

GEROLD ALSMEYER AND ANDREA WINKLER

## METABASINS — A STATE SPACE AGGREGATION FOR HIGHLY DISORDERED ENERGY LANDSCAPES

Glass-forming systems, characterized by a highly disordered energy landscape, have been studied in physics by a simulation-based state space aggregation. Choosing a path-independent approach within the framework of finite Markov chains, this article provides an aggregation procedure which, at an appropriate aggregation level, leads to the definition of certain metastates, called metabasins in Physics (for their properties see the Introduction). Roughly speaking, this will be the finest aggregation such that transitions back to an already visited (meta-)state are very unlikely within a moderate time frame.

### INTRODUCTION

Supercooled liquids of glass forming systems are typical examples of high-dimensional systems with highly disordered energy landscapes and our main concern behind this work. Simulations have shown that many important characteristics of such a system are better described by a process on the set of so-called *metabasins* (MB) than by the more common process on the set of visited minima of the energy landscape (see [12]). Those MB are formed in the following way by aggregation of suitable states of the describing process  $(X_n)_{n \geq 0}$  along a simulated trajectory:

Fixing a reasonable observation time  $T$ , define  $\chi_0 \equiv 0$  and then recursively for  $n \geq 1$

$$\chi_n := \inf \{k > \chi_{n-1} \mid \{X_k, \dots, X_T\} \cap \{X_0, \dots, X_{k-1}\} = \emptyset\}.$$

Then the MB up to  $v := \sup\{n \geq 0 \mid \chi_n \leq T\}$  are chosen as

$$\mathcal{V}_n := \{X_{\chi_n}, \dots, X_{\chi_{n+1}-1}\}, \quad 0 \leq n \leq v.$$

Simulation studies have shown that local sampling within a MB does not affect typical parameters of the process like the diffusion coefficient or the time to reach equilibrium. Dynamical aspects are therefore fully characterized by the MB-valued process. Furthermore, this model reduction by aggregation, as proposed in [12] and [18], offers several advantages (referred to as Properties 1–5 hereafter):

- (1) The probability of a transition from one MB to any other one does not depend on the state at which this MB is entered.
- (2) There are basically no reciprocating jumps between two MB. This is in strong contrast to the unaggregated process where such jumps occur very often: The system falls back to a minimum many times before eventually cresting a high energy barrier and then falling into a new valley, where it will again take many unsuccessful trials to escape. These reciprocating jumps are not only irrelevant for the actual motion on the state space but also complicating the estimation of parameters like the diffusion coefficient or the relaxation time.

---

2000 *Mathematics Subject Classification*. Primary 60J10, 82C44.

*Key words and phrases*. Metastability, metabasins, Markov chain aggregation, disordered systems, exit time, Metropolis algorithm.

- (3) The expected time spent in a MB is proportional to its depth. Thus there is a strong and explicit relation between dynamics and thermodynamics, not in terms of the absolute but the relative energy.
- (4) The energy barriers between any two MB are approximately of the same height, that is, there is an energy threshold  $E_0$  such that, for a small  $\varepsilon$ , it requires a crossing of at least  $E_0 - \varepsilon$  and at most  $E_0 + \varepsilon$  to make a transition from one MB to another. Such systems with  $\varepsilon = 0$  are called trap models (see [3]).
- (5) The sojourn times and the jump distances between successively visited MB (measured in Euclidean distance) form sequences of weakly or even uncorrelated random variables, and are also mutually independent, at least approximately. Therefore, the aggregated process can be well approximated by a continuous time random walk, which in turn simplifies its analysis and thus the analysis of the whole process.

Despite these advantages, the suggested definition of MB has the obvious blemish that it depends on the realization of the considered process and may thus vary from simulation to simulation. To provide a mathematically stringent definition of a *path-independent aggregation* of the state space, which maintains the above properties and is based on the well-established notion of metastable states, is therefore our principal concern here with the main results being Theorem 3.13 and Theorem 3.17. In this endeavor, we will draw on some of the ideas developed by BOVIER in [4] and by SCOPPOLA in [21], most notably her definition of metastable states.

Metastability, a phenomenon of ongoing interest for complex physical systems described by finite Markov processes on very large state spaces, can be defined and dealt with in several ways. It has been derived from a renormalization procedure in [20], by a pathwise approach in [6], and via energy landscapes in [4], the latter being also our approach hereafter. To characterize a supercooled liquid, i.e. a glass forming system at low temperature, via its energy landscape was first done by GOLDSTEIN in 1969 [11] and has by now become a common method. The general task when studying metastability, as well originally raised in physics ([13], [19] or [18]), is to provide mathematical tools for an analysis of the property of thermodynamical systems to evolve in state space along a trajectory of unstable or metastable states with very long sojourn times.

Inspired by simulations of glass forming systems at very low temperatures with the Metropolis algorithm, we will study (as in [21]) finite Markov chains with exponentially small transition probabilities which are determined by an energy function and a parameter  $\beta > 0$ . This parameter can be understood as the inverse temperature and we are thus interested in the behavior of the process as  $\beta \rightarrow \infty$ . We envisage an energy function of highly complex order and without the hierarchical ordering that is typical in spin glass models. A good picture is provided by randomly chosen energies with correlations between neighbors or by an energy landscape that looks like a real mountain landscape. We will show that, towards an aggregation outlined above, the metastable states as defined in [21] are quite appropriate because they have an ordering from a kind of “weak” to a kind of “strong” metastability. Around those states we will define and then study connected valleys [Definition 1.5] characterized by minimal energy barriers. In the limit of low temperatures, any such barrier will determine the speed, respectively probability of a transition between the two valleys it separates. More precisely, in the limit  $\beta \rightarrow \infty$ , the process, when starting in a state  $x$ , will almost surely reach a state with lower barrier earlier than a state with higher barrier [Theorem 2.1]. In the limit of low temperatures, the bottom (minimum) of an entered valley will therefore almost surely be reached before that valley is left again [Proposition 3.4]. As a consequence, the probability for a transition from one valley to another is asymptotically independent of the state where the valley is entered. This is Property 1 above.

Furthermore, since valleys as well as metastable states have a hierarchical ordering, we can build valleys of higher order by a successive merger of valleys of lower order [Proposition 1.12]. Given an appropriate energy landscape, this procedure can annihilate (on the macroscopic scale) the accumulation of reciprocating jumps by merging valleys exhibiting such jumps into a single valley [Subsection 3.2]. Hence, valleys of sufficiently high order will have Property 2.

Beside the macroscopic process [Section 3], which describes the transitions between valleys, one can also analyze the microscopic process [Section 2], that is, the system behavior when moving within a fixed valley. Here we will give a formula for the exit time and connect it with its parameters [Theorem 2.8]. This will confirm Property 3.

Having thus established Properties 1–4 [Theorem 3.13], we will finally proceed to a comparison of our path-independent definition of MB with the path-dependent one given above. It will be shown [Theorem 3.17] that both coincide with high probability under some reasonable conditions on the connectivity of valleys which, in essence, ensure the existence of reasonable path-dependent MB. We will also briefly touch on the phenomenon of quasi-stationarity [Proposition 2.6] which is a large area [17] but to our best knowledge less studied in connection with the aggregation of states of large physical systems driven by energy landscapes.

Let us mention two further publications which, despite having a different thrust, provide definitions of valleys, called basins of attraction or metastates there, to deal with related questions. OLIVIERI & SCOPPOLA [16] fully describe the tube of exit from a domain in terms of which basins of attraction of increasing order are visited during a stay in that domain and for how long these basins are visited. In a very recent publication, BELTRÁN & LANDIM [2], by working with transition rates instead of energies, aim at finding a universal depth (and time scale) for all metastates. However, we rather aim at the *finest aggregation such that transitions back to an already visited metastate are very unlikely within a time frame used in simulations*. This finest aggregation will lead to valleys of very variable depth just as simulations do not exhibit a universal depth or timescale.

## 1. VALLEYS

Let  $X$  be a Markov chain on a finite set  $\mathcal{S}$  with transition matrix  $\mathbf{P} = (p(r, s))_{r, s \in \mathcal{S}}$  and stationary distribution  $\pi$ , and let  $E : \mathcal{S} \rightarrow \mathbb{R}$  be an energy function such that the following conditions hold:

**Irreducibility:**  $\mathbf{P}$  is irreducible with  $p(s, s) > 0$  and  $p(r, s) > 0$  iff  $p(s, r) > 0$  for all  $r, s \in \mathcal{S}$ .

**Transition Probabilities:** There exist parameters  $\beta > 0$  and  $\gamma_\beta > 0$  with  $\gamma_\beta \rightarrow 0, \beta\gamma_\beta \rightarrow \infty$  as  $\beta \rightarrow \infty$  such that

$$e^{-\beta((E(s)-E(r))^+ + \gamma_\beta)} \leq p(r, s) \leq e^{-\beta((E(s)-E(r))^+ - \gamma_\beta)}$$

for all distinct  $r, s \in \mathcal{S}$  with  $p(r, s) > 0$ . Furthermore,

$$p^*(r, s) := \lim_{\beta \rightarrow \infty} p(r, s)$$

exists for all  $r, s \in \mathcal{S}$ , is positive if  $E(r) \geq E(s)$  and = 0 otherwise.

**Reversibility:** The pair  $(\pi, \mathbf{P})$  satisfies the detailed balance equations, i.e.

$$\pi(r)p(r, s) = \pi(s)p(s, r)$$

for all  $r, s \in \mathcal{S}$ .

**Non-Degeneracy:**  $E(r) \neq E(s)$  for all  $r, s \in \mathcal{S}, r \neq s$ .

We are thus dealing with a reversible Markov chain with exponentially small transition probabilities driven by an energy landscape. As an example, which is also the main motivation behind this work, one can think of a Metropolis chain with transition probabilities of the form

$$p(r, s) = \frac{1}{C(r)} e^{-\beta(E(s)-E(r))^+}.$$

Here  $\beta$  is the inverse temperature and  $C(r), r \in \mathcal{S}$ , is a parameter, independent from  $\beta$ , giving the number of neighbors of  $r$ . For  $\gamma_\beta := \max_{r \in \mathcal{S}} \ln(C(r))(\beta + 1)^{-1/2}$  the above conditions are fulfilled. Let us start with the following basic result for the ratios of the stationary distribution.

**Lemma 1.1.** *For any two states  $r, s \in \mathcal{S}$  with  $E(r) > E(s)$ , we have*

$$e^{-\beta(E(r)-E(s)+2\gamma_\beta)} \leq \frac{\pi(r)}{\pi(s)} \leq e^{-\beta(E(r)-E(s)-2\gamma_\beta)}.$$

*Proof.* Reversibility and the assumptions on the transition probabilities imply

$$\begin{aligned} \max\left(\pi(r)e^{-\gamma_\beta\beta}, \pi(s)e^{-\beta(E(r)-E(s))-\gamma_\beta\beta}\right) &\leq \pi(r)p(r, s) \\ &= \pi(s)p(s, r) \\ &\leq \min\left(\pi(r)e^{\gamma_\beta\beta}, \pi(s)e^{-\beta(E(r)-E(s))+\gamma_\beta\beta}\right). \end{aligned}$$

In particular,

$$\begin{aligned} \pi(r)e^{-\gamma_\beta\beta} &\leq \pi(s)e^{-\beta(E(r)-E(s))+\gamma_\beta\beta}, \\ \pi(s)e^{-\beta(E(r)-E(s))-\gamma_\beta\beta} &\leq \pi(r)e^{\gamma_\beta\beta} \end{aligned}$$

and therefore

$$e^{-\beta(E(r)-E(s))-2\gamma_\beta\beta} \leq \frac{\pi(r)}{\pi(s)} \leq e^{-\beta(E(r)-E(s))+2\gamma_\beta\beta}$$

as claimed.  $\square$

Under the stated assumptions, SCOPPOLA [21] has shown the existence of a successive filtration (aggregation)  $\mathcal{S} = M^{(0)} \supset M^{(1)} \supset \dots \supset M^{(n)} = \{s_0\}$ ,  $n \in \mathbb{N}$ , of the state space such that the elements of each set  $M^{(i)}, 1 \leq i \leq n$ , can be called *metastable* in the following sense:

- They arise from the local minima of the energy function or certain modifications of it.
- There is a lower bound on the expected time needed for a transition from  $m_1$  to  $m_2$  for any  $m_1, m_2 \in M^{(i)}$  which increases very fast with  $i$ .
- There exists a constant  $C$  such that

$$\mathbb{P}_m(X_t \notin M^{(i+1)}) \leq e^{-C\beta}$$

for all  $m \in M^{(i)}, 0 \leq i \leq n-1$ , and sufficiently large  $t$ .

This filtration starts with

$$M^{(1)} := \{s \in \mathcal{S} | E(s) < E(r) \text{ for all } r \sim s\},$$

and deletes one local minimum at each step. In fact, the local minimum with minimal activation energy for a transition to another minimum is deleted, see [21] and [20] for further details.

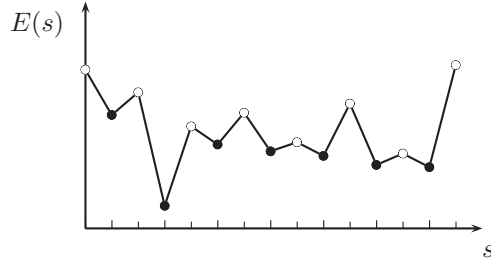


FIGURE 1. Example of an energy landscape with minima shown as black dots (●)

**Example 1.2.** For the simple energy function depicted in Figure 1, a successive application of the algorithm from [21] as illustrated in Figure 2 leads to the following decomposition into subsets of metastable states:

$$\begin{aligned} M^{(1)} &= \{2, 4, 6, 8, 10, 12, 14\}, & M^{(2)} &= \{2, 4, 6, 10, 12, 14\}, \\ M^{(3)} &= \{2, 4, 6, 10, 14\}, & M^{(4)} &= \{2, 4, 10, 14\}, \\ M^{(5)} &= \{4, 10, 14\}, & M^{(6)} &= \{4, 14\}, \\ M^{(7)} &= \{4\}. \end{aligned}$$

Based on the filtration of  $\mathcal{S}$  just described, we now proceed to a definition of a sequence of metastable sets associated with the metastable states which will induce the MB. In order to do so, we must study first minimal paths between two states and maximal energies along such paths.

**Definition 1.3.** (a) For any two distinct states  $r, s \in \mathcal{S}$ , let

$$\begin{aligned} \Gamma(r, s) := \{ & (x_0, \dots, x_k) \mid k \in \mathbb{N}, x_0 = r, x_k = s, \\ & x_i \neq x_j, p(x_i, x_{i+1}) > 0 \text{ for } 0 \leq i \leq k-1, i \neq j \} \end{aligned}$$

be the set of all finite self-avoiding paths from  $r$  to  $s$  having positive probability. For any such path  $\gamma = (\gamma_0, \dots, \gamma_k) \in \Gamma(r, s)$ , let  $|\gamma| := k$  be its length. We further write  $t \in \gamma$  if  $t \in \{\gamma_1, \dots, \gamma_k\}$ .

- (b) A self-avoiding path  $\gamma = (\gamma_1, \dots, \gamma_k)$  from  $r$  to  $s$  is called *minimal* if its maximal energy  $\max_{1 \leq i \leq k} E(\gamma_i)$  is minimal among all  $\gamma' \in \Gamma(r, s)$ . The set of these paths is denoted  $\Gamma^*(r, s)$ .
- (c) The *essential saddle*  $z^*(r, s)$  between  $r$  and  $s$  is then defined as

$$z^*(r, s) := \operatorname{argmax}_{t \in \gamma} E(t) \in \mathcal{S}$$

for any  $\gamma \in \Gamma^*(r, s)$ .

As for (c), it is to be noted that, due to the assumed non-degeneracy of the energy function, the essential saddle is unique, which means that it does not depend on (as it must) which minimal path we choose in the definition of  $z^*(r, s)$ . There may indeed be several minimal paths, every single one thus crossing the saddle at some time. With the help of these notions the valleys can now be defined in a quite concrete way. Let us label the local minima as  $m^{(1)}, \dots, m^{(n)}$ , so that  $M^{(i)} = \{m^{(i)}, \dots, m^{(n)}\}$  for each  $i = 1, \dots, n$ .

**Definition 1.4.** For each  $m \in M^{(i)}$ ,  $1 \leq i \leq n$ , let

$$V_{<}^{(i)}(m) := \left\{ s \in \mathcal{S} \mid E(z^*(s, m)) < E(z^*(s, m')) \text{ for all } m' \in M^{(i)} \setminus \{m\} \right\}.$$

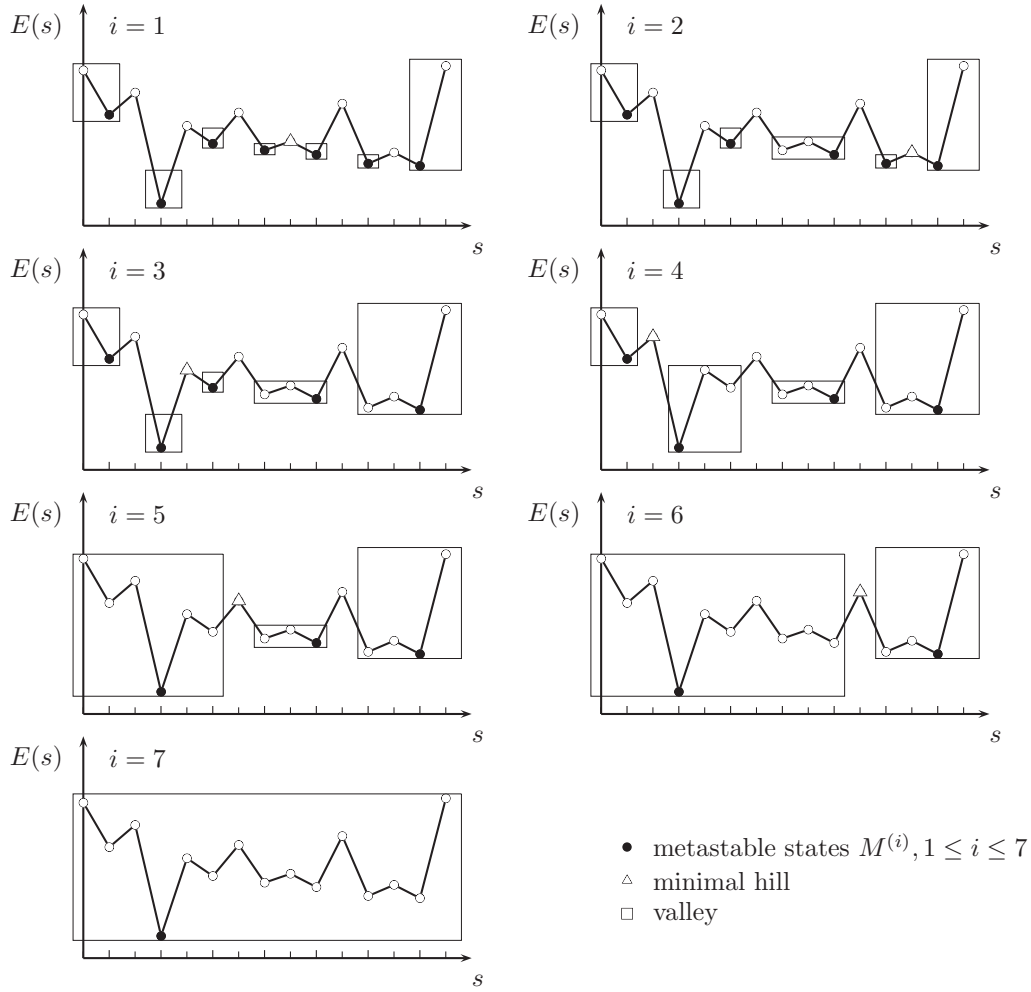


FIGURE 2. Successive application of the algorithm in [21] to the energy landscape in Figure 1. For each step  $i$ , the metastable states as well as the corresponding valleys are shown.

We say that *state  $s$  is attracted by  $m$  at level  $i$* , expressed as  $s \rightsquigarrow m$  at level  $i$ , if

$$E(z^*(s, m)) = \min_{n \in M^{(i)}} E(z^*(s, n))$$

and every minimal path from  $s$  to a state  $m' \in M^{(i)} \setminus \{m\}$  with  $E(z^*(s, m')) = E(z^*(s, m))$  hits  $V_{<}^{(i)}(m)$  at some time. Finally, let

$$l(i) := \inf \{i < j \leq n \mid m^{(i)} \rightsquigarrow m \text{ at level } j \text{ for some } m \in M^{(j)}\}$$

denote the minimal level at which the minimal state  $m^{(i)}$  becomes attracted by a minimal state of superior level.

**Definition 1.5.** (a) *Initialization:* For each  $m \in M^{(1)}$ , define

$$V^{(1)}(m) := \left\{ s \in \mathcal{S} \mid s \rightsquigarrow m \text{ at level } 1 \right\}.$$

as the *valley of order 1 containing  $m$*  and let

$$N^{(1)} := \left( \bigcup_{j=1}^n V^{(1)}(m^{(j)}) \right)^c$$

be the set of non-assigned states at level 1.

(b) *Recursion:* For  $2 \leq i \leq n$  and  $m \in M^{(i)}$ , define

$$V^{(i)}(m) := V^{(i-1)}(m) \cup \left\{ s \in N^{(i-1)} \mid s \rightsquigarrow m \text{ at level } i \right\} \cup \bigcup_{\substack{j:l(j)=i, \\ m^{(j)} \rightsquigarrow m \text{ at level } i}} V^{(j)}(m^{(j)})$$

as the *valley of order  $i$  containing  $m$*  and let

$$N^{(i)} := \left( \bigcup_{j=1}^n V^{(i \wedge j)}(m^{(j)}) \right)^c$$

be the set of non-assigned states at level  $i$ .

Here is a more intuitive description of what the previous two definitions render in a formal way: First, we define, for each level  $i$  and  $m \in M^{(i)}$ , the set  $V_{<}^{(i)}(m)$  of those states  $s$  that are strongly attracted by  $m$  in the sense that  $E(z^*(s, m))$  is strictly smaller than  $E(z^*(s, m'))$  for any other  $m' \in M^{(i)}$ . Then, starting at level one, each valley  $V^{(1)}(m)$ ,  $m \in M^{(1)}$ , is formed from  $V_{<}^{(1)}(m)$  by adjoining all further states  $s$  attracted by  $m$  at this level. This leaves us with a set of non-assigned states, denoted  $N^{(1)}$ . In the next step (level 2), any  $V^{(2)}(m)$  for  $m \in M^{(2)}$  is obtained by adjoining to  $V^{(1)}(m)$  all those  $s \in N^{(1)}$  which are attracted by  $m$  at level 2. Observe that this ensures  $V_{<}^{(2)}(m) \subset V^{(2)}(m)$ . Moreover, if  $m^{(1)}$  is attracted by  $m$  at level 2, then  $V^{(1)}(m^{(1)})$  is merged into  $V^{(2)}(m)$  as well. If no such  $m$  exists (thus  $l(1) > 2$ ), it remains untouched until reaching level  $l(1)$  where its bottom state  $m^{(1)}$  becomes attracted by some  $m' \in M^{(l(1))}$  causing its valley to be merged into  $V^{(l(1))}(m')$ . This procedure continues in the now obvious recursive manner until at level  $n$  all states have been merged into one valley. Obviously, valleys of the same order are pairwise disjoint. Also, valleys once formed at some level can only be merged as a whole and will thus never be ripped apart during the recursive construction. For the energy function depicted in Figure 1, the successively derived valleys of order  $i = 1, \dots, 7$  are shown in Figure 2.

Before proceeding to results on the general shape of valleys, we collect some basic, mostly technical properties of essential saddles which will be useful thereafter.

**Proposition 1.6.** *For any  $r, s, u \in \mathcal{S}$ ,  $0 \leq i \leq n$ ,  $m_1, m_2 \in M^{(i)}$ ,  $m_1 \neq m_2$ , and  $x_1, x_2 \in \mathcal{S}$  with  $x_1 \in V_{<}^{(i)}(m_1)$  and  $x_2 \in V^{(i)}(m_2)$ , we have*

- (a)  $z^*(r, s) = z^*(s, r)$ .
- (b)  $E(z^*(r, s)) \leq E(z^*(r, u)) \vee E(z^*(u, s))$ .
- (c)  $E(z^*(x_2, m_2)) \leq E(z^*(x_2, m'))$  for all  $m' \in M^{(i)}$ .
- (d)  $E(z^*(x_1, m_2)) = E(z^*(m_1, m_2))$ .
- (e)  $E(z^*(x_1, x_2)) \geq E(z^*(m_1, m_2))$ .
- (f)  $z^*(x_1, x_2) \neq x_1$ .

*Proof.* Parts (a) and (b) are obvious.

For (c) we use an induction over  $i$  and note that there is nothing to show when  $i = 1$ . For general  $i$ , we must only verify that  $E(z^*(x_2, m_2)) \leq E(z^*(x_2, m'))$  for all  $m' \in M^{(i)}$  if  $x_2 \in V^{(j)}(m^{(j)})$  for some  $j < i$  such that  $l(j) = i$  and  $m^{(j)} \rightsquigarrow m_2$  at level  $i$  (due to the recursive definition of  $V^{(i)}(m)$ ). But the latter ensures that  $E(z^*(x_2, m^{(j)})) \leq$

$E(z^*(x_2, n))$  for all  $n \in M^{(j)} \supset M^{(i)}$  (inductive hypothesis) as well as  $E(z^*(m^{(j)}, m_2)) \leq E(z^*(m^{(j)}, m'))$  for all  $m' \in M^{(i)}$ . Consequently, for any such  $m'$ ,

$$\begin{aligned} E(z^*(x_2, m_2)) &\leq E(z^*(x_2, m^{(j)})) \vee E(z^*(m^{(j)}, m_2)) \\ &\leq E(z^*(x_2, m^{(j)})) \vee E(z^*(m^{(j)}, m')) \\ &\leq E(z^*(x_2, m^{(j)})) \vee E(z^*(x_2, m')) \\ &= E(z^*(x_2, m')) \end{aligned}$$

as asserted.

For assertion (d), note that  $E(z^*(x_1, m_2)) > E(z^*(x_1, m_1))$ , which in combination with (a) and (b) implies

$$E(z^*(m_1, m_2)) \leq E(z^*(m_1, x_1)) \vee E(z^*(x_1, m_2)) = E(z^*(x_1, m_2))$$

and then further

$$E(z^*(x_1, m_2)) \leq \underbrace{E(z^*(x_1, m_1))}_{< E(z^*(x_1, m_2))} \vee \underbrace{E(z^*(m_1, m_2))}_{\leq E(z^*(x_1, m_2))} \leq E(z^*(x_1, m_2)).$$

So the above must be an identity, i.e.  $E(z^*(x_1, m_2)) = E(z^*(m_1, m_2))$ .

Turning to part (e), we first infer with the help of (c) and (d) that

$$\begin{aligned} (1) \quad E(z^*(x_1, m_1)) &< E(z^*(x_1, m_2)) \\ &= E(z^*(m_1, m_2)) \\ &\leq E(z^*(m_1, x_2)) \vee E(z^*(x_2, m_2)) \\ &= E(z^*(x_2, m_1)) \\ &\leq E(z^*(x_2, x_1)) \vee E(z^*(x_1, m_1)), \end{aligned}$$

thus

$$(2) \quad E(z^*(x_1, m_1)) < E(z^*(x_1, x_2)).$$

Together with the just shown inequality  $E(z^*(m_1, m_2)) \leq E(z^*(x_2, m_1))$  (see (1)) and another use of (c), this yields

$$E(z^*(x_1, x_2)) = E(z^*(x_2, x_1)) \vee E(z^*(x_1, m_1)) \geq E(z^*(x_2, m_1)) \geq E(z^*(m_1, m_2)).$$

Finally, we infer with the help of (2) that

$$E(z^*(x_1, x_2)) > E(z^*(x_1, m_1)) \geq E(x_1)$$

and thus  $z^*(x_1, x_2) \neq x_1$  as claimed in (f).  $\square$

**Remark 1.7.** It is useful to point out the following consequence of the previous proposition. If, for an arbitrary state  $s$  and any two distinct metastable states  $m, n \in M^{(i)}$ , there exists a minimal path  $\gamma$  from  $s$  to  $n$  that hits a state  $r$  with  $E(z^*(r, m)) < E(z^*(r, n))$ , then there is also a minimal path from  $s$  to  $n$  that passes through  $m$ . Namely, if we replace the segment from  $r$  to  $n$  of the former path by the concatenation of two minimal paths from  $r$  to  $m$  and from  $m$  to  $n$ , then the maximal energy of this new path is

$$\begin{aligned} E(z^*(s, n)) \vee E(z^*(r, m)) \vee E(z^*(m, n)) &\leq E(z^*(s, n)) \vee E(z^*(r, m)) \vee E(z^*(r, n)) \\ &= E(z^*(s, n)) \vee E(z^*(r, n)) \\ &= E(z^*(s, n)), \end{aligned}$$

by Proposition 1.6(b), whence the new path has to be minimal from  $s$  to  $n$  as well. This yields two facts:



- (a) A minimal path from  $s$  to  $n$ , where  $s \rightsquigarrow n$  at level  $i$ , hits  $V_{<}^{(i)}(n)$  before it hits any  $r$  with  $E(z^*(r, m)) < E(z^*(r, n))$  for some  $m \in M^{(i)}$ . Otherwise, since the subpath from  $r$  to  $m$  can be chosen to stay in  $\{t \mid E(z^*(t, m)) < E(z^*(t, n))\}$  and thus  $E(z^*(s, m)) = E(z^*(s, n))$ , there would be a path from  $s$  to  $m$  not hitting  $V_{<}^{(i)}(n)$ .
- (b) If  $s \rightsquigarrow n$  at level  $i$  and  $m \in M^{(i)} \setminus \{n\}$  with  $E(z^*(s, n)) = E(z^*(s, m))$ , then a minimal path from  $s$  to  $m$  does not only hit  $V_{<}^{(i)}(n)$  at some time, but in fact earlier than any other valley  $V_{<}^{(i)}(m'), m' \in M^{(i)} \setminus \{n\}$ .

**Lemma 1.8.** *Let  $1 \leq i < j \leq n$ ,  $m = m^{(i)}$  and  $s \in V^{(i)}(m)$ . Then  $s \in V_{<}^{(j)}(m')$  for some  $m' \in M^{(j)}$  implies  $l(i) \leq j$ ,  $m \in V_{<}^{(j)}(m')$  and thus  $V^{(i)}(m) \subset V^{(j)}(m')$ .*

In other words, whenever  $V^{(i)}(m^{(i)})$  contains an element  $s$  which at some higher level  $j$  belongs to some  $V_{<}^{(j)}(m')$ ,  $m' \in M^{(j)}$ , the same must hold true for  $m^{(i)}$  itself implying  $V^{(i)}(m^{(i)}) \subset V^{(j)}(m')$ . Conversely, this guarantees that  $V^{(i)}(m^{(i)})$  will have no common elements with any  $V_{<}^{(j)}(m')$  at levels  $j < l(i)$  where it has not yet been merged into a valley of higher order.

*Proof.* Let us first note that, under the given assumptions,

$$E(z^*(s, m)) \leq E(z^*(s, m')) < E(z^*(s, n))$$

for all  $n \in M^{(j)} \setminus \{m'\}$ , whence

$$E(z^*(m, n)) \leq E(z^*(s, m)) \vee E(z^*(s, n)) = E(z^*(s, n)) \leq E(z^*(s, m)) \vee E(z^*(m, n))$$

entails  $E(z^*(m, n)) = E(z^*(s, n))$  for all such  $n$ . Using this fact, we find that

$$E(z^*(m, m')) \leq E(z^*(s, m)) \vee E(z^*(s, m')) < E(z^*(s, n)) = E(z^*(m, n))$$

for all  $n \in M^{(j)} \setminus \{m'\}$ , which implies  $m \rightsquigarrow m'$  at level  $j$  and thus  $l(i) \leq j$  as well as the other assertions.  $\square$

**Proposition 1.9.** *For every  $m \in M^{(i)}$  and  $1 \leq i \leq n$ ,  $V_{<}^{(i)}(m)$  is connected.*

*Proof.* Pick any  $s \in V_{<}^{(i)}(m)$ , any minimal path from  $s$  to  $m$  and finally any intermediate state  $r$  along this path for which  $r \in V_{<}^{(i)}(m)$  must be verified. For every  $m' \in M^{(i)} \setminus \{m\}$ , we find

$$\begin{aligned} E(z^*(r, m)) &\leq E(z^*(r, s)) \vee E(z^*(s, m)) \\ &= E(z^*(s, m)) \\ &< E(z^*(s, m')) \\ &\leq \underbrace{E(z^*(s, r))}_{< E(z^*(s, m'))} \vee E(z^*(r, m')) \\ &= E(z^*(r, m')), \end{aligned}$$

which shows  $r \in V_{<}^{(i)}(m)$  as required.  $\square$

Note that we have even shown that a minimal path from a state in  $V_{<}^{(i)}(m)$  to  $m$  will never leave this set. We may expect and will indeed show as Proposition 1.11 below that  $V^{(i)}(m)$  is connected as well. The following lemma is needed for its proof.

**Lemma 1.10.** *Given  $1 \leq i \leq n$ ,  $m \in M^{(i)}$  and  $s \rightsquigarrow m$  at level  $i$ , let  $\gamma = (\gamma_1, \dots, \gamma_k) \in \Gamma^*(s, m)$  be a path such that  $E(z^*(\gamma_i, m)) \leq E(z^*(\gamma_i, n))$  for all  $n \in M^{(i)} \setminus \{m\}$ , and which stays in  $V_{<}^{(i)}(m)$  once hitting this set (such a  $\gamma$  exists by Remark 1.7 (a)). Then  $\gamma_j \rightsquigarrow m$  at level  $i$  for each  $j = 1, \dots, k$ .*

*Proof.* There is nothing to prove for  $\gamma_1 = s$  and any  $\gamma_j \in V_{<}^{(i)}(m)$ . So let  $r$  be any other state visited by  $\gamma$ , pick an arbitrary  $n \in M^{(i)} \setminus \{m\}$  with  $E(z^*(r, n)) = E(z^*(r, m))$  and then any minimal path  $\tau$  from  $r$  to  $n$ . Let  $\sigma$  be the subpath of  $\gamma$  from  $s$  to  $r$ . We must show that  $\tau$  hits  $V_{<}^{(i)}(m)$ . First, we point out that the maximal energy  $E(z^*(s, r)) \vee E(z^*(r, n))$  of  $\sigma\tau$ , the concatenation of  $\sigma$  and  $\tau$ , satisfies

$$\begin{aligned} E(z^*(s, n)) &\leq E(z^*(s, r)) \vee E(z^*(r, n)) \leq E(z^*(s, m)) \vee E(z^*(r, m)) \\ &= E(z^*(s, m)) \leq E(z^*(s, n)), \end{aligned}$$

implying  $\sigma\tau \in \Gamma^*(s, n)$  and, furthermore,

$$\begin{aligned} E(z^*(s, r)) \vee E(z^*(r, n)) &= E(z^*(s, m)) \vee E(z^*(r, n)) \\ &= E(z^*(s, m)) \vee E(z^*(r, m)) = E(z^*(s, m)). \end{aligned}$$

Thus  $\sigma\tau$  must hit  $V_{<}^{(i)}(m)$ . But since  $\sigma$  does not hit  $V_{<}^{(i)}(m)$  by assumption, we conclude that  $\tau$  must hit  $V_{<}^{(i)}(m)$ . Since  $\tau \in \Gamma^*(r, n)$  was arbitrary, we infer  $r \rightsquigarrow m$  at level  $i$ .  $\square$

The next two propositions provide information on the shape of the valleys and their nested structure.

**Proposition 1.11.** *For every  $m \in M^{(i)}$  and  $1 \leq i \leq \mathbf{n}$ ,  $V^{(i)}(m)$  is connected.*

*Proof.* We use an inductive argument. If  $i = 1$ , the assertion follows directly from the definition of the level-one valleys because any  $s \in V^{(1)}(m)$ ,  $m \in M^{(1)}$ , may be connected to  $m$  by a minimal path that eventually enters  $V_{<}^{(1)}(m)$  without hitting any other  $V_{<}^{(1)}(n)$  and is therefore completely contained in  $V^{(1)}(m)$  by the previous lemma.

Turning to the inductive step, suppose the assertion holds true up to level  $i - 1$ . Fix any  $m \in M^{(i)}$  and notice that, by the inductive hypothesis,  $V^{(i-1)}(m)$  as well as all  $V^{(j)}(m^{(j)})$  with  $l(j) = i$  and  $m^{(j)} \rightsquigarrow m$  at level  $i$  are connected. Now, since these  $m^{(j)}$  as well as all  $s \in N^{(i-1)}$  attracted by  $m$  at level  $i$  may be connected to  $m$  by minimal paths as assumed in Lemma 1.10, we conclude that  $V^{(i)}(m)$  is also connected.  $\square$

The second proposition shows the nested structure of our construction of valleys.

**Proposition 1.12.** *The following inclusions hold true:*

- (a)  $V^{(1)}(m) \subseteq \dots \subseteq V^{(i)}(m)$  for each  $m \in M^{(i)}$ ,  $1 \leq i \leq \mathbf{n}$ .
- (b)  $V^{(i)}(m) \subseteq V^{(j)}(n)$  for each  $1 \leq i < j \leq \mathbf{n}$ ,  $n \in M^{(j)}$  and  $m \in M^{(i)} \cap V^{(j)}(n)$ .

*Proof.* Since there is nothing to show for (a) we move directly to (b). But if  $m \in M^{(i)} \cap V^{(j)}(n)$ , then the definition of valleys ensures the existence of  $1 \leq k \leq j - i$  and of  $n_1, \dots, n_{k-1} \in M^{(j)} \setminus M^{(i)}$  such that  $n_{p-1} \rightsquigarrow n_p$  at level  $l_p$  for each  $p = 1, \dots, k$  and levels  $i < l_1 < \dots < l_k = j$ , where  $n_0 := m$  and  $n_k := n$ . As a consequence,

$$V^{(i)}(m) \subseteq V^{(l_1)}(n_1) \subseteq \dots \subseteq V^{(l_{k-1})}(n_{k-1}) \subseteq V^{(j)}(n)$$

which proves the asserted inclusion.  $\square$

To finish the analysis of the shape of the valleys we show that they have the following important property: a special class of minimal paths from the inside of any  $V^{(i)}(m)$  to the outside of it must hit its interior  $V_{<}^{(i)}(m)$ . But in order to show this we must first verify that all states attracted by  $m$  at level  $i$  belong to  $V^{(i)}(m)$ .

**Lemma 1.13.** *For each  $1 \leq i \leq \mathbf{n}$  and  $m \in M^{(i)}$ , we have that*

$$\begin{aligned} \left\{ s \in \mathcal{S} \mid s \rightsquigarrow m \text{ at level } i \right\} &\subset V^{(i)}(m) \\ &\subset \left\{ s \in \mathcal{S} \mid E(z^*(s, m)) \leq E(z^*(s, m')) \text{ for all } m' \in M^{(i)} \right\}. \end{aligned}$$

*Proof.* For the second inclusion it suffices to refer to Proposition 1.6(c). The first inclusion being obviously true for  $s \in N^{(i-1)}$ , we turn directly to the case when

$$s \rightsquigarrow n_1 \text{ at level } l_1, \quad n_1 \rightsquigarrow n_2 \text{ at level } l_2, \quad \dots \quad n_{k-1} \rightsquigarrow n_k \text{ at level } l_k$$

with  $k \geq 1$  and  $1 \leq l_1 \leq \dots \leq l_k \leq i-1$ . Here,  $n_1$  denotes the first minimum to which  $s$  is attracted (thus  $s \in V^{(l_1)}(n_1)$ ), while  $n_k$  is the last minimum of this kind in the sequence. We may assume without loss of generality that  $n_j \neq m$  for all  $j$ , for otherwise the assertion is clear.

We show now that  $n_1 \rightsquigarrow m$  at level  $i$  which in turn implies  $n_j \rightsquigarrow m$  at level  $i$  for all  $1 \leq j \leq k$ . As a consequence,  $n_k \in V^{(i)}(m)$  and thus  $s \in V^{(i)}(m)$ . If  $E(z^*(n_1, m)) < E(z^*(n_1, m'))$  for all  $m' \in M^{(i)} \setminus \{m\}$ , the assertion is proved. Hence suppose  $E(z^*(n_1, m)) \geq E(z^*(n_1, m'))$  for some  $m' \in M^{(i)} \setminus \{m\}$ . Then

$$\begin{aligned} E(z^*(s, m')) &\leq E(z^*(s, n_1)) \vee E(z^*(n_1, m')) \\ &\leq E(z^*(s, n_1)) \vee E(z^*(n_1, m)) \\ &\leq E(z^*(s, n_1)) \vee E(z^*(s, m)) \\ &= E(z^*(s, m)) \\ &\leq E(z^*(s, m')), \end{aligned}$$

implies  $E(z^*(s, m)) = E(z^*(s, m'))$  and also that the concatenation of any minimal path  $\gamma$  from  $s$  to  $n_1$  and any minimal path  $\tau$  from  $n_1$  to  $m'$  (with maximal energy  $E(z^*(s, n_1)) \vee E(z^*(n_1, m'))$ ) constitutes a minimal path from  $s$  to  $m'$  and must therefore hit  $V_{<}^{(i)}(m)$ . Note that we can choose  $\gamma$  to stay in  $V^{(l_1)}(n_1)$  since  $s \in V^{(l_1)}(n_1)$  and  $V^{(l_1)}(n_1)$  is connected. Now, if  $\tau$  hits  $V_{<}^{(i)}(m)$ , then  $E(z^*(n_1, m)) = E(z^*(n_1, m'))$  and we are done. Otherwise,  $\gamma$  hits  $V_{<}^{(i)}(m)$  implying  $V^{(l_1)}(n_1) \cap V_{<}^{(i)}(m) \neq \emptyset$ . Now use Lemma 1.8 to conclude  $n_1 \in V_{<}^{(i)}(m)$  and therefore  $n_1 \rightsquigarrow m$  at level  $i$ . This completes the argument for the first inclusion.  $\square$

We provide too further lemmata that will be needed later on.

**Lemma 1.14.** *Let  $m \in M^{(i)}$ ,  $x \rightsquigarrow m$  at level  $i$  and  $y \notin V^{(i)}(m)$ . Then either every minimal path from  $x$  to  $y$  hits the set  $V_{<}^{(i)}(m)$ , or  $E(z^*(x, y)) > E(z^*(x, m))$ .*

*Proof.* Suppose there is a minimal path  $\gamma$  from  $x$  to  $y$  avoiding  $V_{<}^{(i)}(m)$ . Since  $y \notin V^{(i)}(m)$ , it is not attracted by  $m$  at level  $i$  implying the existence of some  $m' \in M^{(i)}$  with  $E(z^*(y, m')) \leq E(z^*(y, m))$  and of some  $\tau \in \Gamma^*(y, m')$  avoiding  $V_{<}^{(i)}(m)$ . Hence, the concatenation  $\gamma\tau$  avoids  $V_{<}^{(i)}(m)$  and must therefore have maximal energy larger than  $E(z^*(x, m))$ . Consequently,

$$\begin{aligned} E(z^*(x, m)) &< E(z^*(x, y)) \vee E(z^*(y, m')) \\ &\leq E(z^*(x, y)) \vee E(z^*(y, m)) \\ &\leq E(z^*(x, y)) \vee E(z^*(x, m)), \end{aligned}$$

and thus  $E(z^*(x, y)) > E(z^*(x, m))$ .  $\square$

In order to state the second lemma, let us define the outer part  $\partial^+V$  of a valley  $V$  to be the set of those states outside  $V$  which are adjacent to a state in  $V$ . With the help of the previous result, we can easily show that  $\partial^+V$  contains only non-assigned states at any level where  $V$  has not yet been merged into a larger valley.

**Lemma 1.15.** *For any  $1 \leq i, j \leq n$  and  $m = m^{(i)}$  with  $l(i) > j$ , the outer part  $\partial^+V$  of the valley  $V := V^{(j \wedge i)}(m)$  is a subset of  $N^{(j)}$  and  $E(z^*(s, m)) = E(s)$  for every  $s \in \partial^+V$ .*

*Proof.* First, let  $s \in \partial^+V$  and suppose that  $s \notin N^{(j)}$ . Then  $s \rightsquigarrow m'$  at level  $k$ , in particular  $s \in V^{(k)}(m')$  for some  $m' \in M^{(k)}$  and  $k \leq j$ . Pick any  $r \in V$  with  $r \sim s$  and note that  $r \in \partial^+V^{(k)}(m')$ . W.l.o.g. we may assume that  $r \rightsquigarrow m$  at level  $j \wedge i$ . Then Lemma 1.14 (with  $x = r$  and  $y = s$ ) ensures that either  $E(z^*(r, s)) > E(z^*(r, m)) \geq E(r)$ , thus  $z^*(r, s) = s$  and  $E(r) < E(s)$ , or  $r \in V_{<}(m)$  and, for some  $n \in M^{(j \wedge i)}$ ,

$$\begin{aligned}
 (3) \quad E(z^*(r, m)) &< E(z^*(r, n)) \\
 &\leq E(z^*(r, s)) \vee E(z^*(s, n)) \\
 &\leq E(z^*(r, s)) \vee E(z^*(s, m)) \\
 &\leq E(z^*(r, s)) \vee E(z^*(r, m)) \\
 &= E(s) \vee E(z^*(r, m)),
 \end{aligned}$$

and thus again  $E(r) < E(s)$ . On the other hand, by the very the same lemma (now with  $x = s$  and  $y = r$ ), we infer  $E(r) > E(s)$  which is clearly impossible. Consequently,  $s$  must be non-assigned at level  $j$  as claimed.

For the second assertion take again  $s \in \partial^+V$  and a minimal path  $\gamma = (s, r, \dots, m)$  from  $s$  to  $m$  with  $r \in V$ . Again, by use of Lemma 1.14, we find either  $E(z^*(r, m)) < E(z^*(r, s)) = E(s)$  or  $r \in V_{<}(m)$ , leading analogously to equation (3) to  $E(z^*(r, m)) < E(s) \vee E(z^*(r, m))$  and thus  $E(z^*(r, m)) < E(s)$ . In conclusion, both cases result in

$$E(z^*(s, m)) = E(s) \vee E(z^*(r, m)) = E(s),$$

finishing the proof.  $\square$

The reader may wonder why valleys are defined here via essential saddles and not via the at first glance more natural overall energy barriers, viz.

$$(4) \quad I(s, m) := \inf_{\gamma \in \Gamma(s, m)} I(\gamma_1, \dots, \gamma_{|\gamma|})$$

with

$$I(s_1, \dots, s_n) := \sum_{i=1}^n (E(s_i) - E(s_{i-1}))^+$$

for a state  $s$  in a valley and the pertinent minimum  $m$ . This latter quantity, also called *activation energy*, is indeed an important parameter in [21]. The reason for our definition becomes clear when regarding the last proposition which shows the nested structure of valleys of increasing order and which may fail to hold when choosing an alternative definition based on the activation energy. Valleys that are formed in one step may then be ripped apart in the next one. This happens, for instance, if there is just one large saddle along the path to a metastable state and several small ones, lower than the essential saddle, along the paths to another minimum such that their total sum is larger than the big saddle. In further support of our approach, it will be seen later that the essential saddles are the critical parameters for the behavior of the aggregated chain (see Theorem 2.4).

In a nutshell, by going from  $(V^{(i)}(m))_{m \in M^{(i)}}$  to  $(V^{(i+1)}(m))_{m \in M^{(i+1)}}$ , some valleys are merged into one (with only the smaller minima retained as metastable states) and additionally those states from  $N^{(i)}$  are added which at level  $i$  were attracted by metastable states now all belonging to the same valley. This induces the following tree-structure on the state space:

- Fix  $\emptyset = s_0$ .
- The first generation of the tree consists of all  $m \in M^{(n-1)} \cup N^{(n-1)}$  and are thus connected to the root.

- The second generation of the tree consists of all  $m \in M^{(n-2)} \cup N^{(n-2)}$ , and  $m$  is connected to the node  $k$  of the first generation for which  $E(z^*(m, k))$  is minimal or to itself (in the obvious sense).
- This continues until in the  $n^{\text{th}}$  generation each state is listed and connected either with its unique point of attraction in the previous generation or with itself.

**Example 1.16.** For the energy function of Example 1.2 and depicted in Figure 1, the described tree is shown in Figure 3. The sets of non-assigned states at the different levels

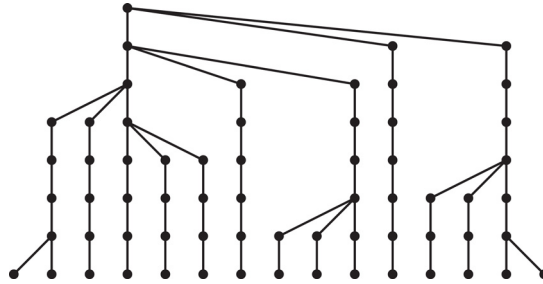


FIGURE 3. The tree belonging to Figure 1

are

$$\begin{aligned}
 N^{(1)} &= \{3, 5, 7, 9, 11, 13\}, & N^{(2)} &= \{3, 5, 7, 11, 13\}, \\
 N^{(3)} &= \{3, 5, 7, 11\}, & N^{(4)} &= \{3, 7, 11\}, \\
 N^{(5)} &= \{7, 11\}, & N^{(6)} &= \{11\}, \\
 N^{(7)} &= \emptyset.
 \end{aligned}$$

At each level  $i$  of such a tree the subtree rooted at a node  $m \in M^{(i)}$  consists of the states in the valley  $V^{(i)}(m)$ . A similar graph-theoretical modeling in order to visualize high dimensional energy landscapes has been used, for example, by OKUSHIMA et al. in [15]. These authors work with saddles of paths as well. In contrast to our approach, every possible path, that is every possible saddle is represented as a node in the tree. But as we will see, in the limit of low temperatures ( $\beta \rightarrow \infty$ ) the essential saddle is all we need.

Now there are two fundamental directions for further investigations:

- (1) **MICROSCOPIC VIEW:** What happens while the process visits a fixed valley  $V$  (of arbitrary level)?

In Section 2, we will show that during each visit of a valley  $V$  its minimum will be reached with probability tending to 1 as  $\beta \rightarrow \infty$ , and we also calculate the expected residence time in  $V$ , establish Property 3 stated in the Introduction and comment briefly on quasi-stationarity.

- (2) **MACROSCOPIC VIEW:** How does the process jump between the valleys?

In Section 3, by drawing on the results of Section 2, we will show that an appropriate aggregated chain is Markovian in the limit as  $\beta \rightarrow \infty$  and calculate its transition probabilities. With this we will finally be able to provide the definition of MB and establish Properties 1, 2 and 4 listed in the Introduction.

## 2. MICROSCOPIC VIEW: FIXING A VALLEY

Based on the provided definition of valleys of different orders, we are now going to study the process when moving in a fixed valley.

**2.1. Trajectories for  $\beta \rightarrow \infty$ .** The first goal in our study of the microscopic process and also the basic result for the subsequent analysis of the macroscopic process deals with the probabilities of reaching certain states earlier than others. From this we will conclude that in the limit  $\beta \rightarrow \infty$  (which is the low temperature limit in the Metropolis Algorithm) the process, when starting somewhere in a valley, will visit its minimum before leaving it.

For  $A \subset \mathcal{S}$  and  $x \in \mathcal{S}$ , we define

$$\tau_A := \inf\{n \geq 1 | X_n \in A\}, \quad \tau_x := \tau_{\{x\}} \quad \text{and} \quad \mathcal{N}(x) := \{y \in \mathcal{S} | p(x, y) > 0\}.$$

Two states  $x, y$  with  $p(x, y) > 0$  are called neighbors ( $x \sim y$ ) and  $\mathcal{N}(x)$  the neighborhood of  $x$ . Hence,  $\mathcal{S}$  may (and will) be viewed as a graph hereafter with edge set  $\{(x, y) | x \sim y\}$ .

**Theorem 2.1.** *Let  $x, y, z \in \mathcal{S}$  be any pairwise distinct states satisfying  $E(z^*(x, z)) > E(z^*(x, y))$  and  $z^*(x, z) \neq x$ . Then there exist nonnegative constants  $K(\beta)$  satisfying  $\sup_{\beta > 0} K(\beta) < \infty$  (and given more explicitly in Proposition 2.4 below) such that*

$$\mathbb{P}_x(\tau_z < \tau_y) \leq K(\beta) e^{-\beta(E(z^*(x, z)) - E(z^*(x, y)) - 7\gamma_\beta)} =: \varepsilon(x, y, z, \beta) \xrightarrow{\beta \rightarrow \infty} 0.$$

Thus in the limit of low temperatures ( $\beta \rightarrow \infty$ ), only the smallest of all possible energy barriers affects the speed of a transition. In particular, we have the following result which is preliminary to the subsequent one.

**Theorem 2.2.** *Given distinct  $x, y \in \mathcal{S}$  and  $m \in M^{(i)}$  such that  $x \rightsquigarrow m$  at level  $i$  and  $y \notin V^{(i)}(m)$ , let  $B := \{z \in \mathcal{S} | E(z^*(x, z)) > E(z^*(x, m))\}$ . Then it holds true that*

$$\begin{aligned} \mathbb{P}_x(\tau_y < \tau_m) &\leq \varepsilon(x, m, y, \beta) \mathbf{1}_B(y) \\ &\quad + \left( \sum_{z: E(z) > E(z^*(x, y))} \varepsilon(x, y, z, \beta) + \sum_{z \in V_{<}^{(i)}(m)} \varepsilon(z, m, y, \beta) \right) \mathbf{1}_{B^c}(y) \\ &=: \tilde{\varepsilon}(x, m, y, \beta) \xrightarrow{\beta \rightarrow \infty} 0. \end{aligned}$$

**Theorem 2.3.** *Given  $m \in M^{(i)}$ ,  $x \in V^{(i)}(m)$  and  $y \notin V^{(i)}(m)$ , let  $k \leq i$  be such that*

$$m_0 := x \rightsquigarrow m_1 \text{ at level } l_1, \quad m_1 \rightsquigarrow m_2 \text{ at level } l_2, \quad \dots \quad m_{k-1} \rightsquigarrow m_k = m \text{ at level } l_k$$

*for suitable  $1 \leq l_1 < \dots < l_k \leq i$ ,  $m_j \in M^{(l_j)}$  for  $j = 1, \dots, k$  determined by the construction in Definition 1.5. Then*

$$\mathbb{P}_x(\tau_y < \tau_m) \leq \sum_{j=1}^k \mathbb{P}_{m_{j-1}}(\tau_y < \tau_{m_j}) \leq \sum_{j=1}^k \tilde{\varepsilon}(m_{j-1}, m_j, y, \beta) \xrightarrow{\beta \rightarrow \infty} 0.$$

For the valleys as defined here this confirms Property 1 stated in the Introduction: If  $\beta$  is sufficiently large, then with high probability the minimum of a valley is visited before this valley is left.

The proof of Theorem 2.1 will be accomplished by a combination of two propositions due to BOVIER et al. [5] for a more special situation. We proceed with a reformulation of the first one in a weaker form and under weaker assumptions. The proof is the same as in the original version that is why we skip it here.

**Proposition 2.4** (compare Theorem 1.8 in [5]). *Let  $x, y, z \in \mathcal{S}$  be pairwise distinct such that  $z^*(x, z) \neq x$ . Then*

$$\begin{aligned} \mathbb{P}_x(\tau_z < \tau_x) &\leq K(\beta) |\mathcal{S}|^{-1} e^{-\beta(E(z^*(x, z)) - E(x) - 2\gamma_\beta)}, \\ \mathbb{P}_x(\tau_y < \tau_x) &\geq |\mathcal{S}|^{-1} e^{-\beta(E(z^*(x, y)) - E(x) + 5\gamma_\beta)}, \end{aligned}$$

*where  $K(\beta) := |\mathcal{S}| \max_{r \in \mathcal{S}} |\mathcal{N}(r)| (|\mathcal{S}| e^{-\beta(\min_{a \neq b, E(a) > E(b)} (E(a) - E(b)) - 2\gamma_\beta)} + 1)$ .*

The second proposition needed to prove Theorem 2.1 is

**Proposition 2.5** (see Corollary 1.6 in [5]). *Given  $I \subset \mathcal{S}$  and distinct  $x, z \in \mathcal{S} \setminus I$ ,*

$$\mathbb{P}_x(\tau_z < \tau_I) = \frac{\mathbb{P}_x(\tau_z < \tau_{I \cup \{x\}})}{\mathbb{P}_x(\tau_{I \cup \{z\}} < \tau_x)}$$

*holds true.*

With this two Propositions, the proof of Theorem 2.1 can now be given quite easily.

*Proof. (of Theorem 2.1)* By first using the previous result and then Proposition 2.4 (with  $K(\beta)$  as defined there), we find

$$\begin{aligned} \mathbb{P}_x(\tau_z < \tau_y) &= \frac{\mathbb{P}_x(\tau_z < \tau_{\{x,y\}})}{\mathbb{P}_x(\tau_{\{z,y\}} < \tau_x)} \\ &\leq \frac{\mathbb{P}_x(\tau_z < \tau_x)}{\mathbb{P}_x(\tau_y < \tau_x)} \\ &\leq K(\beta) e^{-\beta(E(z^*(x,z)) - E(z^*(x,y)) - \tau\gamma_\beta)}. \end{aligned}$$

The argument is completed by noting that  $E(z^*(x,z)) > E(z^*(x,y))$  and  $K(\beta) \geq 0$  converges to

$$|\mathcal{S}| \max_{r \in \mathcal{S}} |\mathcal{N}(r)|$$

as  $\beta \rightarrow \infty$ . □

*Proof. (of Theorem 2.2)* If  $B$  occurs, the asserted bound follows directly from Theorem 2.1. Proceeding to the case when  $B^c$  occurs, i.e.  $E(z^*(x,y)) \leq E(z^*(x,m))$ , we first point out that

$$\begin{aligned} \mathbb{P}_x(\tau_y < \tau_m) &= \mathbb{P}_x(\tau_y < \tau_m, E(X_n) > E(z^*(x,y)) \text{ for some } n \leq \tau_y) \\ &\quad + \mathbb{P}_x(\tau_y < \tau_m, E(X_n) \leq E(z^*(x,y)) \text{ for all } n \leq \tau_y) \\ &=: P_1 + P_2. \end{aligned}$$

For all  $z \in \mathcal{S}$  with  $E(z) > E(z^*(x,y))$ , we have  $z^*(x,z) \neq x$  and  $E(z^*(x,z)) > E(z^*(x,y))$ , for

$$E(z^*(x,z)) \geq E(z) > E(z^*(x,y)) \geq E(x).$$

Therefore, by an appeal to Theorem 2.1,

$$\begin{aligned} P_1 &\leq \mathbb{P}_x(\tau_z < \tau_y \text{ for some } z \text{ with } E(z) > E(z^*(x,y))) \\ &\leq \sum_{z: E(z) > E(z^*(x,y))} \mathbb{P}_x(\tau_z < \tau_y) \\ &\leq \sum_{z: E(z) > E(z^*(x,y))} \varepsilon(x, y, z, \beta). \end{aligned}$$

To get an estimate for  $P_2$ , note that every minimal path from  $x$  to  $y$  must pass through  $V_{<}^{(i)}(m)$  (Lemma 1.14). With this observation and by another appeal to Theorem 2.1, we infer

$$\begin{aligned} \mathbb{P}_x(\tau_y < \tau_m, E(X_n) \leq E(z^*(x,y)) \text{ for all } n \leq \tau_y) &\leq \sum_{z \in V_{<}^{(i)}(m)} \mathbb{P}_z(\tau_y < \tau_m) \\ &\leq \sum_{z \in V_{<}^{(i)}(m)} \varepsilon(z, m, y, \beta), \end{aligned}$$

having further utilized that (by Proposition 1.6(f) and (e))  $z^*(z, y) \neq z$  and

$$\begin{aligned} E(z^*(z, m)) &< E(z^*(z, m')) \\ &\leq E(z^*(z, y)) \vee E(z^*(y, m')) \\ &\leq E(z^*(z, y)) \vee E(z^*(y, m)) \\ &\leq E(z^*(z, y)) \vee E(z^*(z, m)) \\ &= E(z^*(z, y)) \end{aligned}$$

for some  $m' \in M^{(i)} \setminus \{m\}$  with  $E(z^*(y, m)) \geq E(z^*(y, m'))$ , which must exist since  $y \notin V^{(i)}(m)$ .  $\square$

*Proof. (of Theorem 2.3)* We first note that  $y \notin V^{(l_j)}(m_j)$  for all  $1 \leq j \leq k$ . With  $m_0, \dots, m_k$  as stated in the theorem (recall  $m_0 = x$  and  $m_k = m$ ), we obtain

$$\begin{aligned} \mathbb{P}_x(\tau_y < \tau_m) &= \mathbb{P}_x(\tau_{m_1} < \tau_y < \tau_m) + \mathbb{P}_x(\tau_y < \tau_{m_1} \wedge \tau_m) \\ &\leq \mathbb{P}_{m_1}(\tau_y < \tau_m) + \mathbb{P}_x(\tau_y < \tau_{m_1}) \\ &\leq \mathbb{P}_{m_2}(\tau_y < \tau_m) + \mathbb{P}_{m_1}(\tau_y < \tau_{m_2}) + \mathbb{P}_x(\tau_y < \tau_{m_1}) \\ &\quad \vdots \\ &\leq \sum_{j=1}^k \mathbb{P}_{m_{j-1}}(\tau_y < \tau_{m_j}). \end{aligned}$$

Finally use Theorem 2.2 to infer

$$\mathbb{P}_{m_{j-1}}(\tau_y < \tau_{m_j}) \leq \tilde{\varepsilon}(m_{j-1}, m_j, y, \beta)$$

for each  $j = 1, \dots, k$ .  $\square$

**2.2. Quasi-stationarity and exit time.** Naturally, several other questions concerning the behavior of the process when moving in a fixed valley are of interest, and *quasi-stationarity* may appear as one to come up with first. For a given valley  $V$  (of any level), a quasi-stationary distribution  $\nu = (\nu(j))_{j \in V}$  is characterized by the *quasi-invariance*, viz.

$$(5) \quad \mathbb{P}_\nu(X_n = j | \tau_{\mathcal{S} \setminus V} > n) = \nu(j) \quad \text{for all } j \in V,$$

but also satisfies

$$(6) \quad \lim_{n \rightarrow \infty} \mathbb{P}_\mu(X_n = j | \tau_{\mathcal{S} \setminus V} > n) = \nu(j) \quad \text{for all } j \in V$$

if  $\mu$  is an arbitrary distribution with  $\mu(V) = 1$ . The latter property renders uniqueness of  $\nu$ . Since  $\mathcal{S}$  is finite, existence of  $\nu$  follows by an old result due to DARROCH & SENETA [7]. It is obtained as the normalized eigenvector of the Perron-Frobenius eigenvalue  $\lambda = \lambda(V)$  of a modification of  $\mathbf{P}$ . This eigenvalue  $\lambda$  is also the probability for the chain to stay in  $V$  at least one step when started with  $\nu$ , thus  $\mathbb{P}_\nu(\tau_{V^c} > 1) = \lambda$ . As an immediate consequence, one finds that the exit time  $\tau_{V^c}$  has a geometric distribution with parameter  $1 - \lambda$  under  $\mathbb{P}_\nu$ . In the present context, this naturally raises the question how the parameter  $\lambda$  relates to the transition probabilities or the energies of the valley  $V$ . A simple probabilistic argument shows the following basic and intuitively obvious result concerning the eigenvalues associated with the nesting  $V^{(1)}(m) \subset \dots \subset V^{(i)}(m)$  (Proposition 1.12) for any  $1 \leq i \leq \mathbf{n}$  and  $m \in M^{(i)}$ .

**Proposition 2.6.** *Fixing any  $1 \leq i \leq \mathbf{n}$  and  $m \in M^{(i)}$ , let  $\lambda^{(j)} := \lambda(V^{(j)}(m))$  for  $j = 1, \dots, i$ . Then  $\lambda^{(1)} \leq \dots \leq \lambda^{(i)}$ .*



*Proof.* Write  $\nu_j$  as shorthand for the quasi-stationary distribution on  $V^{(j)}(m)$  and  $T_j$  for  $\tau_{\mathcal{S} \setminus V^{(j)}(m)}$ . Plainly,  $T_j \leq T_{j+1}$

$$\begin{aligned}
 (\lambda^{(j)})^n &= \mathbb{P}_{\nu_j}(T_j > n) \leq \mathbb{P}_{\nu_j}(T_{j+1} > n) \\
 &= \int_{\{T_{j+1} > k\}} \mathbb{P}_{X_k}(T_{j+1} > n - k) d\mathbb{P}_{\nu_j} \\
 (7) \qquad &= \mathbb{P}_{\nu_j}(T_{j+1} > k) \mathbb{P}_{\mu_k}(T_{j+1} > n - k),
 \end{aligned}$$

where  $\mu_k(x) := \mathbb{P}_{\nu_j}(X_k = x | T_{j+1} > k)$  for  $x \in V^{(j+1)}$ . Since  $\mathcal{S}$  is finite and by virtue of (6), we have that  $\mu_k \leq 2\nu_{j+1}$  when choosing  $k$  sufficiently large. For any such  $k$ , we find that (7) has upper bound

$$2\mathbb{P}_{\nu_j}(T_{j+1} > k) \mathbb{P}_{\nu_{j+1}}(T_{j+1} > n - k) = 2\mathbb{P}_{\nu_j}(T_{j+1} > k) (\lambda^{(j+1)})^{n-k}.$$

Hence, we finally conclude

$$\lambda^{(j)} \leq \left(2\mathbb{P}_{\nu_j}(T_{j+1} > k) (\lambda^{(j+1)})^{-k}\right)^{1/n} \lambda^{(j+1)}$$

and thereby the assertion upon letting  $n \rightarrow \infty$ .  $\square$

Another question is how long a given valley is visited and thus about its exit time. There is an extensive literature on exit problems for different kinds of stochastic processes. We mention [22, Ch. XI.2] and [8, Ch. 4, §4, Theorem 4.1] as two related to our work. The latter one studies perturbed systems on a continuous space. We can discretize their argument to get, with use of the main theorem in [21], a nice result on the time needed to leave a valley  $V^{(i)}(m)$  for any fixed  $1 \leq i \leq n$  and  $m \in M^{(i)}$ . This result is more explicit than the one in [16, Proposition 4.6].

**Definition 2.7.** For  $1 \leq i \leq n$ ,  $N := N^{(i)}$ , we define the following stopping (entrance/exit) times:

$$\begin{aligned}
 \xi_0^{(i)} &:= \tau_{N^c} \\
 \zeta_n^{(i)} &:= \inf \left\{ k \geq \xi_n^{(i)} | X_k \in N \right\} \\
 \xi_{n+1}^{(i)} &:= \inf \left\{ k \geq \zeta_n^{(i)} | X_k \in N^c \right\}, \quad n \geq 0.
 \end{aligned}$$

The entrance times  $\xi_n^{(i)}$  mark the epochs when a new valley is visited, while the exit times  $\zeta_n^{(i)}$  are the epochs at which a valley is left. The reader should notice that we do not restrict ourselves to valleys of order  $i$  but include those valleys which up to order  $i$  have not yet been absorbed by some larger valley. Exit and entrance times never coincide since there is no way to go from one valley to another without hitting a non-assigned state - crests are always non-assigned (see Lemma 1.15).

In this section, we will focus on  $\zeta_0^{(i)}$  for any fixed  $i$ , thus writing  $\zeta_0 := \zeta_0^{(i)}$  hereafter, but later for the macroscopic process the other times will be needed as well.

For each valley  $V^{(i)}(m)$ ,  $m \in M^{(i)}$ , let us define

$$(8) \quad s_m = s_m^{(i)} := \operatorname{argmin}_{s \in \partial^+ V^{(i)}(m)} E(s) = \operatorname{argmin}_{s \in \partial^+ V^{(i)}(m)} E(z^*(m, s)),$$

where the second equality follows from Lemma 1.15.

**Theorem 2.8.** *Let  $m \in M^{(i)}$ . Then*

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \ln \mathbb{E}_r \zeta_0 = E(s_m) - E(m)$$

for any  $r \in V^{(i)}(m)$ .

For the upper bound, we need a result from [21], which in our notation is:

**Proposition 2.9** (see Main Theorem (iv) in [21]). *For any  $1 \leq i \leq n$ ,  $\beta$  sufficiently large and  $t > 2^{i-1} \exp(\beta(E(s_{m^{(i-1)}}) - E(m^{(i-1)}) + 2i|\mathcal{S}|\gamma_\beta))$ ,*

$$\sup_{x \in M^{(i-1)}} \mathbb{P}_x(\tau_{M^{(i)}} > t) \leq \exp(-\Delta\beta)$$

holds true with a positive constant  $\Delta$ , where  $M^{(0)} = \mathcal{S}$  should be recalled.

This will now be used to show the following result.

**Lemma 2.10.** *Fix  $1 \leq i \leq n$ ,  $m \in M^{(i)}$  and  $r \in V^{(i)}(m)$ . Then, for any  $\beta$  sufficiently large and  $t > 2^i \exp(\beta(E(s_m) - E(m) + 2(i+1)|\mathcal{S}|\gamma_\beta))$ , it holds true that*

$$\mathbb{P}_r(\zeta_0 < (i+1)t) \geq \frac{1}{4}.$$

*Proof.* Let us first note that we can always arrange for  $m$  being equal to  $m^{(i)}$  by sufficiently decreasing the energy function at any  $m' \in M^{(i)} \setminus \{m\}$  so as to make  $E(s_m) - E(m)$  minimal among all states in  $M^{(i)}$ . This affects neither the valley  $V^{(i)}(m)$  and its outer boundary nor the distribution of  $\zeta_0$  when starting in  $m$ , for this distribution does not depend on the energy landscape outside of  $V^{(i)}(m) \cup \partial^+ V^{(i)}(m)$ . When applying the previous proposition, the constant  $\Delta$  may have changed but is still positive which suffices for our purposes. So let  $m = m^{(i)}$  hereafter.

Fix  $t > 2^i \exp(\beta(E(s_m) - E(m) + 2(i+1)|\mathcal{S}|\gamma_\beta))$  and  $T := it$ . Since

$$E(s_m) - E(m) \geq E(s_{m^{(j)}}) - E(m^{(j)})$$

for every  $1 \leq j \leq i$ , we infer

$$\begin{aligned} \mathbb{P}_r(\tau_{M^{(i)}} \leq T) &\geq \mathbb{P}_r(\tau_{M^{(i)}} \leq T, \tau_{M^{(1)}} \leq t) \\ &\geq \mathbb{P}_r(\tau_{M^{(1)}} \leq t) \inf_{x \in M^{(1)}} \mathbb{P}_x(\tau_{M^{(i)}} \leq (i-1)t) \\ &\geq \mathbb{P}_r(\tau_{M^{(1)}} \leq t) \inf_{x \in M^{(1)}} \mathbb{P}_x(\tau_{M^{(2)}} \leq t) \inf_{x \in M^{(2)}} \mathbb{P}_x(\tau_{M^{(i)}} \leq (i-2)t) \\ &\quad \vdots \\ &\geq \prod_{j=1}^i \inf_{x \in M^{(j-1)}} \mathbb{P}_x(\tau_{M^{(j)}} \leq t) \\ &\geq (1 - \exp(-\Delta\beta))^i \\ &\geq \frac{3}{4} \end{aligned}$$

for  $\beta$  sufficiently large. Furthermore, for  $\beta$  so large that  $\mathbb{P}_r(\tau_{M^{(i)}} < \tau_m) \leq 1/4$ , we find that

$$\begin{aligned} \mathbb{P}_r(\tau_{M^{(i)}} \leq T) &= \mathbb{P}_r(\tau_{M^{(i)}} = \tau_m \leq T) + \mathbb{P}_r(\tau_{M^{(i)}} \leq T, \tau_{M^{(i)}} < \tau_m) \\ &\leq \mathbb{P}_r(\tau_m \leq T) + \mathbb{P}_r(\tau_{M^{(i)}} < \tau_m) \\ &\leq \mathbb{P}_r(\tau_m \leq T) + \frac{1}{4}. \end{aligned}$$

By combining both estimates, we obtain

$$\mathbb{P}_r(\tau_m \leq T) \geq \mathbb{P}_r(\tau_{M^{(i)}} \leq T) - \frac{1}{4} \geq \frac{1}{2}.$$

Hence, state  $m$  is hit in time  $T$  with at least probability  $1/2$  when starting in  $r$ . Since  $m = m^{(i)}$ , we further have

$$\mathbb{P}_m(\zeta_0 \leq t) \geq \mathbb{P}_m(\tau_{M^{(i+1)}} \leq t) \geq 1 - \exp(-\Delta\beta) \geq \frac{1}{2}$$

for  $\beta$  sufficiently large. Hence, state  $s_m$  is hit in time  $t$  with at least probability  $1/2$  when starting in  $m$ . By combining the estimates, we finally obtain

$$\begin{aligned} \mathbb{P}_r(\zeta_0 \leq (i+1)t) &\geq \mathbb{P}_r(\zeta_0 \leq T+t | \tau_m \leq T) \mathbb{P}_r(\tau_m \leq T) \\ &\geq \mathbb{P}_r(\tau_m \leq T) \mathbb{P}_m(\zeta_0 \leq t) \\ &\geq \frac{1}{4}, \end{aligned}$$

which proves our claim.  $\square$

*Proof. (of Theorem 2.8)* Using the lemma just shown, we infer

$$\begin{aligned} \mathbb{E}_r(\zeta_0) &\leq (i+1)t \sum_{n \geq 0} (n+1) \mathbb{P}_r(n(i+1)t \leq \zeta_0 < (n+1)(i+1)t) \\ &= (i+1)t \sum_{n \geq 0} (n+1) \left( \mathbb{P}_r(\zeta_0 \geq n(i+1)t) - \mathbb{P}_r(\zeta_0 \geq (n+1)(i+1)t) \right) \\ &= (i+1)t \sum_{n \geq 0} \mathbb{P}_r(\zeta_0 \geq n(i+1)t) \\ &\leq (i+1)t \sum_{n \geq 0} \left( \max_{x \in V} \mathbb{P}_x(\zeta_0 \geq (i+1)t) \right)^n \\ &\leq (i+1)t \sum_{n \geq 0} \left( \frac{3}{4} \right)^n \\ &= 4(i+1)t, \end{aligned}$$

where  $t := 2^i \exp(\beta(E(s_m) - E(m) + 2(i+1)|\mathcal{S}|\gamma_\beta)) + 1$ . Since  $\gamma_\beta \rightarrow 0$ , we get in the limit

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \ln \mathbb{E}_r \zeta_0 \leq E(s_m) - E(m)$$

for all  $r \in V^{(i)}(m)$ .

Turning to the lower bound, define a sequence of stopping times, viz.  $\rho_0 := 0$  and

$$\rho_n := \inf\{k > \rho_{n-1} | X_k = m \text{ or } X_k \in \partial^+ V\}$$

for  $n \geq 1$ . Then  $Z_n := X_{\rho_n}$ ,  $n \geq 0$ , forms a Markov chain the transition probabilities of which when starting in  $m$  can be estimated with the help of Proposition 2.4, namely

$$\begin{aligned} \mathbb{P}(Z_1 \in \partial^+ V^{(i)}(m) | Z_0 = m) &= \mathbb{P}_m(\rho_1 = \zeta_0) \\ &= \mathbb{P}_m(\zeta_0 < \tau_m) \\ &\leq \sum_{s \in \partial^+ V} \mathbb{P}_m(\tau_s < \tau_m) \\ &\leq K(\beta) e^{-\beta(\min_{s \in \partial^+ V} E(z^*(m,s)) - E(m) - 2\gamma_\beta)} \\ &= K(\beta) e^{-\beta(E(s_m) - E(m) - 2\gamma_\beta)} \end{aligned}$$

where  $K(\beta) \rightarrow K \in (0, \infty)$  as  $\beta \rightarrow \infty$ . Further defining  $\nu := \inf\{k \geq 0 | Z_k \in \partial^+ V\}$ , this implies in combination with a geometric trials argument that

$$\mathbb{P}_m(\nu > n) \geq \left(1 - K(\beta) e^{-\beta(E(s_m) - E(m) - 2\gamma_\beta)}\right)^{n-1}.$$

As a consequence,

$$\mathbb{E}_m \zeta_0 = \sum_{n \geq 1} \mathbb{E}_m \underbrace{(\rho_n - \rho_{n-1})}_{\geq 1} \mathbb{1}_{\{\nu \geq n\}} \geq \sum_{n \geq 1} \mathbb{P}_m(\nu \geq n) \geq K(\beta) e^{-\beta(E(s_m) - E(m) - 2\gamma_\beta)}.$$

For arbitrary  $r \in V$ , we now infer

$$\begin{aligned} \mathbb{E}_r \zeta_0 &= \mathbb{E}_r \zeta_0 \mathbb{1}_{\{\zeta_0 \leq \rho_1\}} + \mathbb{E}_r \zeta_0 \mathbb{1}_{\{\zeta_0 > \rho_1\}} \\ &\geq \mathbb{E}_r (\mathbb{E}_r(\zeta_0 \mathbb{1}_{\{\zeta_0 > \rho_1\}} | X_{\rho_1} = m)) \\ &\geq \mathbb{E}_r \mathbb{1}_{\{\zeta_0 > \rho_1\}} \mathbb{E}_m \zeta_0 \\ &\geq \mathbb{P}_r(\zeta_0 > \rho_1) K(\beta) e^{-\beta(E(s_m) - E(m) - 2\gamma_\beta)} \\ &\geq \frac{1}{2} K(\beta) e^{-\beta(E(s_m) - E(m) - 2\gamma_\beta)} \end{aligned}$$

for all sufficiently large  $\beta$ , because  $\lim_{\beta \rightarrow \infty} \mathbb{P}_r(\zeta_0 > \rho_1) = 1$  (Theorem 2.3). Finally, by taking logarithms and letting  $\beta$  tend to  $\infty$ , we arrive at the inequality

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \ln \mathbb{E}_r \zeta_0 \geq E(s_m) - E(m)$$

which completes the proof.  $\square$

In [12],  $E(s_m) - E(m)$ ,  $m \in M^{(i)}$ , is referred to as the depth of the valley  $V(m)$ . Therefore, Property 3 from the Introduction holds true and we can relate thermodynamics of the system (energies) to dynamics of the chain (holding times) in a very precise way. Especially, there is no universal scale for the times spent in different valleys because in general they differ exponentially.

### 3. MACROSCOPIC VIEW: TRANSITIONS BETWEEN VALLEYS

With the help of the nested state space decompositions into valleys of different orders and around bottom states of different stability, we will now be able to provide an appropriate definition of the metabasins (MB) that has been announced and to some extent discussed in the Introduction. We will further define and study macroscopic versions of the original process  $X = (X_n)_{n \geq 0}$ . These are obtained by choosing different levels of aggregation in the sense that they keep track only of the valleys of a chosen level that are visited by  $X$ . The motivation behind this approach is, on the one hand, to exhibit strong relations between properties of the energy landscape and the behavior of  $X$  (as in Theorem 2.8) and, on the other hand, to describe essential features of this process by looking at suitable macroscopic scales.

In the subsequent definition of aggregated versions of  $X$ , we will distinguish between two variants:

- A *time-scale preserving* aggregation that, for a fixed level and each  $n$ , keeps track of the valley the original chain visits at time  $n$  and thus only blinds its exact location within a valley.
- An *accelerated* version that, while also keeping track of the visited valleys, further blinds the sojourn times within a valley by counting a visit just once.

Actually, the definition of these aggregations at a chosen level  $i$  is a little more complicated because their state space, denoted  $\mathcal{S}^{(i)}$  below and the elements of which we call *level  $i$  metastates*, also comprises the non-assigned states at level  $i$  as well as the minima of those valleys that were formed at an earlier level and whose merger is pending at level  $i$  because their minima are not attracted at this level.

**Definition 3.1.** Fix  $1 \leq i \leq n$ , let  $\mathcal{S}^{(i)} := \{m^{(j)} \in M^{(1)} \mid l(j) > i\} \cup N^{(i)}$  and

$$V^{(i)}(s) := \begin{cases} V^{(i)}(m^{(j)}), & \text{if } s = m^{(j)} \text{ for some } j \geq i \\ V^{(j)}(m^{(j)}), & \text{if } s = m^{(j)} \text{ for some } j < i \\ \{s\}, & \text{if } s \in N^{(i)} \end{cases}$$

for  $s \in \mathcal{S}^{(i)}$ . Then define

$$\begin{aligned} \bar{Y}_n^{(i)} &:= \sum_{s \in \mathcal{S}^{(i)}} s \mathbb{1}_{\{X_n \in V^{(i)}(s)\}}, \quad n \geq 0, \\ \text{and } Y_n^{(i)} &:= \bar{Y}_{\sigma_n}^{(i)}, \quad n \geq 0, \end{aligned}$$

where  $\sigma_0 = \sigma_0^{(i)} := 0$  and

$$\sigma_n = \sigma_n^{(i)} := \inf \left\{ k > \sigma_{n-1} \mid \bar{Y}_k^{(i)} \neq \bar{Y}_{k-1}^{(i)} \right\}$$

for  $n \geq 1$ . We call  $\bar{Y}^{(i)} = (\bar{Y}_n^{(i)})_{n \geq 0}$  and  $Y^{(i)} = (Y_n^{(i)})_{n \geq 0}$  the *aggregated chain (AC)* and the *accelerated aggregated chain (AAC) (at level  $i$ )* associated with  $X = (X_n)_{n \geq 0}$ .

So, starting in an arbitrary valley, the original chain stays there for a time  $\zeta_0 = \zeta_0^{(i)}$  (as defined in Definition 2.7) before it jumps via some non-assigned states  $k_1, \dots, k_l$  (staying a geometric time in each of these states) to another valley at time  $\xi_1 = \xi_1^{(i)}$ . There it stays for  $\zeta_1 - \xi_1$  time units before it moves on in a similar manner. By going from  $X$  to its aggregation  $\bar{Y}^{(i)}$  at level  $i$ , we regard the whole valley  $V^{(i)}(s)$  for  $s \in \mathcal{S}^{(i)}$  as one single metastate and therefore give up information about the exact location of  $X$  within a valley.  $\bar{Y}^{(i)}$  is a jump process on  $\mathcal{S}^{(i)}$  with successive sojourn times  $\sigma_{n+1} - \sigma_n$ ,  $n \geq 0$ , which do not only depend on the valley but also on the states of entrance and exit. The AAC then is the embedded chain, viz.

$$\bar{Y}_n^{(i)} = \sum_{j \geq 0} Y_j^{(i)} \mathbb{1}_{\{\sigma_j \leq n < \sigma_{j+1}\}},$$

giving the states only at jumps epochs: starting from the minimum of a first valley it moves to states  $k_1, \dots, k_l \in N^{(i)}$  and then proceeds to the minimum of a second valley, and so on.

Of course, at small temperatures the time spent in a non-assigned state or in a valley around a low order metastable state is very small compared to the time spent in a valley around a metastable state of higher order. Thus, such states can be seen as instantaneous and of little importance for the evolution of the process. We account for them nonetheless for two reasons. First, in the path-dependent definition mentioned in the Introduction and used in Physics, they build small MB of great transitional activity of the process and are thus relevant in view of our goal to provide a definition of MB that conforms as much as possible to a path-dependent one. Second, a complete partitioning of the state space that is an assignment of every  $s \in \mathcal{S}$  to a metastate via a global algorithm fails when merely focusing on  $\{V^{(i)}(m), m \in M^{(i)}\}$  because there is neither an obvious nor natural way how to assign non-assigned states to them.

The *incoherent scattering function* and its associated relaxation time, for  $X$  defined by

$$S(q, n) := \mathbb{E}_\pi \cos(q|X_n - X_0|)$$

(with  $|\cdot|$  being Euclidean distance in phase space) and

$$\tau_q(\varepsilon) := \inf\{n \mid S(q, n) \leq \varepsilon\}, \quad \varepsilon > 0,$$

respectively, may serve as an example which shows the strong relation between the behavior of the original process and its macroscopic versions. For more detailed information on the meaning and relevance of  $S(q, n)$  as a measure of incoherent scattering between the initial state of a glass-forming system and its state  $n$  time steps onward, we refer to the survey by HEUER [12].

**Proposition 3.2.** *For each  $1 \leq i \leq n$  there is a constant  $\Delta(i)$  such that*

$$(9) \quad \sup_{n \geq 0} \mathbb{P}_\pi \left( X_n \neq \bar{Y}_n^{(i)} \right) \leq e^{-\Delta(i)\beta}.$$

*As a consequence, for any given  $\varepsilon > 0$ , the incoherent scattering functions of  $X$  and  $\bar{Y}^{(i)}$  differ by at most  $4\varepsilon$  for  $\beta$  sufficiently large.*

*Proof.* Use Lemma 1.1 to infer

$$\begin{aligned} \mathbb{P}_\pi(X_n \neq \bar{Y}_n^{(i)}) &= \sum_{s \in \mathcal{S}^{(i)} \setminus N^{(i)}} \mathbb{P}_\pi(\bar{Y}_n^{(i)} = s) \sum_{x \in V^{(i)}(s) \setminus \{s\}} \frac{\pi(x)}{\pi(V^{(i)}(s))} \\ &\leq \sum_{s \in \mathcal{S}^{(i)} \setminus N^{(i)}} \mathbb{P}_\pi(\bar{Y}_n^{(i)} = s) \sum_{x \in V^{(i)}(s) \setminus \{s\}} \frac{\pi(x)}{\pi(s)} \\ &\leq \sum_{s \in \mathcal{S}^{(i)} \setminus N^{(i)}} \mathbb{P}_\pi(\bar{Y}_n^{(i)} = s) \sum_{x \in V^{(i)}(s) \setminus \{s\}} e^{-\beta(E(x) - E(s) - 2\gamma_\beta)} \\ &\leq \max_{s \in \mathcal{S}^{(i)} \setminus N^{(i)}} |V^{(i)}(s)| \max_{x \in V^{(i)}(s) \setminus \{s\}} e^{-\beta(E(x) - E(s) - 2\gamma_\beta)}. \end{aligned}$$

This proves equation (9) because  $E(x) > E(s)$  for each  $x \in V^{(i)}(s) \setminus \{s\}$ ,  $s \in \mathcal{S}^{(i)} \setminus N^{(i)}$ . Now let  $\beta$  be so large that  $e^{-\Delta(i)\beta} \leq \varepsilon$  for a given  $\varepsilon > 0$  and observe that

$$\begin{aligned} &\mathbb{E}_\pi \cos(q|X_n - X_0|) \\ &= \int_{\{X_n = \bar{Y}_n^{(i)}, X_0 = \bar{Y}_0^{(i)}\}} \cos(q|X_n - X_0|) d\mathbb{P}_\pi \\ &\quad + \int_{\{X_n \neq \bar{Y}_n^{(i)}\} \cup \{X_0 \neq \bar{Y}_0^{(i)}\}} \cos(q|X_n - X_0|) d\mathbb{P}_\pi \\ &\leq \int_{\{X_n = \bar{Y}_n^{(i)}, X_0 = \bar{Y}_0^{(i)}\}} \cos(q|X_n - X_0|) d\mathbb{P}_\pi + \mathbb{P}_\pi(X_n \neq \bar{Y}_n^{(i)}) + \mathbb{P}_\pi(X_0 \neq \bar{Y}_0^{(i)}) \\ &= \mathbb{E}_\pi \cos(q|\bar{Y}_n^{(i)} - \bar{Y}_0^{(i)}|) - \int_{\{X_n \neq \bar{Y}_n^{(i)}\} \cup \{X_0 \neq \bar{Y}_0^{(i)}\}} \cos(q|\bar{Y}_n^{(i)} - \bar{Y}_0^{(i)}|) d\mathbb{P}_\pi + 2\varepsilon \\ &\leq \mathbb{E}_\pi \cos(q|\bar{Y}_n^{(i)} - \bar{Y}_0^{(i)}|) + 4\varepsilon \end{aligned}$$

and, by a similar argument,

$$\mathbb{E}_\pi \cos(q|\bar{Y}_n^{(i)} - \bar{Y}_0^{(i)}|) \leq \mathbb{E}_\pi \cos(q|X_n - X_0|) + 4\varepsilon.$$

This completes the proof.  $\square$

**3.1. (Semi-)Markov Property.** In general, both aggregated chains are no longer Markovian. Transition probabilities of the AAC not only depend on the current state, i.e. the current valley, but also on the entrance state into that valley, whereas transition probabilities of the AC depend on the current sojourn times which in turn depend on the previous, the present and the next state. On the other hand, since valleys are defined in such a way that asymptotically almost surely (a.a.s.), i.e., with probability tending to one as  $\beta \rightarrow \infty$ , the minimum will be reached from anywhere inside the valley before the valley is left, and since, furthermore, the exit state on the outer boundary a.a.s. equals

the one with the smallest energy, the AAC will be shown below to converge to a certain Markov chain on  $\mathcal{S}^{(i)}$ . Also, the sojourn times depend on the past only via the last and the current state. This means that the AC converges to a semi-Markov chain (for semi-Markov chains see for example [1]):

**Definition 3.3.** Given any nonempty countable set  $\mathcal{S}$ , let  $(M_n, T_n)_{n \geq 0}$  be a bivariate temporally homogeneous Markov chain on  $\mathcal{S} \times \mathbb{N}$ , with transition kernel  $Q(s, \cdot)$  only depending on the first component, viz., for all  $n \geq 0, s \in \mathcal{S}$  and  $t \geq 0$ ,

$$(10) \quad \mathbb{P}(M_{n+1} = s, T_{n+1} \leq t | M_n, T_n) = Q(M_n, \{s\} \times [0, t])$$

holds. Put  $S_n := \sum_{i=0}^n T_i$  for  $n \geq 0$  and  $\nu(t) := \max\{n \geq 0 | S_n \leq t\}$  ( $\max \emptyset := 0$ ) for  $t \geq 0$ . Then  $Z_n := M_{\nu(n)}, n \geq 0$ , is called *semi-Markov chain with embedded Markov chain*  $(M_n)_{n \geq 0}$  and *sojourn or holding times*  $T_0, T_1, \dots$

Note that equation (10) holds iff  $M = (M_n)_{n \geq 0}$  forms a temporally homogeneous Markov chain and the  $(T_n)_{n \geq 0}$  are conditionally independent given  $M$  such that the distribution of  $T_n$  only depends on  $M_{n-1}, M_n$  for  $n \geq 1$  (in a temporally homogeneous manner), and on  $M_0$  for  $n = 0$ . Note further that we have specialized to the case where holding times take values in  $\mathbb{N}$  only (instead of  $(0, \infty)$ ).

Recall from (8) the definition of  $s_m$  for  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and notice that the second equality there entails  $E(z^*(m, s_m)) < E(z^*(m, s))$  for any  $s \in \partial^+ V^{(i)}(m) \setminus \{s_m\}$ . Further recall from our basic assumptions that  $p^*(r, s) = \lim_{\beta \rightarrow \infty} p(r, s)$  exists for all  $r, s \in \mathcal{S}$  and is positive if  $E(r) \geq E(s)$ . The following result, revealing the announced convergence for AAC, confirms in particular that a valley  $V(m), m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , is a.a.s. to be left via  $s_m$ .

**Proposition 3.4.** For each  $1 \leq i \leq n$  and as  $\beta \rightarrow \infty$ , the level  $i$  AAC  $Y^{(i)}$  converges to a Markov chain  $\tilde{Y}^{(i)} = (\tilde{Y}_n^{(i)})_{n \geq 0}$  on  $\mathcal{S}^{(i)}$  with transition probabilities  $\tilde{p}(r, s) = \tilde{p}_i(r, s)$  stated below, that is

$$\lim_{\beta \rightarrow \infty} \mathbb{P}(Y_{n+1}^{(i)} = s | Y_n^{(i)} = r, Y_{n-1}^{(i)} = m_{n-1}, \dots, Y_0^{(i)} = m_0) = \tilde{p}(r, s)$$

for all  $m_0, \dots, m_{n-1}, r, s \in \mathcal{S}^{(i)}$  and  $n \geq 0$ . We have  $\tilde{p}(r, \cdot) := \delta_{s_r}$  if  $r \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , and

$$\begin{aligned} \tilde{p}(r, \cdot) := & \frac{1}{1 - p^*(r, r)} \left( \sum_{s \in \mathcal{N}(r) \cap N^{(i)}} p^*(r, s) \delta_s \right. \\ & \left. + \sum_{s \in \mathcal{S}^{(i)} \setminus N^{(i)}} \left( \sum_{r' \in \mathcal{N}(r) \cap V^{(i)}(s)} p^*(r, r') \right) \delta_s \right), \end{aligned}$$

if  $r \in N^{(i)}$ .

$\tilde{Y}^{(i)} = (\tilde{Y}_n^{(i)})_{n \geq 0}$  is called the *asymptotic jump chain at level  $i$*  hereafter. Note that, typically, it is not irreducible. It may have transient states, not necessarily non-assigned, and its irreducibility classes are of the form  $\{m_1, \dots, m_k, s\}$  for a collection  $m_1, \dots, m_k \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and some  $s \in N^{(i)}$  satisfying  $s = s_{m_1} = \dots = s_{m_k}$ .

*Proof.* Fix  $1 \leq i \leq n$  and write  $Y_n$  for  $Y_n^{(i)}$ . The first step is to verify that, as  $\beta \rightarrow \infty$ ,

$$\mathbb{P}(Y_{n+1} = s | Y_n = r, Y_{n-1} = m_{n-1}, \dots, Y_0 = m_0) = \mathbb{P}_r(Y_1 = s) + o(1)$$

for all  $m_0, \dots, m_{n-1}, r, s \in \mathcal{S}^{(i)}$  and  $n \geq 0$ . If  $r \in N^{(i)}$ , then  $Y_n = X_{\sigma_n}$  and the Markov property of  $X$  provide us with the even stronger result

$$\mathbb{P}(Y_{n+1} = s | Y_n = r, Y_{n-1} = m_{n-1}, \dots, Y_0 = m_0) = \mathbb{P}_r(Y_1 = s).$$

A little more care is needed if  $r \in \mathcal{S}^{(i)} \setminus N^{(i)}$ . For any  $s \in \mathcal{S}^{(i)}$ ,  $x \in V^{(i)}(r)$  and  $n \geq 0$ , we have

$$\begin{aligned} \mathbb{P}(Y_{n+1} = s | Y_n = r, X_{\sigma_n} = x) \\ &= \mathbb{P}_x(Y_1 = s, \tau_r < \sigma_1) + \mathbb{P}_x(Y_1 = s, \tau_r > \sigma_1) \\ &= \mathbb{P}_r(Y_1 = s) \mathbb{P}_x(\tau_r < \sigma_1) + \mathbb{P}_x(Y_1 = s, \tau_r > \sigma_1). \end{aligned}$$

The last two summands can further be bounded by

$$\mathbb{P}_r(Y_1 = s) \mathbb{P}_x(\tau_r < \sigma_1) \leq \mathbb{P}_r(Y_1 = s) \quad \text{and} \quad \mathbb{P}_x(Y_1 = s, \tau_r > \sigma_1) \leq \mathbb{P}_x(\sigma_1 < \tau_r).$$

For the last probability, Theorem 2.3 ensures

$$\mathbb{P}_x(\sigma_1 < \tau_r) \leq \sum_{z \in \partial^+ V^{(i)}(r)} \mathbb{P}_x(\tau_z < \tau_r) \xrightarrow{\beta \rightarrow \infty} 0.$$

Consequently, as  $\beta \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}_r(Y_1 = s) &= \left( 1 - \sum_{z \in \partial^+ V^{(i)}(r)} \tilde{\varepsilon}(x, r, z, \beta) \right) \mathbb{P}_r(Y_1 = s) + o(1) \\ &\leq (1 - \mathbb{P}_x(\sigma_1 < \tau_r)) \mathbb{P}_r(Y_1 = s) + o(1) \\ &\leq \mathbb{P}(Y_{n+1} = s | X_{\sigma_n} = x, Y_n = r) + o(1) \\ &\leq \mathbb{P}_r(Y_1 = s) + o(1), \end{aligned}$$

and therefore

$$\begin{aligned} \mathbb{P}(Y_{n+1} = s | Y_n = r, Y_{n-1} = m_{n-1}, \dots, Y_0 = m_0) \\ &= \sum_{x \in V^{(i)}(r)} \mathbb{P}(Y_{n+1} = s | X_{\sigma_n} = x, Y_n = r) \mathbb{P}(X_{\sigma_n} = x | Y_n = r, Y_{n-1} = m_{n-1}, \dots, Y_0 = m_0) \\ &= \mathbb{P}_r(Y_1 = s) + o(1). \end{aligned}$$

It remains to verify that  $\mathbb{P}_r(Y_1 = s) = \tilde{p}(r, s) + o(1)$  for any  $r, s \in \mathcal{S}^{(i)}$ . If  $r \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , then  $\sigma_1 = \tau_{N^{(i)}}$  and  $Y_1 = X_{\tau_{N^{(i)}}}$ . Since  $E(z^*(r, s_r)) < E(z^*(r, s))$  for each  $s_r \neq s \in N^{(i)} \cap \partial^+ V^{(i)}(r)$ , we now infer with the help of Theorem 2.1

$$\begin{aligned} \mathbb{P}_r(Y_1 \neq s_r) &= \mathbb{P}_r(\tau_s < \tau_{s_r} \text{ for some } s \in N^{(i)} \setminus \{s_r\}) \\ &\leq \sum_{s_r \neq s \in N^{(i)}} \mathbb{P}_r(\tau_s < \tau_{s_r}) \\ &\leq \sum_{s_r \neq s \in N^{(i)}} \tilde{\varepsilon}(r, s_r, s, \beta) \\ &= o(1), \end{aligned}$$

as  $\beta \rightarrow \infty$  and thus  $\mathbb{P}_r(Y_1 \in \cdot) \rightarrow \delta_{s_r} = \tilde{p}(r, \cdot)$  as claimed. If  $r \in N^{(i)}$ , then either  $Y_1 = s \in \mathcal{N}(r) \cap N^{(i)}$ , or  $Y_1 = s \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and  $X_{\sigma_1} = r'$  for some  $r' \in \mathcal{N}(r) \cap V^{(i)}(s)$ . It thus follows that

$$\mathbb{P}_r(Y_1 = s) = \mathbb{P}_r(X_{\sigma_1} = s) = \frac{p(r, s)}{1 - p(r, r)} = \frac{p^*(r, s)}{1 - p^*(r, r)} + o(1)$$

if  $s \in \mathcal{N}(r) \cap N^{(i)}$ , while

$$\mathbb{P}_r(Y_1 = s) = \sum_{r' \in \mathcal{N}(r) \cap V^{(i)}(s)} \mathbb{P}_r(X_{\sigma_1} = r') = \sum_{r' \in \mathcal{N}(r) \cap V^{(i)}(s)} \frac{p^*(r, r')}{1 - p^*(r, r)} + o(1)$$

in the second case.  $\square$



Having shown that  $Y^{(i)}$  behaves asymptotically as a Markov chain, viz. the jump chain  $\tilde{Y}^{(i)}$ , it is fairly easy to verify with the help of the next simple lemma that the augmented bivariate AC  $(\bar{Y}_n^{(i)}, \bar{Y}_{n+1}^{(i)})_{n \geq 0}$  is asymptotically semi-Markovian.

**Lemma 3.5.** *For each  $\beta > 0$ , the sojourn times  $\sigma_{n+1} - \sigma_n$ ,  $n \geq 0$ , of the AC  $\bar{Y}^{(i)}$  are conditionally independent given  $Y^{(i)}$ . The conditional law of  $\sigma_{n+1} - \sigma_n$  depends only on  $(Y_{n-1}^{(i)}, Y_n^{(i)}, Y_{n+1}^{(i)})$  and satisfies*

$$(11) \quad \begin{aligned} & \mathbb{P}(\sigma_{n+1} - \sigma_n \in \cdot \mid Y_{n-1}^{(i)} = x, Y_n^{(i)} = y, Y_{n+1}^{(i)} = z) = Q((x, y, z), \cdot) \\ & := \begin{cases} \text{Geom}(1 - p(y, y)), & \text{if } y \in N^{(i)} \\ \sum_{s \in V^{(i)}(y), s \sim x} \mathbb{P}_s(\sigma_1 \in \cdot \mid Y_1^{(i)} = z) \mathbb{P}_x(X_{\sigma_1} = s), & \text{if } y \notin N^{(i)} \end{cases} \end{aligned}$$

for all  $x, y, z \in \mathcal{S}^{(i)}$  with  $\mathbb{P}(Y_{n-1}^{(i)} = x, Y_n^{(i)} = y, Y_{n+1}^{(i)} = z) > 0$  and  $n \geq 1$ .

*Proof.* The assertions follow easily when observing that, on the one hand, at least one state  $y \in N^{(i)}$  must be visited between two states  $x, z \in \mathcal{S}^{(i)} \setminus N^{(i)}$  (Lemma 1.15) and that, on the other hand, the original chain  $X$  and its aggregation  $\bar{Y}^{(i)}$  coincide at any epoch where a non-assigned state is hit, which renders the Markov property of  $\bar{Y}^{(i)}$  at these epochs. Further details are omitted.  $\square$

In order to formulate the next result, let  $0 = \tilde{\sigma}_0 < \tilde{\sigma}_1 < \dots$  be an increasing sequence of random variables such that its increments  $\tilde{\sigma}_{n+1} - \tilde{\sigma}_n$ ,  $n \geq 0$ , are conditionally independent given the asymptotic jump chain  $\tilde{Y}^{(i)}$ . Moreover, let the conditional law of  $\tilde{\sigma}_{n+1} - \tilde{\sigma}_n$  depend only on  $(\tilde{Y}_{n-1}^{(i)}, \tilde{Y}_n^{(i)}, \tilde{Y}_{n+1}^{(i)})$  and be equal to  $Q((\tilde{Y}_{n-1}^{(i)}, \tilde{Y}_n^{(i)}, \tilde{Y}_{n+1}^{(i)}), \cdot)$ , with  $Q$  as defined in (11). Then  $((\tilde{Y}_n^{(i)}, \tilde{Y}_{n+1}^{(i)}), \tilde{\sigma}_{n+1})_{n \geq 0}$  forms a Markov renewal process and  $(\tilde{Y}_{\tilde{\nu}(n)}^{(i)}, \tilde{Y}_{\tilde{\nu}(n+1)}^{(i)})_{n \geq 0}$  a semi-Markov chain, where  $\tilde{\nu}(n) := \sup\{k \geq 0 \mid \tilde{\sigma}_k \leq n\}$ .

**Proposition 3.6.** *For each  $1 \leq i \leq n$ ,  $((Y_n^{(i)}, Y_{n+1}^{(i)}), \sigma_{n+1})_{n \geq 0}$  converges to the Markov renewal process  $((\tilde{Y}_n^{(i)}, \tilde{Y}_{n+1}^{(i)}), \tilde{\sigma}_{n+1})_{n \geq 0}$  in the sense that*

$$\lim_{\beta \rightarrow \infty} \frac{\mathbb{P}_{y_0}((Y_k^{(i)}, Y_{k+1}^{(i)}) = (y_k, y_{k+1}), \sigma_{k+1} = i_{k+1}, 0 \leq k \leq n)}{\mathbb{P}_{y_0}((\tilde{Y}_k^{(i)}, \tilde{Y}_{k+1}^{(i)}) = (y_k, y_{k+1}), \tilde{\sigma}_{k+1} = i_{k+1}, 0 \leq k \leq n)} = 1$$

for all  $y_0, \dots, y_{n+1} \in \mathcal{S}^{(i)}$ ,  $0 < i_1 < \dots < i_{n+1}$  and  $n \geq 0$  such that the denominator is positive. Furthermore,  $(\bar{Y}_n^{(i)}, \bar{Y}_{n+1}^{(i)})_{n \geq 0}$  is asymptotically semi-Markovian in the sense that

$$\lim_{\beta \rightarrow \infty} \frac{\mathbb{P}_{y_0}((\bar{Y}_k^{(i)}, \bar{Y}_{k+1}^{(i)}) = (y_k, y_{k+1}), 0 \leq k \leq n)}{\mathbb{P}_{y_0}((\tilde{Y}_{\tilde{\nu}(k)}^{(i)}, \tilde{Y}_{\tilde{\nu}(k+1)}^{(i)}) = (y_k, y_{k+1}), 0 \leq k \leq n)} = 1$$

for all  $y_0, \dots, y_{n+1} \in \mathcal{S}^{(i)}$  and  $n \geq 0$  such that the denominator is positive.

*Proof.* The first assertion being obvious by Proposition 3.4, note that it implies, with  $\nu(n) := \sup\{k \geq 0 \mid \sigma_k \leq n\}$ ,

$$\lim_{\beta \rightarrow \infty} \frac{\mathbb{P}_{y_0}((Y_{\nu(k)}^{(i)}, Y_{\nu(k+1)}^{(i)}) = (y_k, y_{k+1}), 0 \leq k \leq n)}{\mathbb{P}_{y_0}((\tilde{Y}_{\tilde{\nu}(k)}^{(i)}, \tilde{Y}_{\tilde{\nu}(k+1)}^{(i)}) = (y_k, y_{k+1}), 0 \leq k \leq n)} = 1$$

for all  $y_0, \dots, y_{n+1} \in \mathcal{S}^{(i)}$  and  $n \geq 0$  such that the denominator is positive. Therefore the second assertion follows when finally noting that

$$Y_{\nu(n)}^{(i)} = \sum_{j \geq 0} Y_j^{(i)} \mathbb{1}_{\{\sigma_j \leq n < \sigma_{j+1}\}} = \bar{Y}_n^{(i)}$$

for each  $n \geq 0$ . □

So we have shown that, although aggregation generally entails the loss of the Markov property, here it leads back to processes of this kind (Markov or semi-Markov chains) in an asymptotic sense at low temperature regimes.

**3.2. Reciprocating Jumps.** As discussed to some extent in the Introduction, we want to find an aggregation level at which reciprocating jumps appear to be very unlikely so as to obtain a better picture of essential features of the observed process. To render precision to this informal statement requires to further specify the term “reciprocating jump” and to provide a measure of likelihood for its occurrence. It is useful to point out first that the original chain  $X$  exhibits two types of reciprocating jumps:

**Intra-valley jumps:** which occur between states inside a valley (starting in a minimum the process falls back to it many times before leaving the valley).

**Inter-valley jumps:** which occur between two valleys (typically, when the energy barrier between these valleys is much lower then the barrier to any other valley).

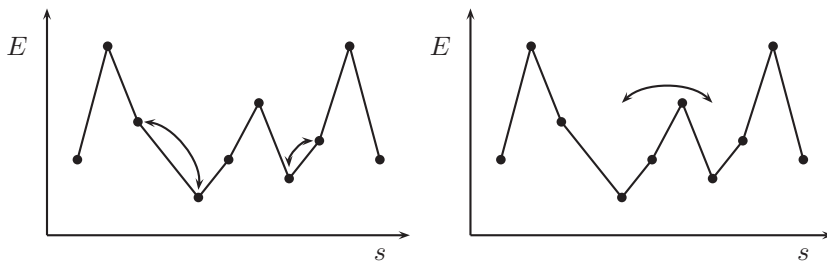


FIGURE 4. Illustration of intra-valley jumps (left panel) versus inter-valley jumps (right panel).

Clearly, intra-valley jumps disappear by aggregating valleys into metastates, while inter-valley jumps may also be viewed as intra-valley jumps for higher order valleys and do occur when transitions between any two of them are much more likely than those to other valleys in which case they should be aggregated into one valley. This motivates the following definition.

**Definition 3.7.** We say the process  $(Y_n^{(i)})_{n \in \mathbb{N}}$  exhibits *reciprocating jumps of order*  $\varepsilon > 0$  if there exists a nonempty subset  $A \subsetneq \mathcal{S}^{(i)} \setminus N^{(i)}$  with the following property: For each  $m_1 \in A$ , there exists  $m_2 \in A$  such that

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \left( \ln \left( \mathbb{P}_{m_1} \left( X_{\xi_1} \in V^{(i)}(m_2) \right) \right) - \ln \left( \mathbb{P}_{m_1} \left( X_{\xi_1} \in V^{(i)}(m) \right) \right) \right) \geq \varepsilon$$

for all  $m \in \mathcal{S}^{(i)} \setminus (N^{(i)} \cup A)$ . In other words, it is exponentially more likely to stay in  $A$  than to leave it (ignoring intermediate visits to non-assigned states).

In view of our principal goal to give a path-independent definition of MBs, we must point out that, by irreducibility, reciprocating jumps *always* occur with positive probability at any nontrivial level of aggregation and can therefore never be ruled out completely. This is in contrast to the path-dependent version by HEUER [12] in which the non-occurrence of reciprocating jumps appears to be the crucial requirement. As a consequence, Definition 3.7 provides an alternative, probabilistic and verifiable criterion for reciprocating jumps to be sufficiently unlikely in a chosen aggregation.

The following proposition contains further information on which valleys are visited successively by providing the probabilities of making a transition from  $V^{(i)}(m)$  to  $V^{(i)}(m')$  for any  $m, m' \in \mathcal{S}^{(i)} \setminus N^{(i)}$ . It is a direct consequence of the asymptotic results in the previous subsection, notably Proposition 3.4.

**Proposition 3.8.** *Let  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ ,  $s_m \in \partial^+ V^{(i)}(m)$  be as defined in (8) (i.e., the state on the outer boundary of  $V^{(i)}(m)$  with minimal energy). Then*

$$\lim_{\beta \rightarrow \infty} \mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m)) = \tilde{p}(s_m, m) = \frac{\sum_{r \in \mathcal{N}(s_m) \cap V^{(i)}(m)} p^*(s_m, r)}{1 - p^*(s_m, s_m)},$$

while

$$\lim_{\beta \rightarrow \infty} \mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m')) = \tilde{p}(s_m, m') + \sum_{n \geq 1} \sum_{r_1, \dots, r_n \in N^{(i)}} \tilde{p}(s_m, r_1) \cdots \tilde{p}(r_{n-1}, r_n) \tilde{p}(r_n, m')$$

for any other  $m' \in \mathcal{S}^{(i)} \setminus N^{(i)}$ .

The reader should notice that, as  $\tilde{p}(s, r) = 0$  whenever  $E(s) < E(r)$ , the last sum actually ranges only over those non-assigned  $r_1, \dots, r_n$  with  $E(s_m) > E(r_1) > \dots > E(r_n) > E(m')$ .

*Proof.* Let us first point out that  $\mathbb{P}_r(X_{\xi_0} \in V^{(i)}(m)) = o(1)$  as  $\beta \rightarrow \infty$  for any  $r \in N^{(i)}$  such that  $E(r) < E(s_m)$ . Namely, since the last property implies  $r \notin \partial^+ V^{(i)}(m)$ , any path from  $r$  into  $V^{(i)}(m)$  must traverse a state  $s \in \partial^+ V^{(i)}(m)$  with  $E(s) \geq E(s_m) > E(r)$ , whence the probability for such a path goes to zero as  $\beta \rightarrow \infty$ . Noting further that  $\mathbb{P}_m(Y_1^{(i)} \neq s_m) = o(1)$  as  $\beta \rightarrow \infty$  by Proposition 3.4, we now infer (with  $\xi_n = \xi_n^{(i)}$ )

$$\begin{aligned} & \mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m)) \\ &= \mathbb{P}_{s_m}(X_{\xi_0} \in V^{(i)}(m)) + o(1) \\ &= \mathbb{P}_{s_m}(Y_1^{(i)} = m) + \sum_{r \in \mathcal{N}(s_m) \cap N^{(i)}} \frac{p(s_m, r)}{1 - p(s_m, s_m)} \mathbb{P}_r(X_{\xi_0} \in V^{(i)}(m)) + o(1) \\ &= \tilde{p}(s_m, m) + o(1). \end{aligned}$$

The expression for  $\tilde{p}(s_m, m)$  in terms of the  $p^*(s_m, r)$  may be read off directly from the formula given in Proposition 3.4. For  $m' \neq m$ ,  $m' \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , we obtain in a similar manner

$$\begin{aligned} & \mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m')) \\ &= \mathbb{P}_{s_m}(Y_1^{(i)} = m') + \sum_{n \geq 1} \sum_{r_1, \dots, r_n \in N^{(i)}} \mathbb{P}_{s_m}(Y_1^{(i)} = r_1, \dots, Y_n^{(i)} = r_n, Y_{n+1}^{(i)} = m') + o(1) \\ &= \tilde{p}(s_m, m') + \sum_{n \geq 1} \sum_{r_1, \dots, r_n \in N^{(i)}} \tilde{p}(s_m, r_1) \cdots \tilde{p}(r_{n-1}, r_n) \tilde{p}(r_n, m') + o(1), \end{aligned}$$

the last line by another appeal to the afore-mentioned proposition.  $\square$

In essence, the previous result tells us that a valley  $V^{(i)}(m')$  is neighbored to  $V^{(i)}(m)$ , that is, reachable with positive probability by the asymptotic jump chain  $\tilde{Y}^{(i)}$  (and thus by  $Y^{(i)}$  at any temperature level  $\beta$ ) without intermediately hitting any other valley, iff there exists at least one (in terms of energies) decreasing path in  $N^{(i)}$  from  $s_m$  to  $m'$ . For any other such pair of valleys, connected by a path through states in  $N^{(i)}$ , the transition probability decreases to zero exponentially in  $\beta$ . If this path can be chosen to be unimodal, here called *uphill-downhill-path*, this can be stated in a very precise way as the next result shows.

**Lemma 3.9.** *Let  $m_0, m_1 \in \mathcal{S}^{(i)} \setminus N^{(i)}$  be two distinct local minima for some  $0 \leq i \leq \mathbf{n}$ . Suppose there exists a minimal path  $\gamma = (\gamma_0, \dots, \gamma_k)$  from  $s_{m_0}$  to  $m_1$  not hitting any other valley but  $V^{(i)}(m_1)$  and such that  $I(\gamma_0, \dots, \gamma_k) = E(z^*(s, m_1)) - E(s_{m_0})$ . Then*

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \ln \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) = -(E(z^*(m_0, m_1)) - E(s_{m_0})).$$

Without assuming the existence of  $\gamma$  as stated, the result remains valid when replacing  $=$  with  $\leq$ .

Note that  $I(\gamma_0, \dots, \gamma_k) = E(z^*(s_{m_0}, m_1)) - E(s_{m_0})$  does indeed imply the already mentioned property that

$$\begin{aligned} E(\gamma_i) &> E(\gamma_{i-1}) \quad \text{for } 1 \leq i \leq j \\ \text{and } E(\gamma_i) &< E(\gamma_{i-1}) \quad \text{for } j+1 \leq i \leq k \end{aligned}$$

if  $\gamma_j = z^*(s_{m_0}, m_1)$ . We call such a path an *uphill-downhill-path* because it first straddles the energy barrier  $E(z^*(s_{m_0}, m_1))$  and then falls down to the local minimum  $m_1$ . The existence of such a path can be found in most 2- or higher dimensional energy landscapes.

*Proof.* With  $\gamma$  as stated, a lower bound for  $\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1))$  is easily obtained as follows:

$$\begin{aligned} \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) &\geq \mathbb{P}_{m_0}(X_{\zeta_0+i} = \gamma_i, 0 \leq i \leq k) \\ &\geq \mathbb{P}_{m_0}(X_{\zeta_0} = s_{m_0}) e^{-\beta I(\gamma_0, \dots, \gamma_k) - \gamma_{\beta} \beta |\mathcal{S}|} \\ &= (1 + o(1)) e^{-\beta(E(z^*(s_{m_0}, m_1)) - E(s_{m_0})) - \gamma_{\beta} \beta |\mathcal{S}|} \\ &= (1 + o(1)) e^{-\beta(E(z^*(m_0, m_1)) - E(s_{m_0})) - \gamma_{\beta} \beta |\mathcal{S}|}. \end{aligned}$$

For an upper bound, which does not require the existence of a  $\gamma$  as claimed, we decompose the event into disjoint sets depending on the number of visits  $N$ , say, to  $m_0$  between 1 and  $\zeta_0 = \zeta_0^{(i)}$ . This leads to

$$\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N = 0) = \mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})$$

and, for  $k \geq 1$ ,

$$\begin{aligned} &\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N = k) \\ &= \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), |\{\tau_{m_0} < n \leq \zeta_0 \mid X_n = m_0\}| = k - 1, \tau_{m_0} < \zeta_0) \\ &= \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N = k - 1) \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0) \\ &\vdots \\ &= \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N = 0) \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0)^k \\ &= \mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0}) \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0)^k. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) &= \sum_{k \geq 0} \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1), N = k) \\ &= \sum_{k \geq 0} \mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0}) \mathbb{P}_{m_0}(\tau_{m_0} < \zeta_0)^k \\ &= \frac{\mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})}{\mathbb{P}_{m_0}(\zeta_0 < \tau_{m_0})}. \end{aligned}$$

By invoking Proposition 2.4, we infer

$$(12) \quad \frac{\mathbb{P}_{m_0}(\xi_1 = \tau_{V^{(i)}(m_1)} < \tau_{m_0})}{\mathbb{P}_{m_0}(\zeta_0 < \tau_{m_0})} \leq \frac{\sum_{r \in V^{(i)}(m_1)} \mathbb{P}_{m_0}(\tau_r < \tau_{m_0})}{\mathbb{P}_{m_0}(\tau_x < \tau_{m_0})} \leq K(\beta) \sum_{r \in V^{(i)}(m_1)} e^{-\beta(E(z^*(m_0, r)) - E(z^*(m_0, x)) - \tau\gamma_\beta)}$$

for all  $x \in V^{(i)}(m_0)^c$ , where

$$K(\beta) = |\mathcal{S}| \left( |\mathcal{S}| \exp \left( -\beta \min_{a \neq b: E(a) > E(b)} (E(a) - E(b)) + 2\gamma_\beta \beta \right) + 1 \right) \max_{r \in \mathcal{S}} |\mathcal{N}(r)|.$$

For any  $r \in V^{(i)}(m_1)$ , we have  $E(z^*(m_0, r)) \geq E(z^*(m_1, r))$  and therefore

$$E(z^*(m_0, m_1)) \leq E(z^*(m_0, r)) \vee E(z^*(r, m_1)) = E(z^*(m_0, r)).$$

Using this in (12), we obtain

$$\frac{\mathbb{P}_{m_0}(\xi_1^{(i)} = \tau_{V^{(i)}(m_1)} < \tau_{m_0})}{\mathbb{P}_{m_0}(\zeta_0 < \tau_{m_0})} \leq K(\beta) |\mathcal{S}| e^{-\beta(E(z^*(m_0, m_1)) - E(z^*(m_0, x)) - \tau\gamma_\beta)}$$

and then, upon choosing  $x = s_{m_0}$  and noting that  $E(z^*(m_0, s_{m_0})) = E(s_{m_0})$ ,

$$\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) \leq K(\beta) |\mathcal{S}| e^{-\beta(E(z^*(m_0, m_1)) - E(s_{m_0}) - \tau\gamma_\beta)}.$$

By combining all previous results, we finally conclude

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \ln \mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) = -(E(z^*(m_0, m_1)) - E(s_{m_0}))$$

as asserted.  $\square$

To summarize, which valleys are visited consecutively depends on (a) their spatial arrangement and (b) the energy barriers between them: A transition from one valley  $V^{(i)}(m_0)$  to another valley  $V^{(i)}(m_1)$  is only possible, if there exists a path from  $s_{m_0}$  to  $V^{(i)}(m_1)$ , not hitting any other valley. This transition is made at small temperatures (i.e. large  $\beta$ ) if the additional energy barrier  $E(z^*(s_{m_0}, m_1)) - E(s_{m_0})$  is sufficiently small or in other words the energy barrier  $E(z^*(s_{m_0}, m_1))$  is approximately of the same height as all other energy barriers, including the barrier  $E(z^*(s_{m_0}, m_0)) = E(s_{m_0})$ .

A result similar to the previous lemma holds true for transitions from  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$  to any  $s \in \partial^+ V(m)$ .

**Lemma 3.10.** *Let  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and  $s \in \partial^+ V^{(i)}(m)$ . Then*

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \ln \mathbb{P}_m(Y_1 = s) = -(E(s) - E(s_m)).$$

*Proof.* For the proof, decompose again the event  $\{Y_1 = s\}$  with respect to the number of visits to  $m$  before  $V(m) = V^{(i)}(m)$  is left (or use Proposition 2.5), giving

$$\mathbb{P}_m(Y_1 = s) = \frac{\mathbb{P}_m(\sigma_1 = \tau_s < \tau_m)}{\mathbb{P}_m(\sigma_1 < \tau_m)}.$$

For an upper bound, we obtain by another appeal to Proposition 2.4 that

$$\mathbb{P}_m(Y_1 = s) = \frac{\mathbb{P}_m(\sigma_1 = \tau_s < \tau_m)}{\mathbb{P}_m(\sigma_1 < \tau_m)} \leq \frac{\mathbb{P}_m(\tau_s < \tau_m)}{\mathbb{P}_m(\tau_{s_m} < \tau_m)} \leq e^{-\beta(E(s) - E(s_m) + o(1))}.$$

For a lower bound, note that, using the strong Markov property,

$$\mathbb{P}_m(\tau_s < \tau_m < \tau_{\partial^+ V(m) \setminus \{s\}}) = \mathbb{P}_m(\tau_s < \tau_m) \mathbb{P}_s(\tau_m < \tau_{\partial^+ V(m) \setminus \{s\}}),$$

where

$$\begin{aligned} \mathbb{P}_s(\tau_m < \tau_{\partial+V(m)\setminus\{s\}}) &\geq \sum_{r \sim s, r \in V(m)} p(s, r) \mathbb{P}_r(\tau_m < \tau_{\partial+V(m)\setminus\{s\}}) \\ &\geq \sum_{r \sim s, r \in V(m)} p(s, r) \mathbb{P}_r(\tau_m < \sigma_1). \end{aligned}$$

Since there has to be at least one state  $r \sim s, r \in V(m)$ , with  $E(r) < E(s)$ , and  $\mathbb{P}_r(\tau_m < \sigma_1) \rightarrow 1$  as  $\beta \rightarrow \infty$ , this last sum is bounded away from 0. Thus,

$$\mathbb{P}_m(\sigma_1 = \tau_s < \tau_m) \geq \mathbb{P}_m(\tau_s < \tau_m < \tau_{\partial+V(m)\setminus\{s\}}) \geq e^{-\beta(E(s)-E(m)+o(1))}.$$

Finally, since

$$\begin{aligned} \mathbb{P}_m(\sigma_1 < \tau_m) &\leq \sum_{r \in \partial+V(m)} \mathbb{P}_m(\tau_r < \tau_m) \\ &\leq \sum_{r \in \partial+V(m)} e^{-\beta(E(r)-E(m)+o(1))} \\ &= e^{-\beta(E(s_m)-E(m)+o(1))}, \end{aligned}$$

we conclude

$$\mathbb{P}_m(Y_1 = s) \geq e^{-\beta(E(s)-E(s_m)+o(1))}$$

and thus the assertion.  $\square$

Now we see for the reciprocating jumps in the accelerated chain:

**Proposition 3.11.** *Fix  $1 \leq i \leq \mathbf{n}$  and  $\varepsilon > 0$ . Then the AAC at level  $i$  exhibits no reciprocating jumps of order  $\varepsilon$  if the following three conditions hold true:*

- (1)  $E(z^*(m_0, m_1)) - E(s_{m_0}) \leq \varepsilon$  for all distinct  $m_0, m_1 \in \mathcal{S}^{(i)} \setminus N^{(i)}$ .
- (2) For each  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , there exist at least two distinct  $m_1, m_2 \in \mathcal{S}^{(i)} \setminus N^{(i)}$ ,  $m \neq m_1, m_2$ , such that  $\mathbb{P}_m(X_{\xi_1} \in V^{(i)}(m_j)) > 0$  for  $j = 1, 2$ .
- (3) For each pair  $m_0, m_1 \in \mathcal{S}^{(i)} \setminus N^{(i)}$  with  $\mathbb{P}_{m_0}(X_{\xi_1} \in V^{(i)}(m_1)) > 0$ , there exists a minimal uphill-downhill-path from  $s_{m_0}$  to  $m_1$  not hitting any valley but  $V^{(i)}(m_1)$ .

The origin of our endeavor to define aggregations with no reciprocating jumps of an order larger than a small  $\varepsilon$  is to obtain an associated process with (almost) decorrelated increments (in Euclidean state space), for this and a proper centering causes the variance up to the  $n$ -th jump to grow with  $n$  instead of  $n^2$ . This is known as diffusive behavior in physics. Without aggregation increments are highly correlated due to the following argument: at any given time, the process is with high probability in a minimum and when leaving it, say by making a positive jump, the next increment is most likely negative because there is a drift back to the minimum. Likewise, the increments of the asymptotic jump chain are neither uncorrelated nor having mean zero since trajectories of  $\tilde{Y}$  on an irreducibility class are almost surely of the form  $m_1 \rightarrow s \rightarrow m_2 \rightarrow s \rightarrow m_3 \rightarrow \dots$ , where  $s \in N^{(i)}$  and  $m_1, m_2, \dots \in \mathcal{S}^{(i)} \setminus N^{(i)}$ . Thus, given the previous increments, it is in general easy to predict the next increment and they do not have mean zero. On the other hand, if we can choose  $\beta$  and  $i$  such that, for any  $m_0, m_1$ , we have  $\mathbb{P}_{m_0}(Y_1 = m_1) \in \{p_{m_0 \pm \varepsilon}\} \cup [0, \varepsilon]$  for  $\varepsilon \ll p_{m_0}$ , we obtain an AAC which behaves roughly like a RW on a graph. Such a RW is diffusive if we assume periodic boundary conditions (or sufficiently large state space compared to the observation time  $n$ ) and an energy landscape  $E$  which is homogeneous enough to ensure zero-mean increments. In particular  $\{m | \mathbb{P}_{m_0}(Y_1 = m) = p_{m_0} \pm \varepsilon\}$  has to comprise at least two states.

**3.3. Metabasins.** A path-independent definition of metabasins can now be given on the basis of the previous considerations.

**Definition 3.12.** A finite Markov chain  $X$  driven by an energy function  $E$  satisfying the assumptions stated at the beginning of Section 1 has *metabasins of order  $\varepsilon > 0$*  if there exists an aggregation level  $i < \mathfrak{n} - 1$  such that the following conditions are fulfilled for each  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ :

- (MB1)  $\sup_{m' \in \mathcal{S}^{(i)} \setminus (N^{(i)} \cup \{m\})} E(z^*(m, m')) - E(s_m) \leq \varepsilon$ .
- (MB2) There are at least two distinct  $m_1, m_2 \in \mathcal{S}^{(i)} \setminus (N^{(i)} \cup \{m\})$  with a minimal uphill-downhill-path from  $s_m$  to  $m_k$  not hitting any other valley but  $V^{(i)}(m_k)$  for  $k = 1, 2$ .

In this case, the valleys  $(V^{(i)}(m))_{m \in \mathcal{S}^{(i)}}$  are called *metabasins (MB) of order  $\varepsilon$* .

The reader should notice that each singleton set  $\{s\}$  consisting of a non-assigned state  $s \in \mathcal{N}^{(i)}$  forms a MB. The conditions (MB1) and (MB2) ensure the good nature of (a) the energy barriers and (b) the spatial arrangement of minima. As already pointed out, this determines which valleys are visited consecutively. Properties of MB which can be concluded from the results of the previous sections are summarized in the next theorem. The reader is reminded of Properties 1–4 stated in the Introduction.

**Theorem 3.13.** *For MB as defined in Definition 3.12 we have*

- (1) *The transition probabilities for jumps between MB do not depend on the point of entrance as  $\beta \rightarrow \infty$  (Property 1).*
- (2) *There are no reciprocating jumps of order  $\varepsilon$  (Property 2).*
- (3) *The expected residence time in a MB depends on  $E$  only via the depth of the MB as  $\beta \rightarrow \infty$  (Property 3).*
- (4) *Regarding only MB pertaining to local minima, the system is a trap model (Property 4).*

*Proof.* (1) follows from Proposition 3.4, (2) from Proposition 3.11, (3) from Theorem 2.8, and (4) directly from the definition.  $\square$

It should not be surprising that the path-dependent definition of MB by HEUER [12] and stated in the Introduction differs from our path-independent one. For example,

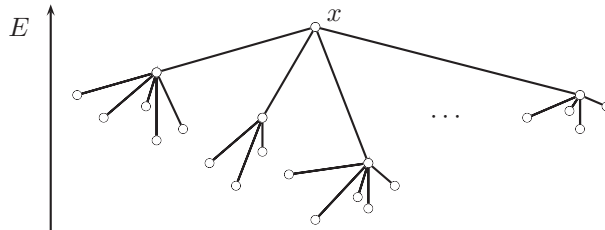


FIGURE 5. Example of an energy landscape with a tree-like structure.

the energy landscape in Figure 5 has no reasonable path-dependent MB because every transition between two branches of the shown tree must pass through the state  $x$ . For a typical trajectory, there will be at most three MB: the states visited before the first occurrence of  $x$ , the states visited between the first and the last occurrence of  $x$ , and the states visited after the last occurrence of  $x$ . The reason for this poor performance is the tree-like structure of the energy landscape or, more generally, the fact that the

connectivity between the branches is too small to allow a selfavoiding walk through more than two branches. This results in a small recurrence time for  $x$  (compared to the number of states visited in between). However, every branch constitutes a MB when using the path-independent definition for sufficiently small  $\varepsilon$ , in which case the AAC forms a Markov chain and, given the Metropolis algorithm, even a RW on the graph.

Having thus exemplified that the two definitions of MB do not necessarily coincide, where the path-independent approach applies to a wider class of energy landscapes, we turn to the question about conditions for them to coincide with high probability. As already pointed out, we have to assume a sufficient connectivity between the metastates to ensure the existence of reasonable path-dependent MB. In terms of this connectivity (for a precise definition see Definition 3.14) and the parameter  $\beta$  and  $\varepsilon$ , our last result, Theorem 3.17 below, provides lower bounds for the probability that both definitions yield the same partition of the state space.

The first step towards this end is to identify and count, for each  $m \in \mathcal{S}^{(i)}$  and a given  $\beta$ , the states  $s \in \mathcal{S}^{(i)}$  for which a transition of  $Y$  from  $m$  to  $s$  is likely. This leads to the announced connectivity parameter.

**Definition 3.14.** Let  $\varepsilon > 0$  and suppose that  $X$  has MB of order  $\varepsilon > 0$  at level  $i$ . Define the *connectivity parameters*  $\eta_i = \eta_{i,\varepsilon}$ ,  $1 \leq i \leq 3$ , by

$$(13) \quad \begin{aligned} \eta_{1,\varepsilon} &:= \min_{n \in N^{(i)}, r \in \mathcal{S}^{(i)}: V^{(i)}(r) \cap \mathcal{N}(n) \neq \emptyset} \left| \left\{ s \in \mathcal{N}(n) \setminus V^{(i)}(r) \mid E(s) \leq E(n) + \varepsilon \right\} \right|, \\ \eta_{2,\varepsilon} &:= \min_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \left| \left\{ s \in \partial^+ V^{(i)}(m) \mid E(s) \leq E(s_m) + \varepsilon \right\} \right|, \\ \eta_{3,\varepsilon} &:= \min_{n \in N^{(i)}} \left| \left\{ s \in \mathcal{S}^{(i)} \mid E(x) \leq E(n) + \varepsilon \text{ for some } x \in V^{(i)}(s) \cap \mathcal{N}(n) \right\} \right|. \end{aligned}$$

$\eta_1$  is the minimal number of neighboring sites of a non-assigned state  $n$  which do not belong to a particular neighboring valley and whose energy is at most  $\varepsilon$  plus the energy of  $n$ .  $\eta_2$  is the minimal number of neighboring sites/valleys of a valley  $V^{(i)}(m)$  whose energy is at most  $\varepsilon$  plus the energy of  $s_m$  and which can be reached via an uphill-path from  $m$ . Finally,  $\eta_3$  is the minimal number of neighboring valleys of a non-assigned state  $n$  which comprise a state with energy of at most  $\varepsilon$  plus the energy of  $n$ .  $\eta_1$  and  $\eta_3$  are always at least 2 and in fact quite large in the very complex energy landscapes of structural glasses. For very small  $\varepsilon$ ,  $\eta_2$  may be 1, but if  $X$  has MB of order  $\varepsilon$  in a high dimensional energy landscape, then  $\eta_2$  can be assumed to be quite large as well.

That transitions to states counted above have reasonable large probabilities is content of the following lemma. Thus, the defined parameters do in fact measure the connectivity of the MB.

**Lemma 3.15.** Let  $\varepsilon > 0$  and suppose that  $X$  has MB of order  $\varepsilon > 0$  at level  $i$  with connectivity parameters defined in (13). Writing  $Y_k$  for  $Y_k^{(i)}$  and  $V(m)$  for  $V^{(i)}(m)$ ,  $m \in \mathcal{S}^{(i)}$ , the following assertions hold true for all sufficiently large  $\beta$ :

- (a) If  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and  $s \in \partial^+ V(m) \cap \{x \mid E(x) - E(s_m) \leq \varepsilon\}$ , or  $m \in N^{(i)}$  and  $s \in \mathcal{S}^{(i)}$  satisfies  $V(s) \cap \{x \in \mathcal{N}(m) \mid E(x) - E(m) \leq \varepsilon\} \neq \emptyset$ , then

$$\mathbb{P}_m(Y_1 = s) \geq e^{-2\beta\varepsilon}.$$

- (b) For any distinct  $m \in N^{(i)}$  and  $s \in \mathcal{S}^{(i)}$ ,

$$\mathbb{P}_m(Y_1 \neq s) \geq \eta_1 e^{-2\beta\varepsilon},$$

- (c) For any distinct  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and  $s \in \mathcal{S}^{(i)}$ ,

$$\mathbb{P}_m(Y_1 \neq s) \geq (\eta_2 - 1) e^{-2\beta\varepsilon},$$



We see that, for  $\varepsilon$  small enough compared to  $\beta$ , transitions with an energy barrier of at most  $\varepsilon$  are still quite likely and thus a jump to a particular valley quite unlikely in the case of high connectivity.

*Proof.* (a) Choose  $\beta_0 > 0$  so large that, for  $\beta \geq \beta_0$ ,  $\gamma_\beta \leq \varepsilon$  and  $\mathbb{P}_m(Y_1 = s) \geq e^{-2\beta(E(s) - E(s_m))}$  for any  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$  and  $s \in \partial^+ V(m)$ , the latter being possible by Lemma 3.10. Then for any such  $m$  and  $s$ , we infer  $\mathbb{P}_m(Y_1 = s) \geq e^{-2\varepsilon\beta}$  provided that additionally  $E(s) \leq E(s_m) + \varepsilon$  holds true. If  $m \in N^{(i)}$ , then  $\mathbb{P}_m(Y_1 = s) \geq e^{-2\varepsilon\beta}$  for any  $s \in \mathcal{S}^{(i)}$  such that  $E(x) \leq E(s) + \varepsilon$  for some  $x \in V(s) \cap \mathcal{N}(m)$ , for

$$\mathbb{P}_m(Y_1 = s) \geq \mathbb{P}_m(X_{\sigma_1} = x) = p(m, x) \geq e^{-\beta((E(x) - E(m))^+ + \gamma_\beta)}.$$

(b) Pick again  $\beta_0$  so large that  $\gamma_\beta \leq \varepsilon$  for all  $\beta \geq \beta_0$ . Then,

$$\begin{aligned} \mathbb{P}_m(Y_1 \neq s) &\geq \sum_{x \in \mathcal{N}(m), x \notin V(s)} p(m, x) \\ &\geq \sum_{x \in \mathcal{N}(m), x \notin V(s), E(x) \leq E(m) + \varepsilon} \exp(-\beta((E(x) - E(m))^+ + \gamma_\beta)) \\ &\geq \eta_1 \exp(-2\beta\varepsilon), \end{aligned}$$

by definition of  $\eta_1$ .

(c) Fix  $\beta_0$  so large that  $\mathbb{P}_m(Y_1 = x) \geq e^{-2\beta(E(x) - E(s_m))}$  for any  $x \in \partial^+ V(m)$  and  $\beta \geq \beta_0$ . In the very same way as in part (b), we then get for all  $\beta \geq \beta_0$

$$\begin{aligned} \mathbb{P}_m(Y_1 \neq s) &\geq \sum_{x \in \partial^+ V(m), x \neq s} \mathbb{P}_m(Y_1 = x) \\ &\geq \sum_{x \in \partial^+ V(m), x \neq s, E(x) \leq E(s_m) + \varepsilon} \exp(-2\beta(E(x) - E(s_m))) \\ &\geq (\eta_2 - 1) \exp(-2\beta\varepsilon), \end{aligned}$$

by definition of  $\eta_2$ . □

The above result motivates that in the case of high connectivity the probability to revisit a particular valley within a fixed time  $T$  is quite small, or in other words, the probability for the AAC to jump along a selfavoiding path is quite high. This is the main step towards the announced theorem and stated below. The observation time  $T$  of course has to be small compared to the cover time of the process.

**Lemma 3.16.** *Let  $\varepsilon > 0$  and suppose that  $X$  has MB of order  $\varepsilon > 0$  at level  $i$  with connectivity parameters defined in (13). Writing  $Y_k$  for  $Y_k^{(i)}$  and  $V(m)$  for  $V^{(i)}(m)$ ,  $m \in \mathcal{S}^{(i)}$ , define*

$$\tau_{V(m)}^{(i)} := \inf\{k \geq 1 | Y_k = m\}.$$

*Then the following assertions hold true for all sufficiently large  $\beta$ :*

(a) *For any  $0 < \delta < 1 - \mathbb{P}_m(Y_2 = m)$  and  $1 \leq T \leq T(m, \beta) + 1$ ,*

$$\mathbb{P}_m\left(\tau_{V(m)}^{(i)} > T\right) \geq \delta,$$

*where*

$$T(m, \beta) := \frac{\ln \delta}{\ln(\min_{m' \neq m} \mathbb{P}_{m'}(Y_1 \neq m)(1 - \mathbf{1}_{\{m' \notin N^{(i)}\}} \delta(m', \beta)))}$$

*and*

$$\delta(m', \beta) := \max_{x \in V(m')} \sum_{z \in \partial^+ V(m')} \tilde{\varepsilon}(x, m', z, \beta).$$

In particular, if  $\delta \leq ((\eta_1 \wedge (\eta_2 - 2))e^{-2\beta\varepsilon})^T$  for some  $T > 0$ , then  $T(m, \beta) \geq T$ .  
 (b) For each  $k \geq 1$  and  $m_0 \in \mathcal{S}^{(i)}$ ,

$$\sum_{m_1, \dots, m_k} \prod_{j=0}^{k-1} \mathbb{P}_{m_j}(Y_1 = m_{j+1}) \geq [\eta_2 \wedge \eta_3]_k e^{-2k\varepsilon\beta}$$

where summation ranges over all pairwise distinct  $m_1, \dots, m_k \in \mathcal{S}^{(i)} \setminus \{m_0\}$  and for  $N \in \mathbb{N}$  we write  $[N]_k := N(N-1) \cdot \dots \cdot (N-k+1)$ .

It should be noticed that  $\mathbb{P}_m(\tau_{V(m)}^{(i)} > 1) = 1$  (the AAC never stays put) and

$$\mathbb{P}_m(\tau_{V(m)}^{(i)} > T) \leq \mathbb{P}_m(\tau_{V(m)}^{(i)} > 2) = 1 - \mathbb{P}_m(Y_2^{(i)} = m)$$

for every  $T \geq 2$  with equality holding only if  $T = 2$ . We thus see that  $\mathbb{P}_m(\tau_{V(m)}^{(i)} > T) \geq \delta$  entails  $\delta < 1 - \mathbb{P}_m(Y_2 = m)$ , the latter being typically large. Furthermore, the bound on the number of self-avoiding path of length  $k$  is very crude and can be improved when knowing more about the spatial arrangement of the metastable states.

*Proof.* (a) Recall from the first part of the proof of Proposition 3.4 that

$$\begin{aligned} \mathbb{P}_m(Y_{n+1} \neq z | Y_n = y, X_{\sigma_n} = x) &\geq \mathbb{P}_y(Y_1 \neq z) \left( 1 - \mathbf{1}_{\{y \notin N^{(i)}\}} \sum_{r \in \partial^+ V(y)} \mathbb{P}_x(\tau_r < \tau_y) \right) \\ &\geq \mathbb{P}_y(Y_1 \neq z) \left( 1 - \mathbf{1}_{\{y \notin N^{(i)}\}} \sum_{r \in \partial^+ V(y)} \tilde{\varepsilon}(x, y, r, \beta) \right) \\ &\geq \mathbb{P}_y(Y_1 \neq z) (1 - \mathbf{1}_{\{y \notin N^{(i)}\}} \delta(y, \beta)) \end{aligned}$$

holds true for all  $y, z \in \mathcal{S}^{(i)}$ ,  $x \in V(y)$  and  $\beta > 0$ . This will now be used repeatedly to show that

$$\mathbb{P}_m(\tau_{V(m)}^{(i)} > T) \geq \left( \min_{m' \neq m} \mathbb{P}_{m'}(Y_1 \neq m) (1 - \mathbf{1}_{\{m' \notin N^{(i)}\}} \delta(m', \beta)) \right)^{T-1}$$

for each  $T > 2$ . Putting  $\mathbf{m}(x) := m'$  if  $x \in V(m')$  for  $m' \in \mathcal{S}^{(i)}$ , we obtain

$$\begin{aligned}
 & \mathbb{P}_m \left( \tau_{V(m)}^{(i)} > T \right) \\
 &= \mathbb{P}_m (Y_1 \neq m, \dots, Y_T \neq m) \\
 &= \sum_{x_1, \dots, x_{T-1} \notin V^{(i)}(m)} \mathbb{P}_m (X_{\sigma_1} = x_1) \prod_{k=1}^{T-2} \mathbb{P}_m (X_{\sigma_{k+1}} = x_{k+1} | X_{\sigma_k} = x_k, Y_k = \mathbf{m}(x_k)) \\
 &\quad \times \mathbb{P}_m (Y_T \neq m | X_{\sigma_{T-1}} = x_{T-1}, Y_{T-1} = \mathbf{m}(x_{T-1})) \\
 &\geq \sum_{x_1, \dots, x_{T-1} \notin V^{(i)}(m)} \mathbb{P}_m (X_{\sigma_1} = x_1) \prod_{k=1}^{T-2} \mathbb{P}_m (X_{\sigma_{k+1}} = x_{k+1} | X_{\sigma_k} = x_k, Y_k = \mathbf{m}(x_k)) \\
 &\quad \times \left( \min_{m' \neq m} (\mathbb{P}_{m'} (Y_1 \neq m) (1 - \mathbb{1}_{\{m' \notin N^{(i)}\}} \delta(m', \beta))) \right) \\
 &\quad \vdots \\
 &\geq \min_{m' \neq m} \left( \mathbb{P}_{m'} (Y_1 \neq m) (1 - \mathbb{1}_{\{m' \notin N^{(i)}\}} \delta(m', \beta)) \right)^{T-1}.
 \end{aligned}$$

But this establishes the asserted inequality when finally observing that the last expression is  $\geq \delta$  iff  $T \leq T(m, \beta) + 1$ .

Having just said that  $T(m, \beta) \geq T$  holds iff

$$\min_{m' \neq m} \left( \mathbb{P}_{m'} (Y_1 \neq m) (1 - \mathbb{1}_{\{m' \notin N^{(i)}\}} \delta(m', \beta)) \right)^T \geq \delta,$$

it suffices to note that, as  $\beta \rightarrow \infty$ ,  $\delta(m', \beta) \rightarrow 0$  holds true if  $m' \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , giving

$$1 - \delta(m', \beta) \geq \frac{\eta_2 - 2}{\eta_2 - 1}$$

for sufficiently large  $\beta$ . Together with Lemma 3.15(b), this further yields

$$\min_{m' \neq m} \left( \mathbb{P}_{m'} (Y_1 \neq m) (1 - \mathbb{1}_{\{m' \notin N^{(i)}\}} \delta(m', \beta)) \right)^T \geq ((\eta_1 \wedge (\eta_2 - 2)) e^{-2\beta\varepsilon})^T$$

and then the assertion.

(c) Here it suffices to notice that, by (a),  $[\eta_2 \wedge \eta_3]_k$  forms a lower bound for the number of self-avoiding paths  $(m_0, \dots, m_k)$  such that  $\mathbb{P}_{m_j} (Y_1 = m_{j+1}) \geq e^{-2\beta\varepsilon}$  for each  $j = 0, \dots, k-1$ .  $\square$

We proceed with the announced result about the relation between path-dependent and path-independent MB. To this end, we fix  $T = \sigma_K$  for some  $K \in \mathbb{N}$ . Let  $\mathcal{V}_k$  for  $k = 1, \dots, v$  denote the random number of MB obtained from  $X_0, \dots, X_T$  as defined in the Introduction. For  $x \in \mathcal{S}$ , we further let  $\mathcal{V}(x)$  denote the MB  $\mathcal{V}_k$  containing  $x$  and put  $\mathcal{V}(x) := \emptyset$  if no such MB exists which is the case iff  $x \notin \{X_0, \dots, X_T\}$ .

**Theorem 3.17.** *Let  $\varepsilon > 0$  and suppose that  $X$  has MB of order  $\varepsilon > 0$  at level  $i$  with connectivity parameters defined in (13). Fix  $K \in \mathbb{N}$ ,  $T = \sigma_K$  and  $0 < \delta \leq ((\eta_1 \wedge (\eta_2 - 1) - 1) e^{-2\beta\varepsilon})^K$ . Then, for each  $0 \leq k < K$  and  $m_0 \in \mathcal{S}^{(i)}$ , there exists  $\beta_0 > 0$  such that for all  $\beta \geq \beta_0$*

$$\text{(a) } \mathbb{P}_{m_0} \left( V_{<}^{(i)}(Y_k) \subseteq \mathcal{V}(Y_k) \right) \geq 1 - (\max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} |V_{<}(m)| + 2) \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m, \beta),$$

where  $V_{<}^{(i)}(s) := \{s\}$  if  $s \in N^{(i)}$ .

$$\text{(b) } \mathbb{P}_{m_0} (\mathcal{V}(Y_j) \subseteq V^{(i)}(Y_j), 0 \leq j < k) \geq 1 - k(\max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m, \beta) + (1 - \delta)).$$

(c) If  $\eta_2 \wedge \eta_3 > K - 1$ , then

$$\mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subseteq V^{(i)}(Y_j), 0 \leq j \leq K-1) \geq [\eta_2 \wedge \eta_3]_K \left(1 - \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m, \beta)\right)^{K-1} e^{-2K\varepsilon\beta}.$$

For the occurring bounds to be significant, two requirements must be met. First,  $K$  must be small compared to the cover time of the AAC and  $\varepsilon$  must be small compared to  $\beta_0$  to ensure  $\exp(-2\beta\varepsilon) \gg 0$ . Second, the connectivity must be high to ensure  $1 - \delta \ll 1$  and  $[\eta_2 \wedge \eta_3]_K e^{-2K\varepsilon\beta} \gg 0$ .

Typically, the inclusions in parts (b) and (c) are strict because of high energy states within a valley that will probably be missed during one simulation run and therefore not belong to any path-dependent MB. On the other hand, since our approach strives to cover the state space as completely as possible by valleys the latter comprise such high energy states whenever they are assignable in the sense described in Section 1.

*Proof.* With  $i$  being fixed, let us write as earlier  $V(m)$  for  $V^{(i)}(m)$ , and also  $V_{<}(m)$  for  $V_{<}^{(i)}(m)$ .

(a) For a given  $0 \leq k < K$ , define

$$\begin{aligned} A_k &:= \{\sigma_k \leq \tau_{Y_k} < \sigma_{k+1}\}, \\ B_k &:= \{\text{for every } x \in V_{<}(m) \text{ exists } \tau_{Y_k} \leq l_x < \sigma_k \text{ such that } X_l = x\}, \\ C_k &:= \{X_l = Y_k \text{ for some } \max_{x \in V_{<}(Y_k)} \tau_x \leq l < \sigma_k\}. \end{aligned}$$

With  $\delta(m, \beta)$  as defined in Lemma 3.16 and using

$$(14) \quad \mathbb{P}_x(\sigma_1 < \tau_m) \leq \sum_{y \in \partial^+ V(m)} \tilde{\varepsilon}(x, m, y, \beta) \leq \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} \delta(m, \beta) =: \delta_{\max}$$

for  $x \in V(m)$ ,  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , we obtain

$$\begin{aligned} &\mathbb{P}_{m_0} \left( V_{<}^{(i)}(Y_k) \subseteq \mathcal{V}(Y_k) \right) \\ &\geq \mathbb{P}_{m_0}(A_k \cap B_k \cap C_k) \\ &= \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(\{X_{\sigma_k} = r\} \cap A_k \cap B_k \cap C_k) \\ &= \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(X_{\sigma_k} = r) \mathbb{P}_{m_0}(A_k | X_{\sigma_k} = r) \mathbb{P}_{m_0}(B_k \cap C_k | \{X_{\sigma_k} = r\} \cap A_k) \\ &= \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(X_{\sigma_k} = r) \mathbb{P}_r(\tau_m < \sigma_1) \\ &\quad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_{<}(m), X_l = m \text{ for some } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1) \\ &\geq \sum_{m \in \mathcal{S}^{(i)}, r \in V(m)} \mathbb{P}_{m_0}(X_{\sigma_k} = r) (1 - \delta_{\max}) \\ &\quad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_{<}(m), X_l = m \text{ for some } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1) \\ &= (1 - \delta_{\max}) \sum_{m \in \mathcal{S}^{(i)}} \mathbb{P}_{m_0}(Y_k = m) \\ (15) \quad &\quad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_{<}(m), X_l = m \text{ for some } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1). \end{aligned}$$

Thus, in order to show that with high probability a path-dependent MB comprises the inner part of a valley, we show that with high probability, when starting in its minimum, the whole inner part will be visited and the process will return to the minimum once more before the valley is left. This is trivial if  $m \in N^{(i)}$  and thus  $V_{<}^{(i)}(m) = \{m\}$ , for then

$$\mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_{<}(m), X_l = m \text{ for some } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1) = 1.$$

More needs to be done if  $m \in \mathcal{S}^{(i)} \setminus N^{(i)}$ , where

$$\begin{aligned} & \mathbb{P}_m \left( \tau_x < \sigma_1 \text{ for every } x \in V_{<}(m), X_l = m \text{ for some } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1 \right) \\ & \geq 1 - \mathbb{P}_m(\tau_x > \sigma_1 \text{ for some } x \in V_{<}(m)) - \mathbb{P}_m(X_l \neq m \text{ for each } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1). \end{aligned}$$

The second probability in the preceding line can further be bounded with the help of (14), viz.

$$\begin{aligned} & \mathbb{P}_m(X_l \neq m \text{ for each } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1) \\ & = \sum_{y \in V_{<}(m)} \mathbb{P}_m(\max_{x \in V_{<}(m)} \tau_x = \tau_y, X_l \neq m \text{ for every } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1) \\ & \leq \sum_{y \in V_{<}(m)} \mathbb{P}_m(\max_{x \in V_{<}(m)} \tau_x = \tau_y) \mathbb{P}_y(\tau_m > \sigma_1) \\ & \leq \delta_{\max}, \end{aligned}$$

while for the first probability, we obtain with the help of Theorem 2.1

$$\begin{aligned} \mathbb{P}_m(\tau_x > \sigma_1 \text{ for some } x \in V_{<}(m)) & \leq \sum_{x \in V_{<}(m)} \mathbb{P}_m(\sigma_1 < \tau_x) \\ & \leq \sum_{x \in V_{<}(m)} \sum_{y \in \partial^+ V(m)} \mathbb{P}_m(\tau_y < \tau_x) \\ (16) \qquad \qquad \qquad & \leq \sum_{x \in V_{<}(m)} \sum_{y \in \partial^+ V(m)} \varepsilon(m, x, y, \beta), \end{aligned}$$

because  $E(z^*(m, y)) > E(z^*(m, x))$  for  $x \in V_{<}(m)$  and  $y \in \partial^+ V(m)$ . The latter can be seen as follows: It has been shown in the proof of Theorem 2.2 that  $E(z^*(x, m)) < E(z^*(x, y))$ . Hence,

$$E(z^*(x, m)) < E(z^*(x, y)) \leq E(z^*(x, m)) \vee E(z^*(y, m)) = E(z^*(y, m))$$

as asserted. Next, we infer

$$E(z^*(x, y)) \leq E(z^*(x, m)) \vee E(z^*(m, y)) \leq E(z^*(x, m)) \vee E(z^*(x, y)) = E(z^*(x, y)),$$

thus  $E(z^*(x, y)) = E(z^*(m, y))$ . Recalling the definition of  $\varepsilon(m, x, y, \beta)$ , the last equality implies  $\varepsilon(m, x, y, \beta) = \varepsilon(x, m, y, \beta)$  which will now be used to further bound the expression in (16), namely

$$\begin{aligned} \sum_{x \in V_{<}(m)} \sum_{y \in \partial^+ V(m)} \varepsilon(m, x, y, \beta) & = \sum_{x \in V_{<}(m)} \sum_{y \in \partial^+ V(m)} \varepsilon(x, m, y, \beta) \\ & = \sum_{x \in V_{<}(m)} \sum_{y \in \partial^+ V(m)} \tilde{\varepsilon}(x, m, y, \beta) \\ & \leq |V_{<}(m)| \delta_{\max} \\ & \leq \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} |V_{<}(m)| \delta_{\max}. \end{aligned}$$

Together with (15) this yields as asserted

$$\begin{aligned}
& \mathbb{P}_{m_0} \left( V_{<}^{(i)}(Y_k) \subseteq \mathcal{V}(Y_k) \right) \\
& \geq (1 - \delta_{\max}) \sum_{m \in \mathcal{S}^{(i)}} \mathbb{P}_{m_0}(Y_k = m) \\
& \quad \times \mathbb{P}_m(\tau_x < \sigma_1 \text{ for every } x \in V_{<}(m), X_l = m \text{ for some } \max_{x \in V_{<}(m)} \tau_x \leq l < \sigma_1) \\
& \geq (1 - \delta_{\max}) \left( 1 - \left( \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} |V_{<}(m)| + 1 \right) \delta_{\max} \right) \\
& \geq 1 - \left( \max_{m \in \mathcal{S}^{(i)} \setminus N^{(i)}} |V_{<}(m)| + 2 \right) \delta_{\max}.
\end{aligned}$$

(b) According to Lemma 3.16, choose  $\beta_0 > 0$  such that

$$(17) \quad \max_{m \in \mathcal{S}^{(i)}} \mathbb{P}_m \left( \tau_{V(m)}^{(i)} \leq K \right) \leq 1 - \delta$$

for each  $\beta \geq \beta_0$ . By using (17) and (14), we now infer

$$\begin{aligned}
& \mathbb{P}_{m_0}(Y_l = Y_k \text{ for some } k+1 \leq l \leq K) \\
& = \sum_{s \in \mathcal{S}} \mathbb{P}_{m_0}(Y_l = Y_k \text{ for some } k+1 \leq l \leq K, X_{\sigma_k} = s) \\
& \leq \sum_{s \in \mathcal{S}} \mathbb{P}_{m_0}(X_{\sigma_k} = s) \left( \mathbb{P}_s(\tau_{V(Y_0)}^{(i)} \leq K - k, X_j = \mathbf{m}(s) \text{ for some } 0 \leq j < \sigma_1) \right. \\
& \quad \left. + \mathbb{1}_{\{s \notin N^{(i)}\}} \mathbb{P}_s(\tau_{V(Y_0)}^{(i)} \leq K - k, X_j \neq \mathbf{m}(s) \text{ for all } 0 \leq j < \sigma_1) \right) \\
& \leq \sum_{s \in \mathcal{S}} \mathbb{P}_{m_0}(X_{\sigma_k} = s) \mathbb{P}_{\mathbf{m}(s)}(\tau_{V(Y_0)}^{(i)} \leq K - k) + \sum_{s \notin N^{(i)}} \mathbb{P}_{m_0}(X_{\sigma_k} = s) \mathbb{P}_s(\sigma_1 < \tau_{\mathbf{m}(s)}) \\
& \leq 1 - \delta + \delta_{\max},
\end{aligned}$$

and finally

$$\begin{aligned}
\mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subset V^{(i)}(Y_j), 0 \leq j < k) & \geq \mathbb{P}_{m_0} \left( \bigcap_{j=0}^{k-1} \{Y_l \neq Y_j, j+1 \leq l \leq K\} \right) \\
& \geq 1 - \sum_{j=0}^{k-1} \mathbb{P}_{m_0}(Y_l = Y_j \text{ for some } j+1 \leq l \leq K) \\
& \geq 1 - k(\delta_{\max} + (1 - \delta)).
\end{aligned}$$

(c) In the following calculation, let  $r_0 = m_0$ ,  $\sum_{m_j}$  range over all  $K$ -vectors  $(m_1, \dots, m_K)$  with pairwise distinct components in  $\mathcal{S}^{(i)} \setminus \{m_0\}$  and, for each  $k < K$ , let  $\sum_{r_1, \dots, r_k}$  range over all  $k$ -vectors  $(r_1, \dots, r_k)$  such that  $r_j \in V(m_j)$  for each  $j = 1, \dots, k$ . As in part (b), use (14) repeatedly to infer

$$\begin{aligned}
& \mathbb{P}_{m_0}(\mathcal{V}(Y_j) \subset V^{(i)}(Y_j), 0 \leq j \leq K-1) \\
& \geq \sum_{m_j} \sum_{r_1, \dots, r_{K-1}} \mathbb{P}_{m_0} \left( \bigcap_{j=0}^{K-1} \{Y_j = m_j, X_{\sigma_j} = r_j, \tau_{m_j} < \sigma_{j+1}\} \cap \{Y_K = m_K\} \right) \\
& = \sum_{m_j} \sum_{r_1, \dots, r_{K-1}} \prod_{j=0}^{K-2} \mathbb{P}_{r_j}(Y_0 = m_j, X_{\sigma_1} = r_{j+1}, \tau_{m_j} < \sigma_1) \\
& \quad \times \mathbb{P}_{r_{K-1}}(\tau_{m_{K-1}} < \sigma_1) \mathbb{P}_{m_{K-1}}(Y_1 = m_K)
\end{aligned}$$

$$\begin{aligned}
 &\geq (1 - \delta_{\max}) \sum_{m_j} \sum_{r_1, \dots, r_{K-1}} \prod_{j=0}^{K-2} \mathbb{P}_{r_j} (Y_0 = m_j, X_{\sigma_1} = r_{j+1}, \tau_{m_j} < \sigma_1) \mathbb{P}_{m_{K-1}} (Y_1 = m_K) \\
 &= (1 - \delta_{\max}) \sum_{m_j} \sum_{r_1, \dots, r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j} (Y_0 = m_j, X_{\sigma_1} = r_{j+1}, \tau_{m_j} < \sigma_1) \\
 &\quad \times \mathbb{P}_{r_{K-2}} (Y_0 = m_{K-2}, Y_1 = m_{K-1}, \tau_{m_{K-2}} < \sigma_1) \mathbb{P}_{m_{K-1}} (Y_1 = m_K) \\
 &= (1 - \delta_{\max}) \sum_{m_j} \sum_{r_1, \dots, r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j} (Y_0 = m_j, X_{\sigma_1} = r_{j+1}, \tau_{m_j} < \sigma_1) \\
 &\quad \times \mathbb{P}_{r_{K-2}} (Y_1 = m_{K-1}, \tau_{m_{K-2}} < \sigma_1) \mathbb{P}_{m_{K-1}} (Y_1 = m_K) \\
 &\geq (1 - \delta_{\max}) \sum_{m_j} \sum_{r_1, \dots, r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j} (Y_0 = m_j, X_{\sigma_1} = r_{j+1}, \tau_{m_j} < \sigma_1) \\
 &\quad \times \mathbb{P}_{m_{K-2}} (Y_1 = m_{K-1}) \mathbb{P}_{r_{K-2}} (\tau_{m_{K-2}} < \sigma_1) \mathbb{P}_{m_{K-1}} (Y_1 = m_K) \\
 &\geq (1 - \delta_{\max})^2 \sum_{m_j} \sum_{r_1, \dots, r_{K-2}} \prod_{j=0}^{K-3} \mathbb{P}_{r_j} (Y_0 = m_j, X_{\sigma_1} = r_{j+1}, \tau_{m_j} < \sigma_1) \\
 &\quad \times \mathbb{P}_{m_{K-2}} (Y_1 = m_{K-1}) \mathbb{P}_{m_{K-1}} (Y_1 = m_K) \\
 &\quad \vdots \\
 &\geq (1 - \delta_{\max})^{K-1} \sum_{m_j} \prod_{j=0}^{K-1} \mathbb{P}_{m_j} (Y_1 = m_{j+1}) \\
 &\geq (1 - \delta_{\max})^{K-1} [\eta_2 \wedge \eta_3]_K e^{-2K\varepsilon\beta},
 \end{aligned}$$

the last line following from Lemma 3.16.  $\square$

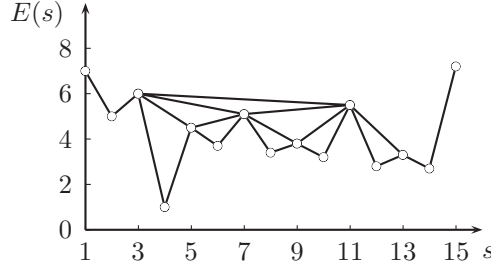


FIGURE 6. 2-dimensional modification of the energy landscape from Example 1.2.

**Example 3.18.** We return to Example 1.2 given in Section 1, but modify the energy landscape by allowing direct transitions between some saddles (see Figure 6) because (MB2) can clearly not be fulfilled in a one-dimensional model. While having no effect on the metastable states  $m \in M^{(i)}$ , valleys change in the way that, for levels  $i \in \{5, 6\}$ , the states  $\{1, 2, 3\}$  do no longer belong to the valley around state 4 and  $\{1, 2\}$  forms its own valley. The energy-differences  $\sup_{m' \in \mathcal{S}^{(i)} \setminus N^{(i)}} E(z^*(m, m')) - E(s_m)$  of the various metastable states  $m$  at each level  $1 \leq i \leq n = 7$  are shown in Figure 7. The supremum

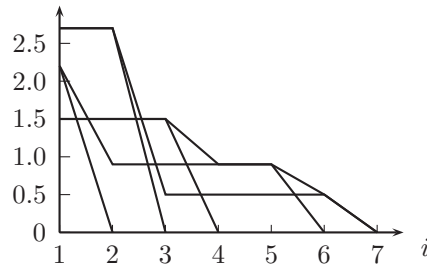


FIGURE 7.  $\sup_{m' \in \mathcal{S}^{(i)} \setminus N^{(i)}} |E(z^*(m, m')) - E(s_m)|$  for the various metastable states in  $\mathcal{S}^{(i)} \setminus N^{(i)}$  in dependence of the level  $1 \leq i \leq n$ .

of these energy differences decreases in  $i$ , and we obtain MB of order 1 for  $i \geq 4$ , and of order 0.5 for  $i \geq 6$ .

To illustrate the behavior, we have run a Metropolis Algorithm on this energy landscape. For initial state  $s = 4$  and  $\beta = 0.75$ , the energies of the trajectories of the original chain as well of the aggregated chain at levels  $i = 3, 4, 5$  are shown in Figure 8. The following observations are worth to be pointed out:

- The number of reciprocating jumps decreases with increasing level of aggregation.
- The deeper the valley, the longer the residence time.
- The motion in state space is well described by the aggregated process.
- Due to the very small size of the state space and a long simulation time, valleys are revisited.

#### ACKNOWLEDGMENT

We are very indebted to Andreas Heuer for sharing his insight about glass forming structures with us and also for his advice and many stimulating discussions that helped to improve the presentation of this article.



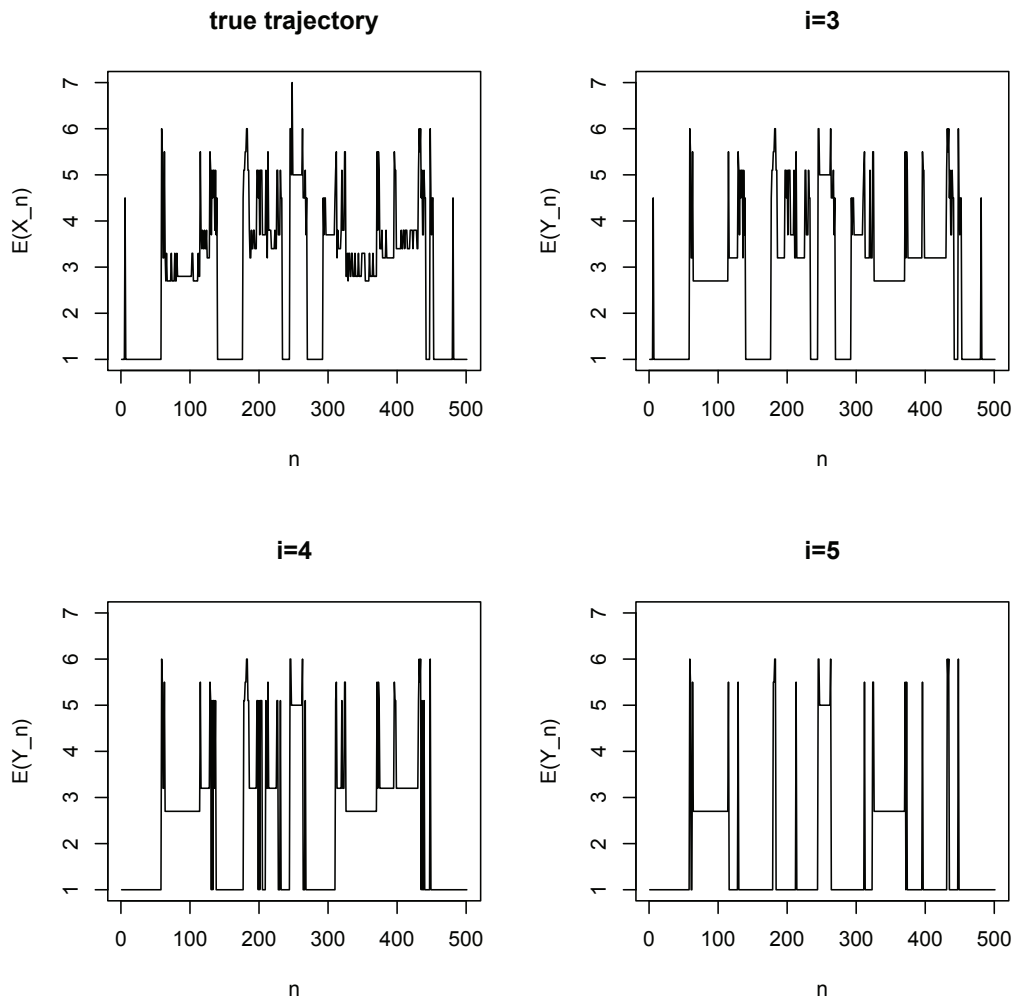


FIGURE 8. Energies of the true trajectory and of the trajectories of the aggregated chain at levels  $i = 3, 4, 5$ .

#### REFERENCES

1. V. S. Barbu and N. Limnios, *Semi-Markov chains and hidden semi-Markov models toward applications: Their use in reliability and DNA analysis*, volume 191 of *Lecture Notes in Statistics*, Springer, New York, 2008.
2. J. Beltrán and C. Landim, *Metastability of reversible finite state markov processes*, *Stochastic Processes and their Applications* **121** (2011), no. 8, 1633 – 1677.
3. J. P. Bouchaud, *Weak ergodicity breaking and aging in disordered systems*, *Journal de Physique I France* **2** (1992), 1705–1713.
4. A. Bovier, *Metastability: a potential theoretic approach*, *International Congress of Mathematicians. Vol. III*, Eur. Math. Soc., Zürich (2006), 499–518.
5. A. Bovier, M. Eckhoff, V. Gaynard, and M. Klein, *Metastability in stochastic dynamics of disordered mean-field models*, *Probab. Theory Related Fields* **119** (2001), no. 1, 99–161.
6. M. Cassandro, A. Galves, E. Olivieri, and M. E. Vares, *Metastable behavior of stochastic dynamics: a pathwise approach*, *J. Statist. Phys.* **35** (1984), no. (5-6), 603–634.
7. J. N. Darroch and E. Seneta, *On quasi-stationary distributions in absorbing discrete-time finite Markov chains*, *J. Appl. Probability* **2** (1965), no. 1, 88–100.
8. M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical*

- Sciences], Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szűcs.
9. G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaft (1912), 456–477.
  10. F. R. Gantmacher, *Applications of the theory of matrices*, Interscience Publishers, Inc., New York, 1959. Translated by J. L. Brenner, with the assistance of D. W. Bushaw and S. Evanusa.
  11. M. Goldstein, *Viscous liquids and the glass transition: A potential energy barrier picture*, The Journal of Chemical Physics **51** (1969), no. 9, 3728–3739.
  12. A. Heuer, *Exploring the potential energy landscape of glass-forming systems: from inherent structures via metabasins to macroscopic transport*, J. Phys.: Condens. Matter **20** (2008), no. 37, 373101.
  13. T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes, *Scaling concepts for the dynamics of viscous liquids near an ideal glassy state*, Phys. Rev. A **40** (1989), no. 2, 1045–1054.
  14. T. M. Liggett, *Interacting particle systems*, volume 276 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, New York, 1985.
  15. T. Okushima, T. Niiyama, K. S. Ikeda, and Y. Shimizu, *Graph-based analysis of kinetics on multidimensional potential-energy surfaces*, Phys. Rev. E **80** (2009), no. 3, 036112.
  16. E. Olivieri and E. Scoppola, *Markov chains with exponentially small transition probabilities: first exit problem from a general domain. II. The general case.*, J. Statist. Phys. **84** (1996), no. 5-6, 987–1041.
  17. P. K. Pollett, *Quasi-stationary distributions: A bibliography*, Version 2010, <http://www.maths.uq.edu.au/~pkp/papers/qsds/qsds.html>.
  18. O. Rubner and A. Heuer, *From elementary steps to structural relaxation: A continuous-time random-walk analysis of a supercooled liquid*, Phys. Rev. E **78** (2008), no. 1, 011504.
  19. S. Sastry, P. G. Debenedetti, and F. H. Stillinger, *Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid*, Nature **393** (1998), 554–557.
  20. E. Scoppola, *Renormalization group for Markov chains and application to metastability*, J. Statist. Phys. **73** (1993), no. 1-2, 83–121.
  21. E. Scoppola, *Metastability for Markov chains: a general procedure based on renormalization group ideas*, Probability and phase transition (Cambridge, 1993), volume 420 of NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. (1994), Kluwer Acad. Publ., Dordrecht, 303–322.
  22. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland, 1st edition, 1987.

INSTITUT FÜR MATHEMATISCHE STATISTIK, UNIVERSITÄT MÜNSTER, EINSTEINSTRASSE 62, 48149 MÜNSTER, GERMANY

INSTITUT FÜR MATHEMATISCHE STATISTIK, UNIVERSITÄT MÜNSTER, EINSTEINSTRASSE 62, 48149 MÜNSTER, GERMANY

*E-mail address:* `andrewinkler@uni-muenster.de`