CONDITIONED, QUASI-STATIONARY, RESTRICTED MEASURES AND ESCAPE FROM METASTABLE STATES

BY R. FERNANDEZ[*], F. MANZO[†,1], F. R. NARDI[‡],
E. SCOPPOLA[†,2] AND J. SOHIER[‡,3]

*University of Utrecht[*], University of Roma Tre[†] and
Eindhoven University of Technology and EURANDOM[‡]*

We study the asymptotic hitting time $\tau^{(n)}$ of a family of Markov processes $X^{(n)}$ to a target set $G^{(n)}$ when the process starts from a "trap" defined by very general properties. We give an explicit description of the law of $X^{(n)}$ conditioned to stay within the trap, and from this we deduce the exponential distribution of $\tau^{(n)}$. Our approach is very broad—it does not require reversibility, the target G does not need to be a rare event and the traps and the limit on n can be of very general nature—and leads to explicit bounds on the deviations of $\tau^{(n)}$ from exponentially. We provide two nontrivial examples to which our techniques directly apply.

1. Introduction.

1.1. *Scope of the paper.* Metastability and related phenomena are associated to systems "trapped" for a long time in some part of their phase space, from which they emerge in a random time—the *exit time*—expected to have an asymptotic exponential distribution. They are the subject of many current studies in the mathematical and mathematical physics literature. This recent activity is motivated, on the one hand, by the confirmation of their widespread occurrence in most evolutions and, on the other hand, on the emergence of a host of new techniques for their rigorous analysis—cycle decomposition [15–17, 27, 29–31, 36], potential theoretic techniques [8, 9, 11, 12], renormalization [33, 34], martingale approach [5–7]. In this article, we focus on an essential component of a metastability analysis: the description of the state of a system trapped in a region A and the estimation of the law of the exit time from this (meta)stable trap to a target set G. These times are

the building blocks for studies of evolutions through arbitrarily complex scenarios involving many of these traps. Our treatment is inspired by the first rigorous paradigm proposed for this type of studies—the *path-wise approach* introduced in [15] and developed, for instance, in [27, 29–31].

In the remaining part of the Introduction, we discuss in detail the main advantages of our approach, but here is a brief summary: (i) We do not make any assumption on the nature of the limit process involved, in particular it applies to fixed-temperature, infinite-volume limits. (ii) The process is not assumed to be reversible, and the traps do not need to be narrowly supported. (iii) (Meta)stable states are described by measures, not just through a few reference configurations. Furthermore, the different natural candidate measures are shown to provide equivalent descriptions within explicit error bounds. (iv) There is no constraint on the exterior G of the trap—in particular, it does neither need to be a rare set, nor to be associated to a stable state. (v) Exit laws are precisely estimated with explicit bounds on deviations from exponentiality. (vi) Our approach relies on a novel type of proof based on controlling the proximity to the quasi-stationary measure. This simplifies proofs and strengthens results.

1.2. *Issues addressed by this approach.* Let us first discuss in some detail the issues we deal with.

General traps. In many visualisations of a trap, authors have in mind some energy profile associated to the invariant measure of the process defining the evolution. A trap corresponds, in this case, to an energy well, and metastability refers to the exit from such a well to a deeper well, leading to stability, that is, corresponding to the asymptotic support of the invariant measure. This point of view is fully exploited, for instance, in the reference book [31]. It is important to keep in mind, however, that this is only one of many possible metastability (or tunneling) scenarios. Indeed, in many simple examples, the invariant measure is not linked to any useful "energy" profile. For instance, any shuffle algorithm for a pack of cards leaves the uniform measure invariant. This measure corresponds to "flat," well-less energy profile. Yet, in this paper we show that for the well-known Top-In-At-Random model, the time it takes an initially well-shuffled pack to attain a particular order is exponentially distributed (see Section 4.2). In this example, the "well" A is entropic in nature. More generally, processes can define free-energy wells. We stress that our setup relies on hypotheses insensitive to the nature of the "well."

Measures associated to traps. States are probability measures, and hence transitions between trapped states have to correspond to transitions between measures (asymptotically) supported by these traps. Nevertheless, in many instances metastable states are associated to individual configurations marking the "bottom of the trap." Such studies are suited to traps that become, asymptotically,

abrupt enough to allow only small fluctuations in the trapped state. A more general scenario should involve "wide traps" corresponding to measures with an extended support and a corresponding thermalization scale. For a given trap, there are three natural candidate measures: the restriction of the invariant measure, the quasi-stationary measure and the empirical measure of the process before exiting the trap. These measures have complementary properties; in particular, the quasi-stationary measure automatically leads to exponential exit times. In this paper, we shed light on the links between these measures in our framework.

General assumptions.   Our approach for metastability includes the following features.

- *No need of reversibility*: Many approaches depend crucially of the reversible character of the invariant measure. This is, in particular, true in the funding literature of the subject by Keilson [23] and Aldous and Brown [1, 2]. Until the recent paper [24], where potential theoretic tools were used to prove the metastable behavior for a nonreversible dynamics, it was also a prerequisite for the application of potential theory. Nevertheless, many examples (we deal with one such example in Section 4.2) show that this hypothesis is not necessary to observe exponential escape times.
- *General nature of the target set G*: Exit times are defined by hitting times to a certain "boundary set" G which, in many instances, is associated to the *saddle points*, or bottlenecks that the system has to cross on its exit trajectories. As such, it is often assumed that the set G is asymptotically negligible. Alternative metastability (and tunneling) studies use G as the support of the stable measure ($=$ bottom of deeper wells). In these cases, the set G has asymptotically full measure, which is in sharp contrast with the previous ones. These two cases show that, for general exit times studies, the set G should simply be associated with the exterior of the well, without any assumption on its asymptotic measure.

The nature of the asymptotic regime.   Metastability, and the exponential escape law, only appear asymptotically in appropriate parameters that gauge the nature and depth of the trap. The type of parametrization determines how the metastable regime is approached. Some approaches, for example, low-temperature limits involve traps that become asymptotically more abrupt. Others, for example, card shuffling of larger packs—keep the geometry of the trap fixed but make the external set G progressively farther from the "bottom" of the trap. In more complicated situations, the limit involves both a change in depth and the complexity of the traps. This happens, for instance, in the study of fixed-temperature spin systems in the thermodynamic limit, for which the changes in the volume lead to a more complex trap scenario with an associated proliferation ("entropy effect") of escape routes (e.g., location of "critical droplets"). Our approach applies equally well to these different limits.

The type of transition probabilities. In many standard metastability studies, the stochastic evolution is itself defined through an energy function (identical or closely related to the one defined by the invariant measure). This is the case, for instance, in the Metropolis algorithm, in which transition probabilities are proportional to the exponential of the positive part of energy differences. A related feature of these models is the fact that the "barriers" between traps—determining, for instance, the mean escape time—are proportional to the energy differences between configurations marking the bottom and the top of energy wells. In the framework of statistical mechanics models, an important family of models excluded by this setup are the cellular automata [18]. It turns out that the parallel character of their dynamics leads to the existence of a large number of paths between any two configurations. The probabilistic barrier characterizing each of these paths cannot be described only by energy differences.

The particular form of the transition probabilities plays no special role in our general approach.

1.3. *Main features of our approach.* Here is how we deal with the issues discussed above.

General traps and asymptotic regime. We consider a family of continuous time irreducible Markov chains $X_t^{(n)}$ on finite state spaces $\mathcal{X}^{(n)}$. The asymptotic regime is associated to $n \to \infty$, but we do not need to specify the nature of the parameters involved. Particular relevant examples are the case where the limit may involve the divergence of the cardinality of the state spaces, and/or some parameter in the transition probabilities such as the inverse temperature. For each n the state space is divided into a *trap* $A^{(n)}$ and its complementary set $G^{(n)} = \mathcal{C}^{(n)} \setminus A^{(n)}$. We do not assume any particular property of $G^{(n)}$; exit times are only determined by the evolution inside $A^{(n)}$ and the structure of $G^{(n)}$ is irrelevant for the exit times. The traps include sets $B^{(n)} \subset A^{(n)}$ of configurations—associated to "bottoms of the well"—which can have arbitrary size as long as they satisfy certain natural assumptions.

More precisely, the definition of trap is contained in three physically natural hypotheses—spelled out in Section 3.1 below—that can be roughly described as follows:

Fast recurrence: For any initial configuration, the process falls within a controlled time $R^{(n)}$ either in $B^{(n)}$ or in $G^{(n)}$. This *recurrence time* acts as a reference time for the whole of the analysis.
Slow escape: Starting from configurations in $B^{(n)}$, the time $\tau^{(n)}$ the process takes to hit $G^{(n)}$ is much larger than the recurrence time $R^{(n)}$.
Fast thermalization: A process started in $B^{(n)}$ achieves "local thermalization" within $B^{(n)}$—in a time much shorter than the escape time $\tau^{(n)}$.

The two time scales—escape and local thermalization—actually define the trap
and the limit process. This definition is sufficiently general to accommodate dif-
ferent types of traps (e.g., energy, entropy- or free-energy driven) and asymptotic
limits (e.g., low temperature, large volume or combinations of both). The scales
do not need to be asymptotically different; a fact that allows the consideration of
less abrupt traps. The recurrence hypothesis controls, in particular, the presence of
complicated landscapes in $A^{(n)} \setminus B^{(n)}$. As discussed below, this control is not nec-
essary if $G^{(n)}$ is a rare set. Furthermore, in concrete examples it is usually not hard
to achieve recurrence at some appropriate time scale, due to the submultiplicative
property of the probability of nonrecurrence

$$\sup_{x \in \mathcal{X}^{(n)}} \mathbf{P}\big(\tau^x_{G^{(n)} \cup B^{(n)}} > kt\big) \leq \sup_{x \in \mathcal{X}^{(n)}} \mathbf{P}\big(\tau^x_{G^{(n)} \cup B^{(n)}} > t\big)^k.$$

Metastable states.   Our approach is based on the fact that, with the above def-
inition of a trap, the three possible notions of trapped state—the restriction of the
invariant measure, the quasi-stationary measure and the empirical measure of the
process before exiting the trap—are largely equivalent and can be used indistinc-
tively. In fact, our proof provides explicit bounds on the asymptotic behavior of
their distances. Therefore, metastable states can be defined by any of them and are
endowed with the physical properties associated to each of them. Mathematically,
this equivalence leads to precise quantitative estimates on the distribution of the
first hitting time of $G^{(n)}$, through notoriously simplified proofs.

General target set $G^{(n)}$.   Except for the recurrence hypothesis, the structure of
the set $G^{(n)}$ plays no role in our theory. It can, therefore, correspond to rare or
nonrare parts of the configuration space. Furthermore, this generality makes our
results applicable to a wide type of phenomena. Indeed, in a truly metastability
scenario the trap corresponds to an asymptotically unlikely family of configura-
tions and the escape is followed by a fall into the true support of the stable phase.
In other cases, however, the escape may be followed by a visit to an equally likely
set of configurations; a situation more appropriately referred to as *tunneling*. This
phenomenon occurs, for instance, if stable states are supported in several asymp-
totically disconnected regions and interest focuses on the way the system migrates
from one of these regions to another one. Hence, the insensitivity to the scenario
after the exit leads to results useful for the study of metastability, tunneling or any
evolution involving traps.

Application to nonreversible dynamics.   Other than the preceding assumptions,
we do not assume any particular property of the transition probabilities, nor do we
assume reversibility. This work is, hence, part of the so far quite reduced circle
of metastability results in the nonreversible case. Until the work [24], lack of re-
versibility was precluding the potential theoretical approach, and relevant publica-
tions are very recent [6, 20]. In particular, the convergence in law of escape times

has been obtained in [7], but assuming that $G^{(n)}$ is a rare event. In [28], the author provides a control on tail distributions that is close to ours, yielding an error of the form $\exp(O(\varepsilon)t)$ instead of our $O(\varepsilon)\exp(-t)$ [see also (ii) in Remark 3.9].

The technique of cycle decompositions [17, 30] applies also to nonreversible dynamics for chains with exponentially small transition probabilities at fixed-volume in the low-temperature limits.

General assumptions on the asymptotic regime. We do not make any assumption on the nature of the limiting process involved. In particular, we cover the case of fixed temperature and infinite volume limit, where it naturally arises an important entropic contribution. There are only few studies that deal with the latter for specific models under reversible dynamics. They consider the case in which the volume grows exponentially with the inverse temperature [10, 21, 22]. In [32], the authors consider the kinetic Ising model at infinite volume, fixed temperature and asymptotically vanishing external magnetic field.

Summary of results. For systems satisfying the three hypotheses above, we prove that the first hitting times to $G^{(n)}$ are asymptotically exponentially distributed, within explicitly bounded error terms. Furthermore, we get explicit control of the variational distance between the quasi-stationary measure and the evolved measure of the process conditioned to stay within $A^{(n)}$. The exponential law is subsequently deduced from the properties of the quasi-stationary measure. This approach leads to stronger control on the distribution of the hitting time than the one obtained in [28]. In fact, our estimations are valid *for every starting point* in $A^{(n)}$, and hence yield a precise description of the empirical measure of the chain up to its first escape from $A^{(n)}$. This approach is different—and more effective—than the one adopted in a precedent paper [19], which was restricted to traps characterized by a single-configuration bottom.

Intuitively, the picture we obtain is the following: in a time of the order of the recurrence time $R^{(n)}$ the process chooses, depending on its starting point, whether it reaches directly $G^{(n)}$ or whether it visits $B^{(n)}$ before. Once in $B^{(n)}$, the process reaches in a very short time a metastable equilibrium within $A^{(n)}$ that is accurately described by the quasi-stationary measure. Due to Markovianness and the long mean exit times, the process requires an exponential time to exit from this temporary equilibrium.

In the second part of the paper, we illustrate the power of our results through applications to two interesting models. Both models have features that put them outside most general theories. The first model is a reversible birth and death process issued from the context of dynamical polymer models. It has the particularity that its energy profile exhibits a double well, but separated by an energy barrier that is only logarithmic. Our second example is the well-known Top-In-At-Random (TIAR) model, which has a uniform invariant measure. This is an example of non-reversible dynamics with a purely entropic barrier. As far as we know, this is the

first time this model has been shown to exhibit exponential behavior of the hitting time to a rare set.

Outline of the paper: in Section 2, we give the main definitions and recall some well known results, and in Section 3 we precise the hypotheses and detail our results. Applications are discussed in Section 4. Finally, Section 5 is devoted to the proofs.

2. Setting. Let $\mathcal{X}^{(n)}$ be a sequence of finite state space depending on a parameter n and $X_t^{(n)}$ be a sequence of *continuous time irreducible Markov chains* on them. We are interested in the asymptotics $n \to \infty$. We should typically think of n as being related to:

1. The cardinality of $|\mathcal{X}^{(n)}|$. This is typically the case for large size physical systems, and our hope would be that the techniques developed in this paper apply to the infinite volume of well-known models issued from statistical physics.

2. The inverse temperature in a Freidlin–Wentzell setup. This issue has been very studied; classical references by now are [31] and [16].

We denote by $Q^{(n)}$ the matrix of transition rates on $\mathcal{X}^{(n)}$ generating the chain $X_t^{(n)}$ and by $(\pi^{(n)}, n \geq 1)$ the corresponding sequence of invariant measures. The family of Markov chains $X_t^{(n)}$ could be equivalently defined in terms of their kernels

$$(2.1) \qquad P^{(n)} = \mathbf{1} + Q^{(n)},$$

and via their continuous time semigroup

$$(2.2) \qquad H_t^{(n)} f := e^{t Q^{(n)}} f = e^{-t} \sum_{k \geq 0} \frac{t^k (P^{(n)})^k}{k!} f.$$

We do not assume reversibility so that the adjoint kernels $(P^{(n)})^*$ defined by

$$(2.3) \qquad \left(P^{(n)}\right)^*(x, y) = \frac{\pi^{(n)}(y)}{\pi^{(n)}(x)} P^{(n)}(y, x)$$

do not necessarily coincide with $P^{(n)}$. We will denote by $X_t^{\leftarrow(n)}$ the corresponding time reversal processes.

Discrete time Markov chains are covered by our techniques and we could also consider more general cases but for an easier presentation of our results, we prefer to restrict our analysis to this setup.

We denote by $\mathbf{P}(\cdot)$ and $\mathbf{E}(\cdot)$ generic probabilities and mean values. We will specify in the events and in the random variables the initial conditions of the process, so that, for instance, for $A^{(n)} \subset \mathcal{X}^{(n)}$ we define the *first hitting time to* $A^{(n)}$ for the chain $X_t^{(n)}$ starting $x \in \mathcal{X}^{(n)}$:

$$(2.4) \qquad \tau_{A^{(n)}}^{(n),x} = \inf\{t \geq 0 : X_t^{(n),x} \in A^{(n)}\}.$$

We also use the small-o notation $o(1)$ to denote a positive quantity which has limit 0 as $n \to \infty$.

We will prove our results keeping in mind the asymptotic regime $n \to \infty$. Nevertheless, we stress that most of our estimates go beyond asymptotics, and more precisely we were able to show our main results with explicit error terms. We choose to write the major part of this article for a fixed value of n (and as a consequence and for lightness of notation, from now on, we drop this explicit dependence from the notations, except in some particular cases like Corollary 3.10), denoting by small Latin letters quantities which should be thought of as asymptotically $o(1)$. For example, the quantity f in (3.3) should be thought of as a sequence $f_n = o(1)$.

For positive sequences a_n, b_n, we will also use the standard notation $a_n \sim b_n$ as soon as $\lim_{n \to \infty} a_n/b_n = 1$, and $a_n \gg b_n$ if $b_n/a_n = o(1)$.

The notation $X = (X_t)_{t \in \mathbf{R}}$ is used for a generic chain on $\mathcal{X}$.

2.1. *Measures and distances.* Let X_t be an irreducible continuous time Markov chain on $\mathcal{X}$ with stationary measure π and transition rates Q, and let $A \subset \mathcal{X}$ be a given set. From now on, we denote by G the complementary of A, and we define the following measures:

– *The invariant measure restricted on A*:

$$(2.5) \qquad \pi_A(\cdot) := \frac{1}{\pi(A)} \pi(\cdot) \mathbf{1}_{\{\cdot \in A\}}.$$

– *The measure of evolution*: for any $t \geq 0$ and $x \in \mathcal{X}$ let

$$(2.6) \qquad \mu_t^x(\cdot) := \mathbf{P}(X_t^x = \cdot),$$

and more generally for any probability measure ν on $\mathcal{X}$:

$$(2.7) \qquad \mu_t^\nu(\cdot) := \mathbf{P}(X_t^\nu = \cdot) := \sum_{x \in \mathcal{X}} \nu(x) \mathbf{P}(X_t^x = \cdot).$$

– *The conditioned evolution measure on A*: for any $t \geq 0$ and $x \in A$ the evolution measure conditioned to A is defined as

$$(2.8) \qquad \tilde{\mu}_{A,t}^x(\cdot) := \mathbf{P}(X_t^x = \cdot | \tau_{A^c}^x > t) = \frac{\mathbf{P}(X_t^x = \cdot, \tau_{A^c}^x > t)}{\mathbf{P}(\tau_{A^c}^x > t)},$$

and more generally for any probability measure ν on $\mathcal{X}$:

$$(2.9) \qquad \tilde{\mu}_{A,t}^\nu(\cdot) := \sum_{x \in \mathcal{X}} \nu(x) \tilde{\mu}_{A,t}^x(\cdot).$$

– *Quasi-stationary measure on A*: the quasi-stationary measure $\mu_A^*(\cdot)$ is a classical notion; it can be defined, for example, in the following way:

$$(2.10) \qquad \mu_A^*(\cdot) := \lim_{t \to \infty} \tilde{\mu}_{A,t}^{\pi_A}(\cdot).$$

We recall the following elementary property about μ_A^*; this is a classical result which can be found, for example, in [35], and which states that starting from μ_A^*, the exit law from A is exactly exponential (no correction is needed). Here and in the rest of the paper, we set $\mathbf{E}[\tau_G^{\mu_A^*}] =: T^*$.

PROPOSITION 2.1.   *For any $t \geq 0$, the following equality holds*:

$$\mathbf{P}\big[\tau_G^{\mu_A^*}/T^* > t\big] = e^{-t}. \tag{2.11}$$

– *The empirical measure on A*: for any $x \in A$, by denoting with $\xi_G^x(y)$ the local time spent in y before τ_G starting at x we can define the empirical measure:

$$\mu_A^{(\mathrm{em})x}(y) = \frac{\mathbf{E}\xi_G^x(y)}{\mathbf{E}\tau_G^x}. \tag{2.12}$$

– *Total variation distance*: given two probability measures π_1 and π_2 on $\mathcal{X}$, we define

$$d_{\mathrm{TV}}(\pi_1, \pi_2) := \frac{1}{2} \sum_{x \in \mathcal{X}} |\pi_1(x) - \pi_2(x)| = \max_{A \subset \mathcal{X}} |\pi_1(A) - \pi_2(A)|$$

and

$$d(t) = \max_{x \in \mathcal{X}} d_{\mathrm{TV}}(\mu_t^x, \pi), \qquad \bar{d}(t) := \max_{x, x'} d_{\mathrm{TV}}(\mu_t^x, \mu_t^{x'}). \tag{2.13}$$

It is well known (see, e.g., [4]) that $\bar{d}$ is submultiplicative; namely, for $m, n > 1$

$$\bar{d}(m + n) \leq \bar{d}(m)\bar{d}(n) \tag{2.14}$$

and

$$\bar{d}(t) \leq d(t) \leq 2\bar{d}(t).$$

For a given set $K \subset \mathcal{X}$, we introduce the quantity

$$\bar{d}_K(t) := \max_{x, x' \in K} d_{\mathrm{TV}}(\mu_t^x, \mu_t^{x'}). \tag{2.15}$$

A submultiplicative property such as (2.14) does not hold in general for the quantity $\bar{d}_K(t)$.

3. Results. We present in this section our main results on the asymptotic vicinity of the conditioned measure and the conditioned invariant one. As a corollary, we can prove exponential behavior of the first hitting time of G with accurate control on the error terms.

3.1. *Hypotheses.* Let $G \subset \mathcal{X}$ be a set representing "goals" for the chain X_t, and define $A := \mathcal{X} \setminus G$, so that τ_G^x represents the first exit time from A and $\tau_G^{\leftarrow x}$ is the first exit time from A for the time reversal process $X_t^{\leftarrow}$. To observe asymptotic exponentiality of the hitting time of G, we need to make a couple of assumptions on the set A; define the quantity

$$(3.1) \qquad f_A(t) := \mathbf{P}(\tau_G^{\pi_A} \leq t) = \frac{1}{2} \sum_{x \in A} \pi_A(x)[\mathbf{P}(\tau_G^x \leq t) + \mathbf{P}(\tau_G^{\leftarrow x} \leq t)].$$

Indeed one readily realizes that $\mathbf{P}(\tau_G^{\pi_A} > t) = \mathbf{P}(\tau_G^{\leftarrow \pi_A} > t)$ using time reversal:

$$\mathbf{P}(\tau_G^{\leftarrow \pi_A} \leq t)$$

$$(3.2) \qquad \begin{aligned} &= \sum_{x,y \in A} \sum_{x_1,\ldots,x_{t-1} \in A} \frac{\pi(x)}{\pi(A)} P^*(x, x_1) \cdots P^*(x_{t-2}, x_{t-1}) P^*(x_{t-1}, y) \\ &= \sum_{y,x \in A} \sum_{x_1,\ldots,x_{t-1} \in A} \frac{\pi(y)}{\pi(A)} P(y, x_{t-1}) P(x_{t-1}, x_{t-2}) \cdots P(x_1, x) \\ &= \mathbf{P}(\tau_G^{\pi_A} \leq t) \end{aligned}$$

and (3.1) follows.

Note that we carried this proof in the discrete time setup for lightness of notation, the continuous time case works in the same way.

We need three natural assumptions on the behavior of the process.

$E(R, f)$ *slow escape*: on a time scale $2R$, starting from the invariant measure restricted to the set A, the process hits G with small probability; concretely, for $f \in (0, 1)$ and $R > 1$, we say that the set A satisfies the hypothesis $E(R, f)$ if

$$(3.3) \qquad\qquad\qquad\qquad f_A(2R) \leq f.$$

If hypothesis $E(R, f)$ is verified, there exists a subset of A of large measure on which the above control holds pointwisely. More precisely, consider an arbitrary $\alpha \in (0, 1)$ and define the following subset of A:

$$(3.4) \qquad B_\alpha := \{x \in A : \tfrac{1}{2}[\mathbf{P}(\tau_G^x \leq 2R) + \mathbf{P}(\tau_G^{\leftarrow x} \leq 2R)] \leq f^\alpha\}.$$

The set B_α is almost of full measure within A, representing in some sense the basin of attraction of the local equilibrium. Indeed using $E(R, f)$, one easily gets that

$$(3.5) \qquad\qquad\qquad\qquad \pi_A(B_\alpha) \geq 1 - (f)^{1-\alpha}$$

since

$$f \geq f_A(2R) \geq \frac{1}{2} \sum_{x \in A \setminus B_\alpha} \pi_A(x)[\mathbf{P}(\tau_G^x \leq 2R) + \mathbf{P}(\tau_G^{\leftarrow x} \leq 2R)]$$

$$\geq (f)^\alpha \pi_A(A \setminus B_\alpha).$$

$T(R, d)$ *fast thermalization in* B_α: for $d \in (0, 1)$, we say that A satisfies the thermalization condition $T(R, d)$ if the process started from any point in B_α loses memory of its initial condition within a time scale R; more precisely

$$(3.6) \qquad\qquad\qquad \bar{d}_{B_\alpha}(R) \le d.$$

$Rc(R, r)$ *fast recurrence to the set* $B_\alpha \cup G$: starting from anywhere inside of the state space, X reaches either G or B_α with high probability within a time R; namely, let $r \in (0, 1)$, we say that A satisfies the hypothesis $Rc(R, r)$ if

$$(3.7) \qquad\qquad\qquad \sup_{x \in \mathcal{X}} \mathbf{P}(\tau^x_{B_\alpha \cup G} > R) < r.$$

Note that both conditions $T(R, d)$ and $Rc(R, r)$ depend on the parameters f and α since they depend on the set B_α. Both for lightness of notation and in view of Proposition 3.2, which states that a control on the mixing time of the dynamics over the whole state space coupled to $E(R, f)$ is sufficient to ensure that conditions $T(R, d)$ and $Rc(R, r)$ hold with explicit parameters, we choose not to write this dependence when refeering to fast thermalization and to fast recurrence.

HYPOTHESIS 3.1 (Hypothesis HpG). When the set A satisfies simultaneously the conditions $E(R, f)$, $Rc(R, r)$ and $T(R, d)$, we say that condition $Hp.G(R, d, f, r)$ is verified.

In words, condition $Hp.G(R, d, f, r)$ for d, f, r small means that the time scale R is large enough so that the process loses memory of its initial condition in B_α, but that within a time $2R$ the trajectories starting from the typical configurations for π_A inside of A still did not hit G, while hitting either B_α or G within the same time scale occurs with high probability. This is a typical situation characterizing metastability.

Before stating our main result, a couple of remarks are in order:

- The recurrence hypothesis Rc can be verified by changing the time scale R. Indeed, if $\tau^{\pi_A}_G$ is much larger than R, our process satisfies hypotheses $E(R, f)$ and $T(R, d)$ within a time $R_+ \gg R$. In this case, $Rc(R_+, r)$ holds with $r = (\sup_{x \in \mathcal{X}} \mathbf{P}(\tau^x_{B_\alpha \cup G} > R))^{R_+/R}$.
- In view of equation (3.5), it would seem natural to replace the thermalization condition $T(R, d)$ by $\bar{d}_A(R) \le d$. We stress that from a physical point of view, this is requiring too much on the whole set of trajectories starting from A and that in most applications there is a region (which is here the set $A \setminus B_\alpha$) in which some trajectories might escape from A within a short time without thermalizing afterward. The typical picture of a well in which trajectories thermalize is here played by the basin of attraction B_α. In different words, with these hypotheses we can apply our results also to cases in which the basin of attraction of the stable state is not known in details.

Nevertheless, in the (degenerate) case where thermalization occurs over the whole set A, condition $Hp.G(R, d, f)$ can be verified. In this spirit, we state the following result.

PROPOSITION 3.2.   *If A satisfies hypothesis $E(R, f)$ and*

$$(3.8) \qquad\qquad \bar{d}_A(R) \leq d,$$

then A satisfies condition $Hp.G(R, d, f, d + f + f^{1-\alpha})$.

We now state our main technical result.

3.2. *Convergence of measures.* Define the following *doubly conditioned evolved measure on A*: for any $t \geq 2R$ and $x \in A$ the evolution measure conditioned to A in $[0, t]$ and to visit B_α within $t - 2R$ is defined as

$$(3.9) \qquad\qquad \hat{\mu}_{A,t}^x(\cdot) := \mathbf{P}(X_t^x = \cdot \,|\, \tau_G^x > t, \tau_{G \cup B_\alpha}^x \leq t - 2R)$$

and

$$(3.10) \qquad\qquad \hat{\mu}_{A,t}^{\pi_A}(\cdot) := \sum_{x \in A} \pi_A(x) \hat{\mu}_{A,t}^x(\cdot).$$

REMARK 3.3.   For any $x \in B_\alpha$ we have $\hat{\mu}_{A,t}^x = \tilde{\mu}_{A,t}^x$ since $\tau_{G \cup B_\alpha}^x = 0$ and

$$(3.11) \qquad\qquad d_{\mathrm{TV}}\big(\hat{\mu}_{A,t}^{\pi_A}, \tilde{\mu}_{A,t}^{\pi_A}\big) \leq \pi_A(A \setminus B_\alpha) < f^\alpha.$$

For $x \in A \setminus B_\alpha$, the measures $\hat{\mu}_{A,t}^x$ and $\tilde{\mu}_{A,t}^x$ can be different, even if the second conditioning $\tau_{G \cup B_\alpha}^x \leq t - 2R$ is an event of large probability when $t > 3R$. In particular, they can be different for x and t such that $\mathbf{P}(\tau_G^x > t) < \mathbf{P}(\tau_{G \cup B_\alpha}^x \leq t - 2R)$.

THEOREM 3.4.   *Under hypothesis $Hp.G(R, d, f, r)$, for every $\alpha \in (0, 1)$, for d and f such that $r + 2f^\alpha < \frac{1}{4}$, we define $\bar{c} := \frac{1}{2} - \sqrt{\frac{1}{4} - c}$. The following vicinity of the doubly conditioned evolution measure and the conditioned invariant one holds*:

$$(3.12) \qquad \sup_{x \in A} \sup_{t \geq 2R} d_{\mathrm{TV}}\big(\hat{\mu}_{A,t}^x, \pi_A\big) < 4\big[\bar{c} + 2f + f^\alpha + d\big] =: \varepsilon_1,$$

$$\sup_{t > 2R} d_{\mathrm{TV}}\big(\tilde{\mu}_{A,t}^{\pi_A}, \pi_A\big) < 4\big[\bar{c} + 2f + f^\alpha + d\big] + f^{1-\alpha}$$

$$(3.13) \qquad\qquad\qquad =: \varepsilon_2 = \varepsilon_1 + f^{1-\alpha}.$$

Moreover,

$$(3.14) \qquad\qquad \sup_{x \in A} \sup_{t \geq 2R} d_{\mathrm{TV}}\big(\hat{\mu}_{A,t}^x, \mu_A^*\big) < \varepsilon_1 + \varepsilon_2.$$

Roughly speaking, Theorem 3.4 states that after time $2R$, the process thermalized in the metastable states in the sense of the doubly conditioned measure; for $x \in B_\alpha$, this result justifies the definition of metastable state for the conditioned evolved measure $\tilde{\mu}_{A,2R}^x$. The important consequence of these results is the fact that the quasi-stationary measures and the conditioned invariant ones are asymptotically close in total variation.

The following is an immediate consequence of Theorem 3.4.

COROLLARY 3.5. *Under the same hypotheses, then the ones of Theorem* 3.4, *for any* $x \in B_\alpha$ *we have*

$$d_{\mathrm{TV}}\big(\mu_A^{(\mathrm{em})x}, \pi_A\big) < \varepsilon_1.$$

Indeed for $x \in B_\alpha$, we have

$$\mu_A^{(\mathrm{em})x}(y) = \frac{\mathbf{E}\xi_G^x(y)}{\mathbf{E}\tau_G^x} = \frac{\int_0^\infty \tilde{\mu}_{A,t}^x(y)\mathbf{P}(\tau_G^x > t)\,dt}{\mathbf{E}\tau_G^x}$$

$$= \pi_A(y) + \frac{\int_0^\infty (\tilde{\mu}_{A,t}^x(y) - \pi_A(y))\mathbf{P}(\tau_G^x > t)\,dt}{\mathbf{E}\tau_G^x}.$$

3.3. *Exponential behaviour.*   A consequence of the results of the previous section is the following asymptotic exponential behavior.

THEOREM 3.6. *Assume that hypothesis HpG is satisfied. There exists* ε *such that* $\varepsilon = O(r + \varepsilon_2)$ *where* ε_2 *has been defined in Theorem* 3.4 *such that the following inequality holds for any* $t \geq 0$:

$$(3.15) \qquad \left| \mathbf{P}\left[\frac{\tau_G^{\pi_A}}{\mathbf{E}[\tau_G^{\mu_A^*}]} \geq t \right] - e^{-t} \right| \leq \varepsilon e^{-t}.$$

A pointwise control also holds starting from B_α; *namely, for every* $t \geq 0$,

$$(3.16) \qquad \sup_{x \in B_\alpha} \left| \mathbf{P}\left[\frac{\tau_G^x}{\mathbf{E}[\tau_G^{\mu_A^*}]} \geq t \right] - e^{-t} \right| \leq \varepsilon e^{-t}.$$

Moreover, for $x \notin B_\alpha$,

$$(3.17) \qquad \left| \mathbf{P}\left[\frac{\tau_G^x}{\mathbf{E}[\tau_G^{\mu_A^*}]} \geq t \right] - \mathbf{P}(\tau_G^x > 2R, \tau_{G \cup B_\alpha}^x \leq R)e^{-t} \right| \leq \varepsilon e^{-t}.$$

Let us mention some straightforward consequences of the strong control on hitting times provided by Theorem 3.6. The first one states an equivalence of time scales in our setup, and directly follows from integrating the relations (3.15) and (3.16) for $t \in \mathbf{R}^+$.

COROLLARY 3.7.   *For any $x \in B_\alpha$, one has*

(3.18) $$\mathbf{E}[\tau_G^{\mu_A^*}] \sim \mathbf{E}[\tau_G^{\pi_A}] \sim \mathbf{E}[\tau_G^x].$$

The second one is an extension of (3.15) to a wide set of starting measures.

COROLLARY 3.8.   *For any probability measure μ such that $\mathrm{Supp}(\mu) \subset B_\alpha$ and any $t \geq 0$*

(3.19) $$\left| \mathbf{P}\left[\frac{\tau_G^\mu}{\mathbf{E}[\tau_G^\mu]} \geq t \right] - e^{-t} \right| \leq \varepsilon e^{-t}.$$

In particular, the equivalence (3.18) *also generalizes to any probability measure μ such that $\mathrm{Supp}(\mu) \subset B_\alpha$, namely*

(3.20) $$\mathbf{E}[\tau_G^{\mu_A^*}] \sim \mathbf{E}[\tau_G^\mu].$$

REMARK 3.9.   The following observations are in order:

– The setting of Theorem 3.6 is quite general; in particular, it is a nonreversible
  setup, and it is noteworthy to realize that the invariant measure of the set of goals
  $\pi(G)$ can be large. Roughly speaking, the goal set does not need to be a *rare
  event*.
– The exponential control on the queues of distribution is exactly of order e^{-t} for
  $t \geq 0$; in particular, we do not lose on the exponential controlling the error as in
  [28], where the control on the queue is of the type $e^{-(1-c)t}$ where c is a small
  constant not depending on t.
– Our results hold pointwise for a large set within the complement of G, and yield
  an explicit control on the errors with respect to the hitting time.
– For $x \in A \setminus B_\alpha$, the picture is intuitively the following: in a time of order R, the
  process either thermalizes in B_α or hits the set G; in the first case, by Markov's
  property, we are back to the case of a process with starting point within B_α.

For practical purposes, we stress that a consequence of Theorem 3.6 can be deduced in the asymptotic regime of Section 2; more precisely, assume that hypotheses $E(R_n, f_n)$, $Rc(R_n, r_n)$ and $T(R_n, d_n)$ hold for a family of Markov processes $(\mathcal{X}^n, X^n)$ with parameters f_n, r_n and d_n which are $o(1)$; then the following result holds.

COROLLARY 3.10.   *Under hypothesis $Hp.G_n$, as $n \to \infty$, the following convergence in law holds*:

(3.21) $$\tau_{G^{(n)}}^{\pi_A} / \mathbf{E}[\tau_{G^{(n)}}^{\pi_A}] \xrightarrow{\mathcal{L}} \mathcal{E}(1),$$

where $\mathcal{E}(1)$ denotes a random variable which is exponentially distributed with mean one.

Furthermore, for any starting measure $\mu^{(n)}$ with support contained in $B_\alpha^{(n)}$, the following convergence in law holds:

$$(3.22) \qquad \tau_{G^{(n)}}^{\mu^{(n)}} / \mathbf{E}\big[\tau_{G^{(n)}}^{\mu^{(n)}}\big] \overset{\mathcal{L}}{\to} \mathcal{E}(1).$$

4. Applications. In this part, we discuss two examples.

Our first example originates from a model for dynamical polymers which was first introduced and analyzed in [13]. The asymptotic exponentiality of the tunneling time was shown in [14] by considering in particular the death and birth process of our first example.

The second case is a model of shuffling cards, well known as Top-In-At-Random (TIAR). This is actually a case where the invariant measure is the uniform one and the "potential barrier" is only entropic.

4.1. *Birth and death processes with logarithmic barrier.*

The model. In this part, $c > 0$ is a constant which may vary from line to line.

The (typically unique) crossing point of the interface of the dynamical polymer model with the wall in the delocalized phase follows the law of the birth and death process on $\{0, \ldots, n\}$ with invariant measure

$$\pi(x) := Z^{-1} \frac{1}{(x \vee 1)^{3/2}((n-x) \vee 1)^{3/2}},$$

where $Z = Z(n)$ is a normalization constant. We refer to [14], Section 1.2, for details.

We note that it is easy to show that

$$(4.1) \qquad Z(n) \sim cn^{3/2}.$$

We consider the Metropolis dynamics associated to this invariant measure, which means that the birth rate $b(\cdot)$ of the chain is given by

$$(4.2) \qquad b(x) = \min\left\{1, \frac{\pi(x+1)}{\pi(x)}\right\},$$

so that $Q(x, x+1) = b(x)$. Moreover, the death rate $d(\cdot)$ is given by

$$(4.3) \qquad Q(x, x-1) = d(x) = \min\left\{1, \frac{\pi(x-1)}{\pi(x)}\right\}.$$

This birth and death process has been introduced in [14] to describe the evolution of the (typically unique) crossing location of the pinning line by a polymer chain in the delocalized regime.

From (4.2) and (4.3), one readily observes that for $x \le n/2$, the process exhibits a small drift toward $\{0\}$ since $b(x) < 1$ and $d(x) = 1$, while for $x > n/2$ there is a

small drift toward $\{n\}$ since $b(x) = 1$ and $d(x) < 1$. Otherwise stated, we are considering a one-dimensional random walk under a potential given by a symmetric double well with minima in 0 and n and with a small logarithmic barrier at $\{n/2\}$.

It is shown in [14] that the mean time for the process starting from $\{0\}$ to reach $\{n/2\}$, and hence $\{n\}$, is of order $n^{5/2}$; however, one directly observes that the ratio $\frac{\pi(0)}{\pi(n/2)} = \frac{n^{3/2}}{8}$. This entails that this model is a case where the metastable behavior is much more delicate with respect to the Metropolis case where the tunneling time is of order of the ratio of the value of the invariant measure between the bottom and the top of the barrier; see [31], Chapter 6, Theorem 6.23, for details about this well-known phenomenon; as a consequence, rough arguments based on reversibility have to fail in this model, see in particular point (ii) in [31], Chapter 6, Theorem 6.23. The main result of this part is the following.

THEOREM 4.1.   *Define $G := \{n/2, n/2 + 1, \ldots, n\}$, $\alpha \in (0, 1)$ and n large enough such that $n^\alpha < n/2$. HpG holds with $n^{5/2-\varepsilon} \ll R$ for small enough ε.*

PROOF.   Our aim is to exhibit a sequence $(R_n)_{n \geq 0}$ such that the hypothesis of Theorem 3.6 are satisfied. In fact, we will show that the stronger hypothesis of Proposition 3.2 actually hold. We define the set

(4.4) $B_\alpha := \{0, \ldots, n^\alpha\}.$

Our first step is to show that for any $x \in \{0, \ldots, n/2 - 1\}$, the following equivalence holds as $n \to \infty$:

(4.5) $\mathbf{P}(\tau_0^x > \tau_{n/2}^x) \sim \left(\frac{2x}{n}\right)^{5/2}.$

We first recall the classical notion of resistance between two states x and y:

(4.6) $R(x, y) = (\pi(x)P(x, y))^{-1}.$

In this one-dimensional setup, it is a classical result that the resistances are linked to the law of first hitting to a set through the basic relation (see, e.g., [26], Chapter 2)

(4.7) $\mathbf{P}[\tau_0^x > \tau_y^x] = R(x, y),$

where $x < y$ and $R(x, y)$ is obtained by summation from the $R(k, k+1)$, namely

(4.8) $R(x, y) = \sum_{k=x}^{y-1} R(k, k+1).$

In our specific case, it is an easy computation to realize that, as $x \to \infty$,

(4.9) $R(x, x+1) = \left(\frac{\pi(x)}{1 + (x/(x+1))^{3/2}}\right)^{-1} \sim x^{3/2} + (x+1)^{3/2}.$

Indeed, representing the Markov chain like a series of resistances, we deduce that

$$R(0, x) \sim c x^{5/2}$$

and the probability in (4.5) representing the potential at point x is given by

$$\mathbf{P}(\tau_0^x > \tau_{n/2}^x) = V(x) = \frac{R(0, x)}{R(0, n/2)} \sim \left(\frac{2x}{n}\right)^{5/2}.$$

From (4.5), one deduces that there exists a constant $c > 0$ such that

(4.10) $$\mathbf{P}(\tau_0^{n^\alpha} < \tau_{n/2}^{n^\alpha}) \geq 1 - c(2n^{\alpha-1})^{5/2}.$$

Then we define two sequences of stopping times (τ_i) and (σ_i) in the following way:

$$\tau_0 := \inf\{t : X_t = 0\}, \qquad \sigma_i := \inf\{t > \tau_{i-1} : X_t = n^\alpha\},$$

$$\tau_i := \inf\{t > \sigma_i : X_t = 0\}$$

and, for a given $T > 0$, we introduce

$$\nu(T) := \max\{i : \tau_i < T \wedge \tau_{n/2}^{n^\alpha}\}.$$

For any $x \leq n^\alpha$ and $N > 1$, making use of (4.10), we can write

$$\mathbf{P}(\tau_{n/2}^x > T) \geq \mathbf{P}(\tau_{n/2}^x > T, \nu(T) < N)$$

(4.11) $$= \mathbf{P}(\tau_{n/2}^x > T \mid \nu(T) < N)\mathbf{P}(\nu(T) < N)$$

$$\geq \left(1 - (2n^{\alpha-1})^{5/2}\right)^N \mathbf{P}(\nu(T) < N).$$

On the other hand, we can show that

(4.12) $$\mathbf{E}(\nu(T)) \leq \frac{T}{\mathbf{E}(\tau_{n^\alpha}^0)}.$$

Equation (4.12) is obtained in the following way: we consider the martingale (M_k) defined by

(4.13) $$M_k := \sum_{i=1}^{k} [\tau_{n^\alpha}^0 - \mathbf{E}(\tau_{n^\alpha}^0)].$$

Since $T \geq \nu(T)\tau_{n^\alpha}^0$, we get the inequality

(4.14) $$M^{\nu(T)} \leq T - \nu(T)\mathbf{E}(\tau_{n^\alpha}^0).$$

Then we apply Doob's optional-stopping theorem to M at time $\nu(T)$, and (4.12) follows.

Now we prove the equivalence:

(4.15) $$\mathbf{E}(\tau_{n^\alpha}^0) \sim c n^{5\alpha/2}.$$

We make use of the equivalences (4.9) and (4.1) and we write

(4.16)
$$\mathbf{E}(\tau_{n^\alpha}^0) = \sum_{k=0}^{n^\alpha-1} \pi(k) R_{n^\alpha}^k = \sum_{k=0}^{n^\alpha-1} \pi(k) \sum_{i=k}^{n^\alpha-1} R_i^{i+1}$$

$$\sim c \sum_{k=0}^{n^\alpha-1} \pi(k) \sum_{i=k}^{n^\alpha-1} i^{3/2} \sim c \sum_{i=0}^{n^\alpha-1} \sum_{k=0}^{i} i^{3/2} \frac{n^{3/2}}{(1 \vee k)^{3/2}(n-k)^{3/2}}.$$

Since uniformly on $k \in [0, n^\alpha]$, we have

(4.17)
$$\frac{n^{3/2}}{(n-k)^{3/2}} \sim 1,$$

we deduce from (4.16) that

(4.18)
$$\mathbf{E}(\tau_{n^\alpha}^0) \sim c \sum_{i=0}^{n^\alpha-1} i^{3/2} \sum_{k=0}^{i} \frac{1}{(1 \vee k)^{3/2}} \sim c \sum_{i=0}^{n^\alpha-1} i^{3/2} \sim c n^{5\alpha/2},$$

and hence we recover (4.15).

Now we prove that the conditions of Proposition 3.2 hold.

Mixing condition. Fix $\varepsilon > 0$ sufficiently small and consider $R = R_n = n^{(5/2)-\varepsilon}$ and $N = n^{(1-\alpha)(5/2)-\varepsilon/2}$. We first prove that $d_A(R_n) \to 0$ by using its estimate in terms of coupling time (see, e.g., [25], Chapter 5, Theorem 5.2, for this classical result):

(4.19)
$$d(R_n) \le \mathbf{P}(\tau_0^{n/2} > R_n) \le \frac{\mathbf{E}\tau_0^{n/2}}{R_n}.$$

We show the bound

(4.20)
$$\mathbf{E}[\tau_0^{n/2}] \le cn^2.$$

Indeed, we have the inequalities

$$\mathbf{E}[\tau_0^{n/2}] = \sum_{k=0}^{n/2} \pi(k) R_0^k \le c \sum_{k=0}^{n/2} \frac{1}{(1 \vee k)^{3/2}} \sum_{i=0}^{k} i^{3/2}$$

$$\le c \sum_{i=0}^{n/2} i^{3/2} \sum_{k=i}^{n/2} \frac{1}{(1 \vee k)^{3/2}} \sim c \sum_{i=0}^{n/2} i^{3/2} i^{-1/2} \sim cn^2,$$

from which (4.20) follows.

Hence, the quantity in the right-hand side of (4.19) vanishes as $n \to \infty$, and (3.8) holds with $R = R_n$.

Slow escape. On the other hand, combining Markov's inequality and (4.15), we get

$$(4.21) \qquad \mathbf{P}\big(\nu(T) < N\big) \geq 1 - \frac{\mathbf{E}(\nu(T))}{N} \geq 1 - c\frac{T}{n^{5\alpha/2}N} = 1 - cn^{-\varepsilon/2}.$$

Finally, we combine (4.11) and (4.21) to deduce that, for any $x \in \mathcal{B}_\alpha$,

$$(4.22) \qquad \mathbf{P}\big(\tau_{n/2}^x > T\big) \geq \big(1 - cn^{-\varepsilon/2}\big)\big(1 - n^{-\varepsilon/2}\big) \sim 1 - Cn^{-\varepsilon/2}.$$

We finally conclude

$$\sum_x \pi(x)\mathbf{P}\big(\tau_{n/2}^{(n),x} \leq 2T_n\big) \leq \sum_{x \leq n^\alpha} \pi(x)Cn^{-\varepsilon/2} + \sum_{x > n^\alpha} \pi(x)$$

$$\leq Cn^{-\varepsilon/2} + C'n^{-\alpha/2} \to 0,$$

which achieves the proof of $E(R_n, f)$; thus we can apply Theorem 3.6 to deduce Theorem 4.1. $\quad\square$

4.2. *Top-In-At-Random, an entropic barrier.* In this part, we apply the results of the Section 3.2 to the Top-In-At-Random model. The state space is $\mathcal{X}^{(n)} = \{\text{permutations of } (1, \ldots, n)\}$ and it is called the "deck of n cards." The dynamics is the following: at each step, we take the upper card and put it in a random position; more precisely, we update the configuration $x = (x_1, \ldots, x_n)$ to a new configuration $x^k = (x_2, \ldots, x_k, x_1, x_{k+1}, \ldots, x_n)$, with k uniform in $\{1, \ldots, n\}$. We call the transition form x to x^k a "shuffle." By symmetry, it is straightforward to realize that the invariant measure of this dynamics is uniform on $\mathcal{X}^{(n)}$.

This model is well known, it is historically the first model which was shown to exhibit cutoff [3]. Here, we will only use the fact that in the same paper the authors show that $R_n \sim n \log(n)$.

We consider $G := \{(1, \ldots, n)\}$ the ordered configuration, and we define a projection

$$(4.23) \qquad\qquad\qquad \sigma(x) = \max\{i \leq n, x_i > x_{i+1}\}.$$

It is easy to see that $\sigma(x)$ measures the minimal number of shuffles to reach G starting from x; in particular, $\sigma(x) = 0$ if and only if $x = G$.

LEMMA 4.2. *The process $\sigma(X_t^x)$ is still Markovian with transition probabilities given by*

$$(4.24) \quad P_{i,j} := \mathbf{P}\big[\sigma(X_{t+1}^x) = j | \sigma(X_t^x) = i\big] = \begin{cases} 1/n, & \text{if } j = i - 1, \\ i/n, & \text{if } j = i, \\ 1/n, & \text{if } j = i + 1, \ldots, n. \end{cases}$$

PROOF.   Let $y = X_t^x$ be the current configuration and k be the new position for the first card, so that $X_{t+1}^x = y^k$. Let $\sigma := \sigma(y)$. If the first card is inserted above σ, namely if $k < \sigma$, then $\sigma(y^k) = \sigma$. There are $\sigma - 1$ available choices.

Otherwise, consider the stack formed by the last $n - \sigma$ cards of y plus the card $y(1)$. Each of the possible positions $k \in \{\sigma, \dots, n\}$ corresponds to a different $\sigma(x^k) \in \{\sigma - 1, \dots, n - 1\}$. To see this, it is sufficient to observe that $k \longrightarrow \sigma(x^k)$ is invertible in $\{\sigma, \dots, n\}$, the inverse application being $\{\sigma - 1, \dots, n - 1\}$. Let $k^* := \max\{i \in \{\sigma, \dots, n\}; y(i) < y(1)\}$ be the "natural position" of the card $x(1)$ into this stack [notice that $k^* = \sigma$ if $y(1) < y(\sigma + 1)$].

We get the value $\sigma(y^k) = \sigma' \in \{\sigma - 1, \dots, n - 1\}$ by choosing

$$(4.25) \qquad k(\sigma') = \begin{cases} k^*, & \text{if } \sigma' = \sigma - 1, \\ \sigma', & \text{if } \sigma - 1 < \sigma' < k^*, \\ \sigma' + 1, & \text{if } \sigma' \geq k^*. \end{cases}$$

Thus, under the condition $k \geq \sigma$, $\sigma(X_{t+1}^x)$ takes the values in $\{\sigma - 1, \dots, n - 1\}$ with uniform probability.   □

We denote by $\sigma^i(t)$ the Markov process with transition matrix P defined in (4.24).

To show that HpG holds, we need to get good estimates on the hitting times τ_0^k of $\{0\}$ for the projected chain σ^k.

Let $\xi_j^i(k) := |\{t \leq \tau_j^i, \sigma^i(t) = k\}|$ be the local time spent in k before hitting j when starting from i. We define $\xi_j^i := \xi_j^i(i)$.

Since in the downward direction only one-step transitions are allowed, we have

$$\mathbf{E}[\tau_0^{n-1}] = \mathbf{E}\left[\sum_{k=1}^{n-1} \xi_0^{n-1}(k)\right]$$

$$(4.26) \qquad = \sum_{k=1}^{n} \mathbf{E}[\xi_0^k] \mathbf{P}[\tau_k^{n-1} < \tau_0^{n-1}]$$

$$= \sum_{k=1}^{n} \mathbf{E}[\xi_0^k],$$

where we used the strong Markov property at time τ_k^0 in the second identity.

Since

$$\mathbf{P}[\xi_j^i > n] = \mathbf{P}[\tau_i^i < \tau_j^i] \mathbf{P}[\xi_j^i > n - 1]$$

$$(4.27) \qquad = \mathbf{P}[\tau_i^i < \tau_j^i]^n,$$

ξ_j^i is a geometric variable with mean $\mathbf{P}[\tau_j^i < \tau_i^i]^{-1}$.

Let I be a subset of $\{0, \ldots, n-1\}$. We recall the standard "renewal" identity

$$(4.28) \qquad \mathbf{P}[\tau_I^i < \tau_j^i] = \frac{\mathbf{P}[\tau_I^i < \tau_{j \cup i}^i]}{\mathbf{P}[\tau_{I \cup j}^i < \tau_i^i]}.$$

Our main estimate is contained in the following lemma.

LEMMA 4.3.   *For any* $k \in [1, n]$,

$$(4.29) \qquad \mathbf{P}[\tau_0^k < \tau_k^k] \leq \frac{(n-1)(n-k-1)!}{n!}.$$

PROOF.   By (4.28),

$$
\begin{aligned}
(4.30) \qquad \mathbf{P}[\tau_0^k < \tau_k^k] &= P_{k,k-1} \mathbf{P}[\tau_0^{k-1} < \tau_k^{k-1}] \\
&= P_{k,k-1} \frac{\mathbf{P}[\tau_0^{k-1} < \tau_{k,k-1}^{k-1}]}{\mathbf{P}[\tau_{0,k}^{k-1} < \tau_{k-1}^{k-1}]}.
\end{aligned}
$$

Since the downward moves are single-step, for $k \geq 1$, we have

$$(4.31) \qquad \mathbf{P}[\tau_0^{k-1} < \tau_{k,k-1}^{k-1}] = \mathbf{P}[\tau_0^{k-1} < \tau_{k-1}^{k-1}].$$

On the other hand,

$$
\begin{aligned}
(4.32) \qquad \mathbf{P}[\tau_{0,k}^{k-1} < \tau_{k-1}^{k-1}] &= \mathbf{P}[\tau_0^{k-1} < \tau_{k-1}^{k-1}] + \mathbf{P}[\tau_k^{k-1} < \tau_{k-1}^{k-1} < \tau_0^{k-1}] \\
&\geq \mathbf{P}[\tau_0^{k-1} < \tau_{k-1}^{k-1}] + 1 - P_{k-1,k-2} - P_{k-1,k-1},
\end{aligned}
$$

where we used again the fact that downhill moves are unitary.

Let $E_k := \mathbf{P}[\tau_0^k < \tau_k^k]^{-1}$. Plugging (4.31), (4.32) into (4.30), and using (4.24) we get

$$(4.33) \qquad E_k \geq n + (n-k)E_{k-1}.$$

Since $E_1 = p_{1,0}^{-1} = n$, we prove inductively that

$$(4.34) \qquad E_k \geq n \sum_{j=1}^{k} \frac{(n-j-1)!}{(n-k-1)!}.$$

Indeed, (4.34) holds for $k = 1$ and, by (4.33), (4.34),

$$
\begin{aligned}
(4.35) \qquad E_{k+1} &\geq n\left(1 + (n-k-1)\sum_{j=1}^{k} \frac{(n-j-1)!}{(n-k-1)!}\right) \\
&= \frac{n}{(n-k-2)!}\left((n-k-2)! + \sum_{j=1}^{k}(n-j-1)!\right).
\end{aligned}
$$

By taking only the largest term in (4.34), (4.29) immediately follows. $\square$

LEMMA 4.4. *The invariant measure of σ^i is*

(4.36)
$$\bar{\mu}(k) = \begin{cases} \dfrac{1}{n!}, & \text{if } k = 0, \\[2mm] \dfrac{n-k}{(n-k+1)!}, & \text{if } k > 0. \end{cases}$$

PROOF.   Direct computation shows that, for $k \in \{0, \ldots, n\}$,

(4.37)
$$\bar{\mu}(\{0, \ldots, k\}) = \frac{1}{(n-k)!},$$

and, therefore,

(4.38)
$$\begin{aligned} \sum_{i=0}^{n} \bar{\mu}(i) P_{i,j} &= \frac{1}{n} \sum_{i=0}^{j+1} \bar{\mu}(i) + \frac{j-1}{n} \bar{\mu}(j) \\ &= \frac{1}{n} \frac{1}{(n-j-1)!} + \frac{j-1}{n} \bar{\mu}(j) \\ &= \frac{1}{n} \left(\frac{1}{(n-j-1)!} \frac{(n-j+1)!}{n-j} + (j-1) \right) \bar{\mu}(j) \\ &= \frac{(n-j+1)+(j-1)}{n} \bar{\mu}(j) = \bar{\mu}(j). \end{aligned} \qquad \square$$

THEOREM 4.5.   *HpG holds in the TIAR model with $n \log n \ll R_n \ll (n-2)!$.*

PROOF.   We get

(4.39)
$$\begin{aligned} \sum_{x \in \mathcal{X}_n} \mu(x) \mathbf{P}(\tau_G^x < R) &= \sum_{k=0}^{n-1} \bar{\mu}(k) \mathbf{P}(\tau_0^k < R) \\ &\leq \sum_{k=1}^{n-1} \bar{\mu}(k) \mathbf{P}(\xi_0^k < R) + \bar{\mu}(0) \\ &= \sum_{k=1}^{n-1} \bar{\mu}(k) (1 - \mathbf{P}[\tau_k^k < \tau_0^k]^R) + \bar{\mu}(0) \\ &\leq \sum_{k=1}^{n-1} \bar{\mu}(k) R \mathbf{P}[\tau_0^k < \tau_k^k] + \bar{\mu}(0), \end{aligned}$$

where we used (4.27) in the third identity, combined with the fact that for $r > 1$ and for any $x \in (0, 1)$, one has the inequality

(4.40)
$$1 - x^r \leq r(1-x).$$

By using Lemma 4.3 and Lemma 4.4, we get

$$\text{r.h.s. of (4.39)} \le R \sum_{k=1}^{n-1} \frac{n-k}{(n-k+1)!} \frac{(n-1)(n-k-1)!}{n!} + \frac{1}{n!}$$

$$(4.41) \qquad\qquad = \frac{R(n-1)}{n!} \sum_{j=2}^{n} \frac{1}{j} + \frac{1}{n!}$$

$$= \frac{Rn \log n}{n!}(1 + o(1)). \qquad\qquad\qquad \square$$

5. Proofs of Section 3.2.

5.1. *Control on evolution measures for $t \le 2R$.*   We first prove that, starting from π_A, the law of X stays close to π_A at least until time $2R$.

LEMMA 5.1.   *If $E(R, f)$ holds, then for $t \le 2R$*

$$(5.1) \qquad\qquad\qquad d_{\mathrm{TV}}(\mu_t^{\pi_A}, \pi_A) \le f.$$

PROOF.   We make use of the invariance of π and we write

$$d_{\mathrm{TV}}(\mu_t^{\pi_A}, \pi_A) = \frac{1}{2\pi(A)} \sum_{y \in \mathcal{X}} \left| \sum_{x \in A} \pi(x) \mathbf{P}(X_t^x = y) - \pi(y) \mathbf{1}_{\{y \in A\}} \right|$$

$$= \frac{1}{2\pi(A)} \sum_{y \in \mathcal{X}} \left| \sum_{x \in A} \pi(x) \mathbf{P}(X_t^x = y) - \sum_{z \in \mathcal{X}} \pi(z) \mathbf{P}(X_t^z = y) \mathbf{1}_{\{y \in A\}} \right|$$

$$\le \frac{1}{2\pi(A)} \left[\sum_{y \in G} \sum_{x \in A} \pi(x) \mathbf{P}(X_t^x = y) + \sum_{y \in A} \sum_{z \in G} \pi(z) \mathbf{P}(X_t^z = y) \right].$$

Recalling the definition of the time reversal process $X^\leftarrow$

$$\pi(z) \mathbf{P}(X_t^z = y) = \pi(y) \mathbf{P}(X_t^{\leftarrow y} = z),$$

and noting that $\sum_{y \in G} \mathbf{P}(X_t^x = y) \le \mathbf{P}(\tau_G^x \le t)$ (and that the same inequality holds for the time reversal process), we get

$$d_{\mathrm{TV}}(\mu_t^{\pi_A}, \pi_A) \le f_A(t) \le f$$

as soon as $t \le 2R$. $\square$

LEMMA 5.2.   *If $E(R, f)$ and $T(R, d)$ hold, then for any $t \in [R, 2R]$, we have*

$$(5.2) \qquad\qquad\qquad \sup_{x \in B_\alpha} d_{\mathrm{TV}}(\mu_t^x, \mu_t^{\pi_A}) \le d + f^{1-\alpha}.$$

PROOF.   We directly get

$$\sup_{x \in B_\alpha} d_{\mathrm{TV}}\big(\mu_t^x, \mu_t^{\pi_A}\big)$$

(5.3)
$$= \sup_{x \in B_\alpha} \frac{1}{2} \sum_{y \in \mathcal{X}} \Big| \sum_{z \in A} \frac{\pi(z)}{\pi(A)} \mathbf{P}(X_t^z = y) - \mathbf{P}(X_t^x = y) \Big|$$

$$\leq \sup_{x \in B_\alpha} \sup_{z \in B_\alpha} \frac{1}{2} \sum_{y \in \mathcal{X}} \big| \mathbf{P}(X_t^z = y) - \mathbf{P}(X_t^x = y) \big| + \frac{\pi(A \setminus B_\alpha)}{\pi(A)}$$

$$\leq d + f^{1-\alpha}. \qquad \qquad \square$$

PROOF OF PROPOSITION 3.2.   Let us assume now that $\bar{d}_A(R) \leq d$. We deduce

(5.4)
$$\sup_{x \in A} d_{\mathrm{TV}}\big(\mu_t^{\pi_A}, \mu_t^x\big)$$

$$= \sup_{x \in A} \frac{1}{2} \sum_{y \in \mathcal{X}} \Big| \sum_{z \in A} \frac{\pi(z)}{\pi(A)} \mathbf{P}(X_t^z = y) - \mathbf{P}(X_t^x = y) \Big| \leq \bar{d}_A(t) \leq d$$

for $t = R$. By triangular inequality, combining Lemma 5.1 and equality (5.4), we obtain that

(5.5)
$$\sup_{x \in A} d_{\mathrm{TV}}\big(\pi_A, \mu_R^x\big) \leq d + f.$$

Hence, we get that

(5.6)
$$\sup_{x \in \mathcal{X}} \mathbf{P}\big(\tau_{B_\alpha \cup G}^x > R\big) = \sup_{x \in A} \mathbf{P}\big(\tau_{B_\alpha \cup G}^x > R\big)$$

$$\leq \sup_{x \in A}\big(1 - \mathbf{P}\big(\tau_{B_\alpha}^x \leq R\big)\big)$$

$$\leq \sup_{x \in A}\big(1 - \mu_R^x(B_\alpha)\big) \leq d + f + f^{1-\alpha},$$

which proves Proposition 3.2. $\square$

5.2. *Proof of Theorem* 3.4.   The main point is the proof of (3.12). Indeed, (3.13) follows from noting that (3.12) implies

$$\sup_{t > 2R} d_{\mathrm{TV}}\big(\hat{\mu}_{A,t}^{\pi_A}, \pi_A\big) < \varepsilon_1,$$

and then we combine (3.11) and the triangular inequality.

On the other hand, (3.14) will follow from (3.13), from the triangular inequality and from the convergence

$$\mu_A^*(y) = \lim_{t \to \infty} \tilde{\mu}_{A,t}^{\pi_A}(y),$$

which, from (3.14), implies

$$d_{\mathrm{TV}}(\pi_A, \mu_A^*) < \varepsilon_2.$$

The strategy to prove (3.12) is to use the recurrence in the set B_α; let us first define for each $s \geq 0$ and $x \in \mathcal{X}$

(5.7) $$\sigma^x(s) := \inf\{s' > s : X_{s'}^x \in B_\alpha \cup G\}.$$

We will follow the general guideline:

1. We first control the distance between the conditioned and the restricted measures for starting points $z \in B_\alpha$ and $t \in [R, 2R]$.

2. Then we prove estimates on the distribution of τ_G^x in Lemma 5.4.

3. We complete the proof combining the previous ingredients and strong Markov property at time $\sigma(t - 2R)$.

4. Finally, we complete the proof for every starting point in A.

Note that these ingredients are frequently used in metastability in order to control the loss of memory with respect to initial conditions, and consequently to deduce exponential behavior. We refer to [19] for similar results in finite state space regime, where the basin B_α is replaced by a single metastable state and the occurrences to it determine the loss of memory.

5.2.1. *Control of the thermalization.* We state the following result.

LEMMA 5.3.   *For $t \in [R, 2R]$, any $z \in B_\alpha$ and any $y \in A$ let*

(5.8) $$g_t^z(y) := \mathbf{P}(X_t^z = y, \tau_G^z > t) - \pi_A(y),$$

and define

(5.9) $$g := \sup_{t \in [R,2R]} \sup_{z \in B_\alpha} \frac{1}{2} \sum_{y \in A} |g_t^z(y)|.$$

The following inequality holds:

(5.10) $$g \leq d + f + f^{1-\alpha} + 2f^\alpha.$$

PROOF.   We first note that

(5.11) $$\mathbf{P}(X_t^z = y, \tau_G^z > t) = \mu_t^z(y) - \mathbf{P}(X_t^z = y, \tau_G^z \leq t),$$

and since for any $z \in B_\alpha$, $\sum_{y \in A} \mathbf{P}(X_t^z = y, \tau_G^z \leq t) \leq \mathbf{P}(\tau_G^z \leq t) \leq 2f^\alpha$, Lemma 5.3 will follow by triangular inequality once we prove that

(5.12) $$\sup_{t \in [R,2R]} \sup_{z \in B_\alpha} d_{\mathrm{TV}}(\mu_t^z, \pi_A) \leq d + f^{1-\alpha} + f.$$

Combining Lemma 5.1, Lemma 5.2 and the triangular inequality, for $t \in [R, 2R]$, $z \in B_\alpha$, we obtain (5.12).   □

5.2.2. *Control of the recurrence time.*

LEMMA 5.4. *Assume that d and f are small enough for the quantity*

$$(5.13) \qquad \bar{c} := \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - (r + 2f^\alpha)}$$

to be well defined. Let $t \geq 2R$ and $t' \leq R$; for any $x \in B_\alpha$, one has the inequalities

$$(5.14) \qquad 1 \leq \frac{\mathbf{P}(\tau_G^x > t - t')}{\mathbf{P}(\tau_G^x > t)} \leq 1 + 2\bar{c},$$

$$(5.15) \qquad 1 \leq \frac{\mathbf{P}(\tau_G^{\pi_A} > t - t')}{\mathbf{P}(\tau_G^{\pi_A} > t)} \leq 1 + 2\bar{c}.$$

PROOF. The lower bound is trivial. As far as the upper bound is concerned, by monotonicity it is sufficient to prove that

$$(5.16) \qquad \frac{\mathbf{P}(\tau_G^x > t - R)}{\mathbf{P}(\tau_G^x > t)} \leq 1 + 2\bar{c}.$$

Define $\sigma := \sigma^x(t - 2R)$ [see definition (5.7)] and apply Markov's property to get

$$\begin{aligned}
\mathbf{P}&\big(\tau_G^x \in (t - R, t)\big) \\
&= \mathbf{P}\big(\tau_G^x \in (t - R, t), \sigma > t - R\big) + \mathbf{P}\big(\tau_G^x \in (t - R, t), \sigma \leq t - R\big) \\
&\leq \mathbf{P}\big(\tau_G^x > t - 2R\big)\Big[\sup_{z \in \mathcal{X}} \mathbf{P}(\tau_{G \cup B_\alpha}^z \geq R) + \sup_{z \in B_\alpha} \mathbf{P}(\tau_G^z \leq 2R)\Big] \\
&\leq \mathbf{P}\big(\tau_G^x > t - 2R\big)[r + 2f^\alpha],
\end{aligned}$$

where we made use of the definition of B_α and of inequality (3.7). Defining

$$c := r + 2f^\alpha,$$

we get the inequality

$$(5.17) \qquad \mathbf{P}\big(\tau_G^x > t - R\big) \leq \mathbf{P}\big(\tau_G^x > t\big) + c\mathbf{P}\big(\tau_G^x > t - 2R\big),$$

from which we deduce

$$(5.18) \qquad 1 - c\frac{\mathbf{P}(\tau_G^x > t - 2R)}{\mathbf{P}(\tau_G^x > t - R)} \leq \frac{\mathbf{P}(\tau_G^x > t)}{\mathbf{P}(\tau_G^x > t - R)}.$$

For $i \geq 1$, define the quantity

$$(5.19) \qquad y_i := \frac{\mathbf{P}(\tau_G^x > iR)}{\mathbf{P}(\tau_G^x > (i - 1)R)}.$$

Equation (5.18) entails that the sequence $(y_i)_{i \geq 1}$ satisfies the recursive inequality

$$1 - \frac{c}{y_{i-1}} \leq y_i.$$

This implies that for every $i \geq 1$, one has

$$(5.20) \qquad\qquad\qquad\qquad y_i \geq 1 - \bar{c}.$$

Let us show (5.20) by induction; if there exists $i \geq 1$ such that $y_i \geq 1 - \bar{c}$, then for every $j \geq i$, one has $y_j \geq 1 - \bar{c}$. Indeed, making use of equation (5.2.2), we get

$$y_{i+1} \geq 1 - \frac{c}{y_i} \geq 1 - \frac{c}{1-\bar{c}} = 1 - \bar{c}.$$

For the base of the induction, we note that

$$y_1 = \mathbf{P}(\tau_G^x > R) \geq 1 - \mathbf{P}(\tau_G^x \leq 2R) \geq 1 - \sup_{z \in B_\alpha} \mathbf{P}(\tau_G^z \leq 2R)$$

$$(5.21)$$

$$\geq 1 - 2f^\alpha > 1 - c > 1 - \bar{c}.$$

In particular, this implies that for every $t = kR, k \geq 2$

$$(5.22) \qquad\qquad \frac{\mathbf{P}(\tau_G^x > t - R)}{\mathbf{P}(\tau_G^x > t)} = \frac{1}{y_k} \leq \frac{1}{1-\bar{c}} \leq 1 + 2\bar{c},$$

since $\bar{c} < \frac{1}{2}$, which entails the claim of equation (5.16) in the case when t/R is an integer.

Assume now that t/R is not an integer, and define $k = \lfloor t/R \rfloor$, so that $t = kR + t_0$ with $t_0 < R$ and define

$$y_i(t_0) := \frac{\mathbf{P}(\tau_G^x > iR + t_0)}{\mathbf{P}(\tau_G^x > (i-1)R + t_0)}.$$

For each $t_0 < R$, we have the same recursive inequality for $y_i(t_0)$ and again for the base of the induction we have

$$y_1(t_0) := \frac{\mathbf{P}(\tau_G^x > R + t_0)}{\mathbf{P}(\tau_G^x > t_0)} \geq \mathbf{P}(\tau_G^x > 2R)$$

so that the induction still holds; this completes the proof of (5.16) for every $t \geq 2R$.

The proof of (5.15) follows the same way, noting that one uses condition (3.1) instead of (3.4) to initialize the recurrence. $\quad\square$

5.2.3. *Proof of Theorem* 3.4 *for starting points* $z \in B_\alpha$. As before, considering the time $\sigma^x = \sigma^x(t - 2R)$, the following equality holds:

$$\mathbf{P}(X_t^x = y, \tau_G^x > t)$$

$$= \sum_{z \in B_\alpha} \int_{t-2R}^{t-R} \mathbf{P}(X_s^x = z, \tau_G^x > s, \sigma^x \in (s, s + ds)) \mathbf{P}(X_{t-s}^z = y, \tau_G^z > t - s)$$

$$+ \mathbf{P}(X_t^x = y, \tau_G^x > t, \sigma^x > t - R).$$

By the definition of $g_t^z(y)$, we get

$$\mathbf{P}(X_t^x = y, \tau_G^x > t)$$

$$= \sum_{z \in B_\alpha} \int_{t-2R}^{t-R} \mathbf{P}(X_s^x = z, \tau_G^x > s, \sigma^x \in (s, s+ds)) [\pi_A(y) + g_{t-s}^z(y)]$$

$$+ \mathbf{P}(X_t^x = y, \tau_G^x > t, \sigma^x > t - R)$$

so that

$$\sum_{y \in A} |\tilde{\mu}_t^x(y) - \pi_A(y)|$$

$$\leq \sum_{y \in A} \left| \frac{\mathbf{P}(\tau_G^x > \sigma^x, \sigma^x \in [t-2R, t-R])}{\mathbf{P}(\tau_G^x > t)} - 1 \right| \pi_A(y)$$

(5.23)

$$+ \sum_{y \in A} \sum_{z \in B_\alpha} \int_{t-2R}^{t-R} \frac{\mathbf{P}(X_s^x = z, \tau_G^x > t - 2R, \sigma^x \in (s, s+ds))}{\mathbf{P}(\tau_G^x > t)} |g_{t-s}^z(y)|$$

$$+ \frac{\mathbf{P}(\tau_G^x > t - 2R, \sigma^x > t - R)}{\mathbf{P}(\tau_G^x > t)} =: \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

By using the Markov property, the estimates (5.14), (5.10) and the recursion in $B_\alpha \, Rc(R, r)$, we complete the proof by estimating the three terms, I, II and III of the r.h.s. of (5.23). Note that in what follows, we will make repeated use of the monotonicity property.

The first term can be estimated by

$$\mathrm{I} \leq 2\bar{c} \vee r(1 + 2\bar{c}) \leq 2\bar{c} \vee 2r \leq 2\bar{c}.$$

Indeed, recalling that $t \geq 2R$, one can apply Lemma 5.4 to get

$$\frac{\mathbf{P}(\tau_G^x > \sigma^x, \sigma^x \in [t-2R, t-R])}{\mathbf{P}(\tau_G^x > t)} \leq \frac{\mathbf{P}(\tau_G^x > t - 2R)}{\mathbf{P}(\tau_G^x > t)} \leq 1 + 2\bar{c}.$$

On the other hand, making use of the Markov property at time $t - 2R$,

$$\frac{\mathbf{P}(\tau_G^x > \sigma^x, \sigma^x \in [t-2R, t-R])}{\mathbf{P}(\tau_G^x > t)}$$

$$\geq \frac{\mathbf{P}(\tau_G^x > t - R, \sigma^x \in [t-2R, t-R])}{\mathbf{P}(\tau_G^x > t)}$$

$$\geq \frac{\mathbf{P}(\tau_G^x > t - R)}{\mathbf{P}(\tau_G^x > t)} - \frac{\mathbf{P}(\tau_G^x > t - 2R, \sigma^x > t - R)}{\mathbf{P}(\tau_G^x > t)}$$

$$\geq 1 - \frac{\mathbf{P}(\tau_G^x > t - 2R)}{\mathbf{P}(\tau_G^x > t)} \sup_{z \in \mathcal{X}} \mathbf{P}(\tau_{B_\alpha \cup G}^z > R)$$

$$\geq 1 - r(1 + 2\bar{c}).$$

To deal with II, we use Lemma 5.3 to get

$$\text{II} \leq g \frac{\mathbf{P}(\tau_G^x > t - 2R)}{\mathbf{P}(\tau_G^x > t)} \leq (2f^\alpha + f^{1\alpha} + f + d)(1 + 2\bar{c})$$

$$\leq 4(2f + f^\alpha + d),$$

and similarly for III

$$\text{III} \leq r(1 + 2\bar{c}) \leq 2\bar{c}.$$

Finally, we prove (3.13); from (3.5), we obtain that for $t \geq 2R$

$$d_{\text{TV}}(\tilde{\mu}_{A,t}^{\pi_A}, \pi_A) = \frac{1}{2} \sum_{y \in \mathcal{X}} \left| \sum_{x \in A} \pi_A(x) \left[\tilde{\mu}_{A,t}^x(y) - \pi_A(y) \right] \right|$$

$$\leq 10\bar{c}\pi_A(B_\alpha) + \pi_A(B_\alpha^c)$$

$$\leq 10\bar{c} + f^{1-\alpha},$$

and finally we can conclude by making use of Remark 3.3.

5.2.4. *Proof of Theorem* 3.4 *for starting points* $x \in A \setminus B_\alpha$. We consider $t \geq 2R$, $x \in A \setminus B_\alpha$, $y \in A$ and we make use of Markov's property to get the equality

$$\mathbf{P}(X_t^x = y, \tau_G^x > t, \tau_{G \cup B_\alpha}^x < t - 2R)$$

$$(5.24) \qquad = \int_0^{t-2R} \sum_{z \in B_\alpha} \mathbf{P}(X_s^x = z, \tau_G^x > s, \tau_{G \cup B_\alpha}^x \in ds)$$

$$\times \mathbf{P}(X_{t-s}^z = y, \tau_G^z > t - s).$$

Since for $z \in B_\alpha$, $\tilde{\mu}_{A,t}^z$ and $\hat{\mu}_{A,t}^z$ coincide (see Remark 3.3), making use of (3.12) for starting points $z \in B_\alpha$, which we already proved in the previous section, we get that for $s \in [0, t - 2R]$, the quantity $f_s^z(y)$ defined by

$$(5.25) \qquad f_s^z(y) = \tilde{\mu}_{t-s,A}^z(y) - \pi_A(y)$$

satisfies

$$(5.26) \qquad \sup_{s \leq t - 2R} \sup_{z \in B_\alpha} \sum_{y \in A} |f_s^z(y)| < \varepsilon_1.$$

On the other hand, as in (5.24), we get

$$(5.27) \qquad \mathbf{P}(\tau_G^x > t, \tau_{G \cup B_\alpha}^x < t - 2R)$$

$$= \int_0^{t-2R} \sum_{z \in B_\alpha} \mathbf{P}(X_s^x = z, \tau_G^x > s, \tau_{G \cup B_\alpha}^x \in ds)\mathbf{P}(\tau_G^z > t - s).$$

Combining (5.24) and (5.27), we can write

$$(5.28) \quad \mathbf{P}\big(X_t^x = y, \tau_G^x > t, \tau_{G \cup B_\alpha}^x < t - 2R\big) - \pi_A(y)\mathbf{P}\big(\tau_G^x > t, \tau_{G \cup B_\alpha}^x < t - 2R\big)$$

$$= \int_0^{t-2R} \sum_{z \in B_\alpha} \mathbf{P}\big(X_s^x = z, \tau_G^x > s, \tau_{G \cup B_\alpha}^x \in ds\big)$$

(5.29)

$$\times \big(\pi_A(y)\mathbf{P}\big(\tau_G^z > t - s\big) + f_s^z(y)\mathbf{P}\big(\tau_G^z > t - s\big)\big)$$

$$- \pi_A(y) \int_0^{t-2R} \sum_{z \in B_\alpha} \mathbf{P}\big(X_s^x = z, \tau_G^x > s, \tau_{G \cup B_\alpha}^x \in ds\big)$$

(5.30)

$$\times \mathbf{P}\big(\tau_G^z > t - s\big)$$

$$(5.31) \qquad = \int_0^{t-2R} \sum_{z \in B_\alpha} \mathbf{P}\big(X_s^x = z, \tau_G^x > s, \tau_{G \cup B_\alpha}^x \in ds\big) f_s^z(y)\mathbf{P}\big(\tau_G^z > t - s\big).$$

From the last equality and (5.27), we deduce that for any $t \geq 2R$, $x \in A \setminus B_\alpha$ and $y \in A$,

$$(5.32) \qquad \big|\hat{\mu}_{t,A}^x(y) - \pi_A(y)\big| \leq \sup_{s \leq t-2R} \sup_{z \in B_\alpha} \big|f_s^z(y)\big|,$$

from which (3.12) follows from considering (5.26).

5.3. *Proof of Theorem* 3.6. We define

$$(5.33) \qquad\qquad T^* = \mathbf{E}\big[\tau_G^{\mu_A^*}\big].$$

For $y \in A$, we consider the quantity

$$(5.34) \qquad\qquad \delta(y) := 2d_{\mathrm{TV}}\big(\hat{\mu}_{A,2R}^y, \mu_A^*\big),$$

and recalling (3.14), we have

$$\delta := \sup_{y \in A} \delta(y) \leq \varepsilon_1 + \varepsilon_2.$$

We first show that the recurrence time is asymptotically negligible with respect to T^*.

LEMMA 5.5. *There exists a constant $C > 0$ such that the following inequality holds*:

$$(5.35) \qquad\qquad \frac{R}{T^*} \leq C(f + \delta).$$

PROOF. By Proposition 2.1, we have

$$(5.36) \qquad\qquad \mathbf{P}\big[\tau_G^{\mu_A^*} > 2R\big] = e^{-2R/T^*}.$$

On the other hand, by making use of (3.14) and of $E(R, f)$,

$$\mathbf{P}[\tau_G^{\mu_A^*} > 2R] = \sum_{z \in A} \mu_A^*(z)\mathbf{P}[\tau_G^z > 2R]$$

(5.37)

$$= \mathbf{P}[\tau_G^{\pi_A} > 2R] + \sum_{z \in A}[\mu_A^*(z) - \pi_A(z)]\mathbf{P}[\tau_G^z > 2R]$$

(5.38) $$\geq \mathbf{P}[\tau_G^{\pi_A} > 2R] - d_{\text{TV}}(\pi_A, \mu_A^*) \geq 1 - f - \delta,$$

which proves the claim. □

LEMMA 5.6.   *The following holds for any $k \geq 0$:*

(5.39) $$\sup_{y \in A} \mathbf{P}(\tau_G^y > 2kR) \leq e^{-k(2R/T^*)}\left(1 + O\left(\delta + r + \frac{R}{T^*}\right)\right).$$

PROOF.   Making use of Markov's property at time $2R$, we have for any $t \geq 2R$ and $y \in A$

(5.40) $$\mathbf{P}(\tau_G^y > t) = \mathbf{P}(\tau_G^y > 2R, \tau_{B_\alpha \cup G}^y \leq R)\sum_{z \in A}\hat{\mu}_{A,2R}^y(z)\mathbf{P}(\tau_G^z > t - 2R)$$

(5.41) $$+ \sum_{z \in A}\mathbf{P}(\tau_G^y > 2R, \tau_{B_\alpha \cup G}^y > R, X_{2R}^y = z)\mathbf{P}(\tau_G^z > t - 2R).$$

By using Proposition 2.1 and the hypothesis $Rc(R, r)$, we have the following estimate:

$$\mathbf{P}(\tau_G^y > t) \leq \mathbf{P}(\tau_G^y > 2R, \tau_{B_\alpha \cup G}^y \leq R)$$

$$\times \left[e^{-(t-2R)/T^*} + \sum_{z \in A}(\hat{\mu}_{A,2R}^y(z) - \mu_A^*(z))\mathbf{P}(\tau_G^z > t - 2R)\right]$$

$$+ r\sup_{z \in A}\mathbf{P}(\tau_G^z > t - 2R).$$

In particular, for $t = 2(k + 1)R$, we get

(5.42)
$$|\mathbf{P}(\tau_G^y > (k + 1)2R) - \mathbf{P}(\tau_G^y > 2R, \tau_{B_\alpha \cup G}^y \leq R)e^{-(t-2R)/T^*}|$$
$$\leq (\delta + r)\sup_{z \in A}\mathbf{P}(\tau_G^z > k2R)$$

and

(5.43) $$\sup_{y \in A}\mathbf{P}(\tau_G^y > 2(k + 1)R) \leq e^{-k(2R/T^*)} + (\delta + r)\sup_{z \in A}\mathbf{P}(\tau_G^z > 2kR).$$

From (5.43), we prove by recurrence that

(5.44) $$\sup_{y \in A}\mathbf{P}(\tau_G^y > 2kR) \leq (\delta + r)^k + \frac{e^{-k(2R/T^*)}}{e^{-2R/T^*} - (\delta + r)}.$$

Indeed, by Lemma 5.5, as soon as R/T^*, δ and r are small enough, (5.44) holds for $k = 0$; assume (5.44) to be true for a given $k \geq 0$. Using (5.43), we get that

$$\sup_{y \in A} \mathbf{P}(\tau_G^y > 2(k+1)R)$$

(5.45)

$$\leq e^{-(k+1)(2R/T^*)} + (\delta + r)\left((\delta + r)^k + \frac{e^{-k(2R/T^*)}}{e^{-2R/T^*} - (\delta + r)}\right)$$

(5.46) $$\leq (\delta + r)^{k+1} + e^{-k(2R/T^*)}\left(\frac{e^{-2R/T^*} - (\delta + r) + (\delta + r)}{e^{-2R/T^*} - (\delta + r)}\right),$$

which closes the recursion.

Now we note that (5.44) implies Lemma 5.6.   □

PROOF OF THEOREM 3.6.   For $x \in A \setminus B_\alpha$, (3.17) directly follows from (5.42) and from (5.44). Note that one gets from a statement on a time scale $2R$ to a statement on a generic time scale $t \geq 0$ in a standard way and by using Lemma 5.5.

To get the statement (3.16), we note that for $x \in B_\alpha$, we have

(5.47) $$\left|\mathbf{P}[\tau_G^x > 2R, \tau_{B_\alpha \cup G}^x \leq R] - 1\right| \leq r + f^\alpha,$$

and hence we are done.

Finally, in the same way, making use of the slow escape property $E(R, f)$, we get that

(5.48) $$\mathbf{P}(\tau_G^{\pi_A} > t) = e^{-t/T^*}\left(1 + O\left(\delta + \frac{R}{T^*} + f\right)\right).$$ □

REFERENCES

[1] ALDOUS, D. J. and BROWN, M. (1992). Inequalities for rare events in time-reversible Markov chains. I. In *Stochastic Inequalities* (*Seattle, WA, 1991*). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **22** 1–16. IMS, Hayward, CA. MR1228050

[2] ALDOUS, D. J. and BROWN, M. (1993). Inequalities for rare events in time-reversible Markov chains. II. *Stochastic Process. Appl.* **44** 15–25. MR1198660

[3] ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–348. MR0841111

[4] ALDOUS, D. and DIACONIS, P. (1987). Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** 69–97. MR0876954

[5] BELTRÁN, J. and LANDIM, C. (2011). Metastability of reversible finite state Markov processes. *Stochastic Process. Appl.* **121** 1633–1677. MR2811018

[6] BELTRÁN, J. and LANDIM, C. (2012). Tunneling and metastability of continuous time Markov chains II, the nonreversible case. *J. Stat. Phys.* **149** 598–618. MR2998592

[7] BENOIS, O., LANDIM, C. and MOURRAGUI, M. (2013). Hitting times of rare events in Markov chains. *J. Stat. Phys.* **153** 967–990. MR3131505

[8] BIANCHI, A. and GAUDILLIÈRE, A. (2011). Metastable states, quasi-stationary and soft measures, mixing time asymprtotics via variational principles. Available at arXiv:1103.1143.

[9] BOVIER, A. (2006). Metastability: A potential theoretic approach. In *International Congress of Mathematicians*, *Vol. III* 499–518. Eur. Math. Soc., Zürich. MR2275693

[10] BOVIER, A., DEN HOLLANDER, F. and SPITONI, C. (2010). Homogeneous nucleation for Glauber and Kawasaki dynamics in large volumes at low temperatures. *Ann. Probab.* **38** 661–713. MR2642889

[11] BOVIER, A., ECKHOFF, M., GAYRARD, V. and KLEIN, M. (2000). Metastability and small eigenvalues in Markov chains. *J. Phys. A* **33** L447–L451. MR1804034

[12] BOVIER, A., ECKHOFF, M., GAYRARD, V. and KLEIN, M. (2001). Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theory Related Fields* **119** 99–161. MR1813041

[13] CAPUTO, P., MARTINELLI, F. and TONINELLI, F. L. (2008). On the approach to equilibrium for a polymer with adsorption and repulsion. *Electron. J. Probab.* **13** 213–258. MR2386733

[14] CAPUTO, P., LACOIN, H., MARTINELLI, F., SIMENHAUS, F. and TONINELLI, F. L. (2012). Polymer dynamics in the depinned phase: Metastability with logarithmic barriers. *Probab. Theory Related Fields* **153** 587–641. MR2948687

[15] CASSANDRO, M., GALVES, A., OLIVIERI, E. and VARES, M. E. (1984). Metastable behavior of stochastic dynamics: A pathwise approach. *J. Stat. Phys.* **35** 603–634. MR0749840

[16] CATONI, O. (1999). Simulated annealing algorithms and Markov chains with rare transitions. In *Séminaire de Probabilités*, *XXXIII*. *Lecture Notes in Math.* **1709** 69–119. Springer, Berlin. MR1767994

[17] CATONI, O. and CERF, R. (1995/1997). The exit path of a Markov chain with rare transitions. *ESAIM Probab. Stat.* **1** 95–144 (electronic). MR1440079

[18] CIRILLO, E. N. M. and NARDI, F. R. (2003). Metastability for a stochastic dynamics with a parallel heat bath updating rule. *J. Stat. Phys.* **110** 183–217. MR1966327

[19] FERNANDEZ, R., MANZO, F., NARDI, F. R. and SCOPPOLA, E. (2014). Asymptotically exponential hitting times and metastability: A pathwise approach without reversibility. Available at arXiv:1406.2637.

[20] GAUDILLIÈRE, A. and LANDIM, C. (2014). A Dirichlet principle for nonreversible Markov chains and some recurrence theorems. *Probab. Theory Related Fields* **158** 55–89. MR3152780

[21] GAUDILLIÈRE, A. and NARDI, F. R. (2010). An upper bound for front propagation velocities inside moving populations. *Braz. J. Probab. Stat.* **24** 256–278. MR2643566

[22] GAUDILLIÈRE, A., DEN HOLLANDER, F., NARDI, F. R., OLIVIERI, E. and SCOPPOLA, E. (2009). Ideal gas approximation for a two-dimensional rarefied gas under Kawasaki dynamics. *Stochastic Process. Appl.* **119** 737–774. MR2499857

[23] KEILSON, J. (1979). *Markov Chain Models—Rarity and Exponentiality*. *Applied Mathematical Sciences* **28**. Springer, New York. MR0528293

[24] LANDIM, C. (2014). Metastability for a nonreversible dynamics: The evolution of the condensate in totally asymmetric zero range processes. *Comm. Math. Phys.* **330** 1–32. MR3215575

[25] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[26] LYONS, R. and PERES, Y. (2015). *Probability on Trees and Networks*. Cambridge Univ. Press, Cambridge. In preparation; available at http://mypage.iu.edu/~rdlyons.

[27] MANZO, F., NARDI, F. R., OLIVIERI, E. and SCOPPOLA, E. (2004). On the essential features of metastability: Tunnelling time and critical configurations. *J. Stat. Phys.* **115** 591–642. MR2070109

[28] OLIVEIRA, R. I. (2013). Mean field conditions for coalescing random walks. *Ann. Probab.* **41** 3420–3461. MR3127887

[29] OLIVIERI, E. and SCOPPOLA, E. (1995). Markov chains with exponentially small transition probabilities: First exit problem from a general domain. I. The reversible case. *J. Stat. Phys.* **79** 613–647. MR1327899

[30] OLIVIERI, E. and SCOPPOLA, E. (1996). Markov chains with exponentially small transition probabilities: First exit problem from a general domain. II. The general case. *J. Stat. Phys.* **84** 987–1041. MR1412076

[31] OLIVIERI, E. and VARES, M. E. (2005). *Large Deviations and Metastability. Encyclopedia of Mathematics and Its Applications* **100**. Cambridge Univ. Press, Cambridge. MR2123364

[32] SCHONMANN, R. H. and SHLOSMAN, S. B. (1998). Wulff droplets and the metastable relaxation of kinetic Ising models. *Comm. Math. Phys.* **194** 389–462. MR1627669

[33] SCOPPOLA, E. (1993). Renormalization group for Markov chains and application to metastability. *J. Stat. Phys.* **73** 83–121. MR1247859

[34] SCOPPOLA, E. (1994). Metastability for Markov chains: A general procedure based on renormalization group ideas. In *Probability and Phase Transition* (*Cambridge*, 1993). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **420** 303–322. Kluwer Academic, Dordrecht. MR1283189

[35] SENETA, E. (2006). *Nonnegative Matrices and Markov Chains*. Springer, New York. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544]. MR2209438

[36] TROUVÉ, A. (1992). Partially parallel simulated annealing: Low and high temperature approach of the invariant measure. In *Applied Stochastic Analysis* (*New Brunswick*, *NJ*, 1991). *Lecture Notes in Control and Inform. Sci.* **177** 262–278. Springer, Berlin. MR1169933

R. FERNANDEZ
UNIVERSITY OF UTRECHT
P.O. BOX 80010
3508 TA UTRECHT
THE NETHERLANDS
E-MAIL: R.Fernandez1@uu.nl

F. MANZO
E. SCOPPOLA
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ROMA TRE
LARGO SAN MURIALDO 1
00146 ROMA
ITALY
E-MAIL: manzo.fra@gmail.com
        scoppola@mat.uniroma3.it

F. NARDI
DEPARTMENT OF MATHEMATICS
  AND COMPUTER SCIENCE
EINDHOVEN UNIVERSITY OF TECHNOLOGY
METAFORUM (MF) 4.076
P.O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: f.r.nardi@tue.nl

J. SOHIER
DEPARTMENT OF MATHEMATICS
  AND COMPUTER SCIENCE
EINDHOVEN UNIVERSITY OF TECHNOLOGY
METAFORUM (MF) 4.118
P.O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: jsohier@tue.nl