Journal of Statistical Physics (2019) 174:1239–1262
https://doi.org/10.1007/s10955-019-02233-3

Exact Results on the First Hitting via Conditional Strong Quasi-Stationary Times and Applications to Metastability

F. Manzo[1] · E. Scoppola[1]

Received: 18 July 2018 / Accepted: 17 January 2019 / Published online: 24 January 2019

Abstract

In the setting of non-reversible Markov chains on finite or countable state space, exact results on the distribution of the first hitting time to a given set G are obtained. A new notion of "conditional strong quasi stationary time" is introduced to describe the local relaxation time. This time is defined via a generalization of the strong stationary time. Rarity of the target set G is not required and the initial distribution can be completely general. The results clarify the the role played by the initial distribution on the exponential law; they are used to give a general notion of metastability and to discuss the relation between the exponential distribution of the first hitting time and metastability.

Keywords First hitting · Strong stationary time · Metastability

1 Introduction

The distribution of the first hitting time τ_G to a goal set G, is a classical topic widely discussed in the literature. In particular (see for instance [1–3,14]) it is well known that under suitable "rarity hypotheses" for the target set G and for a suitable starting measure α, the hitting time τ_G^α is approximately exponential.

In this paper we study the distribution of the hitting time in the general setting of non-reversible, ergodic Markov chains, starting from an arbitrary initial distribution α and we prove an exact (non-asymptotic) representation formula for $\mathbb{P}\left(\tau_G^\alpha > t\right)$ in terms of a new notion of "conditional strong quasi-stationary time" (CSQST in the following). We apply this representation formula to obtain a very detailed control on the exponential approximation of the exit law in terms of the ratio between a local relaxation time, related to the CSQST, and the hitting time τ_G. In particular, we get new information of the role of the starting distribution.

Heuristically, the system, under "rarity hypothesis" on the target set G, before reaching the target, thermalizes to a local equilibrium. From then on, G is reached after many trials, which give the exponential behavior of τ_G. This means that, for the exponential behavior of

✉ E. Scoppola
scoppola@mat.uniroma3.it

[1] Dipartimento di Matematica e Fisica, University of Roma "Roma Tre" Largo San Murialdo, 1 - 00146 Roma, Italy

τ_G, it is sufficient that the time needed to reach G is much longer than the time needed to relax to the local equilibrium.

The same scheme can be applied to describe metastable situations: here the system spends a long time in a set $A = G^c$, thermalizing on a "short" time-scale to a local equilibrium, often described by the *quasi-stationary measure*

$$\mu^*(\cdot) := \lim_{t\to\infty} \mathbb{P}\left(X_t = \cdot \mid t < \tau_G\right). \tag{1.1}$$

Only after a "long" time spent in A the system leaves this set and then rapidly reaches a stable state contained in G.

This time comparison is very common in the literature, see for instance [7]. Rarity hypotheses, as well as metastability hypotheses, are often given in terms of the ratio between the following two different time-scales: the "short" time-scale characterizing the approach to the local equilibrium, and the "long" time-scale characterizing the arrival to G. The precise definition of the short and long time-scales, however, vary according to the methods used by different authors and to the different regimes at issue.

In some of these regimes, hitting times to "renewal points" are a very powerful tool to describe the behavior of the chain. In particular, renewal ideas have been used in metastability literature (see e.g. [7,16]). Suppose that, asymptotically in some parameter of the system, μ^* concentrates on a single point m and the invariant measure π concentrates in G. In this case, the metastability hypothesis can be given in terms of a time comparison (see [12] for a discussion) as

$$\sup_{x\notin m\cup G} \frac{\mathbb{E}\tau^x_{m\cup G}}{\mathbb{E}\tau^m_G} \ll 1 \tag{1.2}$$

The idea is that if we observe the system on a time scale larger than the local relaxation time

$$R := \sup_{x\notin m\cup G} \mathbb{E}\tau^x_{m\cup G},$$

the process behaves like a two state chain, since all other points $x \notin m \cup G$ decay rapidly to m or G.

In this simplified scheme we see that the hitting time τ^α_G depends on the starting distribution α on A in a crucial way, since there can exist starting states in A "close" to the target G, i.e., states from which G is reached in a short time with large probability; for these starting states the law of τ_G is not expected to be exponential. It is natural to consider the initial time scale R in which the system makes its choice and relaxes either to m or to G. The notion of local relaxation time that we give in this paper is inspired by the time $\tau_{m\cup G}$, but, unlike all other choices in the literature, it leads to an exact representation formula for the hitting time in the general case: non reversible, non recurrent, for any initial state.

The key idea is to replace the hitting time to the metastable state m with a sort of "hitting time to the quasi-stationary measure", obtained via a generalization of the notion of strong stationary time (see [4,15]): We define a *conditional strong quasi-stationary time* (CSQST) τ^α_* satisfying

$$\mathbb{P}(X^\alpha_t = y,\ \tau^\alpha_* = t) = \mu^*(y)\mathbb{P}(\tau^\alpha_* = t < \tau^\alpha_G) \qquad \forall\ y \notin G, \quad \forall t \geq 0.$$

This means that the law of the process at the CSQST, τ^α_*, conditioned to remain in the complement of G, is the quasi-stationary one; under the conditioning $\tau^\alpha_* < \tau^\alpha_G$, the variables $X^\alpha_{\tau^\alpha_*}$ and τ^α_* are independent. Here and below α denotes the initial distribution.

CSQST's are the central object of this paper and give a very powerful description of the approach to the local equilibrium. The point is to use this CSQST in the decomposition

$$\mathbb{P}(\tau_G^\alpha > t) = \mathbb{P}(\tau_G^\alpha > t\ ;\ \tau_*^\alpha \le t) + \mathbb{P}(\tau_{*,G}^\alpha > t) \tag{1.3}$$

where $\tau_{*,G}^\alpha := \tau_G^\alpha \wedge \tau_*^\alpha$ takes the role of $\tau_{m\cup G}^\alpha$, the local relaxation time. Equation (1.3), has some interesting features that we will exploit in this paper in order to obtain bounds on the exponential approximation:

1. It is an exact formula, that does not require reversibility and holds for any initial state.
2. The conditional quasi-stationary property of τ_* allows to give exponential bounds on the first term in the r.h.s. of (1.3). Since $\mathbb{P}(\tau_G^{\mu^*} > t) = \lambda^t$ (see (1.7) for the definition of λ), and the event in the first term implies a visit to μ^*, its probability is proportional to λ^t.
3. The role of the starting measure α is explicit and will be related to the long-time asymptotics of the chain. Even when the system does not exhibit exponential behavior or α is outside the "basin of attraction of μ^*, Eq. (1.3) always holds. This exact decomposition formula, when looked on long times t, produces rich asymptotic results especially on the dependence of the asymptotic exponential law on the initial distribution. We will show that, in the long period, the main effect of the initial state can be encoded in a time-shift and that the first term in (1.3) asymptotically goes like $\lambda^{t+\delta_\alpha}$, where the time-shift δ_α is a real constant, possibly negative. To our knowledge, no other result in the literature gives a comparable control on the initial state.
4. Exponential behavior emerges when the second term $\mathbb{P}(\tau_{*,G}^\alpha > t)$ is negligible with respect to the first one in (1.3); we will show that it is natural to express the metastable hypothesis in terms of the comparison between the times $\tau_{*,G}^\alpha$ and $\tau_G^{\mu^*}$. The first represents the local relaxation time, to be compared with the hitting time. In this way, we get a probabilistic interpretation of the exponential approximation.

Let us mention that the idea of using a strong time that somehow catches the arrival to the quasi-stationary measure is not new in the literature; in [11], for a birth-and-death process starting from 0, in a particular regime, the authors construct what they call a *"strong quasi-stationary time"* for this purpose. Although the motivations are similar, our approach is different, our notion of conditional Strong Quasi Stationary Time is completely general and its existence does not require any additional assumptions besides ergodicity of the stochastic matrix outside G.

More difficult is to replace G by the stationary measure π. This problem is not discussed in the present paper.

As a final remark we have to note that, at least at the current stage, our result is not a computational tool, since the quantities appearing in our representation formula are in general difficult to calculate. Indeed, their computation requires a detailed control of the h-transform of the process (see the local chain $\widetilde{X}_t$ below). This is in some sense the common downside of very general results. The advantage is that general connections between dynamical quantities come to light. It would be interesting to determine a viable set of sufficient conditions ensuring the control of this local chain $\widetilde{X}_t$.

The paper is organized as follows.

- In Sect. 1.1.2 we introduce a local chain $\widetilde{X}_t$ on $A := G^c$ related to the Doob transform of P. This construction is useful in order to
  - determine the dependence of $\mathbb{P}(\tau_G^\alpha > t)$ on the initial distribution α in terms of the time shift δ_α;

  - control the distribution of the CSQST τ_*^α and the terms in (1.3). In particular its first term can be rewritten as $\mathbb{P}(\tau_G^\alpha > t\ ;\ \tau_*^\alpha \le t) = \lambda^{t+\delta_\alpha}\left(1 - \tilde{s}^{\tilde{\alpha}}(t)\right)$, where $\tilde{s}^{\tilde{\alpha}}(t)$ is the separation from stationarity for the law of the local Markov chain $\widetilde{X}_t^{\tilde{\alpha}}$ at time t and $\tilde{\alpha}$ is a measure on A induced by α.
  - obtain rough estimates of $\mathbb{P}(\tau_G^\alpha > t)$ in terms of $\tilde{s}^{\tilde{\alpha}}(t)$.

- In Sect. 1.2 we state our main results:

  - in Theorem 1 the existence of minimal CSQST;
  - in Theorem 2 the sub-multiplicative property of the distribution of $\tau_{*,G}^\alpha$
  - in Theorem 3 the exact representation formula for the hitting time τ_G^α;
  - in Theorem 4, under the additional hypothesis of time separation, we give sharp explicit estimates on the distribution of τ_G^α.

- In Sect. 2 we collect the proofs of our main results. We introduce an auxiliary process, the *tracking process*, to provide a construction of the CSQST, which is discussed in Sect. 2.2. In Sect. 2.5 we prove the representation formula for the hitting time.
  We use again the tracking process to construct the *ephemeral measure* in Sect. 2.3 describing the process before the CSQST. Even if the tracking process has transition probabilities depending on the initial distribution α and on time, the ephemeral measure, constructed with it, has a nice semigroup property that turns out to be the main ingredient in the proof of submultiplicativity of the distribution of $\tau_{*,G}^\alpha$, the local relaxation time.
- We give in Sect. 3 a simple example where the CSQST is explicitly constructed in terms of a sequence of hitting times. This example is also useful to discuss the relation between metastability and exponential distribution of the decay time.

1.1 General Setting, Definitions and Preliminary Remarks

1.1.1 Hypotheses and Notation

We collect in this subsection definitions and notation used in the paper.

- *Process* we will consider a discrete time Markov chain $\{X_t\}_{t\in\mathbb{N}}$ on a finite state space $\mathcal{X}$. Our results can be extended to the case of countable state space but for the sake of simplicity we consider the finite case. We denote by $P(x, y)$ the transition matrix and by $\mu_t^x(\cdot)$ the measure at time t, starting at x, i.e., $\mu_t^x(y) = \mathbb{P}(X_t^x = y) = P^t(x, y)$, for any $y \in \mathcal{X}$. More generally, given an initial distribution α on $\mathcal{X}$

$$\mu_t^\alpha(y) = \mathbb{P}(X_t^\alpha = y) = \sum_{x\in\mathcal{X}} \alpha(x) P^t(x, y)$$

  Starting conditions (starting state x or starting measure α) will be denoted by a superscript in random variables (i.e., X_t^x, X_t^α, τ^x, ...).
  Let $G \subset \mathcal{X}$ be a target set and τ_G its first hitting time

$$\tau_G := \min\{t \ge 0\ ;\ X_t \in G\}.$$

- *Ergodicity* we will study the process $\{X_t\}_{t\in\mathbb{N}}$ up to time τ_G, so it is not restrictive to consider G as a set of absorbing states. We assume ergodicity on $A := \mathcal{X}\setminus G$. More precisely, denoting by $[P]_A$ the sub-stochastic matrix obtained from P by restriction to A

$$[P]_A(x,y) = P(x,y) \qquad \forall x, y \in A, \qquad \text{with} \quad \sum_{y \in A} [P]_A(x,y) \le 1,$$

we suppose $[P]_A$ a *primitive matrix*, i.e., there exists an integer n such that $\big([P]_A\big)^n$ has strictly positive entries.

– *Separation* given two measures ν_1 and ν_2 on $\mathcal{X}$ their *separation* is defined by

$$sep(\nu_1, \nu_2) := \max_{y \in \mathcal{X}} \left[1 - \frac{\nu_1(y)}{\nu_2(y)}\right] \tag{1.4}$$

– *Scalar product* given two functions $a(x)$ and $b(x)$ on $A = \mathcal{X} \backslash G$ we define their *scalar product* as

$$a \cdot b := \sum_{x \in A} a(x) b(x).$$

– *Strong Stationary Time* (see [5] and [4]): a randomized stopping time τ_π^α is a *Strong Stationary Time (SST)* for the Markov chain X_t^α with starting distribution α and stationary measure π, if for any $t \ge 0$ and $y \in \mathcal{X}$

$$\mathbb{P}\left(X_t^\alpha = y, \tau_\pi^\alpha = t\right) = \pi(y) \mathbb{P}\left(\tau_\pi^\alpha = t\right).$$

This is equivalent to say

$$\mathbb{P}\left(X_t^\alpha = y \middle| \tau_\pi^\alpha \le t\right) = \pi(y) \tag{1.5}$$

If τ_π^α is a strong stationary time then

$$\mathbb{P}(\tau_\pi^\alpha > t) \ge sep(\mu_t^\alpha, \pi), \qquad \forall t \ge 0 \tag{1.6}$$

since by multiplying (1.5) by $\mathbb{P}\left(\tau_\pi^\alpha \le t\right)$ we get for any y:

$$\mathbb{P}\left(\tau_\pi^\alpha \le t\right) \pi(y) = \mathbb{P}\left(X_t^\alpha = y, \ \tau_\pi^\alpha \le t\right) \le \mathbb{P}\left(X_t^\alpha = y\right), \qquad \forall t \ge 0$$

When (1.6) holds with the equal sign for any t, the strong stationary time is *minimal*.

– *Quasi-stationary measure on A* by the Perron–Frobenius theorem, there exists $\lambda < 1$ such that λ is the spectral radius of $[P]_A$ and there exists a unique probability measure that is a non negative left eigenvector of $[P]_A$ corresponding to λ, i.e.,

$$\mu^* [P]_A = \lambda \mu^* \tag{1.7}$$

μ^* coincides with the quasi-stationary measure mentioned in the introduction. We get immediately

$$\mathbb{P}\left(\tau_G^{\mu^*} > t\right) = \lambda^t.$$

Moreover, the quasi-stationary measure μ^* satisfies the following equation (see [9]):

$$\mu^*(\cdot) = \lim_{t \to \infty} \mathbb{P}\left(X_t^x = \cdot \mid t < \tau_G^x\right) \qquad \forall x \in A. \tag{1.8}$$

– *Hitting distribution* starting from μ^*, the *hitting distribution* to G is defined for any $y \in G$ as

$$\begin{aligned} \omega(y) &:= \mathbb{P}\Big(X_{\tau_G}^{\mu^*} = y\Big) = \sum_{t \ge 1} \mathbb{P}\Big(X_t^{\mu^*} = y, \tau_G^{\mu^*} = t\Big) \\ &= \sum_{s \ge 0} \sum_{z \in A} \lambda^s \mu^*(z) P(z,y) = \frac{\sum_{z \in A} \mu^*(z) P(z,y)}{1 - \lambda}. \end{aligned} \tag{1.9}$$

– *Conditional strong quasi-stationary time* a randomized stopping time τ_*^α is a *conditional strong quasi-stationary time* (CSQST) if for any $y \in A$, and $t \geq 0$

$$\mathbb{P}(X_t^\alpha = y,\ \tau_*^\alpha = t) = \mu^*(y)\mathbb{P}(\tau_*^\alpha = t < \tau_G^\alpha). \tag{1.10}$$

or, in other words, for any $y \in A$ and $t \geq 0$

$$\mathbb{P}\left(X_t^\alpha = y, \tau_*^\alpha = t \mid t < \tau_G^\alpha\right) = \mu^*(y)\mathbb{P}\left(\tau_*^\alpha = t \mid t < \tau_G^\alpha\right) \tag{1.11}$$

which is equivalent to

$$\mathbb{P}\left(X_{\tau_*^\alpha}^\alpha = y \middle| \tau_*^\alpha < \tau_G^\alpha\right) = \mu^*(y). \tag{1.12}$$

1.1.2 The Local Chain $\widetilde{X}_t$ on A

In this subsection we construct an ergodic Markov chain $\widetilde{X}_t$ on A, that we call the *local chain*.

Many dynamics have been used in the literature to describe the local behavior of the process X_t on A. Examples are the reflected process or the conditioned process (see for instance [6,13]).

We use here a local chain $\widetilde{X}_t$ constructed by means of the right eigenvector of $[P]_A$ corresponding to λ. This construction is related to the Doob h-transform of $[P]_A$ (see for instance [15]). This chain $\widetilde{X}_t$ is also related to the "reversed chain" in Darroch–Seneta, introduced in [9] while considering the large time asymptotics.

The construction is the following: by the Perron-Frobenius theorem there exists a unique positive right eigenvector γ of $[P]_A$ corresponding to λ, i.e.,

$$[P]_A\gamma = \lambda\gamma \qquad \text{with normalization} \qquad \mu^* \cdot \gamma = 1. \tag{1.13}$$

This eigenvector is related to the asymptotic ratios of the survival probabilities (see eg [8])

$$\lim_{t\to\infty} \frac{\mathbb{P}(\tau_G^x > t)}{\mathbb{P}(\tau_G^y > t)} = \frac{\gamma(x)}{\gamma(y)} \qquad x, y \in A.$$

For any $x, y \in A$, define the stochastic matrix

$$\widetilde{P}(x, y) := \frac{\gamma(y)}{\gamma(x)} \frac{P(x, y)}{\lambda}. \tag{1.14}$$

Notice that $\widetilde{P}$ is a primitive matrix. Let ν be its invariant measure

$$\sum_{x\in A} \nu(x)\widetilde{P}(x, y) = \nu(y) = \sum_{x\in A} \nu(x)\frac{\gamma(y)}{\gamma(x)} \frac{P(x, y)}{\lambda}$$

it is easy to see that

$$\gamma(x) = \frac{\nu(x)}{\mu^*(x)}, \qquad \forall x \in A$$

For the chain $\widetilde{X}_t$ we define

$$\tilde{s}^x(t, y) := 1 - \frac{\widetilde{P}^t(x, y)}{\nu(y)}$$

$$\tilde{s}^x(t) = sep(\tilde{\mu}_t^x, \nu) = \sup_{y\in A} \tilde{s}^x(t, y), \quad \tilde{s}(t) := \sup_{x\in A} \tilde{s}^x(t).$$

Note that $\tilde{s}^x(t) \in [0, 1]$. Moreover, since $\widetilde{P}$ is a primitive matrix, it is well known (see for instance [5], Lemma 3.7) that $\tilde{s}(t)$ has the sub-multiplicative property:

$$\tilde{s}(t+u) \leq \tilde{s}(t)\tilde{s}(u).$$

This implies in particular an exponential decay in time of $\tilde{s}(t)$.

The relation between the local chain and the original chain X_t on $\mathcal{X}$ is given by the definition (1.14) and more generally by

$$\widetilde{P}^t(x, y) = \frac{\gamma(y)}{\gamma(x)} \frac{P^t(x, y)}{\lambda^t} \qquad \forall t \geq 0. \tag{1.15}$$

1.1.3 Preliminary Remarks

We can use this last relation to obtain a *rough estimate about the absorption time* τ_G^α. We give here this simple calculation in order to point out the dependence on the initial distribution α of the distribution of τ_G^α by means of a *time shift* defined by

$$\delta_\alpha := \log_\lambda \left(\alpha \cdot \gamma\right) \tag{1.16}$$

Note that $\delta_\alpha \in \Big[\log_\lambda \left(\min_{x\in A} \gamma(x)\right), \log_\lambda \left(\max_{x\in A} \gamma(x)\right)\Big]$ and $\delta_{\mu^*} = 0$.

We will show that it is natural to associate to every initial measure α the following measure $\tilde{\alpha}$ for the local chain $\widetilde{X}_t$:

$$\tilde{\alpha}(x) := \frac{\alpha(x)\gamma(x)}{\alpha \cdot \gamma}.$$

Indeed,

$$\begin{aligned}
\mathbb{P}(\tau_G^\alpha > t) &= \sum_{y\in A}\sum_{x\in A} \alpha(x) P^t(x, y) \\
&= \sum_{y\in A}\sum_{x\in A} \alpha(x)\gamma(x)\lambda^t \mu^*(y) \frac{\widetilde{P}^t(x, y)}{\nu(y)} \\
&= \lambda^t \sum_{x\in A} \alpha(x)\gamma(x) \sum_{y\in A} \mu^*(y)(1 - \tilde{s}^x(t, y)) \\
&= \lambda^{t+\delta_\alpha}\Big(1 - \sum_{y\in A} \mu^*(y)\tilde{s}^{\tilde{\alpha}}(t, y)\Big) \geq \lambda^{t+\delta_\alpha}\Big(1 - \tilde{s}^{\tilde{\alpha}}(t)\Big)
\end{aligned} \tag{1.17}$$

with

$$\tilde{s}^{\tilde{\alpha}}(t, y) := \sum_{x\in A} \tilde{\alpha}(x)\tilde{s}^x(t, y) \qquad \text{and} \qquad \tilde{s}^{\tilde{\alpha}}(t) := \sup_{y\in A} \tilde{s}^{\tilde{\alpha}}(t, y) \tag{1.18}$$

Note that from (1.17) we obtain for any initial distribution α

$$\lambda^{t+\delta_\alpha}(1 - \tilde{s}^{\tilde{\alpha}}(t)) \leq 1 \qquad \forall t \geq 0. \tag{1.19}$$

To obtain an upper bound on $\mathbb{P}(\tau_G^\alpha > t)$, we can consider the minimal strong stationary time $\tilde{\tau}_\nu^x$ (see [4]) such that

$$\mathbb{P}(\widetilde{X}_t^x = y,\ \tilde{\tau}_\nu^x = t) = \nu(y)\mathbb{P}(\tilde{\tau}_\nu^x = t)$$

with

$$\mathbb{P}(\tilde{\tau}_\nu^x > t) = \tilde{s}^x(t).$$

Similarly, we have

$$\begin{aligned}\mathbb{P}(\tau_G^\alpha > t) &= \sum_{y\in A}\sum_{x\in A}\alpha(x)\gamma(x)\lambda^t\frac{\mu^*(y)}{\nu(y)}\mathbb{P}(\widetilde{X}_t^x = y, \tilde{\tau}_\nu^x \le t)\\ &\quad+\sum_{y\in A}\sum_{x\in A}\alpha(x)\gamma(x)\lambda^t\frac{\mu^*(y)}{\nu(y)}\mathbb{P}(\widetilde{X}_t^x = y, \tilde{\tau}_\nu^x > t)\\ &\le \sum_{y\in A}\sum_{x\in A}\alpha(x)\gamma(x)\lambda^t\frac{\mu^*(y)}{\nu(y)}\nu(y)\mathbb{P}(\tilde{\tau}_\nu^x \le t)\\ &\quad+\frac{1}{\min_y\gamma(y)}\sum_{x\in A}\alpha(x)\gamma(x)\lambda^t\mathbb{P}(\tilde{\tau}_\nu^x > t)\\ &= \lambda^{t+\delta_\alpha}\Big[1+\tilde{s}^{\tilde{\alpha}}(t)\big(\frac{1}{\min_y\gamma(y)}-1\big)\Big].\end{aligned}$$

This quantity could be much larger that 1, since $\frac{1}{\min_y\gamma(y)} \ge 1$, and so this estimate from above on the distribution of τ_G^α is quite rough. However, we have to note that this factor is independent of time so that, for large t, due to the exponential decay of $\tilde{s}(t)$, and so of $\tilde{s}^{\tilde{\alpha}}(t)$, the estimate is not trivial.

It is interesting to notice that for sufficiently large t, say $t \ge \min\{n \in \mathbb{N} : \; n+\delta_\alpha \ge 0\}$, the separation $\tilde{s}^{\tilde{\alpha}}(t)$ has a straightforward meaning for the X_t process: it is related to the separation between the measure μ_t^α and the evolution starting from the quasi-stationary measure corrected with a time-shift δ_α, namely the measure

$$\mu_{t+\delta_\alpha}^{\mu^*}(y) := \begin{cases}\lambda^{t+\delta_\alpha}\mu^*(y) & \text{if } y\in A\\ 1-\lambda^{t+\delta_\alpha}\omega(y) & \text{if } y \in G\end{cases},$$

where ω is the hitting distribution defined in (1.9). More precisely, by the definition of the process $\widetilde{X}_t$ (see (1.15)) we have for any $y \in A$:

$$\tilde{s}^{\tilde{\alpha}}(t,y) := 1-\sum_{x\in A}\tilde{\alpha}(x)\frac{\tilde{P}^t(x,y)}{\nu(y)} = 1-\frac{\mu_t^\alpha(y)}{\mu_{t+\delta_\alpha}^{\mu^*}(y)}. \tag{1.20}$$

This means that

$$\sum_{x\in A}\tilde{\alpha}(x)\frac{\widetilde{P}^t(x,y)}{\nu(y)} = \frac{\mu_t^\alpha(y)}{\mu_{t+\delta_\alpha}^{\mu^*}(y)}, \qquad y\in A$$

so that the convergence to equilibrium of the local chain $\widetilde{X}_t^{\tilde{\alpha}}$ controls the convergence of the chain X_t^α to the evolution starting from the quasi-stationary measure corrected with a time-shift, $\mu_{t+\delta_\alpha}^{\mu^*}$, as far as its permanence in the set A is concerned. This is the reason why the local chain $\widetilde{X}_t$ is crucial in our discussion.

1.2 Main Results

We collect in this section our main results on conditional strong quasi stationary times (CSQST) and their application to control the distribution of the hitting time τ_G.

From the definition of CSQST (1.10) we can prove the following:

Proposition 1 *For any initial distribution α on A and for any conditional strong quasi stationary time τ_*^α we have for any $t \geq 0$:*

$$\mathbb{P}(\tau_*^\alpha \leq t < \tau_G^\alpha) = \sum_{u \leq t} \lambda^{t-u} \mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha) \leq \lambda^{t+\delta_\alpha}(1 - \tilde{s}^{\tilde{\alpha}}(t)).$$

Proposition 1 suggests a new notion of minimality.

Definition 1 For any initial distribution α on A a conditional strong quasi stationary time τ_*^α is *minimal* if for any $t \geq 0$:

$$\mathbb{P}(\tau_*^\alpha \leq t < \tau_G^\alpha) = \lambda^{t+\delta_\alpha}(1 - \tilde{s}^{\tilde{\alpha}}(t)).$$

The existence of minimal conditional strong quasi-stationary times is given by the following Theorem.

Theorem 1 *For any initial distribution α on A, there exists a* minimal conditional strong quasi stationary time τ_*^α *such that for any $t > 0$*

$$\mathbb{P}(\tau_*^\alpha = t < \tau_G^\alpha) = \lambda^{t+\delta_\alpha}(\tilde{s}^{\tilde{\alpha}}(t-1) - \tilde{s}^{\tilde{\alpha}}(t)).$$

Note that in particular for a minimal conditional strong quasi stationary time we have for $t \geq 0$

$$\mathbb{P}(\tau_*^\alpha > t,\ \tau_*^\alpha < \tau_G^\alpha) = \sum_{u > t} \lambda^{u+\delta_\alpha}(\tilde{s}^{\tilde{\alpha}}(u-1) - \tilde{s}^{\tilde{\alpha}}(u)) \leq \lambda^{t+\delta_\alpha} \tilde{s}^{\tilde{\alpha}}(t).$$

Let τ_*^α be a minimal CSQST and define

$$\tau_{*,G}^\alpha = \tau_G^\alpha \wedge \tau_*^\alpha.$$

This time plays the role of *local relaxation time* or *metastability time*, like $\tau_{m \cup G}$ in the metastable hypothesis (1.2). It is a sub-multiplicative time:

Theorem 2 *If τ_*^α is a minimal CSQST, then for any positive u and v*

$$\sup_\alpha \mathbb{P}\Big(\tau_{*,G}^\alpha > u + v\Big) \leq \sup_\alpha \mathbb{P}\Big(\tau_{*,G}^\alpha > u\Big) \sup_\alpha \mathbb{P}\Big(\tau_{*,G}^\alpha > v\Big). \tag{1.21}$$

The local relaxation time $\tau_{*,G}^\alpha$ is a key ingredient in the following representation formula:

Theorem 3 *For any initial distribution α on A, if τ_*^α is a minimal conditional strong quasi stationary time, we have, for any $t \geq 0$*

$$\mathbb{P}\Big(\tau_G^\alpha > t\Big) = \lambda^{t+\delta_\alpha}(1 - \tilde{s}^{\tilde{\alpha}}(t)) + \mathbb{P}\Big(\tau_{*,G}^\alpha > t\Big) \tag{1.22}$$

Moreover, for any $y \in G$, we have

$$\mathbb{P}\Big(X_{\tau_G^\alpha}^\alpha = y\Big) = \mathbb{P}\Big(\tau_G^\alpha < \tau_*^\alpha,\ X_{\tau_G^\alpha}^\alpha = y\Big) + \omega(y)\mathbb{P}\Big(\tau_G^\alpha > \tau_*^\alpha\Big), \tag{1.23}$$

where ω is the hitting distribution starting from μ^* (see Eq. (1.9)).

This theorem provides a control on the convergence to an exponential distribution for the hitting time τ_G and on the hitting distribution and it gives a probabilistic interpretation of the errors in the exponential approximation of $\mathbb{P}\Big(\tau_G^\alpha > t\Big)$ in terms of conditional strong quasi-stationary times.

In order to obtain a multiplicative bound on the exponential distribution, it is useful to rewrite Eq. (1.22) as

$$\frac{\mathbb{P}\left(\tau_G^\alpha > t\right)}{\lambda^{t+\delta_\alpha}} - 1 = -\tilde{s}^{\tilde{\alpha}}(t) + \lambda^{-t-\delta_\alpha}\mathbb{P}\left(\tau_{*,G}^\alpha > t\right). \tag{1.24}$$

The first error term $-\tilde{s}^{\tilde{\alpha}}(t)$ decays exponentially fast in t and it will be easy to deal with; the second error term $\lambda^{-t-\delta_\alpha}\mathbb{P}\left(\tau_{*,G}^\alpha > t\right)$ will decay faster than the leading term only under suitable metastability hypotheses.

Let τ_*^α be a minimal CSQST. Define the *local relaxation time scale* as

$$R := \sup_\alpha \mathbb{E}\left(\tau_{*,G}^\alpha\right) \tag{1.25}$$

and the *relaxation time*

$$T := (1-\lambda)^{-1} = \mathbb{E}\left(\tau_G^{\mu^*}\right) \tag{1.26}$$

Definition 2 We call *time-separation of rate* a, with $a > 0$, the condition

$$\frac{R}{T} = \left(1 - \frac{1}{T}\right)^T e^{-a} = \lambda^T e^{-a}. \tag{1.27}$$

Note that $e^{-1/\lambda} \le \lambda^T \le e^{-1}$ become strict bounds when T is large, as in metastable situations.

Theorem 4 *Under the time-separation hypothesis of rate a, given in Definition 2, for any initial measure* α *on* A *and for any positive integer* n*,*

$$\left|\frac{\mathbb{P}\left(\tau_G^\alpha > nT\right)}{\lambda^{nT+\delta_\alpha}} - 1\right| < e^{-an}\lambda^{-\delta_\alpha}\frac{e^{1/\lambda}}{1-e^{-a}}. \tag{1.28}$$

Note that the time-separation hypothesis (1.27) does not exclude the existence of starting states $x \in A$ from which the process reaches G in a very short time with high probability. When the starting distribution α is concentrated on such states, we expect to have λ^{δ_α} very small. This implies that Theorem 4 provides a sharp result, in the case of small n, only if the parameter a in (1.27) is sufficiently large and λ^{δ_α} is not too small. More precisely, if the support of the starting measure α is contained in a "basin of attraction of the metastable state" defined for instance as (see [13])

$$B := \left\{x \in A :\ \mathbb{P}(\tau_G^x > 2R) > 3/4\right\},$$

we can give a very rough estimate $\lambda^{\delta_\alpha} \ge 1/4$.

Indeed by using the trivial estimate

$$\begin{aligned}\mathbb{P}(\tau_G^\alpha > nT) &= \sum_{y\in A}\mathbb{P}(X_{2R}^\alpha = y)\mathbb{P}(\tau_G^y > nT - 2R) \ge \lambda^{nT-2R}\mathbb{P}(\tau_*^\alpha \le 2R < \tau_G^\alpha)\\ &= \lambda^{nT-2R}\Big[\mathbb{P}(\tau_G^\alpha > 2R) - \mathbb{P}(\tau_{*,G}^\alpha > 2R)\Big],\end{aligned}$$

by Theorem 4 and the Markov inequality we get

$$\lambda^{\delta_\alpha} = \lim_{n\to\infty}\frac{\mathbb{P}(\tau_G^\alpha > nT)}{\lambda^{nT}} \ge \lambda^{-2R}\Big[\mathbb{P}(\tau_G^\alpha > 2R) - \mathbb{P}(\tau_{*,G}^\alpha > 2R)\Big] \ge 1/4.$$

In many applications it is interesting to study the behavior of the process on an intermediate time-scale S, say $R \ll S \ll T$. The process has an *early exponential behavior* if Eq. (1.28) holds when replacing T with S. In [12] the early exponential behavior of the first hitting time is proved in a particular case, with a particular starting configuration. In our setting, we can study the early behavior starting from a general measure α under time-separation hypothesis with large rate a.

2 Proofs

2.1 Tracking Process

In order to prove the existence of a minimal CSQST, we introduce an auxiliary *tracking process*. The construction is inspired to [5] and [10], where the existence of strong stationary times is proved. The idea is to duplicate the state space into two layers and to define a process on this larger state space with a jump probability from one layer to the other one. In order to have a general construction, we introduce first a control function to define the jump rate.

Definition 3 Let $\mathbb{Z}_{\geq -1}$ denotes integers larger or equal to -1. The function $m(t) : \mathbb{Z}_{\geq -1} \longrightarrow [0, 1]$ is a *control function for the process starting at* α if it is a monotonic decreasing function with

$$m(t) \geq \tilde{s}^{\tilde{\alpha}}(t) \text{ for } t \geq 0, \qquad m(-1) = 1.$$

Given a control function $m(t)$ for every $z \in A$, we define the following *jump probabilities* for any $t \geq 0$

$$J^{\alpha}(t, z) := \frac{m(t-1) - m(t)}{m(t-1) - \tilde{s}^{\tilde{\alpha}}(t, z)}, \tag{2.1}$$

with the convention $0/0 = 0$. Since $m(t-1) \geq m(t) \geq s^{\tilde{\alpha}}(t, z)$, we have $J^{\alpha}(t, z) \in [0, 1]$ for any $z \in A$ and any t. For any $t \geq 0$ and any $z \in G$ we define

$$J^{\alpha}(t, z) \equiv J^{\alpha}(t, G) := 1.$$

Definition 4 On the state space $\boldsymbol{\mathcal{X}} := \mathcal{X} \times \{0, 1\}$, consider the transition matrix

$$\begin{aligned} Q_t^{\alpha}((y, 0), (z, 0)) &:= P(y, z)\Big(1 - J^{\alpha}(t, z)\Big), \\ Q_t^{\alpha}((y, 0), (z, 1)) &:= P(y, z) J^{\alpha}(t, z), \\ Q_t^{\alpha}((y, 1), (z, e)) &:= P(y, z) \mathbb{1}_{\{e=1\}}, \end{aligned} \tag{2.2}$$

where $e \in \{0; 1\}$; also, consider the initial distribution $\boldsymbol{\alpha}$ on the two layers of $\boldsymbol{\mathcal{X}}$, defined, for any $x \in A$, as

$$\begin{aligned} \boldsymbol{\alpha}(x, 0) &:= \alpha(x)\big(1 - J^{\alpha}(0, x)\big) = \alpha(x) - \lambda^{\delta_{\alpha}} \mu^*(x)(1 - m(0)), \\ \boldsymbol{\alpha}(x, 1) &:= \alpha(x) J^{\alpha}(0, x) = \lambda^{\delta_{\alpha}} \mu^*(x)(1 - m(0)). \end{aligned} \tag{2.3}$$

We define the *tracking process* $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$ via

$$\mathbb{P}\left(\bigcap_{u=0}^{t} (\boldsymbol{X}_u^{\boldsymbol{\alpha}} = \boldsymbol{y}_u)\right) = \boldsymbol{\alpha}(\boldsymbol{y}_0) \prod_{u=0}^{t-1} Q_u^{\alpha}(\boldsymbol{y}_u, \boldsymbol{y}_{u+1})$$

with $\boldsymbol{y}_u \in \boldsymbol{\mathcal{X}}$ for any $u \leq t$.

By (2.2), (2.3) it is immediate to see that the marginal distribution of $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$ on $\mathcal{X}$ corresponds to the distribution of X_t^α, so that we can study each event defined for the process X_t^α in terms of sets of paths of the process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$. For this reason, with an abuse of notation, we denote with the same symbol $\mathbb{P}$ the probability of events defined in terms of the process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$.

Notice that unlike the process defined in [5] and [10] for the strong stationary times, in our definition of the jump rates we use the separation $\tilde{s}^{\tilde{\alpha}}$ for the process $\widetilde{X}_t^{\tilde{\alpha}}$, defined on $\widetilde{\mathcal{X}}$.

We want also to note that the starting measure α appears as a parameter in the definition of the transition matrix Q (see (2.1), (2.2)), the process is time-inhomogeneous and Markov property does not hold. However, we can get rid of this dependence and recover a sort of semigroup property by considering a suitable conditioning of the process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$. We will clarify this point, that represents a crucial ingredient in our approach, in Sect. 2.3.

We will be interested to the process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$ up to its first hitting to the set $\mathcal{X} \times \{1\}$, i.e. for $t \leq \tau_1^{\boldsymbol{\alpha}}$ with

$$\tau_1^{\boldsymbol{\alpha}} := \tau_{\mathcal{X}\times\{1\}}^{\boldsymbol{\alpha}} = \min\{t \geq 0\,;\ \boldsymbol{X}_t^{\boldsymbol{\alpha}} = (y, 1) \text{ for some } y \in \mathcal{X}\}, \tag{2.4}$$

indeed, we prove that $\tau_1^{\boldsymbol{\alpha}}$ is a conditional strong quasi-stationary time.

This construction of a CSQST is quite implicit, for it requires the knowledge of the separation $\tilde{s}^{\tilde{\alpha}}(t)$ at any time, which in general is very hard to obtain. In this paper we use CSQST as a theoretical tool and we are not concerned with their explicit construction. However, it is well-known that in some systems the separation can be estimated with the distribution of a hitting time to a suitable halting state (see e.g. [15]). In the example in Sect. 3, we exploit this idea to construct explicitly a (non-minimal) CSQST.

2.2 Conditional Strong Quasi Stationary Times (CSQST)

In this section we prove Proposition 1 and Theorem 1.

Let us start by proving Proposition 1: by the definition of CSQST we have for any $y \in A$

$$\mathbb{P}(X_t^\alpha = y,\ \tau_*^\alpha = u) = \mu^*(y)\mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha), \qquad \text{for any } \ u \geq 0 \tag{2.5}$$

If (2.5) hods for any $u \in [0, t]$ then we have:

$$\begin{aligned}
\mathbb{P}(X_t^\alpha = y,\ \tau_*^\alpha \leq t) &= \sum_{u\leq t}\sum_{z\in A}\mathbb{P}(X_u^\alpha = z,\ \tau_*^\alpha = u)P^{t-u}(z, y)\\
&= \sum_{u\leq t}\sum_{z\in A}\mu^*(z)\mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha)P^{t-u}(z, y)\\
&= \mu^*(y)\sum_{u\leq t}\lambda^{t-u}\mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha)
\end{aligned} \tag{2.6}$$

and by summing over $y \in A$:

$$\mathbb{P}(\tau_*^\alpha \leq t < \tau_G^\alpha) = \sum_{u\leq t}\lambda^{t-u}\mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha). \tag{2.7}$$

Moreover for any $y \in A$ we have

$$\begin{aligned}
\mu_t^\alpha(y) \geq \mathbb{P}(\tau_*^\alpha \leq t, X_t^\alpha = y) &= \sum_{u\leq t}\sum_{z\in A}\mathbb{P}(\tau_*^\alpha = u, X_u^\alpha = z)P^{t-u}(z, y)\\
= \lambda^t\sum_{u\leq t}\lambda^{-u}\mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha)\mu^*(y)
\end{aligned}$$

so that by (1.20)

$$\frac{\mu_t^\alpha(y)}{\lambda^t \mu^*(y)} = \lambda^{\delta_\alpha}(1 - \tilde{s}^{\tilde{\alpha}}(t, y)) \geq \sum_{u \leq t} \lambda^{-u} \mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha) = \mathbb{P}(\tau_*^\alpha \leq t < \tau_G^\alpha)\lambda^{-t}$$

since this holds for any $y \in A$ we get

$$\mathbb{P}(\tau_*^\alpha \leq t < \tau_G^\alpha) = \sum_{u \leq t} \lambda^{t-u} \mathbb{P}(\tau_*^\alpha = u < \tau_G^\alpha) \leq \lambda^{t+\delta_\alpha}(1 - \tilde{s}^{\tilde{\alpha}}(t)). \tag{2.8}$$

□

We prove a stronger version of Theorem 1. Indeed by choosing the control function $m(t) = \tilde{s}^{\tilde{\alpha}}(t)$, Theorem 1 immediately follows by:

Theorem 5 *For any initial distribution* α *on* A *and for any control function* $m(t)$ *there exists* a conditional strong quasi stationary time τ_*^α *such that*

$$\mathbb{P}(\tau_*^\alpha \leq t < \tau_G^\alpha) = \lambda^{t+\delta_\alpha}(1 - m(t)) \qquad \textit{for all } \; t \geq 0$$

with

$$\mathbb{P}(\tau_*^\alpha = t < \tau_G^\alpha) = \lambda^{t+\delta_\alpha}(m(t-1) - m(t)).$$

To prove Theorem 5 consider now the tracking process defined in Sect. 2.1 and the hitting time

$$\tau_1^{\boldsymbol{\alpha}} := \tau_{\mathcal{X}\times\{1\}}^{\boldsymbol{\alpha}} = \min\{t \geq 0\,;\; \boldsymbol{X}_t^{\boldsymbol{\alpha}} = (y, 1) \text{ for some } y \in \mathcal{X}\}.$$

We will prove that $\tau_1^{\boldsymbol{\alpha}}$ satisfies the following condition for any $t \geq 0$:

$$\mathcal{C}(t) := \begin{cases} \mathbb{P}(X_t^\alpha = y,\ \tau_1^{\boldsymbol{\alpha}} = t) = \mu^*(y)\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = t < \tau_G^\alpha) & \text{for any } y \in A \\ \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = t < \tau_G^\alpha) = \lambda^{t+\delta_\alpha}\big(m(t-1) - m(t)\big) \end{cases}$$

If $\mathcal{C}(u)$ is verified for any $u \leq t$, we can conclude

$$\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \leq t < \tau_G^\alpha) = \sum_{u \in [0,t]} \lambda^{t-u} \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u < \tau_G^\alpha) = \lambda^{t+\delta_\alpha}\big(1 - m(t)\big).$$

In order to prove that $\tau_1^{\boldsymbol{\alpha}}$ satisfies $\mathcal{C}(t)$ for all $t \geq 0$ we proceed by induction on t. For $t = 0$, by the definition of the initial distribution $\boldsymbol{\alpha}$ in definition 4 we immediately verify $\mathcal{C}(0)$. Indeed, for $y \in A$ we get

$$\begin{aligned} \mathbb{P}(X_0^\alpha = y,\ \tau_1^{\boldsymbol{\alpha}} = 0) = \mathbb{P}(\boldsymbol{X}_0^{\boldsymbol{\alpha}} = (y, 1)) &= \lambda^{\delta_\alpha}\mu^*(y)\big(1 - m(0)\big) \\ &= \mu^*(y)\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = 0 < \tau_G^\alpha) \end{aligned}$$

with

$$\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = 0 < \tau_G^\alpha) = \sum_{x \in A} \mathbb{P}(\boldsymbol{X}_0^{\boldsymbol{\alpha}} = (x, 1)) = \lambda^{\delta_\alpha}\big(1 - m(0)\big).$$

To prove the induction step we use the following:

Lemma 1 *If for any* $u \in [0, t]$ *and for any* $y \in A$ *we have*

$$\mathbb{P}\big(X_u^\alpha = y,\ \tau_1^{\boldsymbol{\alpha}} = u\big) = \mu^*(y)\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u < \tau_G^\alpha)$$

then, for any $z \in A$,

$$\mathbb{P}\big(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (z, 1)\big) = \mu^*(z)\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \leq t < \tau_G^\alpha).$$

Proof Note first that under the hypothesis of the Lemma, by (2.6) we get

$$\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \le t < \tau_G^{\alpha}) = \sum_{u \le t} \lambda^{t-u} \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u < \tau_G^{\alpha}). \tag{2.9}$$

We have

$$\begin{aligned}
\mathbb{P}\big(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (z,1)\big) &= \sum_{u \le t} \sum_{y \in A} \mathbb{P}\big(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (z,1),\ \tau_1^{\boldsymbol{\alpha}} = u,\ X_u^{\alpha} = y\big) \\
&= \sum_{u \le t} \sum_{y \in A} \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u,\ X_u^{\alpha} = y) P^{t-u}(y,z) \\
&= \sum_{u \le t} \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u) \sum_{y \in A} \mu^*(y) P^{t-u}(y,z) \\
&= \mu^*(z) \sum_{u \le t} \lambda^{t-u} \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u < \tau_G^{\alpha}) = \mu^*(z) \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \le t < \tau_G^{\alpha}).
\end{aligned}$$

□

Suppose now that $\mathcal{C}(u)$ holds for $u \in [0,t]$. By using then Lemma 1 we get

$$\begin{aligned}
&\mathbb{P}\big(X_{t+1}^{\alpha} = y, \tau_1^{\boldsymbol{\alpha}} = t+1\big) \\
&\quad = \sum_{z \in \mathcal{X}} \mathbb{P}\big(\boldsymbol{X}_{t+1}^{\boldsymbol{\alpha}} = (y,1) | \boldsymbol{X}_t^{\boldsymbol{\alpha}} = (z,0)\big) \mathbb{P}\big(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (z,0)\big) \\
&\quad = \sum_{z \in \mathcal{X}} P(z,y) J^{\alpha}(t+1,y) \big[\mu_t^{\alpha}(z) - \mathbb{P}(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (z,1))\big] \\
&\quad = J^{\alpha}(t+1,y) \left[\mu_{t+1}^{\alpha}(y) - \sum_{z \in \mathcal{X}} \mu^*(z) P(z,y) \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \le t < \tau_G^{\alpha})\right].
\end{aligned} \tag{2.10}$$

Since $\mathcal{C}(u)$ holds for $u \in [0,t]$ we have

$$\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \le t < \tau_G^{\alpha}) = \sum_{u \le t} \lambda^{t-u} \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} = u < \tau_G^{\alpha}) = \lambda^{t+\delta_\alpha}(1 - m(t))$$

Recalling that, by (1.20),

$$\mu_{t+1}^{\alpha}(y) = \lambda^{t+1+\delta_\alpha} \mu^*(y) [1 - \tilde{s}^{\tilde{\alpha}}(t+1,y)],$$

we obtain

$$\mathbb{P}\big(X_{t+1}^{\alpha} = y, \tau_1^{\boldsymbol{\alpha}} = t+1\big) = J^{\alpha}(t+1,y) \lambda^{t+1+\delta_\alpha} \mu^*(y) \big[1 - \tilde{s}^{\tilde{\alpha}}(t+1,y) - (1 - m(t))\big].$$

By using the definition of $J^{\alpha}(t+1,y)$, we get

$$\mathbb{P}\big(X_{t+1}^{\alpha} = y, \tau_1^{\boldsymbol{\alpha}} = t+1\big) = [m(t) - m(t+1)] \lambda^{t+1+\delta_\alpha} \mu^*(y)$$

so that, by summing on $y \in A$

$$\mathbb{P}\big(\tau_1^{\boldsymbol{\alpha}} = t+1 < \tau_G^{\alpha}\big) = [m(t) - m(t+1)] \lambda^{t+1+\delta_\alpha},$$

we show that $\mathcal{C}(t+1)$ holds, concluding the proof of Theorem 5 and so of Theorem 1. □

2.3 Ephemeral Measure

In this section we describe the behavior of the process before $\tau_1^{\boldsymbol{\alpha}}$. We call this behavior "ephemeral" since in metastable situations $\tau_1^{\boldsymbol{\alpha}}$ is typically much smaller than τ_G.

Consider the tracking process before $\tau_1^{\boldsymbol{\alpha}}$, more precisely, the conditioned measure on $\mathcal{X} \times \{0\}$ obtained by the process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$ conditioned to the layer $\{0\}$:

Definition 5 The measure

$$\Phi_t^{\boldsymbol{\alpha}}(x) := \mathbb{P}\left(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (x,0) \,|\tau_1^{\boldsymbol{\alpha}} > t\right). \tag{2.11}$$

is called the *ephemeral measure*.

With a slight abuse of notation, we consider this ephemeral measure either as a measure on $\boldsymbol{\mathcal{X}}$ (with support in $A \times \{0\}$) or as a measure in A. Recalling that $\boldsymbol{\alpha}(x,0) = \alpha(x)\Big(1 - J^{\alpha}(0,x)\Big)$ we get for the ephemeral measure:

$$\begin{aligned}\Phi_t^{\boldsymbol{\alpha}}(x) = &\frac{1}{\mathbb{P}\left(\tau_1^{\boldsymbol{\alpha}} > t\right)} \sum_{x_0 \in A} \alpha(x_0)\Big(1 - J^{\alpha}(0,x_0)\Big) \\ &\times \sum_{x_1,\ldots,x_{t-1}} \left[\prod_{s=1}^{t-1} P(x_{s-1},x_s)\Big(1 - J^{\alpha}(s,x_s)\Big)\right] P(x_{t-1},y)\Big(1 - J^{\alpha}(t,y)\Big)\end{aligned}$$

We will prove the following "Markov-like" properties for the tracking process and for the hitting time $\tau_1^{\boldsymbol{\alpha}}$:

Proposition 2 *Consider the tracking process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$ starting at $\boldsymbol{\alpha}$ and with control function*

$$m(t) = \tilde{s}^{\tilde{\alpha}}(t) \quad \forall t \geq 0 \tag{2.12}$$

then for any $x \in A$

$$\Phi_{t+u}^{\boldsymbol{\alpha}}(x) = \Phi_u^{\Phi_t^{\boldsymbol{\alpha}}}(x)$$

Proposition 3 *Consider the tracking process $\boldsymbol{X}_t^{\boldsymbol{\alpha}}$ with control function $m(t) = \tilde{s}^{\tilde{\alpha}}(t), \quad \forall t \geq$ 0, then*

$$\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t+u) = \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)\mathbb{P}(\tau_1^{\Phi_t^{\boldsymbol{\alpha}}} > u)$$

From this Proposition the submultiplicativity property of Theorem 2 easily follows (see Sect. 2.4).

To prove these propositions, we introduce two technical lemmas to obtain the crucial property on the jump rates given in Lemma 4.

Recalling from Sect. 1.1.2 that

$$\tilde{\alpha}(x) = \frac{\alpha(x)\gamma(x)}{\alpha \cdot \gamma},$$

we denote by $\tilde{\Phi}_t^{\tilde{\alpha}}$ the measure

$$\tilde{\Phi}_t^{\tilde{\alpha}}(x) = \frac{\Phi_t^{\boldsymbol{\alpha}}(x)\gamma(x)}{\Phi_t^{\boldsymbol{\alpha}} \cdot \gamma}.$$

Lemma 2 *For all t and α there exist real numbers K and H such that for all x and $y \in A$,*

$$\tilde{\Phi}_t^{\tilde{\alpha}}(y) = K \sum_{x \in A} \tilde{\alpha}(x) \tilde{P}^t(x, y) - H\nu(y).$$

Proof By the CSQST property of $\tau_1^{\boldsymbol{\alpha}}$, we see that

$$\begin{aligned} \Phi_t^{\boldsymbol{\alpha}}(y) &= \frac{\mathbb{P}\left(X_t^\alpha = y\right) - \mathbb{P}\left(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (y, 1)\right)}{\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)} \\ &= K' \mu_t^\alpha(y) - H' \mu^*(y) \end{aligned}$$

with $K' = 1/\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)$ and $H' = \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} \le t < \tau_G^\alpha)/\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)$. By plugging this equation into the definition of $\tilde{\Phi}_t^{\tilde{\alpha}}$, we get

$$\frac{\Phi_t^{\boldsymbol{\alpha}}(y)\gamma(y)}{\Phi_t^{\boldsymbol{\alpha}} \cdot \gamma} = \frac{\sum_{x\in A} \alpha(x) \left(K' P_{x,y}^t - H'\mu^*(y)\right) \gamma(y)}{\Phi_t^{\boldsymbol{\alpha}} \cdot \gamma}, \tag{2.13}$$

by using (1.15) and $\nu(x) = \mu^*(x)\gamma(x)$, we get

$$\text{(r.h.s. of 2.13)} = K'' \sum_{x\in A} \alpha(x) \frac{\gamma(x)}{\gamma(y)} \lambda^t \tilde{P}_{x,y}^t \gamma(y) - H\nu(y),$$

where $K'' = \frac{K'}{\Phi_t^{\boldsymbol{\alpha}} \cdot \gamma}$ and $H = \frac{H'}{\Phi_t^{\boldsymbol{\alpha}} \cdot \gamma}$ do not depend on y. Since $\tilde{\alpha}(x) = \frac{\alpha(x)\gamma(x)}{\alpha \cdot \gamma}$, we immediately get the thesis with $K = \lambda^t K'' \alpha \cdot \gamma$. □

With the help of Lemma 2, we can prove the following iteration formula for the separation $\tilde{s}^{\tilde{\alpha}}(t, y)$:

Lemma 3 *For all t and α there exist real numbers U and V such that for all u and $z \in A$*

$$\tilde{s}^{\tilde{\alpha}}(t + u, z) = U \tilde{s}^{\tilde{\Phi}_t^{\tilde{\alpha}}}(u, z) + V$$

Proof

$$1 - \tilde{s}^{\tilde{\Phi}_t^{\tilde{\alpha}}}(u, z) = \frac{\sum_{x\in A} \tilde{\Phi}_t^\alpha(x) \widetilde{P}_{x,z}^u}{\nu(z)}. \tag{2.14}$$

By Lemma 2,

$$\begin{aligned} \text{(r.h.s of 2.14)} &= \frac{\sum_{x\in A} \left(K \tilde{\mu}_t^{\tilde{\alpha}}(x) - H\nu(x)\right) \widetilde{P}_{x,z}^u}{\nu(z)} \\ &= K \frac{\tilde{\mu}_{t+u}^{\tilde{\alpha}}(z)}{\nu(z)} - H = K\left(1 - \tilde{s}^{\tilde{\alpha}}(t + u, z)\right) - H \end{aligned}$$

and the thesis follows immediately with $U = 1/K$ and $V = 1 - \frac{1+H}{K}$. □

A corollary of this result is the following

Lemma 4 *For any initial measure α on A , any time t and u and $z \in \mathcal{X}$*

$$J^\alpha(t + u, z) = J^{\Phi_t^{\boldsymbol{\alpha}}}(u, z)$$

Proof By direct computation, if $z \in A$,

$$J^{\alpha}(t+u,z) = \frac{\tilde{s}^{\tilde{\alpha}}(t+u-1) - \tilde{s}^{\tilde{\alpha}}(t+u)}{\tilde{s}^{\tilde{\alpha}}(t+u-1) - \tilde{s}^{\tilde{\alpha}}(t+u,z)} = \frac{\tilde{s}^{\tilde{\Phi}_t^{\tilde{\alpha}}}(u-1) - \tilde{s}^{\tilde{\Phi}_t^{\tilde{\alpha}}}(u)}{\tilde{s}^{\tilde{\Phi}_t^{\tilde{\alpha}}}(u-1) - \tilde{s}^{\tilde{\Phi}_t^{\tilde{\alpha}}}(u,z)} = J^{\Phi_t^{\boldsymbol{\alpha}}}(u,z);$$

while, for $z \in G$, we have $J^{\alpha}(t+u,z) = J^{\Phi_t^{\boldsymbol{\alpha}}}(u,z) = 1$. □

Proof of Proposition 3 By the Markov property and Lemma 4,

$$\begin{aligned}
&\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t+u) \\
&= \sum_{x_t \in A} \mathbb{P}(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (x_t, 0)) \sum_{x_{t+1},\ldots,x_{t+u}} \prod_{v=1}^{u} P(x_{t+v-1}, x_{t+v}) \Big(1 - J^{\alpha}(t+v, x_{t+v})\Big) \\
&= \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t) \sum_{x_t \in A} \Phi_t^{\boldsymbol{\alpha}}(x_t) \sum_{x_{t+1},\ldots,x_{t+u}} \prod_{v=1}^{u} P(x_{t+v-1}, x_{t+v}) \Big(1 - J^{\Phi_t^{\boldsymbol{\alpha}}}(v, x_{t+v})\Big) \\
&= \mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)\mathbb{P}(\tau_1^{\Phi_t^{\boldsymbol{\alpha}}} > u)
\end{aligned}$$

□

Proof of Proposition 2 With the same expansion we write

$$\begin{aligned}
&\Phi_{t+u}^{\boldsymbol{\alpha}}(x_{t+u}) \\
&= \sum_{x_t \in A} \frac{\mathbb{P}(\boldsymbol{X}_t^{\boldsymbol{\alpha}} = (x_t, 0))}{\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t+u)} \sum_{x_{t+1},\ldots,x_{t+u-1}} \prod_{v=1}^{u} P(x_{t+v-1}, x_{t+v}) \Big(1 - J^{\alpha}(t+v, x_{t+v})\Big) \\
&= \frac{\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)}{\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t+u)} \sum_{x_t \in A} \Phi_t^{\boldsymbol{\alpha}}(x_t) \sum_{x_{t+1},\ldots,x_{t+u-1}} \prod_{v=1}^{u} P(x_{t+v-1}, x_{t+v}) \Big(1 - J^{\Phi_t^{\boldsymbol{\alpha}}}(v, x_{t+v})\Big) \\
&= \frac{\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t)\mathbb{P}(\tau_1^{\Phi_t^{\boldsymbol{\alpha}}} > u)}{\mathbb{P}(\tau_1^{\boldsymbol{\alpha}} > t+u)} \Phi_u^{\Phi_t^{\boldsymbol{\alpha}}}(x_{t+u})
\end{aligned}$$

and by Proposition 3 we conclude the proof. □

2.4 Submultiplicativity of $\sup_{\alpha} \mathbb{P}(\tau_{*,G}^{\boldsymbol{\alpha}} > t)$

In this section we prove Theorem 2.

Let τ_*^{α} be a minimal CSQST, and $\tau_{*,G}^{\alpha} = \tau_*^{\alpha} \wedge \tau_G^{\alpha}$ the associated local relaxation time. We prove that the function $f(t) := \sup_{\alpha} \mathbb{P}(\tau_{*,G}^{\alpha} > t)$ is submultiplicative, i.e. that for any $t, u > 0$, $f(t+u) \le f(t)f(u)$. This fundamental property implies that $\mathbb{P}(\tau_{*,G}^{\alpha} > t)$ has an exponential bound, allowing in the next section to estimate the error terms in (1.22).

We start by observing that it is sufficient to study a particular realization of a minimal CSQST, since

$$\begin{aligned}
\mathbb{P}(\tau_{*,G}^{\alpha} > t) &= \sum_{y \in A} \mathbb{P}\left(X_t^{\alpha} = y;\ \tau_*^{\alpha} > t\right) = \sum_{y \in A} \left(\mu_t^{\alpha}(y) - \mu^*(y)\mathbb{P}(\tau_*^{\alpha} \le t < \tau_G^{\alpha})\right) \\
&= 1 - \mu_t^{\alpha}(G) - \lambda^{t+\delta_{\alpha}}(1 - \tilde{s}^{\tilde{\alpha}}(t))
\end{aligned}$$

does not depend on the choice of the minimal CSQST.

Consider a particular realization of the minimal CSQST, namely, the time $\tau_1^{\boldsymbol{\alpha}}$ defined in Sect. 2.1. Applying now Proposition 3 we immediately complete the proof.

2.5 Representation Formula for τ_G^α with τ_*^α

In this section we prove Theorem 3. Equation (1.22) is an immediate consequence of the definition of minimal CSQST and of Theorem 1.

To prove the final statement of the theorem on the hitting distribution note that for any $y \in G$ we have

$$\mathbb{P}\Big(X^\alpha_{\tau^\alpha_G} = y\Big) = \mathbb{P}\Big(\tau^\alpha_G < \tau^\alpha_*,\ X^\alpha_{\tau^\alpha_G} = y\Big) + \mathbb{P}\Big(\tau^\alpha_G > \tau^\alpha_*,\ X^\alpha_{\tau^\alpha_G} = y\Big)$$

The second term in the r.h.s. can be written as

$$\begin{aligned}&\sum_{t=0}^{\infty}\sum_{z\in A}\mathbb{P}\Big(\tau^\alpha_G > t = \tau^\alpha_*,\ X^\alpha_t = z\Big)\mathbb{P}\Big(X^z_{\tau^z_G} = y\Big)\\ &\quad = \sum_{t=0}^{\infty}\sum_{z\in A}\mu^*(z)\mathbb{P}\Big(\tau^\alpha_G > t = \tau^\alpha_*\Big)\mathbb{P}\Big(X^z_{\tau^z_G} = y\Big)\\ &\quad = \omega(y)\mathbb{P}\Big(\tau^\alpha_G > \tau^\alpha_*\Big)\end{aligned}$$

so that (1.23) holds.

2.6 Under the Time-Separation Hypothesis

In this section we prove Theorem 4.

By Theorem 3 we have

$$\frac{\mathbb{P}\Big(\tau^\alpha_G > nT\Big)}{\lambda^{nT+\delta_\alpha}} - 1 = -\tilde{s}^{\tilde{\alpha}}(nT) + \frac{\mathbb{P}(\tau^\alpha_{*,G} > nT)}{\lambda^{nT+\delta_\alpha}}.$$

By applying Theorem 2 and the Markov inequality we have

$$\mathbb{P}(\tau^\alpha_{*,G} > nT) \le \left(\sup_\alpha \mathbb{P}(\tau^\alpha_{*,G} > T)\right)^n \le \left(\frac{R}{T}\right)^n. \tag{2.15}$$

We can prove the upper bound:

$$\frac{\mathbb{P}\Big(\tau^\alpha_G > nT\Big)}{\lambda^{nT+\delta_\alpha}} - 1 \le \left(\frac{R}{T\lambda^T}\right)^n \lambda^{-\delta_a} = \frac{e^{-an}}{\alpha\cdot\gamma}.$$

As for the lower bound, notice that, by Theorem 1 and by the minimality of τ^α_*,

$$\mathbb{P}(\tau^\alpha_{*,G} = t) \ge \mathbb{P}(\tau^\alpha_* = t < \tau^\alpha_G) = \lambda^{t+\delta_\alpha}\left(\tilde{s}^{\tilde{\alpha}}(t-1) - \tilde{s}^{\tilde{\alpha}}(t)\right),$$

so that

$$\begin{aligned}\tilde{s}^{\tilde{\alpha}}(t) &= \sum_{u>t}\tilde{s}^{\tilde{\alpha}}(u-1) - \tilde{s}^{\tilde{\alpha}}(u) \le \sum_{u>t}\lambda^{-u-\delta_\alpha}\mathbb{P}(\tau^\alpha_{*,G} = u)\\ &= \sum_{u>t}\lambda^{-u-\delta_\alpha}\left(\mathbb{P}(\tau^\alpha_{*,G} > u-1) - \mathbb{P}(\tau^\alpha_{*,G} > u)\right)\\ &= \lambda^{-t-\delta_\alpha}\mathbb{P}(\tau^\alpha_{*,G} > t) + \frac{1-\lambda}{\lambda}\sum_{u>t}\lambda^{-u-\delta_\alpha}\mathbb{P}(\tau^\alpha_{*,G} > u).\end{aligned} \tag{2.16}$$

Thus, the total error in (1.22) can be bounded as

$$\mathbb{P}(\tau_{*,G}^{\alpha} > t) - \lambda^{t+\delta_\alpha} \tilde{s}^{\tilde{\alpha}}(t) \geq -\frac{1-\lambda}{\lambda}\lambda^t \sum_{u>t} \lambda^{-u}\mathbb{P}(\tau_{*,G}^{\alpha} > u). \tag{2.17}$$

Again by Markov inequality and submultiplicativity,

$$\sum_{u>nT} \lambda^{-u}\mathbb{P}(\tau_{*,G}^{\alpha} > u) \leq T \sum_{k\geq n} \lambda^{-(k+1)T}\left(\frac{R}{T}\right)^k = \frac{T\lambda^{-T+1}}{1-\frac{R}{T\lambda^T}}\left(\frac{R}{T\lambda^T}\right)^n,$$

where we used the fact that the sum is convergent since $\frac{R}{T\lambda^T} < 1$ by hypothesis.

Thus, we obtain the lower bound

$$\begin{aligned}\frac{\mathbb{P}\left(\tau_G^{\alpha} > nT\right)}{\lambda^{nT+\delta_\alpha}} - 1 &\geq -\frac{1-\lambda}{\lambda}\lambda^{-\delta_a} \sum_{u>nT} \lambda^{-u}\mathbb{P}(\tau_{*,G}^{\alpha} > u) \\ &\geq -\lambda^{-\delta_a}\frac{\lambda^{-T}}{1-e^{-a}}e^{-an}\end{aligned}$$

and the thesis immediately follows.

3 An Example: The Rim

As explained in the introduction, metastability is associated to the existence of two asymptotically-separated time-scales: a "short" time-scale in which the system relaxes to a sort of "apparent equilibrium" and a "long" time-scale that characterizes the arrival to the invariant measure. We introduce here a simple model, inspired by a similar example introduced in [10] to show how the conditional Strong Time language can be used to formalize this picture.

In this model, each term in the representation formula can be computed, so that we can illustrate the meaning of the terms of the representation formula (1.22) in an explicit case.

Let $n \in \mathbb{N}$ be an integer parameter. The state-space of the model is $\mathbb{T}_n \cup G$, where $\mathbb{T}_n$ denotes the 1-dimensional discrete torus of lenght 4^n(labeled from 0 to $4^n - 1$) and G is a single absorbing state. The graph is illustrated in Fig. 1.

All transition probabilities will be invariant under rotations by multiples of 4 , so that, according to the heuristic definition of metastability given above, we should compare the time to diffuse onto the ring to the time needed to take one of the spokes. If the former time is much shorter than the latter, the system somehow thermalizes before undergoing the transition to equilibrium; if not, we cannot talk about metastability even when the arrival time to G is exponentially-distributed.

The transition probabilities, as we will show later on, are chosen to keep as simple as possible the construction of the CSQST. On the same graph, all choices with similar symmetries would allow the construction of the CSQST, but the construction would not be as simple.

Let $\mathbb{T}_n^0$ be the subset of the multiples of 4 in $\mathbb{T}_n$, $\mathbb{T}_n^1$ denote the subset of the odd numbers in $\mathbb{T}_n$ and $\mathbb{T}_n^2 = \mathbb{T}_n \setminus \left(\mathbb{T}_n^0 \cup \mathbb{T}_n^1\right)$ denote the remaining subset.

Let $\lambda \in (0, 1)$ be a real parameter that will correspond to the largest eigenvalue of the matrix $[P]_A$. The non-null elements of the transition matrix P are:

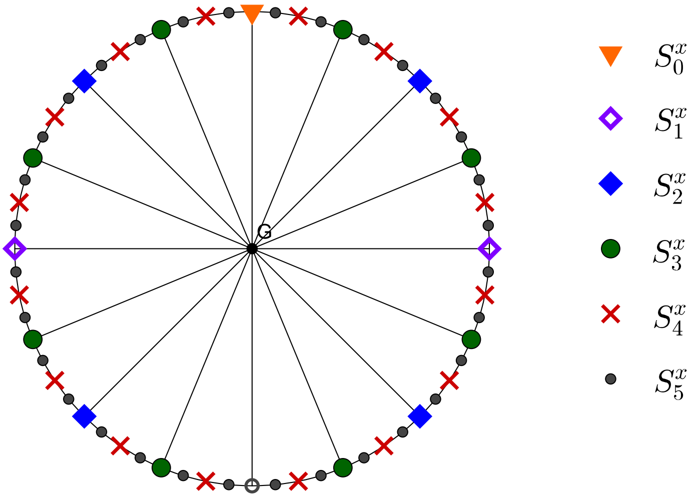

Fig. 1 The graph and the subsets S_k^x when $n = 3$

If $x \in \mathbb{T}_n^0$ (that is, a multiple of 4),

$$P_{x,y} = \begin{cases} \frac{\lambda}{2} & \text{if } y = x \\ \frac{\lambda^2(2-\lambda)}{32-32\lambda+4\lambda^2} & \text{if } y \in \mathbb{T}_n \text{ and } |x-y| = 1 \\ \frac{8-12\lambda+4\lambda^2}{8-8\lambda+\lambda^2} & \text{if } y = G \end{cases}$$

If $x \in \mathbb{T}_n^1$ (that is, an odd number),

$$P_{x,y} = \begin{cases} \frac{\lambda}{2} & \text{if } y = x \\ \frac{8-8\lambda+\lambda^2}{4(2-\lambda)} & \text{if } y \in \mathbb{T}_n^0 \text{ and } |x-y| = 1 \\ \frac{\lambda^2}{4(2-\lambda)} & \text{if } y \in \mathbb{T}_n^2 \text{ and } |x-y| = 1 \end{cases}$$

If $x \in \mathbb{T}_n^2$ (that is, an even number but not a multiple of 4),

$$P_{x,y} = \begin{cases} \frac{\lambda}{2} & \text{if } y = x \\ \frac{2-\lambda}{4} & \text{if } y \in \mathbb{T}_n \text{ and } |x-y| = 1 \end{cases}$$

Moreover, $P_{G,G} = 1$.
The Perron-Frobenius theorem and a direct computation allows to prove the following:

Proposition 4 *λ is the largest eigenvalue of the sub-markovian matrix $[P]_A$, associated to the left eigenvector (normalized to 1)*

$$\mu_x^* = \begin{cases} 4^{-n}\frac{8-8\lambda+\lambda^2}{2-\lambda} & \textit{if } x \in \mathbb{T}_n^0 \\ 4^{-n}\lambda & \textit{if } x \in \mathbb{T}_n^1 \\ 4^{-n}\frac{\lambda^2}{(2-\lambda)} & \textit{if } x \in \mathbb{T}_n^2 \end{cases}$$

and to the right eigenvector

$$\gamma_x = \begin{cases} \frac{2-\lambda}{8-8\lambda+\lambda^2} & \text{if } x \in \mathbb{T}_n^0 \\ \frac{1}{\lambda} & \text{if } x \in \mathbb{T}_n^1 \\ \frac{2-\lambda}{\lambda^2} & \text{if } x \in \mathbb{T}_n^2 \end{cases}$$

with normalization such that $\sum_x \gamma_x \mu_x^* = 1$.

Remark 1 If $x \in \mathbb{T}_n^1$ and $y = x \pm 1$, then $P_{x,y} = 4^{n-1}\mu_y^*$, while $P_{x,x} = 2 \cdot 4^{n-1}\mu_x^*$. Therefore, starting from the uniform distribution α on $\mathbb{T}_n^1$:

$$\mu_1^\alpha(y) \equiv P_{\alpha,y} = \sum_{x \in \mathbb{T}_n^1} 2 \cdot 4^{-n} P(x,y) = \mu_y^*$$

In order to construct the CSQST, we define a family of sets S_k^x recursively:
Let x be a starting configuration, for $k \in \{1, \ldots, 2n-1\}$

$$\begin{aligned} \tau^x(0) &:= \tau^x_{\mathbb{T}_n^0} \\ S_0^x &:= \left\{ X^x_{\tau^x(0)} \right\} \\ S_k^x &:= \left\{ y \in \mathbb{T}_n \; ; \; y \pm 2^{2n-k-1} \in S_{k-1}^x \right\} \\ \tau^x(k) &:= \inf \left\{ t > \tau^x(k-1) \; ; \; X_t^x \in S_k^x \right\} \end{aligned}$$

where the symbol $\pm$ denotes the sum/difference modulo 4^n (see Figure 1).

Remark 2 (1) For every $k \in \{2, \ldots, 2n-2\}$, between each two consecutive elements of S_{k-1}^x we put two elements of S_k^x.
(2) $|S_k^x| = 2^k$.
(3) The sets S_k^x are stochastic only because S_0^x is stochastic. As we will see in what follows, due to the symmetry of the model, we are mainly interested in the case $x = 0$. In this case $S_0^x = 0$.
(4) To each element y of S_k^x is associated a unique "parent" $g(y)$ in S_{k-1}^x such that $|y - g(y)| = 2^{2n-k-1}$. Each parent has two offsprings and for every $k \geq 1$, if $X^x_{\tau^x(k)} = y$, then $X^x_{\tau^x(k-1)} = g(y)$.

With these definitions we can state our main result on this model.

Theorem 6 *The time*

$$\tau_\star^x := \tau^x(2n-1) + 1$$

is a CSQST.

Proof We first consider the case $x = 0$ and we start by proving inductively that for each $k \in \{0, \ldots, 2n-1\}$, $X^x_{\tau^0(k)}$ is independent of $\tau^0(k)$ and uniformly distributed on S_k^0: for $y \in S_k^0$,

$$\mathbb{P}\left(X_t^0 = y, \; \tau^0(k) = t\right) = 2^{-k}\mathbb{P}\left(\tau^0(k) = t\right). \tag{3.1}$$

□

Indeed, for $k = 0$, $S_0^0 = 0$ is a singleton and there is nothing to prove. For $k \geq 1$,

$$\mathbb{P}\left(X_t^0 = y,\ \tau^0(k) = t\right) = \sum_{s<t} \mathbb{P}\left(X_t^0 = y,\ \tau^0(k) = t,\ X_s^0 = g(y),\ \tau^0(k-1) = s\right)$$

by symmetry we have

$$\mathbb{P}\left(X_t^0 = y,\ \tau^0(k) = t\right) = \frac{1}{2} \sum_{s<t} \mathbb{P}\left(\tau^0(k) = t,\ X_s^0 = g(y),\ \tau^0(k-1) = s\right)$$

and by the inductive hypothesis we get

$$= \frac{1}{2} \sum_{s<t} \mathbb{P}\left(\tau^0(k) = t \middle| X_s^0 = g(y),\ \tau^0(k-1) = s\right) 2^{-k+1} \mathbb{P}\left(\tau^0(k-1) = s\right)$$

By symmetry we can ignore the conditioning $X_s^0 = g(y)$ obtaining

$$= \frac{1}{2} \sum_{s<t} \mathbb{P}\left(\tau^0(k) = t,\ \tau^0(k-1) = s\right) 2^{-k+1} = 2^{-k} \mathbb{P}\left(\tau^0(k) = t\right)$$

Now, we observe that S_{2n-1}^0 is the set $\mathbb{T}_n^1$ of the odd numbers, from which it is not allowed to move to G, and by recalling Remark 1 we immediately obtain for any $y \in \mathbb{T}_n$:

$$\mathbb{P}\left(X_t^0 = y,\ \tau_\star^0 = t\right) = \mu_y^* \mathbb{P}\left(\tau_\star^0 = t\right)$$

Since in our example $\tau_\star^0 = t$ entails $\tau_G^0 > t$, we can condition the last two formulae and see that $\tau_\star^0$ is a CSQST.

If we now consider an arbitrary starting point x, due to the definition of S_0^x and τ_0^x, this is equivalent to start from some point in $\mathbb{T}_n^0$, defined below as S_0^x, depending on x. As noted above this point S_0^x is random if $x \notin \mathbb{T}_n^0$. However, due to the symmetry of the model, every starting point in $\mathbb{T}_n^0$ is equivalent to 0. This concludes the proof of the theorem.

We want to discuss now the application of our representation formula to this particular model.

Let us call $\mathrm{Opp}(x) := S_0^x + 2^{2n-1}$ the point opposite to S_0^x. In order to reach this point, the process has to visit every set S_k^x, so that we get the estimates

$$\mathbb{P}\left(\tau_\star^x > t\right) \leq \mathbb{P}\left(\tau_{\mathrm{Opp}(x)}^x \geq t\right) \quad \text{and} \quad \mathbb{P}\left(\tau_{\star,G}^x > t\right) \leq \mathbb{P}\left(\tau_{\mathrm{Opp}(x),G}^x \geq t\right).$$

Standard diffusive bounds show that $\mathbb{P}\left(\tau_{\mathrm{Opp}(x)}^x \leq 4^{2n-1}\right)$ is larger than a constant $c \in (0, 1]$. By dividing the time t into intervals of length 4^{2n-1}, we get

$$\mathbb{P}\left(\tau_{\mathrm{Opp}(x),G}^x \geq t\right) \leq \left(\sup_{y \in \mathbb{T}^n} \mathbb{P}\left(\tau_{\mathrm{Opp}(x),G}^x \geq 4^{2n-1}\right)\right)^{4^{-2n+1}t} \leq (1-c)^{4^{-2n+1}t} \tag{3.2}$$

In order to compute the other terms in the representation formula, we consider the local chain on $\mathbb{T}_n$

$$\widetilde{P}_{x,y} = \frac{\gamma(y)}{\gamma(x)} \frac{P_{x,y}}{\lambda} = \begin{cases} \frac{1}{2} & \text{if } x = y \\ \frac{1}{4} & \text{if } |x-y| = 1 \end{cases}$$

Clearly, the local process $\widetilde{P}_{x,y}^t$ is a lazy random walk on the ring. It is easy to control the convergence to equilibrium for this process in separation distance. Indeed, by standard diffusive estimates we get that $\tilde{s}^x(4^{2n-1}) \leq b$ for some constant $b \in (0, 1)$.

Since $\sup_{x\in\mathbb{T}_n} \tilde{s}^x(t)$ is submultiplicative, we get

$$\widetilde{s}^x(t) \leq b^{4^{-2n+1}t}. \tag{3.3}$$

From these estimates, we see that the error terms in (1.22) can be estimated by $\gamma_x\lambda^t b^{4^{-2n+1}t} + (1-c)^{4^{-2n+1}t}$ and, when the time needed to diffuse onto the ring is smaller than the mean time $1/(1-\lambda)$ to reach G, they decay faster than the leading term $\gamma_x\lambda^t$.

It is useful to compare the estimate given by Theorem 3 with a direct computation of $\mathbb{P}\left(\tau_G^x > t\right)$. To this end, let us introduce the projection operator $p : \mathbb{T}_n \cup G \to \{0, 1, 2, G\}$ defined by $p(G) = G$ and $x \in \mathbb{T}_n^{p(x)}$. We notice that the projection $p(P^t)$ is itself a Markov process with transition matrix

$$\bar{P} = \begin{pmatrix} P_{0,0} & 2P_{0,1} & 0 & P_{0,G} \\ P_{1,0} & P_{1,1} & P_{1,2} & 0 \\ 0 & 2P_{2,1} & P_{2,2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

this means that $P_{i,j} = P_{x,y}$ with $x \in \mathbb{T}_n^i$, $y \in \mathbb{T}_n^j$ for any $i, j \in \{0, 1, 2, G\}$. Thus, $\mathbb{P}\left(\tau_G^x > t\right) = \bar{\mathbb{P}}\left(\tau_G^{p(x)} > t\right)$, where $\bar{\mathbb{P}}$ denotes the probability for the Markov chain with transition matrix $\bar{P}$.

The largest eigenvalue of the restricted matrix is again λ and the quasi-stationary measure is the projection of μ^*:

$$\begin{cases} \bar{\mu}_0^* = \frac{1}{4}\frac{8-8\lambda+\lambda^2}{2-\lambda} \\ \bar{\mu}_1^* = \frac{\lambda}{2} \\ \bar{\mu}_2^* = \frac{\lambda^2}{4(2-\lambda)} \end{cases}$$

Remark 3 $\bar{\mu}_i^* = \bar{P}_{1,i}$. Hence, when starting from 1, the projected chain reaches equilibrium at time 1.

Thus,

$$\begin{aligned}
\bar{\mathbb{P}}\left(\tau_G^1 > t\right) &= \lambda^{t-1} = \gamma_1\lambda^t \\
\bar{\mathbb{P}}\left(\tau_G^2 > t\right) &= \bar{P}_{2,2}^t + \sum_{s=0}^{t-1} \bar{P}_{2,2}^s \bar{P}_{2,1}\lambda^{t-s-2} = \left(\frac{\lambda}{2}\right)^t + \sum_{s=0}^{t-1}\left(\frac{\lambda}{2}\right)^s \frac{2-\lambda}{2}\lambda^{t-s-2} \\
&= \left(\frac{\lambda}{2}\right)^t + \frac{2-\lambda}{2}\lambda^{t-2}2(1-2^{-t}) = \gamma_2\lambda^t\left(1-2^{-t}\left(1-\frac{1}{\gamma_2}\right)\right), \\
\bar{\mathbb{P}}\left(\tau_G^0 > t\right) &= \bar{P}_{0,0}^t + \sum_{s=0}^{t-1} \bar{P}_{0,0}^s \bar{P}_{0,1}\lambda^{t-s-2} = \left(\frac{\lambda}{2}\right)^t + \frac{\lambda^2}{4}\gamma_0 4\lambda^{t-2}(1-2^{-t}) \\
&= \gamma_0\lambda^t\left(1-2^{-t}\left(1-\frac{1}{\gamma_0}\right)\right).
\end{aligned}$$

We see that, due to the symmetry of this system, the distribution of τ_G^x can be approximated with an exponential distribution much before the diffusive time on the ring. In other words, the hitting time has an exponential behavior even before the metastable time.

Acknowledgements We thank Amine Asselah, Nils Berglund, Pietro Caputo, Frank den Hollander, Roberto Fernandez and Alexandre Gaudillière for many fruitful discussions. This work was partially supported by the A*MIDEX project (n. ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR). E.S. has been supported by the PRIN 20155PAWZB Large Scale Random Structures.

References

1. Aldous, D.: Markov chains with almost exponential hitting times. Stoch. Proc. Appl. **13**, 305–310 (1982)
2. Aldous, D., Brown, M.: Inequalities for rare events in time reversible Markov chains I. In: Shaked, M., Tong, Y.L. (eds.) Stochastic Inequalities. vol. 22, pp. 1–16. Lecture Notes of the Institute of Mathematical Statistics (1992)
3. Aldous, D., Brown, M.: Inequalities for rare events in time reversible Markov chains II. Stoch. Proc. Appl. **44**, 15–25 (1993)
4. Aldous, D., Diaconis, P.: Shuffling cards and stopping times. Am. Math. Monthly **93**, 333–348 (1986)
5. Aldous, D., Diaconis, P.: Strong uniform times and finite random walks I. Adv. Appl. Math. **8**, 66–97 (1987)
6. Bianchi, A., Gaudilliere, A.: Metastable states, quasi-stationary and soft masures, mixing time asymptotics via variational principles. arXiv:1103.1143 (2011)
7. Bovier, A., den Hollander, F.: Metastability, a Potential-Theoretic Approach. Springer, New York (2010)
8. Collet, P., Martínez, S., San Martín, J.: Quasi-stationary Distributions: Markov Chains, Diffusions and Dynamical Systems. Springer, New York (2012)
9. Darroch, J.N., Seneta, E.: On quasi-stationary distributions in absorbing discrete-time finite Markov chains. J. Appl. Prob. **2**, 88–100 (1965)
10. Diaconis, P., Fill, J.A.: Strong stationary times via a new form of duality. Ann. Probab. **18**(4), 1483–1522 (1990)
11. Diaconis, P., Miclo, L.: On times to quasi-stationary for birth and death processes. J. Theor. Probab. **22**(3), 558–586 (2009)
12. Fernández, R., Manzo, F., Nardi, F.R., Scoppola, E.: Asymptotically exponential hitting times and metastability: a pathwise approach without reversibility. Electron. J. Probab. **20**, 1–37 (2015)
13. Fernández, R., Manzo, F., Nardi, F.R., Scoppola, E., Sohier, J.: Conditioned, quasi-stationary, restricted measures and escape from metastable states. Ann. Appl. Prob. **26**, 760–793 (2016)
14. Keilson, J.: Markov Chain Models-Rarity and Exponentiality. Springer, New York (1979)
15. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. AMS (2009)
16. Olivieri, E., Vares, M.E.: Large deviations and metastability Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge (2005)