of a timed automaton [4]. For technical reasons we consider only timed region graphs with bounded clocks. We will justify this assumption in section 4.1.

Timed region graphs. Let X be a finite set of variables called *clocks*. Clocks have non-negative values bounded by a constant M. A *rectangular constraint* has the form $x \sim c$ where $\sim \in \{\leq, <, =, >, \geq\}$, $x \in X$, $c \in \mathbb{N}$. A *diagonal constraint* has the form $x - y \sim c$ where $x, y \in X$. A guard is a finite conjunction of rectangular constraints. A *zone* is a set of clock vectors $\mathbf{x} \in [0, M]^X$ satisfying a finite conjunction of rectangular and diagonal constraints. A *region* is a zone which is minimal for inclusion (e.g. the set of points (x_1, x_2, x_3, x_4) which satisfy the constraints $0 = x_2 < x_3 - 4 = x_4 - 3 < x_1 - 2 < 1$). Regions of $[0, 1]^2$ are depicted in Fig. 3.1.

As we work by analogy with finite graphs, we introduce timed region graphs which are roughly timed automata without labels on transitions and without initial and final states. Moreover we consider a state space decomposed in regions. Such a decomposition in regions are quite standard for timed automata and does not affect their behaviours (see e.g. [14,7]).

A timed region graph is a tuple $(X, Q, \mathbb{S}, \Delta)$ such that

– X is a finite set of clocks.
– Q is a finite set of *locations*.
– $\mathbb{S}$ is the set of *states* which are couples of a location and a clock vector ($\mathbb{S} \subseteq Q \times [0, M]^X$). It admits a region decomposition $\mathbb{S} = \cup_{q \in Q} \{q\} \times \mathbf{r}_q$ where for each $q \in Q$, $\mathbf{r}_q$ is a region.
– Δ is a finite set of *transitions*. Any transition $\delta \in \Delta$ goes from a *starting location* $\delta^- \in Q$ to an *ending location* $\delta^+ \in Q$; it has a set $\mathfrak{r}(\delta)$ of clocks to reset when firing δ and a fleshy guard $\mathfrak{g}(\delta)$ to satisfy to fire it. Moreover, the set of clock vectors that satisfy $\mathfrak{g}(\delta)$ is projected on the region $\mathbf{r}_{\delta^+}$ when the clocks in $\mathfrak{r}(\delta)$ are resets.

Runs of the timed region graph. A *timed transition* is an element (t, δ) of $\mathbb{A} =_{def} [0, M] \times \Delta$. The *time delay* t represents the time before firing the transition δ.

Given a state $s = (q, \mathbf{x}) \in \mathbb{S}$ (i.e $\mathbf{x} \in \mathbf{r}_q$) and a timed transition $\alpha = (t, \delta) \in \mathbb{A}$ the *successor* of s by α is denoted by $s \triangleright \alpha$ and defined as follows. Let $\mathbf{x}'$ be the clock vector obtained from $\mathbf{x} + (t, \ldots, t)$ by resetting clocks in $\mathfrak{r}(\delta)$ ($x'_i = 0$ if $i \in \mathfrak{r}(\delta)$, $x'_i = x_i + t$ otherwise). If $\delta^- = q$ and $\mathbf{x} + (t, \ldots, t)$ satisfies the guard $\mathfrak{g}(\delta)$ then $\mathbf{x}' \in \mathbf{r}_{\delta^+}$ and $s \triangleright \alpha = (\delta^+, \mathbf{x}')$ else $s \triangleright \alpha = \perp$. Here and in the rest of the paper $\perp$ represents every undefined state.

We extend the successor action $\triangleright$ to words of timed transitions by induction: $s \triangleright \varepsilon = s$ and $s \triangleright (\alpha \boldsymbol{\alpha}') = (s \triangleright \alpha) \triangleright \boldsymbol{\alpha}'$ for all $s \in \mathbb{S},\ \alpha \in \mathbb{A},\ \boldsymbol{\alpha}' \in \mathbb{A}^*$.

A *run* of the timed region graph $\mathcal{G}$ is a word $s_0 \alpha_0 \cdots s_n \alpha_n \in (\mathbb{S} \times \mathbb{A})^{n+1}$ such that $s_{i+1} = s_i \triangleright \alpha_i \neq \perp$ for all $i \in \{0, \ldots, n-1\}$ and $s_n \triangleright \alpha_n \neq \perp$; its reduced version is $[s_0, \alpha_0 \ldots \alpha_n] \in \mathbb{S} \times \mathbb{A}^{n+1}$ (for all $i > 0$ the state s_i is determined by its preceding states and timed transition and thus is a redundant information). In the following we will use without distinction extended and reduced version of runs. We denote by $\mathcal{R}_n$ the set of runs of length n ($n \geq 1$).

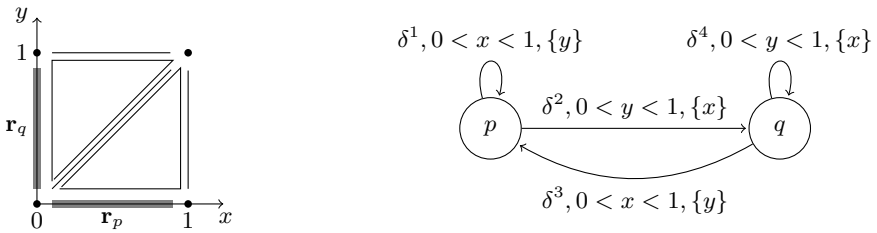


Fig. 1. The running example. Right: $\mathcal{G}^{\mathtt{ex1}}$; left: Its state space (in gray).

Example 1. Let $\mathcal{G}^{\mathtt{ex1}}$ be the timed region graph depicted on Fig. 3.1 with $\mathbf{r}_p$ and $\mathbf{r}_q$ the region described by the constraints $0 = y < x < 1$ and $0 = x < y < 1$ respectively. Successor action is defined by $[p, (x, 0)] \triangleright (t, \delta^1) = [p, (x + t, 0)]$ and

$[p,(x,0)] \triangleright (t,\delta^2) = [q,(0,t)]$ if $x+t<1$; $[q,(0,y)] \triangleright (t,\delta^3) = [p,(t,0)]$ and $[q,(0,y)] \triangleright (t,\delta^4) = [q,(0,y+t)]$ if $y+t<1$. An example of run of $\mathcal{G}^{\texttt{ex1}}$ is $(p,(0.5,0))(0.4,\delta^1)(p,(0.9,0))(0.8,\delta^2)(q,(0,0.8))(0.1,\delta^3)(p,(0.1,0))$.

Integrating over states and runs; volume of runs. It is well known (see [4]) that a region is uniquely described by the integer parts of clocks and by an order on their fractional parts, e.g. in the region $\mathbf{r}^{\text{ex}}$ given by the constraints $0 = x_2 < x_3 - 4 = x_4 - 3 < x_1 - 2 < 1$, the integer parts are $\lfloor x_1 \rfloor = 2, \lfloor x_2 \rfloor = 0, \lfloor x_3 \rfloor = 4, \lfloor x_4 \rfloor = 3$ and fractional parts are ordered as follows $0 = \{x_2\} < \{x_3\} = \{x_4\} < \{x_1\} < 1$. We denote by $\gamma_1 < \gamma_2 < \cdots < \gamma_d$ the fractional parts different from 0 of clocks of a region $\mathbf{r}_q$ (d is called the dimension of the region). In our example the dimension of $\mathbf{r}^{\text{ex}}$ is 2 and $(\gamma_1,\gamma_2) = (x_3 - 4, x_1 - 2)$. We denote by $\mathit{\Gamma}_q$ the simplex $\mathit{\Gamma}_q = \{\boldsymbol{\gamma} \in \mathbb{R}^d \mid 0 < \gamma_1 < \gamma_2 < \cdots < \gamma_d < 1\}$. The mapping $\phi_{\mathbf{r}} : \mathbf{x} \mapsto \gamma$ is a natural bijection from the d dimensional region $\mathbf{r} \subset \mathbb{R}^{|X|}$ to $\mathit{\Gamma}_q \subset \mathbb{R}^d$. In the example the pre-image of a vector (γ_1,γ_2) is $(\gamma_2+2, 0, \gamma_1+4, \gamma_1+3)$.

Example 2 (Continuing example 1). The region $\mathbf{r}_p = \{(x,y) \mid 0 = y < x < 1\}$ is 1-dimensional, $\phi_{\mathbf{r}_p}(x,y) = x$ and $\phi^{-1}_{\mathbf{r}_p}(\gamma) = (\gamma, 0)$.

Now, we introduce simplified notation for sums of integrals over states, transitions and runs. We define the integral of an integrable[3] function $f : \mathbb{S} \to \mathbb{R}$ (over states):

$$\int_{\mathbb{S}} f(s)ds = \sum_{q \in Q} \int_{\mathit{\Gamma}_q} f(q, \phi^{-1}_{\mathbf{r}_q}(\boldsymbol{\gamma}))d\boldsymbol{\gamma}.$$

where $\int .d\boldsymbol{\gamma}$ is the usual integral (w.r.t. the Lebesgue measure). We define the integral of an integrable function $f : \mathbb{A} \to \mathbb{R}$ (over timed transitions):

$$\int_{\mathbb{A}} f(\alpha)d\alpha = \sum_{\delta \in \mathit{\Delta}} \int_{[0,M]} f(t,\delta)dt$$

and the integral of an integrable function $f : \mathcal{R}_n \to \mathbb{R}$ (over runs) with the convention that $f[s,\boldsymbol{\alpha}] = 0$ if $s \triangleright \alpha = \bot$:

$$\int_{\mathcal{R}_n} f[s,\boldsymbol{\alpha}]d[s,\boldsymbol{\alpha}] = \int_{\mathbb{S}} \int_{\mathbb{A}} \dots \int_{\mathbb{A}} f[s,\boldsymbol{\alpha}]d\alpha_1 \dots d\alpha_n ds$$

To summarize, we take finite sums over finite discrete sets Q, $\mathit{\Delta}$ and take integrals over dense sets $\mathit{\Gamma}_q$, $[0,M]$. More precisely, all the integrals we define have their corresponding measures[4] which are products of counting measures on discrete sets $\mathit{\Sigma}$, Q and Lebesgue measure over subsets of $\mathbb{R}^m$ for some $m \geq 0$

[3] A function $f : \mathbb{S} \to \mathbb{R}$ is integrable if for each $q \in Q$ the function $\boldsymbol{\gamma} \mapsto f(q, \phi^{-1}_{\mathbf{r}_q}(\boldsymbol{\gamma}))$ is Lebesgue integrable. A function $f : \mathbb{A} \to \mathbb{R}$ is integrable if for each $\delta \in \mathit{\Delta}$ the function $t \mapsto f(t,\delta)$ is Lebesgue integrable.

[4] We refer the reader to [15] for an introduction to measure and probability theory.

(e.g. Γ_q, $[0, M]$). We denote by $\mathfrak{B}(\mathbb{S})$ (resp. $\mathfrak{B}(\mathbb{A})$) the set of measurable subsets of $\mathbb{S}$ (resp. $\mathbb{A}$).

The volume of the set of n-length runs is defined by:

$$\texttt{Vol}(\mathcal{R}_n) = \int_{\mathcal{R}_n} 1d[s, \boldsymbol{\alpha}] = \int_{\mathbb{S}} \int_{\mathbb{A}^n} \mathbf{1}_{s \triangleright \boldsymbol{\alpha} \neq \perp} d\boldsymbol{\alpha} ds$$

Remark 1. The use of reduced version of runs is crucial when dealing with integrals (and densities in the following). Indeed the following integral on the extended version of runs is always null since variables are linked ($s_{i+1} = s_i \triangleright \alpha_i$ for $i = 0..n-2$): $\int_{\mathbb{A}} \int_{\mathbb{S}} \dots \int_{\mathbb{A}} \int_{\mathbb{S}} 1_{s_0 \alpha_0 \cdots s_{n-1} \alpha_{n-1} \in \mathcal{R}_n} ds_0 d\alpha_0 \dots ds_{n-1} d\alpha_{n-1} = 0$.

3.2 SPOR on timed region graphs

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *stochastic process over runs* (SPOR) of a timed region graph $\mathcal{G}$ is a sequence of random variables $(Y_n)_{n \in \mathbb{N}} = (S_n, A_n)_{n \in \mathbb{N}}$ such that:

C.1) For all $n \in \mathbb{N}$, $S_n : (\Omega, \mathcal{F}, P) \to (\mathbb{S}, \mathfrak{B}(\mathbb{S}))$ and $A_n : (\Omega, \mathcal{F}, P) \to (\mathbb{A}, \mathfrak{B}(\mathbb{A}))$.

C.2) The initial state S_0 has a probability density function (PDF) $p_0 : \mathbb{S} \to \mathbb{R}^+$ i.e. for every $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$, $P(S_0 \in \mathcal{S}) = \int_{s \in \mathcal{S}} p_0(s) ds$ (in particular $P(S_0 \in \mathbb{S}) = \int_{s \in \mathbb{S}} p_0(s) ds = 1$).

C.3) Probability on every timed transition only depends on the current state: for every $n \in \mathbb{N}$, $\mathcal{A} \in \mathfrak{B}(\mathbb{A})$, for almost every[5] $s \in \mathbb{S}$, $y_0 \cdots y_n \in (\mathbb{S} \times \mathbb{A})^n$,

$$P(A_n \in \mathcal{A} | S_n = s, Y_n = y_n, \dots, Y_0 = y_0) = P(A_n \in \mathcal{A} | S_n = s),$$

moreover this probability is given by a conditional PDF $p(.|s) : \mathbb{A} \to \mathbb{R}^+$ such that $P(A_n \in \mathcal{A} | S_n = s) = \int_{\alpha \in \mathcal{A}} p(\alpha|s) d\alpha$ and $p(\alpha|s) = 0$ if $s \triangleright \alpha = \perp$ (in particular $P(A_n \in \mathbb{A} | S_n = s) = \int_{\alpha \in \mathbb{A}} p(\alpha|s) d\alpha = 1$).

C.4) States are updated deterministically knowing the previous state and transition: $S_{n+1} = S_n \triangleright A_n$.

The Markovian properties C.3) and C.4) permits to define probability density functions for portion of runs $Y_i \cdots Y_{i+n-1}$ knowing the value of S_i : for $\boldsymbol{\alpha} \in \mathbb{A}^n$ and $s_0 \in \mathbb{S}$ we defined $p_n(\boldsymbol{\alpha}|s_0)$ by the following chain rule

$$p_n(\boldsymbol{\alpha}|s_0) = p(\alpha_0|s_0) p(\alpha_1|s_1) \dots p(\alpha_{n-1}|s_{n-1}). \tag{3}$$

where for each $j = 1..n-1$ the state updates are defined by $s_j = s_{j-1} \triangleright \alpha_{j-1}$; then $p_n(.|s)$ is satisfies

$$P((S_i, A_i) \cdots (S_{i+n-1}, A_{i+n-1}) \in R | S_i = s) = \int_{\mathbb{A}^n} p_n(\boldsymbol{\alpha}|s) 1_{[s, \boldsymbol{\alpha}] \in R} d\boldsymbol{\alpha}.$$

[5] A property *prop* (like "f is positive", "well defined"...) on a set B holds *almost everywhere* when the set where it is false has measure (volume) 0: $\int_B \mathbf{1}_{b \not\models prop} db = 0$.

The PDF for $Y_0 \cdots Y_{n-1}$ is $p_n(s, \boldsymbol{\alpha}) =_{def} p_0(s)p_n(\boldsymbol{\alpha}|s)$ i.e

$$P(Y_0 \cdots Y_{n-1} \in R) = \int_{\mathcal{R}_n} p_n[s, \boldsymbol{\alpha}] 1_{[s,\boldsymbol{\alpha}]\in R} d[s, \boldsymbol{\alpha}].$$

The SPOR $(Y_n)_{n\in\mathbb{N}}$ is called *stationary* whenever for all $i, n \in \mathbb{N}$, $Y_i \cdots Y_{i+n-1}$ has the same PDF as $Y_0 \cdots Y_{n-1}$ which is p_n.

Probability measure on infinite runs and ergodicity. Given a measurable sets of runs $R \in \mathfrak{B}[(\mathbb{S} \times \mathbb{A})^{n+1}]$ with $n \geq 0$, one can extend it into a set of bi-infinite runs $R_\infty \subseteq (\mathbb{S} \times \mathbb{A})^{\mathbb{Z}}$ as follows: $R_\infty = \{(y_i)_{i\in\mathbb{Z}} \in (\mathbb{S} \times \mathbb{A})^{\mathbb{Z}} | y_0 \cdots y_n \in R\}$.

Let σ be the shift map on bi-infinite runs i.e. $\sigma((y_i)_{i\in\mathbb{Z}}) = (y'_i)_{i\in\mathbb{Z}}$ with $y'_i = y_{i-1}$.

A probability measure on μ is called shift invariant if $\mu(\sigma(A)) = \mu(A)$ for every μ-measurable set of bi-infinite runs A.

Let Y be a *stationary* SPOR then by a classical extension theorem due to Kolmogorov one can define a shift invariant measure P_Y of bi-infinite runs such that $P_Y(R_\infty) = P(Y_0 \cdots Y_n \in R)$.

We characterize in (4) below, the probability $P_Y(A)$ of a shift invariant set A (i.e. $\sigma(A) = A$) in terms of probability on its finite factors. For every shift invariant set of bi-infinite runs A and $n \in \mathbb{N}$ we denote by A_n the set of finite factor of runs of A of length n: $A_n = \{y_0 \cdots y_n | (y_i)_{i\in\mathbb{Z}} \in A\}$. The set A is P_Y measurable iff $A_n \in \mathfrak{B}[(\mathbb{S} \times \mathbb{A})^{n+1}]$ for every n and in that case the following holds

$$P_Y(A) = \lim_{n\to+\infty} P(Y_0 \cdots Y_n \in A_n). \tag{4}$$

A stochastic process Y is *ergodic* whenever it is stationary and every shift-invariant measurable set A has probability $P_Y(A)$ equal to 0 or 1. Definition of ergodicity for general probability measures can be found in [15].

In the following $E_Y(.)$ denote the expectation w.r.t. P_Y (for a stationary SPOR Y):

$$E_Y(f) = \int_{y\in(\mathbb{S}\times\mathbb{A})^{\mathbb{Z}}} f(y) dP_Y(y). \tag{5}$$

When f depends only on a finite number of coordinates then $E_Y(f)$ is conveniently describe with the PDFs p_n. Two examples that appears in Proposition-definition 4 below are:

$$E_Y(\log p_n[S_0, A_0 \cdots A_n]) = \int_{\mathcal{R}_n} p_n[s, \boldsymbol{\alpha}] \log_2 p_n[s, \boldsymbol{\alpha}] d[s, \boldsymbol{\alpha}] \tag{6}$$

$$E_Y(\log p(A_0|S_0)) = \int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log_2 p(\alpha|s) d\alpha ds \tag{7}$$

Simulation according to a SPOR. Given a SPOR Y, a run $(s_0, \boldsymbol{\alpha}) \in \mathcal{R}_n$ can be chosen randomly w.r.t. Y with a linear number of the following operations: random pick according to p_0 or $p(.|s)$ and computing of a successor. Indeed it suffices to pick s_0 according to p_0 and for $i = 0..n-1$ to pick α_i according to $p(.|s_i)$ and to make the update $s_{i+1} = s_i \triangleright \alpha_i$.

3.3 Entropy

In this sub-section, we define entropy for timed region graphs and SPOR. The first one is inspired by [7] and the second one by [25].

Entropy of a timed region graph

Proposition-definition 3 *Given a timed region graph $\mathcal{G}$, the following limit exists and defined the entropy of $\mathcal{G}$:*

$$\mathcal{H}(\mathcal{G}) = \lim_{n\to\infty} \frac{1}{n} \log_2(\texttt{Vol}(\mathcal{R}_n)).$$

When $\mathcal{H}(\mathcal{G}) > -\infty$, the timed region graph is *thick*, the volume behaves w.r.t. n like an exponent: $\texttt{Vol}(\mathcal{R}_n) \approx 2^{n\mathcal{H}}$. When $\mathcal{H}(\mathcal{G}) = -\infty$, the timed region graph is *thin*, the volume decays faster than any exponent: $\forall \rho > 0,\ \texttt{Vol}(\mathcal{R}_n) << \rho^n$.

Entropy of a SPOR

Proposition-definition 4 *If Y is a stationary SPOR, then*

$$E_Y(-\log p_n[S_0, A_0 \cdots A_n])/n \to_{n\to\infty} E_Y(-\log p(A_0|S_0))$$

which can be re-written as

$$-\frac{1}{n}\int_{\mathcal{R}_n} p_n[s,\boldsymbol{\alpha}] \log_2 p_n[s,\boldsymbol{\alpha}] d[s,\boldsymbol{\alpha}] \to_{n\to\infty} -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log_2 p(\alpha|s) d\alpha ds.$$

This limit is called the entropy of Y, denoted by $H(Y)$.

Proof.

$$\begin{aligned} E_Y(-\log p_n[S_0, A_0 \cdots A_n])/n &= E[-\log p_0(S_0) \prod_{i=0}^{n} p(A_i|S_i)]/n \\ &= E_Y(-\log p_0(S_0))/n - \sum_{i=0}^{n} E[\log p(A_i|S_i)]/n \\ &= E_Y(-\log p_0(S_0))/n - E_Y(\log p(A_0|S_0)) \text{ by stationarity} \end{aligned}$$

This quantity tends to $E_Y(-\log p(A_0|S_0))$ when n tends to $+\infty$. □

Proposition 2. *Let $\mathcal{G}$ be a timed region graph and Y be a stationary SPOR on $\mathcal{G}$. Then the entropy of Y is upper bounded by that of $\mathcal{G}$: $H(Y) \leq \mathcal{H}(\mathcal{G})$.*

Proof. The proof follows from the following fact: for all $n \in \mathbb{N}$,

$$-\int_{\mathcal{R}_n} p_n[s,\boldsymbol{\alpha}] \log_2 p_n[s,\boldsymbol{\alpha}] d[s,\boldsymbol{\alpha}] \leq \log_2(\texttt{Vol}(\mathcal{R}_n)) \tag{8}$$

We need some definitions and properties concerning Kullback-Leibler divergence before proving this fact.

The Kullback-Leibler-divergence[6] (KL-divergence) from a PDF p_n to another p'_n is

$$D(p_n||p'_n) = \int_{\mathcal{R}_n} p_n[s,\boldsymbol{\alpha}] \log_2 \frac{p_n[s,\boldsymbol{\alpha}]}{p'_n[s,\boldsymbol{\alpha}]} d[s,\boldsymbol{\alpha}].$$

The KL-divergence is always positive with equality to 0 if and only if p_n and p'_n are equal almost everywhere (see e.g. [17] chapter 8). It permits to measure how far a probability distribution is from another.

Now we can prove (8). The KL-divergence from an arbitrary distribution p_n to the uniform distribution $[s,\boldsymbol{\alpha}] \mapsto 1/\mathtt{Vol}(\mathcal{R}_n)$ is $\log_2(\mathtt{Vol}(\mathcal{R}_n)) - h(p_n) \geq 0$ with equality if and only if p_n is uniform almost everywhere. □

The main contribution of this article is a construction of an ergodic SPOR Y^* for which the equality $H(Y^*) = \mathcal{H}(\mathcal{G})$ holds i.e. a timed analogue of the Shannon-Parry Markov Chain recalled in section 2.

4 Maximal entropy SPOR and quasi uniform sampling

In this section $\mathcal{G}$ is a timed region graph satisfying the technical condition below (section 4.1). We present an ergodic SPOR Y^* for which the upper bound on entropy is reached $H(Y^*) = \mathcal{H}(\mathcal{G})$ (Theorem 4).

4.1 Technical assumptions

In this section we explain and justify several technical assumptions on the timed region graph $\mathcal{G}$ we make in the following.

Bounded delays. If the delays were not bounded the sets of runs $\mathcal{R}_n$ would have infinite volumes and thus a quasi uniform random generation cannot be achieved.

Fleshy transitions. We consider timed region graphs whose transitions are *fleshy* [7]: there is no constraints of the form $x = c$ in their guards. Non fleshy transitions yield a null volume and are thus useless. Delete them reduces the size of the timed region graph considered and ensures that every path has a positive volume (see [7,12] for more justifications and details).

Strong connectivity of the set of locations. We will consider only timed region graph which are strongly connected i.e. locations are pairwise reachable. This condition (usual in the discrete case we generalize) is not restrictive since the set of locations can be decomposed in strongly connected components and then a maximal entropy SPOR can be designed for each components.

Thickness. In the maximal entropy approach we adopt, we need that the entropy is finite $\mathcal{H}(\mathcal{G}) > -\infty$. This is why we restrict our attention to *thick* timed

[6] this notion has a lot of names such as relative entropy, Kullback-Leibler distance, KLIC...

region graph. The dichotomy between thin and thick timed region graphs was characterized precisely in [12] where it appears that thin timed region graph are degenerate. The key characterization of thickness is the existence of a forgetful cycle [12]. When the locations are strongly connected, existence of such a forgetful cycle ensures that the state space $\mathbb{S}$ is strongly connect i.e. for all $s, s' \in \mathbb{S}$ there exists $\boldsymbol{\alpha} \in \mathbb{A}^*$ such that $s \triangleright \boldsymbol{\alpha} = s'$.

Weak progress cycle condition. In [7] the following assumption (known as the *progress cycle condition*) was made: for some positive integer constant D, on each path of D consecutive transitions, all the clocks are reset at least once.

Here we use a weaker condition: for a positive integer constant D, a timed region graph satisfies the *D weak progress condition* (D-WPC) if on each path of D consecutive transitions at most one clock is not reset during the entire path.

The timed region graph on Fig. 3.1 does not satisfy the progress cycle condition (e.g. x is not reset along δ^1) but satisfies the 1-WPC.

4.2 Main theorems

Here we give the two main theorems of the paper. There proofs are given in section 6.

Theorem 4. *There exists a positive real ρ and two functions $v, w : \mathbb{S} \mapsto \mathbb{R}$ positive almost everywhere such that the following equations define the PDF of an ergodic SPOR Y^* with maximal entropy: $H(Y^*) = \mathcal{H}(\mathcal{G})$.*

$$p_0^*(s) = w(s)v(s); \quad p^*(\alpha|s) = \frac{v(s \triangleright \alpha)}{\rho v(s)}. \tag{9}$$

Objects ρ, v, w are spectral attributes of an operator Ψ defined in the next section.

An ergodic SPOR satisfies an asymptotic equipartition property (AEP) (see [17] for classical AEP and [1] which deals with the case of non necessarily Markovian stochastic processes with density). Here we give our own AEP. It strongly relies on the pointwise ergodic theorem (see [15]) and on the Markovian property satisfied by every SPOR (conditions C.3 and C.4).

Theorem 5 (AEP for SPOR). *If Y is an ergodic SPOR then*

$$P_Y[\{s_0\alpha_0 s_1\alpha_1 \cdots \mid -(1/n)\log_2 p_n[s_0, \alpha_0 \cdots \alpha_n] \to_{n\to+\infty} H(Y)\}] = 1$$

This theorem applied to the maximal entropy SPOR Y^* means that long runs have a high probability to have a quasi uniform density:

$$p_n^*[s_0, \alpha_0 \cdots \alpha_n] \approx 2^{-nH(Y^*)} \approx 1/\mathtt{Vol}(\mathcal{R}_n) \text{ (since } H(Y^*) = H(\mathcal{G})).$$

4.3 The operator Ψ.

The maximal entropy SPOR is a lifting to the timed setting of the Shannon-Parry Markov-chain of a finite strongly connected graph. The definition of this

chain is based on the Perron-Frobenius theory applied to the adjacency matrix M of the graph. This theory ensures that there exists both a positive eigenvector v of M for the spectral radius[7] ρ (i.e. $Mv = \rho v$) and a positive eigenvector w of the transposed matrix $M^\top$ for ρ (i.e. $M^\top w = \rho w$). The initial probability distribution on the states Q of the Markov-chain is given by $p_i = v_i w_i$ for $i \in Q$ and the transition probability matrix P is given by $P_{ij} = v_j M_{ij}/(\rho v_i)$ for $i, j \in Q$. The timed analogue of M is the operator Ψ introduced in [7]. To define ρ,v and w, we will use the theory of positive linear operators (see e.g. [20]) instead of the Perron-Frobenius theory used in the discrete case.

The operator Ψ of a timed region graph is defined by:

$$\forall f \in L_2(\mathbb{S}),\ \forall s \in \mathbb{S},\ \Psi f(s) = \int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \text{ (with } f(\perp) = 0), \tag{10}$$

where $L_2(\mathbb{S})$ is the Hilbert space of square integrable functions from $\mathbb{S}$ to $\mathbb{R}$ with the scalar product $\langle f, g\rangle = \int_{\mathbb{S}} f(s)g(s)ds$ and associated norm $||f||_2 = \sqrt{\langle f, f\rangle}$.

Proposition 3. *The operator Ψ defined in (10) is a positive continuous linear operator on $L_2(\mathbb{S})$.*

Proof. We describe first a change of coordinates $(\gamma, t) \mapsto (\gamma', \sigma)$ which roughly permits to write the integral of a function evaluated on $s \triangleright \alpha$ by an integral of a function over states s'. This change of coordinate will be needed to pass from equation (12) to equation (13) below. For a real y we denote by $\{y\}$ its fractional part. Let $\delta \in \Delta$, d the dimension of $\mathbf{r}_{\delta^-}$ and d' the dimension of $\mathbf{r}_{\delta^+}$. Let t, γ, γ' such that $(\delta^-, \gamma) \triangleright (t, \delta) = (\delta^+, \gamma')$. Modulo a permutation of coordinate that only depends on δ we have $(\{\gamma_1 + t\}, \ldots \{\gamma_d + t\}, \{t\}) = (\gamma', \sigma)$ for some $\sigma \in [0, 1]^{d+1-d'}$. Indeed, the coordinate $\{t\}$ corresponds to clocks null in δ^- and not null in δ^+ if such a clock exists (it is a coordinate of σ otherwise); a coordinate γ_i of γ corresponding to a non resetting clock yields a new coordinate $\{\gamma_i + t\}$ of γ'; a coordinate γ_i of γ corresponding to a resetting clock yields a coordinate $\{\gamma_i + t\}$ of σ.

The change of coordinates $(\gamma, t) \mapsto (\gamma', \sigma)$ from the set $G(\delta) =_{def} \{(\gamma, t) \in \Gamma_{\delta^-} \times [0, M] \mid (\delta^-, \gamma) \triangleright (t, \delta) \neq \perp\}$ to its image denoted by $G'(\delta)$ is linear with a Jacobian equals to 1.

Now we prove the following inequality for every $f \in L_2(\mathbb{S})$

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \leq |\Delta| \int_{s'} f(s')^2 ds' = |\Delta|.||f||_2 \tag{11}$$

By definition:

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \int_{(\gamma,t) \in G(\delta)} f((\delta^-, \gamma) \triangleright (t, \delta))^2 d\gamma dt \tag{12}$$

[7] Recall from linear algebra (resp. spectral theory) that the spectrum of a matrix (resp. of an operator) Ψ is the set $\{\lambda \in \mathbb{C} \text{ s.t. } \Psi - \lambda Id \text{ is not invertible.}\}$. The spectral radius ρ of Ψ is the radius of the smallest disc centered in 0 which contains all the spectrum.

which after the change of coordinates $(\gamma, t) \mapsto (\gamma', \sigma)$ yields:

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \int_{(\gamma', \sigma) \in G'(\delta)} f(\delta^+, \gamma')^2 d\gamma' d\sigma \tag{13}$$

If we denote by $g_\delta(\gamma') = \texttt{Vol}(\{\sigma \mid (\gamma', \sigma) \in G'(\delta)\})$ we can simplify the last integral:

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \int_{\gamma' \in \Gamma_{\delta^+}} f(\delta^+, \gamma')^2 g_\delta(\gamma') d\gamma'$$

The coordinates of σ belong to $[0, 1]$ and thus for every $\gamma' \in \Gamma_{\delta^+}$, the set $\{\sigma \mid (\gamma', \sigma) \in G'(\delta)\}$ is included in a hypercube of side 1. We deduce that $g_\delta(\gamma') \leq 1$ for every $\gamma' \in \Gamma_{\delta^+}$ and obtain the expected inequality (11):

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \leq |\Delta|. \sum_{q' \in Q} \int_{\gamma' \in \Gamma_{q'}} f(q', \gamma')^2 d\gamma' = |\Delta|. ||f||_2$$

Now we can prove that Ψ acts on $L_2(\mathbb{S})$. We have

$$[\Psi(f)(s)]^2 = \left(\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \right)^2 \leq \texttt{Vol}(\mathbb{A}) \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha$$

by virtue of the Cauchy-Schwartz inequality on $L_2(\mathbb{A})$ applied to the constant function 1 and the function $\alpha \mapsto f(s \triangleright \alpha)$ (this latter function is defined and integrable for almost every s by Fubini's theorem and inequality (11)).

We will conclude the proof by bounding for all $f \in L_2(\mathbb{S})$ the operator norm $||\Psi(f)||_2 = (\int_{\mathbb{S}} \Psi(f)(s)^2 ds)^{\frac{1}{2}}$ by $|\Delta| \texttt{Vol}(\mathbb{A})^{\frac{1}{2}} ||f||_2$:

$$\begin{aligned} ||\Psi(f)||_2 = \int_{\mathbb{S}} \Psi(f)(s)^2 ds &= \int_{\mathbb{S}} \left(\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \right)^2 ds \\ &\leq \texttt{Vol}(\mathbb{A}) \int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \\ &\leq |\Delta| \texttt{Vol}(\mathbb{A}) ||f||_2 \end{aligned}$$

The last inequality comes from inequality (11). □

The adjoint operator Ψ^* (acting also on $L_2(\mathbb{S})$) is the analogue of $M^\top$. It is formally defined by the equation:

$$\forall f, g \in L_2(\mathbb{S}),\ \langle \Psi f, g \rangle = \langle f, \Psi^* g \rangle. \tag{14}$$

When elevated to a power n greater than D the constant occurring in the weak progress cycle condition, the operator Ψ and Ψ^* are easier to describe and behave very similarly to matrices as we will see now.

4.4 Kernels and matrix notation.

An operator Ψ is said to be an *Hilbert-Schmidt integral operator* (HSIO) if there exists a function $k \in L^2(\mathbb{S} \times \mathbb{S})$ (called the *kernel*) such that

$$\forall f \in L_2(\mathbb{S}),\ \forall s \in \mathbb{S},\ \Psi f(s) = \int_{s' \in \mathbb{S}} k(s, s') f(s') ds'.$$

With HSIOs, analogy with matrices is strengthen and easier to use e.g. when Ψ has a kernel k then Ψ^* has the kernel: $k^*(s, s') = k(s', s)$ (it is a direct analogue of matrix transposition). Moreover HSIOs have the good property to be *compact*. The compactness of Ψ^D was the key technical point used in [7] to prove a theorem similar to our Theorem 6 below. Here the following proposition implies that Ψ^D and $(\Psi^*)^D$ are Hilbert-Schmidt integral operator.

Proposition 4. *For every $n \geq D$ there exists a function $k_n \in L_2(\mathbb{S} \times \mathbb{S})$ such that: $\Psi^n(f)(s) = \int_{\mathbb{S}} k_n(s, s') f(s') ds'$ and $\Psi^{*n}(f)(s) = \int_{\mathbb{S}} k_n(s', s) f(s') ds'$.*

This proposition is a straightforward corollary of the following more precise lemma 1 used also in the proof of irreducibility of Ψ and Ψ^* (Proposition 6). To state this lemma we recall from [12] the definition of the reachability relation and adopt a matrix notation. For $q, q' \in Q$, we denote by $\texttt{Reach}(n, q, q')$ the set of couple (γ, γ') such that (q', γ') is reachable in n step from (q, γ); formally: $\texttt{Reach}(n, q, q') = \{(\gamma, \gamma') \in \Gamma_p \times \Gamma_q \mid \exists \boldsymbol{\alpha} \in \mathbb{A}^n,\ (q, \gamma) \triangleright \boldsymbol{\alpha} = (q', \gamma')\}$.

It is convenient to adopt the following matrix notation: each function f of $L^2(\mathbb{S})$ is represented by a row vector $\mathbf{f}$ of functions $f_q \in L^2(\Gamma_q)$. The operator Ψ is represented as a $Q \times Q$ matrix $[\Psi]$ for which each element $[\Psi]_{q,q'}$ is an operator from $L^2(\Gamma_{q'})$ to $L^2(\Gamma_q)$. Action of $[\Psi]$ on $\mathbf{f}$ is given by the following formula:

$$\forall i \in Q,\ ([\Psi]\mathbf{f})_i = \sum_{j \in Q} [\Psi]_{ij} f_j.$$

With this matrix notation the matrix for Ψ^* is simply defined by: for all $i, j \in Q$, $[\Psi^*]_{ij} = ([\Psi]_{ji})^*$.

Now we can state the technical lemma describing the kernels of the operators $[\Psi^n]_{ij}$ for $n \geq D$.

Lemma 1. *For every $i, j \in Q$, and $n \geq D$, the operator $[\Psi^n]_{ij} : L_2(\Gamma_i) \to L_2(\Gamma_j)$ has a kernel $k_{ij} \in L_2(\Gamma_{\mathbf{r}} \times \Gamma_{\mathbf{r}'})$ positive almost everywhere in $\texttt{Reach}(n, i, j)$ continuous and piecewise polynomial. Such functions k_{ij} are computable.*

Proof. We first introduce a notation for the successor of a vector γ by a delay vector $\boldsymbol{t} \in [0, M]^n$. Let $\pi = \delta_1 \cdots \delta_n$ be a path from a location q to a location q', $\gamma \in \Gamma_q$, $\boldsymbol{t} \in [0, M]^n$ and $\boldsymbol{\alpha} = (t_1, \delta_1) \ldots (t_n, \delta_n)$. If there exists γ' such that $(q, \gamma) \triangleright \boldsymbol{\alpha} = (q', \gamma')$ then we define $\gamma \triangleright_\pi \mathbf{t} = \gamma'$ else we define $\gamma \triangleright_\pi \mathbf{t} = \bot$

We denote by $P_\pi(\gamma)$ the polytope of delay vector $\mathbf{t}$ that can be read from γ along π i.e. $P_\pi(\gamma) = \{\mathbf{t} \mid \gamma \triangleright_\pi \mathbf{t} \neq \bot\}$. We denote by $\texttt{Reach}(\pi) = \{(\gamma, \gamma \triangleright_\pi \mathbf{t}) \mid \gamma \triangleright_\pi \mathbf{t} \neq \bot\}$. We also define an operator Ψ_π as follows.

$$\Psi_\pi f(\gamma) = \int_{P_\pi(\gamma)} f(\gamma \triangleright_\pi \mathbf{t}) d\mathbf{t}. \tag{15}$$

Then Ψ^n can be decomposed into a sum of operators Ψ_π as follows:

$$[\Psi^n]_{ij} f(\gamma) = \sum_{\pi | \pi \text{ goes from } i \text{ to } j \text{ and } |\pi| = n} \Psi_\pi f(\gamma).$$

Now it suffices to prove that if π is a path leading from i to j and with $|\pi| = n \geq D$ then Ψ_π has a kernel k_π which is piecewise polynomial and non-zero in $\texttt{Reach}(\pi)$.

The idea of the proof is to operate a change of coordinate which transforms several time delays of $\mathbf{t}$ into the vector γ'. Let d'be the dimension of the ending region $\mathbf{r}_j$. In $\mathbf{r}_j$, there are d' clocks non zero which fractional parts are pairwise different and which corresponds to coordinates of γ'. We sort them as follows $y^1 < \cdots < y^{d'}$. By the D weak progress condition, only one clock is not reset during π, this must be the oldest and thus the greatest: $y^{d'}$. If $y^{d'}$ was not reset along π its value is of the form $y^{d'} = x + \sum_{i=1}^n t_i$ where x is a clock possibly null of the starting region $\mathbf{r}_p$ (in this case we pose $i_{d'} = 1$), otherwise it is of the form $y^{d'} = \sum_{i=i_{d'}}^n t_i$ where $i_{d'} - 1 \in \{1, \dots n-1\}$ is the index of the transition where $y^{d'}$ was reset for the last time. Similarly for the other clocks we define $i_1 > i_2 > \cdots > i_d$ where for each $l \in \{1, \dots, d'-1\}$, $i_l - 1$ is the index of the transition where y^l was reset for the last time. We have thus $y^l = \sum_{i=i_l}^n t_i$.

We denote by I the set of indices $\{i_1, \dots, i_{d'}\}$ and $\bar{I} = \{1, \dots, n\} \setminus I$. The function which maps $\mathbf{t}_I = (t_{i_1}, \dots, t_{i_{d'}})$ to $(y_1, \dots, y_{d'})$ is a linear change of coordinates which preserves the volume. More precisely we have $(y_1, \dots, y_{d'})^\top = M(t_{i_1}, \dots, t_{i_{d'}})^\top + \boldsymbol{b}$ where M is an upper triangular matrix with only 1 on the diagonal and where $\boldsymbol{b}$ is of the form $(0, \dots, 0, x)^\top$ or $(0, \dots, 0, 0)^\top$ depending whether there is a non reset clock or not.

We write vectors $\mathbf{t} \in \mathbb{R}^n$ as $\mathbf{t} = (\mathbf{t}_{\bar{I}}, \mathbf{t}_I)_I$ to say that $\mathbf{t}_{\bar{I}}$ regroups the coordinate of $\mathbf{t}$ whose indices are in $\bar{I}$ and $\mathbf{t}_I$ regroup the coordinate of $\mathbf{t}$ whose indices are in I. The function which maps $\mathbf{t} = (\mathbf{t}_{\bar{I}}, \mathbf{t}_I)_I$ to $(\mathbf{t}_{\bar{I}}, \gamma')_I$ is a change of coordinate whose jacobian is 1 (γ' is obtained from $(y_1, \dots, y_{d'})$ by a translation by the constant vector $(\lfloor y_1 \rfloor, \dots, \lfloor y_{d'} \rfloor)$ followed by a permutation of coordinates).

Now let us consider the domains of integration before and after the change of coordinates. The old domain of integration is $P_\pi(\gamma) = \{\mathbf{t} \mid \gamma \triangleright_\pi \mathbf{t} \neq \bot\}$, this domain is a polytope. We denote by P the new domain of integration i.e. $(\mathbf{t}_{\bar{I}}, \gamma')_I \in P$ iff $(\mathbf{t}_{\bar{I}}, \mathbf{t}_I)_I \in P_\pi(\gamma)$.

When we fix $(\gamma, \gamma') \in \texttt{Reach}(\pi)$ we denote by $P_\pi(\gamma, \gamma')$ the set of vectors $\mathbf{t}_{\bar{I}}$ such that $(\mathbf{t}_{\bar{I}}, \gamma')_I \in P$. This corresponds intuitively to the set of timed vectors which leads from γ to γ'. Applying the change of coordinate in (15) we get

$$\Psi_\pi f(\gamma) = \int_{\Gamma_j} \int_{P_\pi(\gamma, \gamma')} \mathbf{1}_{\mathbf{t}_{\bar{I}} \in P_\pi(\gamma, \gamma')} d\mathbf{t}_{\bar{I}} f(\gamma') d\gamma'$$

We obtain the expected form of Ψ_π by defining the kernel as

$$k_\pi(\gamma, \gamma') = \texttt{Vol}[P_\pi(\gamma, \gamma')].$$

It remains to prove that this kernel is piecewise polynomial and non null when $(\gamma, \gamma') \in \texttt{Reach}(\pi)$. We have $(\gamma, \gamma') \in \texttt{Reach}(\pi)$ if and only if the set $P_\pi(\gamma, \gamma')$

is non empty. In this case $P_\pi(\gamma, \gamma')$ is moreover an open polytope (a polytope involving strict inequalities) as a section of the open polytope $P_\pi(\gamma)$, its volume is thus positive and so is $k_\pi(\gamma, \gamma')$.

The polytope $P_\pi(\gamma, \gamma')$ can be defined by a conjunction of inequalities of the following form: $\sum_{i \in \bar{I}} a_i t_i + \sum_{i=1}^{\dim(\mathbf{r}_i)} b_i \gamma_i + \sum_{i=1}^{\dim(\mathbf{r}_j)} c_i \gamma'_i > e$ with $a_i,\ b_i,\ c_i,\ e \in \mathbb{N}$. The volume of such a polytope (when integrating the t_i) can be shown to be piecewise polynomial and continuous in γ_i and γ'_j. □

4.5 Spectral radius and entropy

As in the discrete case, the entropy is equal to the logarithm of the spectral radius (Theorem 6 below). This was the main theorem of [7]. We must prove this theorem in our setting since the functional space of [7] was different from ours and assumptions on the model were somewhat more restrictive.

The following proposition ensure some regularity for eigenfunctions which permits to adapt the results of [7] to our settings.

Proposition 5. *For each eigenvalue $\lambda \neq 0$, each solution f of the eigenfunction equation $\Psi f = \lambda f$ (resp $\Psi^* f = \lambda f$) is continuous and bounded*[8]*.*

Proof. Let f be a solution of the eigenfunction equation $\Psi f = \lambda f$. Lemma 1 implies that Ψ^D is a kernel operator with a kernel k_D piecewise polynomial (and thus bounded on $\mathbb{S}^2$). The function f satisfies for almost every s:

$$\Psi^D f(s) = \lambda^D f(s) = \int k_D(s, s') f(s') ds'.$$

Thus f is bounded almost everywhere by $\lambda^{-D} \sup(k_D) \int |f(s')| ds'$.

We have describe precisely the form of k_D in the proof of Lemma 1: for every $i, j \in Q$ the kernel k_{ij} of $[\Psi^D]_{ij}$ is a sum of kernel of the form (see the proof of Lemma 1):

$$k_\pi(\gamma, \gamma') = \texttt{Vol}[P_\pi(\gamma, \gamma')].$$

The function $\gamma \mapsto \int k_\pi(\gamma, \gamma') f(\gamma') d\gamma'$ is continuous since the domain of integration $(\gamma' \mid (\gamma, \gamma') \in \texttt{Reach}(\pi))$ depends continuously on γ and the integrand is continuous w.r.t γ and bounded almost everywhere. By summing over all path π, we obtain that $f : s \mapsto \lambda^{-D} \int k_D(s, s') f(s') ds'$ is continuous and bounded. A similar proof can be written for Ψ^* since it has the kernel $k^*_D(s', s) = k_D(s, s')$. □

Theorem 6 (adapted from [7] to $L_2(\mathbb{S})$). *The spectral radius ρ is a positive eigenvalue (i.e. $\rho > 0$ and $\exists v \in L_2(\mathbb{S})$ s.t. $\Psi v = \rho v$) and $\mathcal{H}(\mathcal{G}) = \log_2(\rho)$.*

[8] To be more formal, f as an element of L_2 is a class of functions pairwise equal almost everywhere, it admits a unique representative which is continuous and bounded.

Proof. We adapt to the functional space $L_2(\mathbb{S})$ the proof of the main theorem of [7].

Proof of $\mathcal{H}(\mathcal{G}) \leq \log_2 \rho$: The so called Gelfand formula gives

$$\rho = \lim_{n\to\infty} ||\Psi^n||_2^{\frac{1}{n}}.$$

As $V_n = \Psi^n(1)$ we have

$$||V_n||_2 = ||\Psi^n 1||_2 \leq ||\Psi^n||_2 ||1||_2$$

and thus

$$\limsup_{n\to\infty} \frac{\log(||V_n||_2)}{n} \leq \log_2 \rho.$$

Recall that

$$\mathcal{H}(\mathcal{G}) = \limsup_{n\to\infty} \frac{\log(\texttt{Vol}(\mathcal{R}_n))}{n}$$

and that

$$\texttt{Vol}(\mathcal{R}_n) = \int_{\mathbb{S}} V_n(s)ds = ||V_n||_1.$$

Thus

$$\mathcal{H}(\mathcal{G}) = \limsup_{n\to\infty} \frac{\log(||V_n||_1)}{n}.$$

It remains to prove that

$$\limsup_{n\to\infty} \frac{\log(||V_n||_1)}{n} \leq \limsup_{n\to\infty} \frac{\log(||V_n||_2)}{n}.$$

This come from the Cauchy-Schwartz inequality:

$$||V_n||_1 \leq ||V_n||_2 ||1||_2 \leq ||V_n||_2 \sqrt{\texttt{Vol}(\mathbb{S})}.$$

Proof of ρ is a positive eigenvalue for Ψ and Ψ^*:

By the preceding part of the proof and using the hypothesis $\mathcal{H} > -\infty$ we have $\rho \geq 2^{\mathcal{H}} > 0$. A necessary condition for a positive spectral radius to be an eigenvalue is the compactness of some power A^n of the operator A where $A = \Psi$ or Ψ^*. This is ensured by proposition 4 as HSIOs are compacts operator. Thus there exists v such that $\Psi v = \rho v$ and w such that $\Psi^* w = \rho w$.

Proof of $\log_2 \rho = \mathcal{H}(\mathcal{G})$:

Proposition 5 ensures that the eigenfunction v define above is continuous and bounded (everywhere). Let C be an upper bound for v i.e a positive constant such that $\forall s \in \mathbb{S},\ |v(s)| < C$. We have that

$$\forall s \in \mathbb{S},\ n \in \mathbb{N},\ \rho^n |v(s)| = |\Psi^n v(s)| \leq \Psi^n |v|(s) \leq C\Psi^n 1(s) = CV_n(s). \quad (16)$$

where the first inequality is a variant of the so called triangular inequality, it can be proven as follows: let v^+ and v^- be the positive and the negative part of v then

$$|\Psi^n v(s)| = |\Psi^n v^+(s) - \Psi^n v^-(s)| \leq |\Psi^n v^+(s)| + |\Psi^n v^-(s)| = \Psi^n v^+(s) + \Psi^n v^-(s) = \Psi^n |v|(s).$$

Now we integrate (16) wrt. s and obtain $0 < \rho^n ||v||_1 \leq C||V_n||_1 = C\texttt{Vol}(\mathcal{R}_n)$. Taking $\liminf_{n\to\infty} \frac{1}{n}\log(.)$ in this latter inequality we obtain:

$$\log_2 \rho \leq \liminf_{n\to\infty} \frac{1}{n}\log(\texttt{Vol}(\mathcal{R}_n)) \leq \limsup_{n\to\infty} \frac{1}{n}\log(\texttt{Vol}(\mathcal{R}_n)) = \mathcal{H}(\mathcal{G}) \leq \log_2 \rho \tag{17}$$

where the last inequality comes from the previous part of the proof. Thus all inequalities of (17) are equalities and we conclude that $\log_2 \rho = \mathcal{H}(\mathcal{G})$. □

4.6 A Perron-Frobenius like theorem for timed region graph

The following theorem defines $v,\ w$ used in the definition of the maximal entropy SPOR (9).

Theorem 7. *There exists a unique eigenfunction (up to a scalar constant) v of Ψ (resp. w of Ψ^*) for the eigenvalue ρ which is positive almost everywhere. Any non-negative eigenfunction of Ψ (resp. Ψ^*) is collinear to v (resp. w).*

Eigenfunctions v and w are chosen such that $\langle w, v\rangle = 1$.

Expressing ρ, v, w as solution of integral equations with kernels. It is worth mentioning that for any $n \geq D$, the objects ρ, v (resp. w) are solutions of the eigenvalue problem $\int_{\mathbb{S}} k_n(s,s')v(s')ds' = \rho^n v(s)$ with v non negative (resp. $\int_{\mathbb{S}} k_n(s',s)w(s')ds' = \rho^n w(s)$ with w non negative); unicity of v (resp. w) up to a scalar constant is ensured by Theorem 7. The matrix notation, where we denote by $k_{n,q,q'}$ the kernel of $[\Psi^n]_{qq'}$, gives a system of integral equations for v and ρ:

$$\sum_{q'\in Q} \int_{\Gamma_{q'}} k_{n,q,q'}(\gamma,\gamma')v_{q'}(\gamma')d\gamma' = \rho^n v_q(\gamma), \quad \text{for } q \in Q \tag{18}$$

and another system for w and ρ:

$$\sum_{q\in Q} \int_{\Gamma_q} k_{n,q,q'}(\gamma,\gamma')w_q(\gamma)d\gamma = \rho^n w_{q'}(\gamma), \quad \text{for } q' \in Q. \tag{19}$$

Further computability issues for $\rho,\ v$ and w are discussed in the conclusion.

Proof of theorem 7. The proof of Theorem 7 is based on theorem 11.1 condition e) of [20] (recalled in Theorem 8 below) which is a generalization of Perron-Frobenius Theorem to positive linear operators. The main hypothesis to prove is the irreducibility of Ψ whose analogue in the discrete case is the irreducibility of the adjacency matrix M. Recall from section 2 that M is irreducible if for all states i, j there exists $n \geq 1$ such that $M^n_{ij} > 0$ (this is equivalent to the strong connectivity of the graph).

The operator Ψ is said to be *irreducible* if the following condition holds: if $\Psi f \leq af$ for some $a > 0$ and a non-negative non-null $f \in L_2$ implies that f is

quasi-interior which means that $\langle f, g\rangle > 0$ for every non-negative and non null $g \in L_2(\mathbb{S})$.

The irreducibility of Ψ and Ψ^* is essentially due to the strong connectivity of the state space $\mathbb{S}$ which is traduced by the positivity of kernels between every two locations q, q'. More precisely we will use lemma 1 above and lemma 2 just below which traduces the thickness of the timed region graph (see section 4.1 and [12]).

Lemma 2. *There exists an $n \geq D$ such that $\texttt{Reach}(n, q, q') = \Gamma_q \times \Gamma_{q'}$.*

Proof. This lemma is a direct consequence of results of [12]. The following assertions and definitions (slightly adapted to our notation) can be found in [12]. A path π from q to q' is called forgetful if $\texttt{Reach}(\pi) = \Gamma_{\mathbf{r}} \times \Gamma_{\mathbf{r}'}$ where $\texttt{Reach}(\pi)$ is the reachability relation restrained to π defined in the proof of lemma 1. Every path which contains a forgetful cycle is forgetful. If $\mathcal{G}$ is thick it contains a forgetful cycle f (with $|f| > 0$). Let $l \in \tilde{Q}$ such that f leads from l to l and π, π' such that π leads from q to l and π' leads from l to q'. Such paths exist by strong connectivity of the set of location. Let $m \geq D$, the path $\pi f^D \pi'$ is forgetful and leads from q to q' and thus $\texttt{Reach}(D|f| + |\pi| + |\pi'|, q, q') = \Gamma_q \times \Gamma_{q'}$ □

Now we can state the irreducibility of Ψ and Ψ^*

Proposition 6. *Ψ and Ψ^* are irreducible.*

Proof. Let $f \in L_2$ non-negative non-null and $a > 0$ such that $\Psi f \leq af$. Let $g \in L_2(\mathbb{S})$ be non negative and non null; we show that $\langle f, g\rangle > 0$. There are $i, j \in Q$ such that g_i, f_j are non negative and non null. By lemma 1 and 2 there exists an n such that $[\Psi^n]_{ij}$ has a kernel k_{ij} positive almost everywhere and thus $[\Psi^n]_{ij} f_j(s) = \int_{s'} k_{ij}(s, s') f(s') ds' > 0$ for almost every s. We are done since

$$a^n \langle f, g\rangle \geq \langle \Psi^n f, g\rangle \geq \langle [\Psi^n]_{ij} f_j, g_i\rangle > 0.$$

This also prove the irreducibility of Ψ^* since $k^*_{ij} = k_{ji}$. □

The conclusion of theorem 6 furnishes the hypotheses of theorem 11.1 condition e) of [20] (Theorem 8 below). We defined the cone K to be the subset of $L_2(\mathbb{S})$ of non-negative functions. It satisfies $\Psi(K) \subseteq K$, it is minihedral ([20]6.1 example d)) and is reproducing i.e. all functions of $f \in L_2(\mathbb{S})$ can be written as $f = f^+ - f^-$ with $f^-, f^+ \in K$. The conclusion of this latter theorem achieve the proof of our theorem.

Theorem 8 ([20], theorem 11.1 condition e)). *Suppose that $\Psi K \subseteq K$, Ψ has a normalized eigenfunction $v \in K$ with corresponding eigenvalue ρ, K is reproducing and minihedral, the operator Ψ is irreducible and the operator Ψ^* has an eigenfunction w in K^* which correspond to the eigenvalue ρ. Then the eigenvalue is simple and there is no other normalized eigenfunction different from ρ in K.*

5 Examples

5.1 Running example completed

Consider again the timed region graph depicted in Fig. 3.1. The matrix notation of (10) is:

$$[\Psi]\begin{pmatrix} f_{\mathbf{r}^1} \\ f_{\mathbf{r}^2} \end{pmatrix} = \begin{pmatrix} \gamma \mapsto \int_\gamma^1 f_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 f_{\mathbf{r}^2}(\gamma')d\gamma' \\ \gamma \mapsto \int_0^1 f_{\mathbf{r}^1}(\gamma')d\gamma' + \int_\gamma^1 f_{\mathbf{r}^2}(\gamma')d\gamma' \end{pmatrix}$$

We can deduce that operators Ψ and Ψ^* are HSIO with matrices of kernels:

$$k = \begin{pmatrix} \mathbf{1}_{0<\gamma\leq\gamma'<1} & \mathbf{1}_{0<\gamma'<1} \\ \mathbf{1}_{0<\gamma'<1} & \mathbf{1}_{0<\gamma\leq\gamma'<1} \end{pmatrix};\ k^* = \begin{pmatrix} \mathbf{1}_{0<\gamma'\leq\gamma<1} & \mathbf{1}_{0<\gamma'<1} \\ \mathbf{1}_{0<\gamma'<1} & \mathbf{1}_{0<\gamma'\leq\gamma<1} \end{pmatrix}.$$

Eigenfunctions and spectral radius v, w, ρ are solutions of the eigenvalue equations $[\Psi]\mathbf{v} = \rho\mathbf{v}$ and $[\Psi^*]\mathbf{w} = \rho\mathbf{w}$ which written in the form of (18) and (19) (for $n = 1$) yield

$$\begin{aligned}
\rho v_{\mathbf{r}^1}(\gamma) &= \int_\gamma^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 v_{\mathbf{r}^2}(\gamma')d\gamma'; \\
\rho v_{\mathbf{r}^2}(\gamma) &= \int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_\gamma^1 v_{\mathbf{r}^2}(\gamma')d\gamma'; \\
\rho w_{\mathbf{r}^1}(\gamma) &= \int_0^\gamma w_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 w_{\mathbf{r}^2}(\gamma')d\gamma'; \\
\rho w_{\mathbf{r}^2}(\gamma) &= \int_0^1 w_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^\gamma w_{\mathbf{r}^2}(\gamma')d\gamma'.
\end{aligned}$$

We differentiate one time the equations and obtain:

$$\rho v'_{\mathbf{r}^i}(\gamma) = -v_{\mathbf{r}^i}(\gamma); \rho w'_{\mathbf{r}^i}(\gamma) = w_{\mathbf{r}^i}(\gamma)\ (i \in \{1, 2\}).$$

Thus the functions are of the form $v_{\mathbf{r}^i}(\gamma) = v_{\mathbf{r}^i}(0)e^{-\gamma/\rho}$, $w_{\mathbf{r}^i}(\gamma) = w_{\mathbf{r}^i}(0)e^{\gamma/\rho}$. Remark that $\rho v_{\mathbf{r}^1}(0) = \int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 v_{\mathbf{r}^2}(\gamma')d\gamma' = \rho v_{\mathbf{r}^2}(0)$ and thus $v_{\mathbf{r}^1} = v_{\mathbf{r}^2}$ (we can divide by $\rho > 0$ since we can prove that $\mathcal{H} > -\infty$ using theory of [12]). Similarly $w_{\mathbf{r}^1}(1) = w_{\mathbf{r}^2}(1)$ and thus $w_{\mathbf{r}^1} = w_{\mathbf{r}^2}$. The constant ρ satisfies the condition $v_{\mathbf{r}^1}(0) = 2\int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma'/\rho = 2v_{\mathbf{r}^1}(1) = 2v_{\mathbf{r}^1}(0)e^{-1/\rho}$. Thus we obtain: $\rho = \frac{1}{\ln(2)}$; $\begin{pmatrix} v_{\mathbf{r}^1}(\gamma) \\ v_{\mathbf{r}^2}(\gamma) \end{pmatrix} = C\begin{pmatrix} 2^{-\gamma} \\ 2^{-\gamma} \end{pmatrix}$; $\begin{pmatrix} w_{\mathbf{r}^1}(\gamma) \\ w_{\mathbf{r}^2}(\gamma) \end{pmatrix} = C'\begin{pmatrix} 2^{\gamma} \\ 2^{\gamma} \end{pmatrix}$ with C and C' two positive constants.

Finally the maximal entropy SPOR for $\mathcal{G}^{\mathtt{ex1}}$ is given by:

$$\begin{aligned}
&p_0^*(p, (\gamma, 0)) = p_0^*(q, (0, \gamma)) = \frac{1}{2} \text{ for } \gamma \in (0, 1); \\
&p^*(t, \delta^1|p, (\gamma, 0)) = p^*(t, \delta^4|q, (0, \gamma)) = \frac{2^{-t}}{\rho} \text{ for } \gamma \in (0, 1), t \in [0, 1-\gamma); \\
&p^*(t, \delta^2|p, (\gamma, 0)) = p^*(t, \delta^3|q, (0, \gamma)) = \frac{2^{\gamma-t}}{\rho} \text{ for } \gamma \in (0, 1), t \in (0, 1).
\end{aligned}$$

5.2 Our favorite example

The timed region graph depicted on Fig. 5.2 with $\mathbf{r}_p = \{(x,y) \mid 0 = y < x < 1\}$ and $\mathbf{r}_q = \{(x,y) \mid 0 = x < y < 1\}$ is the underlying structure of a timed automaton introduced by Asarin and Degorre in [7]. With these author we have illustrated the concept of thickness [12] and of generating function [6] on it. This example is closely related to the class of alternating permutations as we have shown in [11].

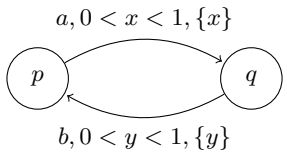


Fig. 2. A timed graph whose operator is self adjoint: $\mathcal{G}^{\mathsf{ex2}}$

The operators Ψ and Ψ^* are equal, indeed they are HSIO with the same matrices of kernels[9]:

$$k = k^* = \begin{pmatrix} 0 & \mathbf{1}_{0<\gamma'<1-\gamma<1} \\ \mathbf{1}_{0<\gamma'<1-\gamma<1} & 0 \end{pmatrix}$$

The maximal entropy SPOR is given by the following PDFs:

$$p_0^*(p,(\gamma,0)) = p_0^*(q,(0,\gamma)) = \cos^2\left(\frac{\pi}{2}\gamma\right) \text{ for } \gamma \in (0,1);$$

$$p^*(t,a|p,(\gamma,0)) = p^*(t,b|q,(0,\gamma)) = \frac{\pi}{2}\frac{\cos(\frac{\pi}{2}t)}{\cos(\frac{\pi}{2}\gamma)}\mathbf{1}_{t<1-\gamma} \text{ for } \gamma \in (0,1), t \in [0,1-\gamma);$$

6 Proof of main theorems (section 4.2)

We give the proof of theorem 4 in several steps

6.1 Proof of Y is a SPOR

The function v and w are defined up to a scaling constant in Theorem 7 and are chosen such that $\int_{\mathbb{S}} p_0^*(s) = \langle v, w\rangle = 1$. The function v is positive almost everywhere and $v(s \triangleright \alpha) = 0$ when $s \triangleright \alpha = \perp$ thus $p(\alpha|s)$ is defined for almost every $s \in \mathbb{S}$, $\alpha \in \mathbb{A}$ and equals 0 when $s \triangleright \alpha = \perp$. Finally we have $\int_{\mathbb{A}} \frac{v(s\triangleright\alpha)}{\rho v(s)} d\alpha = \frac{\Psi v(s)}{\rho v(s)} = 1$ since v is an eigenfunction for ρ. □

[9] Such a self adjoint operator (i.e. $\Psi = \Psi^*$) in a Hilbert space has nice properties.

6.2 Proof of Y^* is stationary

First we remark that a SPOR is stationary whenever the probability on states remains the same:

Proposition 7. *a SPOR is stationary if and only if S_1 has the PDF p_0 and thus the same law as S_0.*

Proof. The only if part is straightforward. For the other part let Y be a SPOR such that S_1 has the PDF p_0. We show by recurrence that S_n has the PDF p_0 ($n \geq 1$). For this, let us suppose that S_n has the PDF p_0 for some $n \geq 1$. We show that S_{n+1} has the same law has S_1 and thus has the PDF p_0. For every measurable set of states $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$,

$$\begin{aligned} P(S_{n+1} \in \mathcal{S}) &= \int_{\mathbb{S}} \int_{\mathbb{A}} p_0(s) p(\alpha|s) P(S_n \triangleright A_n \in \mathcal{S} | S_n = s, A_n = \alpha) d\alpha ds \\ &= \int_{\mathbb{S}} \int_{\mathbb{A}} p_0(s) p(\alpha|s) 1_{s \triangleright \alpha \in \mathcal{S}} d\alpha ds \\ &= P(S_1 \in \mathcal{S}) \end{aligned}$$

We have shown that for all $i \geq n$, S_i has the PDF p_0. Now we remind from (3) that the PDF of $Y_i \cdots Y_{i+n-1}$ knowing that $S_i = s$ is $p_n(\boldsymbol{\alpha}|s)$. We conclude that $Y_i \cdots Y_{i+n-1}$ has the expected PDF $p_n(s, \boldsymbol{\alpha}) = p_0(s) p_n(\boldsymbol{\alpha}|s)$.

□

We applied this proposition to show that Y^* is stationary

Proof. For every measurable set of states $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$,

$$\begin{aligned} P(S_1 \in \mathcal{S}) &= P(S_0 \triangleright A_0 \in \mathcal{S}) \\ &= \int_S \int_{\mathbb{A}} p_0(s) p(\alpha|s) 1_{s \triangleright \alpha \in \mathcal{S}} d\alpha ds \\ &= \int_{\mathbb{S}} \int_{\mathbb{A}} v(s) w(s) \frac{v(s \triangleright \alpha)}{\rho v(s)} 1_{s \triangleright \alpha \in \mathcal{S}} d\alpha ds \\ &= \int_S w(s) \int_{\mathbb{A}} v(s \triangleright \alpha) 1_{s \triangleright \alpha \in \mathcal{S}} d\alpha ds / \rho \\ &= \langle w, \Psi(v 1_{\mathcal{S}}) \rangle / \rho \\ &= \langle \Psi^* w, v 1_{\mathcal{S}} \rangle / \rho \quad \text{By definition of } \Psi^* \text{ see (14)} \\ &= \langle w, v 1_{\mathcal{S}} \rangle \text{ since } w \text{ is an eigenfunction of } \Psi^* \text{ for } \rho \\ &= \int_S p_0(s) 1_{s \in \mathcal{S}} ds \\ &= P(S_0 \in \mathcal{S}). \end{aligned}$$

□

6.3 Proof of $H(Y^*) = \mathcal{H}(\mathcal{G})$

$$
\begin{aligned}
H(Y^*) &= -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log_2 p(\alpha|s) d\alpha ds \\
&= -\int_{\mathbb{S}} v(s)w(s) \int_{\mathbb{A}} \frac{v(s \triangleright \alpha)}{\rho v(s)} \log_2 \frac{v(s \triangleright \alpha)}{\rho v(s)} d\alpha ds \\
&= -\frac{1}{\rho} \int_{\mathbb{S}} w(s) \int_{\mathbb{A}} v(s \triangleright \alpha)[\log_2 v(s \triangleright \alpha) - \log_2(\rho v(s))] d\alpha ds \\
&= -\frac{1}{\rho} \langle w, \Psi(v \log_2 v) \rangle + \frac{1}{\rho} \langle w \log_2 v, \Psi v \rangle + \frac{\log_2 \rho}{\rho} \langle w, \Psi v \rangle \\
&= -\frac{1}{\rho} \langle \Psi^* w, v \log_2 v \rangle + \langle w \log_2 v, v \rangle + \log_2(\rho) \langle w, v \rangle \text{ since } v \text{ is an eigenfunction of } \Psi \text{ for } \rho \\
&= -\langle w, v \log_2 v \rangle + \langle w \log_2 v, v \rangle + \log_2(\rho) \text{ since } \langle w, v \rangle = 1 \text{ and } w \text{ is an eigenfunction of } \Psi^* \text{ for } \rho \\
&= \log_2(\rho) = \mathcal{H}(\mathcal{G}).
\end{aligned}
$$

6.4 Ergodicity of Y^*

We first introduce a "stochastic" operator φ which is the continuous analogue of a stochastic matrix. We then prove an ergodic property on φ (proposition 10) which permits to prove the ergodicity of Y^*.

Operator φ and its conjugate φ^*. Let $L_2(v^2 ds)$ be the space of function f such that $fv \in L_2(\mathbb{S})$. The dual space of $L_2(v^2 ds)$ is isomorphic to $L_2(ds/v^2)$. The norm on $L_2(v^2 ds)$ is $||f||_{L_2(v^2 ds)} = ||fv||_2$.

Let $\varphi : L_2(v^2 ds) \to L_2(v^2 ds)$ be the linear operator defined by $\varphi(f) = \Psi(vf)/v$. One can see that $\varphi^*(f) = v\Psi^*(f/v)$.

We have construct this operator by analogy with the transition probability matrix of the Shannon Parry Markov-chain:$P_{i,j} = M_{i,j} v_j / \rho v_i$ with M the adjacency matrix of the graph is the analogue of Ψ (see section 2).

The operators φ^k ($k \geq 0$) are associated with the conditional PDFs $p_k(\boldsymbol{\alpha}|s) = p_k(\boldsymbol{\alpha})/p_0(s)$ defined in (3):

$$
\varphi^k(f)(s) = \int_{\boldsymbol{\alpha} \in \mathbb{A}^k} p_k(\boldsymbol{\alpha}|s) f(s \triangleright \boldsymbol{\alpha}) d\boldsymbol{\alpha}. \tag{20}
$$

The analogy between the operator φ and the matrix of a finite Markov-chain can also be applied to the eigenfunctions. The eigenfunctions for the spectral radius of φ (which is 1) are the constant functions while the eigenfunctions for the adjoint (analogous with the transposed matrix) are all collinear to the stationary PDF on states p_0^*. First we have an existence lemma

Lemma 3. $\varphi(1) = 1$, $\varphi^*(p_0^*) = p_0^*$.

Then we have uniqueness given in the following proposition. This is an application of Theorem 8 we have already use to prove Theorem 7.

Proposition 8. *The spectral radius of φ is 1. It is a simple eigenvalue of φ for which 1 is an eigenfunction ($\varphi(1) = 1$). Every positive eigenfunction of φ are constant (i.e. collinear to 1). p_0^* is an eigenfunction of φ^* for the spectral radius 1 which is a simple eigenvalue ($\varphi^*(p_0^*) = p_0^*$). Every positive eigenfunction of φ^* are collinear to p_0^*.*

Proposition 9. *Some power φ^p ($p \in \mathbb{N}$) has a spectral gap, i.e. the spectral radius of φ^p is a simple eigenvalue (here all the eigenfunctions for the spectral radius 1 are constant) and the rest of the spectrum of φ^p belongs to the disc $C_\lambda = \{z||z| \leq \lambda\}$ for some λ strictly lower than the spectral radius.*

Proof. φ is a compact operator with spectral radius 1. A well known results in spectral theory asserts that there is only a countable number of point in the spectrum of a compact operator and that all nonzero points of the spectrum are isolated and are eigenvalues.

We can apply the theorem at the beginning of section 3.4 of [24]. This theorem states that there exists $p \in \mathbb{N}$ such that every eigenvalue ω of modulus 1 satisfies $\omega^p = 1$ and thus φ^p has only one eigenvalue of modulus 1 which is its spectral radius. The other eigenvalue ω^p of φ^p are such that $\omega^p < \beta$ for some $\beta < 1$ since there is no accumulation point other than 0.

Proposition 8 just above guarantees that the constant function 1 is the single eigenfunction for φ and thus for φ^p, this eigenfunction is positive and therefore φ^p has a spectral gap β. □

With such a spectral gap, iterates of φ on a function f converge in direction toward the positive eigenfunction 1 of φ (i.e. converge to a constant) provided f is not orthogonal to the eigenfunction p_0^* of φ^*:

Lemma 4. *For all $f \in L_2(v^2 ds)$ such that $E_Y(f) = \int_{\mathbb{S}} f(s) p_0^*(s) ds \neq 0$ the following holds*

$$||\varphi^{pk}(f) - E_Y(f)||_{L_2(v^2 ds)} \to_{k \to +\infty} 0.$$

Proof. This is ensured by Theorem 15.4 of [20] whose hypothesis is the existence of a gap for φ^p (Proposition 9). □

Proposition 10. *Let $f \in L_2(v^2 ds)$ such that $E_Y(f) \neq 0$ and $g_n(s) = \frac{1}{n} \sum_{k=1}^{n} \varphi^k(f)(s) - E_Y(f)$ then*

$$||g_n||_{L_2(v^2 ds)} \to_{n \to +\infty} 0.$$

Proof. We have

$$||g_n||_{L_2(v^2 ds)} \leq \sum_{i=1}^{p} \frac{1}{n} \sum_{k=0}^{n-1} ||\varphi^{pk+i}(f)(s) - E_Y(f)||_{L_2(v^2 ds)}.$$

Now it suffices to remark that for all $i \in \{1, \ldots, p\}$ the sequence $||\varphi^{pk+i}(f) - E_Y(f)||_{L_2(v^2 ds)} \to_{k \to +\infty}$ converges to 0 and thus so does its Cesaro mean. This convergence follows from Lemma 4 applied to φ^i since $\varphi^{pk+i}(f) = \varphi^{pk}(\varphi^i f)$ and $E_Y(\varphi^i(f)) = E_Y(f) > 0$. □

The following lemma roughly states that, asymptotically Y^* is independent from its past which is a sufficient condition for ergodicity.

Lemma 5. *Let R be a measurable subset of $(\mathbb{S} \times \mathbb{A})^{m+1}$ ($m \in \mathbb{N}$) then*

$$\frac{1}{n}\sum_{k=1}^{n} P(Y_0^* \cdots Y_m^* \in R \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in R) \to_{n\to\infty} P(Y_0^* \cdots Y_m^* \in R)^2.$$

Proof. For all $k \in \mathbb{N}$ we have:

$$P(Y_0^* \cdots Y_m^* \in R \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in R) = \int_R p_m[s, \boldsymbol{\alpha}] P(Y_{m+k}^* \cdots Y_{2m+k}^* \in R | S_m = s \triangleright \boldsymbol{\alpha}) d[s, \boldsymbol{\alpha}].$$

We use stationarity and characterization of φ^k (20) and obtain:

$$P(Y_{m+k}^* \cdots Y_{2m+k}^* \in R | S_m = s) = P(Y_k^* \cdots Y_{k+m}^* \in R | S_0 = s) = \varphi^k(f)(s)$$

with $f(s) = P(Y_0^* \cdots Y_m^* \in R | S_0 = s)$. We have also that

$$P(Y_0^* \cdots Y_m^* \in R)^2 = \int_R p_m[s, \boldsymbol{\alpha}] P(Y_0^* \cdots Y_m^* \in R) d[s, \boldsymbol{\alpha}].$$

We will use the proposition 10 with

$$g_n(s) = \frac{1}{n}\sum_{k=1}^{n} \varphi^k(f)(s) - P(Y_0^* \cdots Y_m^* \in R) = \frac{1}{n}\sum_{k=1}^{n} \varphi^k(f)(s) - E_Y(f).$$

We can end the proof with the following inequalities:

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{k=1}^{n} P(Y_0^* \cdots Y_m^* \in R \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in R) - P(Y_0^* \cdots Y_m^* \in R)^2 \right| \\
&\leq \int_{\mathbb{S}} \int_{A_\omega} p_m[s, \boldsymbol{\alpha}] |g_n(s \triangleright \boldsymbol{\alpha})| d\boldsymbol{\alpha} ds \\
&\leq \int_{\mathbb{S}} \varphi^m(|g_n|) p(s) ds = \int_{\mathbb{S}} \varphi^m(|g_n|) v(s) w(s) ds \\
&\leq ||w||_\infty \int_{\mathbb{S}} \varphi^m(|g_n|) v(s) ds \text{ (since } w \text{ is bounded by Proposition 5)} \\
&\leq ||w||_\infty ||\varphi^m(|g_n|) v||_2 \sqrt{\texttt{Vol}(\mathbb{S})} \text{ (by Cauchy Schwartz inequality)} \\
&= ||w||_\infty ||\varphi^m(|g_n|)||_{L_2(v^2 ds)} \sqrt{\texttt{Vol}(\mathbb{S})} \\
&\leq ||w||_\infty ||\varphi^m||_{L_2(v^2 ds)} ||g_n||_{L_2(v^2 ds)} \sqrt{\texttt{Vol}(\mathbb{S})} \to_{n\to+\infty} 0 \text{ (by Proposition 10).}
\end{aligned}$$

□

Now we can achieve the proof that Y^* is ergodic.

Proof. Consider a shift invariant set A. We will show that $P(A) \in \{0,1\}$. We suppose that $P_Y(A) < 1$ and show that $P_Y(A) \leq P_Y(A)^2$. These inequalities imply that $P_Y(A) = 0$.

Using (4), for every ϵ, there exists an $m \in \mathbb{N}$ such that $P(Y_0^* \cdots Y_m^* \in A_{m+1}) \in [P(A), P(A)+\epsilon]$. By set inclusion we have:

$$P(A) \leq P(Y_0^* \cdots Y_m^* \in A_{m+1} \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in A_{m+1})$$

Taking the Cesaro average (i.e. summing over k and dividing by n) we obtain:

$$P(A) \leq \frac{1}{n} \sum_{k=1}^{n} P(Y_0^* \cdots Y_m^* \in A_{m+1} \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in A_{m+1}).$$

Taking the limit and using lemma 5 we obtain:

$$P(A) \leq P(Y_0^* \cdots Y_m^* \in A_{m+1})^2 \leq (P(A)+\epsilon)^2.$$

When ϵ tends to 0, we obtain the required inequality. □

This last paragraph has achieved the proof of Theorem 4.

6.5 Proof of the asymptotic equipartition property (Theorem 5)

The pointwise ergodic theorem states that with probability 1 an infinite runs r satisfies $\frac{1}{n}\sum_{k=0}^{n-1} f(\sigma^k(r)) \to_{n\to+\infty} E_Y(f)$ where σ is the shift map i.e. $\sigma(y_0 y_1 \cdots) = y_1 y_2 \cdots$ and f is such that $E_Y(|f|) < +\infty$.

Here we define f by $f(s_0\alpha_0 s_1 \alpha_1 \cdots) = -\log p(\alpha_0|s_0)$ and thus

$$\frac{1}{n}\sum_{k=0}^{n-1} f(\sigma^k(r)) = -\frac{1}{n}\sum_{k=0}^{n-1} \log_2 p(\alpha_k|s_k) = -\frac{1}{n}[\log_2 p_n[s_0, \alpha_0 \cdots \alpha_{n-1}] - \log_2 p_0(s_0)].$$

On the other side $E_Y(f) = E_Y(-\log p(A_0|S_0)) = -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log p(\alpha|s) d\alpha ds$ which is equal to $H(Y)$ by definition.

It remains to show that $E_Y(|f|) < +\infty$. Indeed $E_Y(|f|) = E_Y(f) + 2E_Y(f^-) = H(Y) + 2E_Y(f^-)$ where $f^- = (|f| - f)/2$ is the negative part of f and

$$E_Y(f^-) \leq \max_{\mathbb{R}^+}(x \mapsto -x\log_2 x)\mathtt{Vol}(\mathbb{A}) < +\infty.$$

□

7 Conclusion and perspectives

In this article, we have proved the existence of an ergodic stochastic process over runs of a timed region graph $\mathcal{G}$ with maximal entropy, provided $\mathcal{G}$ has finite entropy ($\mathcal{H} > -\infty$) and satisfies the D weak progress condition.

The next question is to know how simulation can be achieved in practice. Symbolic computation of ρ and v have been proposed in [7] for subclasses of

deterministic TA. In the same article, an iterative procedure is also given to estimate the entropy $\mathcal{H} = \log_2(\rho)$. We think that approximations of ρ, v and w using an iterative procedure on Ψ and Ψ^* would give a SPOR with entropy as close to the maximum as we want. A challenging task for us is to determine an upper bound on the convergence rate of such an iterative procedure.

Connection with information theory is clear if we consider as in [5], a timed regular language as a source of timed words. A SPOR is in this approach a stochastic source of timed words. It would be very interesting to lift compression methods (see [22,17]) from untimed to timed setting.

Acknowledgements.

I thank Eugene Asarin, Aldric Degorre and Dominique Perrin for sharing motivating discussions.

References

1. Paul H. Algoet and Thomas M. Cover. A sandwich proof of the Shannon-McMillan-Breiman theorem. *The annals of probability*, 16(2):899–909, 1988.
2. Rajeev Alur and Mikhail Bernadsky. Bounded model checking for GSMP models of stochastic real-time systems. In *HSCC'06*, volume 3927 of *LNCS*, pages 19–33.
3. Rajeev Alur, Costas Courcoubetis, and David L. Dill. Model-checking for probabilistic real-time systems. In *ICALP'91*, LNCS 510, pages 115–126, 1991.
4. Rajeev Alur and David L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126:183–235, 1994.
5. Eugene Asarin, Nicolas Basset, Marie-Pierre Béal, Aldric Degorre, and Dominique Perrin. Toward a timed theory of channel coding. In *FORMATS'12*, LNCS 7595.
6. Eugene Asarin, Nicolas Basset, Aldric Degorre, and Dominique Perrin. Generating functions of timed languages. In *MFCS*, LNCS 7464, pages 124–135, 2012.
7. Eugene Asarin and Aldric Degorre. Volume and entropy of regular timed languages: Analytic approach. In *FORMATS'09*, LNCS 5813, pages 13–27, 2009.
8. Eugene Asarin and Aldric Degorre. Volume and entropy of regular timed languages: Discretization approach. In *CONCUR'09*, LNCS 5710, pages 69–83, 2009.
9. Christel Baier, Nathalie Bertrand, Patricia Bouyer, Thomas Brihaye, and Marcus Größer. Probabilistic and topological semantics for timed automata. In *FSTTCS'07*, LNCS 4855, pages 179–191, New Delhi, India, 2007. Springer.
10. Nicolas Basset. A maximal entropy stochastic process for a timed automaton. 2013. To appear in ICALP'2013.
11. Nicolas Basset. Counting and generating permutations using timed languages. 2013. submitted for publication, long version available: `http://hal-upec-upem.archives-ouvertes.fr/hal-00820373`.
12. Nicolas Basset and Eugene Asarin. Thin and thick timed regular languages. In *FORMATS'11*, LNCS 6919, pages 113–128, 2011.
13. Mikhail Bernadsky and Rajeev Alur. Symbolic analysis for GSMP models with one stateful clock. In *HSCC'07*, LNCS 4416, pages 90–103, 2007.
14. Nathalie Bertrand, Patricia Bouyer, Thomas Brihaye, and Nicolas Markey. Quantitative model-checking of one-clock timed automata under probabilistic semantics. In *QEST*, pages 55–64. IEEE Computer Society, 2008.

15. Patrick Billingsley. *Probability and measure*, volume 939. Wiley, 2012.
16. Patricia Bouyer, Thomas Brihaye, Marcin Jurdziński, and Quentin Menet. Almost-sure model-checking of reactive timed automata. In *QEST'12*, pages 138–147, 2012.
17. Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.
18. Alexandre David, Kim G. Larsen, Axel Legay, Marius Mikucionis, Danny Bøgsted Poulsen, Jonas van Vliet, and Zheng Wang. Statistical model checking for networks of priced timed automata. In *FORMATS'11*, LNCS 6919, pages 80–96, 2011.
19. Jean-Francois Kempf, Marius Bozga, and Oded Maler. As soon as probable: Optimal scheduling under stochastic uncertainty. In *TACAS*, pages 385–400, 2013.
20. M.A. Krasnosel'skij, E.A. Lifshits, and A.V. Sobolev. *Positive Linear Systems: the Method of Positive Operators*. Heldermann Verlag, Berlin, 1989.
21. Douglas Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.
22. M. Lothaire. *Applied Combinatorics on Words (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, New York, NY, USA, 2005.
23. W. Parry. Intrinsic Markov chains. *Transactions of the American Mathematical Society*, pages 55–66, 1964.
24. H.H. Schaefer and M.P.H. Wolff. *Topological vector spaces*, volume 3. Springer Verlag, 1999.
25. C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.