

# ON THE POLYNOMIAL CONVERGENCE RATE TO NONEQUILIBRIUM STEADY STATES

BY YAO LI

*University of Massachusetts, Amherst*

We consider a stochastic energy exchange model that models the 1-D microscopic heat conduction in the nonequilibrium setting. In this paper, we prove the existence and uniqueness of the nonequilibrium steady state (NESS) and, furthermore, the polynomial speed of convergence to the NESS. Our result shows that the asymptotic properties of this model and its deterministic dynamical system origin are consistent. The proof uses a new technique called the induced chain method. We partition the state space and work on both the Markov chain induced by an “active set” and the tail of return time to this “active set.”

**1. Introduction.** As a ubiquitous process, heat conduction has been studied for over two hundred years. However, from a mathematical point of view, many microscopic aspects of heat conduction in solids and gas are still unclear. For example, the derivation of macroscopic heat conduction laws like Fourier’s law from microscopic Hamiltonian dynamics is a well-known challenge in statistical mechanics for the past over a century. Over the last several decades, numerous mathematical models of 1-D microscopic heat conduction have been proposed and studied. Some of these models have purely deterministic dynamics [10, 12, 20, 35], while others are defined by stochastic differential equations [3, 11, 27, 32–34] or Markov jump processes [15, 23, 24, 36]. These models give mathematical frameworks for studying nonequilibrium phenomena including basic properties of nonequilibrium steady states (NESS), thermal conductivity, local thermodynamic equilibrium (LTE), fluctuation theorems, and eventually Fourier’s law.

This paper focuses on fundamental properties of nonequilibrium phenomena including the existence and uniqueness of the NESS and, furthermore, the polynomial-speed convergence to the NESS, for a class of 1-D microscopic heat conduction models. The model we study is a stochastic energy exchange model that is inspired by the KMP model introduced in [19], in which a chain of  $N$  sites are coupled with two heat baths. Each site carries a certain amount of energy. An exponential clock is associated with each pair of adjacent sites. When the clock rings, these two sites exchange energy in a “random halves” fashion. The energy exchange with the bath follows a similar rule. Different from the model in [19], the

---

Received March 2017; revised April 2018.

*MSC2010 subject classifications.* Primary 60J25, 82C05; secondary 37N05, 60G07, 82C35.

*Key words and phrases.* Microscopic heat conduction, polynomial convergence rate, Markov jump process, induced chain method.

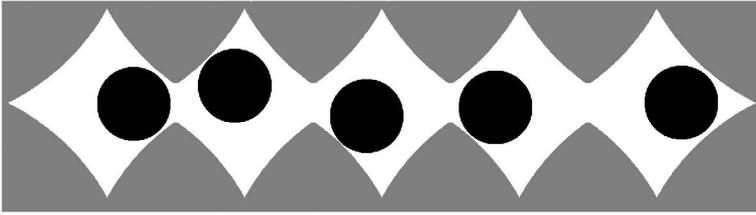


FIG. 1. *Locally confined particle system.*

rate of an exponential clock here is energy-dependent. In this paper, the clock rate between two adjacent sites, called the *stochastic energy exchange rate*, depends on the square root of the minimum of two site energies. We refer Section 2.2 for the precise description of the model.

The motivation of letting the stochastic energy exchange rate depend on the minimum of site energy comes from the study of a deterministic dynamical system heat conduction model, called the locally confined particle system. Introduced in [4], the locally confined particle system is a chain of locally confining cells in  $\mathbb{R}^2$  like in Figure 1. An identical rigid disk-shaped moving particle is contained in each cell. A particle cannot pass through the “bottleneck” between adjacent cells but can collide with its neighbors. Therefore, kinetic energy can be exchanged by particle-particle collisions. Studying such a purely deterministic dynamical system, especially in the nonequilibrium setting, is very challenging. Only very limited rigorous results are known. On the other hand, it is well known that chaotic billiard systems like the Lorentz gas have many stochastic properties due to the quick decay of correlation [1, 5–7]. Hence a natural approach is to only record the kinetic energy of each particle and approximate this model by the stochastic energy exchange model described above. Let  $E_i$  and  $E_{i+1}$  be kinetic energies of adjacent particles. Assume the geometry of the cell allows a particle to be able to completely “hide” from its neighbors, that is, neighboring particles cannot collide with a particle when it is located at some area of its cell. Our numerical simulations in [21] show that when starting from a fixed energy configuration  $(E_i, E_{i+1})$ , the first particle-particle collision time is well approximated by an exponential distribution. Heuristically, this is an expected result because for a sufficiently chaotic dynamical system, the rescaled “return time” and “hitting time” to asymptotically small set both converge to the same exponential distribution [17]. The numerical simulation in [21] further shows that the rate of the exponential distribution for the first collision time can be approximated by  $\sim \sqrt{\min\{E_i, E_{i+1}\}}$  when one of  $E_i$  and  $E_{i+1}$  is sufficiently small. The heuristic reason of this rate is that when one particle is sufficiently slow and out the reach of its neighbors, the next particle-particle collision time should be primarily determined by the kinetic energy of the slow particle.

We remark that at a certain time rescaling limit, the locally confined particle system may have a different stochastic energy exchange rate. In a nonrigorous study

of the locally confined particle system [13, 14], a rate function  $\sim \sqrt{E_i + E_{i+1}}$  is obtained at a certain rare interaction limit and time rescaling limit. Assuming this rate of interaction, the mixing rate is known to be exponential [15, 23, 36]. Without any time rescaling limit, it is a simple mathematical fact that the speed of mixing in the locally confined particle system cannot be faster than  $t^{-2}$ , provided particles can “hide” from their neighbors. (See Lemma 3.1 of [21].) This  $\sim t^{-2}$  speed of mixing is also one of the main result of this paper. Therefore, the approximate interaction rate  $\sim \sqrt{\min\{E_i, E_{i+1}\}}$  computed in [21] is consistent with the asymptotical dynamics of the locally confined particle system at its original time scale.

In this paper, among other results, we proved that when the stochastic rate of energy exchange between two sites are  $\sim \sqrt{\min\{E_i, E_{i+1}\}}$ , the Markov chain generated by our model has  $\sim t^{-2}$  rate of mixing and  $\sim t^{-2}$  rate of contraction. These results completely match our analytical and numerical results about the locally confined particle system in [21]. As shown in the proof later in this paper, the main source of the slow-speed mixing comes from the rate  $\sim \sqrt{\min\{E_i, E_{i+1}\}}$ . When one site acquires a very low amount of energy from an energy exchange, the rates of two corresponding clocks become very low and cannot be “rescued” by other “faster clocks.” Hence the next energy exchange at this site will not happen within a long time period, which obviously slows down the speed of mixing and convergence. We remark that this is also consistent with the mechanism of slow-speed mixing phenomenon of the locally confined particle system.

In addition to the ergodicity, the quantitative property of the NESS is also of great interest. Our result shows the absolute continuity of the NESS with respect to the Lebesgue measure. In addition, we obtain the tail of the first passage time to a certain uniform reference set. This helps us to show that the tail of the marginal distribution of NESS with respect to each site is  $\geq E^{-1/2}$  when  $E \ll 1$ . Since the explicit formulation of the NESS usually cannot be given, a detailed study on the properties of the NESS will rely heavily on numerical simulations. We will write a separate paper to numerically study the NESS, the long-range correlation, and the thermal conductivity of the generalized KMP model studied in this paper.

Despite the straightforward heuristic argument, a rigorous proof of the slow-speed mixing of a Markov process is known to be difficult. To the best of our knowledge, our result is the first polynomial convergence result in nonequilibrium settings. We prove that the Markov process generated by the generalized KMP model has a mixing rate  $\sim t^{-2}$  and a convergence rate  $\sim t^{-1}$  to the NESS. The closest related results we know are the slower-than-exponential convergence to the NESS in [8, 9, 37, 38] and the polynomial convergence to the equilibrium in [25]. In addition to the upper bound of convergence, we also showed that the speed of convergence to NESS has a lower bound  $t^{-1-\gamma}$  for any  $\gamma > 0$ . This further confirms the polynomial ergodicity.

The method of proving the polynomial-speed convergence to the steady-state is called the induced chain method. Since the source of slow convergence is the low-energy site, we partition the phase into two parts: the “active” set and the “inactive”

set, where the “active” set means all site energies are above a certain threshold. Then we work on the Markov chain induced by the “active” set. Different from the model in [25], where the “active” set satisfies a Doeblin-type condition, in this paper we need some extra work to show the stochastic stability of the induced chain. The induced chain method consists of three steps. We first show that the induced chain admits a uniform reference set, on which trajectories can be coupled with strictly positive probability. Then we control the first passage time to this uniform reference set for the induced chain, this is done by constructing a Lyapunov function as we have done in [25]. Last, we show that the time duration of one step of the induced chain has a polynomial tail by a technical construction of Lyapunov functions. A global Lyapunov function is obtained from a “tower construction” of local Lyapunov functions with respect to nearest neighbor interactions. The three steps above imply that the first passage time of the full system to the uniform reference set has a polynomial tail. Then we can apply results from discrete renewal theory and prove the polynomial-speed convergence and mixing.

We remark that a common way of proving polynomial-speed convergence is to construct a Lyapunov function [16, 25]. However, in this model such a construction is too complicated to be practical. One needs to consider both the lack of tightness and the possible inactive clocks in the construction of a Lyapunov function. By using the induced chain method, we can treat these two problems separately, and eventually give the tail of the first passage time to a uniform reference set. This method is useful in proving the polynomial (or subexponential) convergence of other models. In addition, we believe the induced chain technique can be extended into a hierarchy of finitely many induced chains and be applied to a wider range of problems.

The paper is organized in the following way: Section 2 introduces the model and states the main result. The main strategy of proof, that is, the induced chain method, is introduced in Section 3. Estimations for the time duration of one step of the induced chain are given in Section 4. The first entry time to the uniform reference set of the induced chain is done in Section 5. Finally, we complete the whole proof in Section 6.

## 2. Model and result.

2.1. *Stochastic approximation of deterministic dynamics.* We start with a short review of the locally confined particle system and its stochastic approximation. Consider a chain of cells in  $\mathbb{R}^2$  that are formed by finitely many piecewise  $C^3$  curves. A rigid disk-shaped *moving particle* is confined in each cell, as shown in Figure 1. Two adjacent cells are connected by a “bottleneck” opening such that particles cannot pass the opening but can collide with each other. A particle moves freely until it collides with the cell boundary or its neighbor particles. In addition, we assume each cell forms a strongly chaotic billiard table. In the absence of other particles, the billiard map of one particle is exponentially mixing. We refer to [4]

for the ergodicity of the locally confined particle system under suitable conditions and [6] for major results of dynamic billiards.

In the locally confined particle system, a particle has a quick decay of correlation due to frequent collisions with the cell boundary. Therefore, it is natural to simplify the model by assuming that the geometry within a cell is forgotten by the particle. More precisely, we only record the kinetic energy carried by a particle and assume that the time to the next energy exchange is exponentially distributed. The rate of this exponential distribution is called the *stochastic energy exchange rate*, denoted by  $R(E_i, E_{i+1})$ , where  $E_i$  and  $E_{i+1}$  are kinetic energies of particles. If in addition, we assign a suitable rule for the energy redistribution in a particle-particle collision, a Markov jump process is obtained.

In [21], we numerically showed that the time to the next particle-particle collision always has an exponential tail that depends on the energy configuration of particles. This further supports the idea of approximating the locally confined particle system by a Markov jump process. The rate  $R(E_i, E_{i+1})$  can be computed numerically as the slope of the exponential tail of the waiting time to the next particle-particle collision. Our numerical result in [21] showed that  $R(E_i, E_{i+1}) \sim \sqrt{\min\{E_i, E_{i+1}\}}$  when at least one of  $E_i$  or  $E_{i+1}$  is sufficiently small.

As stated in the [Introduction](#), the slow convergence phenomenon comes from the slow clock rate when one of  $E_i$  or  $E_{i+1}$  is very small. To preserve this qualitative property of the model, it is sufficient to let the stochastic energy exchange rate be  $\sqrt{\min\{E_i, E_{i+1}\}}$  for all small  $\min\{E_i, E_{i+1}\}$ . Hence we assume  $R(E_i, E_{i+1}) = \min\{K, \sqrt{\min\{E_i, E_{i+1}\}}\}$  in this paper for the sake of simplicity. We idealize the energy exchange in a particle-particle collision by choosing it as a “random halves” energy redistribution. Our numerical simulation shows that this is a reasonable choice, as the amount of redistributed energy has a strictly positive probability density function. In addition, the system is coupled with two heat baths and we prescribe a similar rule for the energy exchange with the heat bath. This gives rise to the stochastic energy exchange model as will be described in the next subsection.

*2.2. Description of the stochastic energy exchange model.* Now we give a precise description of the stochastic energy exchange model. Consider a chain of  $N$  lattice sites connected to two heat baths at the ends. The energy at each site is denoted by  $E_1, \dots, E_N$ , respectively. The temperature of heat baths are  $T_L$  and  $T_R$ . An exponential clock is associated with each pair of adjacent sites. The rate of the clock depends on the energy at both sites, denoted by  $R(E_i, E_{i+1})$ . When the clock rings, the energy at two sites are pooled together and redistributed randomly as

$$(2.1) \quad (E'_i, E'_{i+1}) = (p(E_i + E_{i+1}), (1 - p)(E_i + E_{i+1})),$$

where  $p$  is a uniform random variable distributed on  $(0, 1)$  that is independent of everything else. In addition, an exponential clock is associated with the first (resp., last) site and the left (resp., right) heat bath, whose rate is  $R(T_L, E_1)$  [resp.,  $R(E_N, T_R)$ ] for the same rate function used above. When the clock rings, the energy at the first (resp., last) site exchanges energy with an exponential random variable:

$$E'_1 = p(E_1 + X_L) \quad (\text{resp. } E'_N = p(E_N + X_R)),$$

where  $p$  is a uniform random variable on  $(0, 1)$  that is independent of everything else,  $X_L$  (resp.,  $X_R$ ) is an exponential random variable with mean  $T_L$  (resp.,  $T_R$ ). All exponential clocks are assumed to be mutually independent. We remark that the uniform random variable  $p$  is chosen to simplify the proof. Our method works for other choices of  $p$  with uniformly positive and bounded density on  $(0, 1)$ .

The rate  $R(E_i, E_{i+1})$ , called the *stochastic energy exchange rate* between sites  $i$  and  $i + 1$ , has the following form:

$$R(E_i, E_{i+1}) = \min\{K, \sqrt{\min(E_i, E_{i+1})}\},$$

where  $K \gg 1$  is a sufficiently large constant. As explained above, the stochastic energy exchange rate is assumed to be the square root of the minimum of site energies when either of the site energies is sufficiently small. The maximum of stochastic energy exchange rate is set as  $K < \infty$  for technical reasons. Without such an assumption, the Lyapunov function is not in the domain of the infinitesimal generator of the Markov chain, which imposes certain technical complexity [28]. As the aim of this paper is to show that the property of the rate function at low energy leads to polynomial rate of convergence to NESS, we choose to cap the energy exchange rate by  $K$ .  $K$  is assumed to be sufficiently large so that it will not significantly affect the dynamics at any “normal configuration.” In particular, we assume  $K \gg T_L, T_R$ .

It is easy to see from the description that this model generates a Markov jump process  $\mathbf{E}_t = (E_1(t), \dots, E_N(t))$  on  $\mathbb{R}_+^N$ . For any measurable function  $f$ , the infinitesimal generator of  $\mathbf{E}_t$  is

$$\begin{aligned} &\mathcal{L}f(E_1, \dots, E_N) \\ &= \sum_{i=1}^{N-1} R(E_i, E_{i+1}) \\ &\quad \times \int_0^1 \{f(E_1, \dots, E_{i-1}, p(E_i + E_{i+1}), (1-p)(E_i + E_{i+1}), E_{i+2}, \dots, E_N) \\ &\quad - f(E_1, \dots, E_N)\} dp \\ &\quad + R(T_L, E_1) \\ &\quad \times \int_0^\infty \int_0^1 \left\{ \frac{1}{T_L} e^{-s/T_L} f(p(E_1 + s), E_2, \dots, E_N) - f(E_1, \dots, E_N) \right\} dp ds \end{aligned}$$

$$\begin{aligned}
 &+ R(E_N, T_R) \\
 &\times \int_0^\infty \int_0^1 \left\{ \frac{1}{T_R} e^{-s/T_R} f(E_1, \dots, E_{N-1}, p(E_N + s)) - f(E_1, \dots, E_N) \right\} dp ds.
 \end{aligned}$$

We denote the transition kernel of  $\mathbf{E}_t$  by  $P^t(\mathbf{E}, \cdot)$ , where  $\mathbf{E} \in \mathbb{R}_+^N$ . The left and right operator generated by  $P^t$  are

$$(P^t \zeta)(\mathbf{E}) = \int_{\mathbb{R}_+^N} P^t(\mathbf{E}, d\mathbf{x}) \zeta(\mathbf{x})$$

for a measurable function  $\zeta(\mathbf{E})$  on  $\mathbb{R}_+^N$ , and

$$(\mu P^t)(B) = \int_{\mathbb{R}_+^N} P^t(\mathbf{E}, B) \mu(d\mathbf{E})$$

for a probability measure  $\mu$  on  $\mathbb{R}_+^N$ . We also use the notation  $\mathbb{P}_{\mathbf{E}}$  and  $\mathbb{E}_{\mathbf{E}}$  for conditional probability and conditional expectation with respect to the initial condition  $\mathbf{E}_0 = \mathbf{E}$ .

In this paper, we will use energy exchange events frequently. The event that the exponential clock between  $E_{i-1}$  and  $E_i$  rings at a time  $t$  (resp., at a stopping time  $\tau$ ) is denoted by  $C_i(t)$  [resp.,  $C_i(\tau)$ ] for all  $i = 1, \dots, N + 1$ . In addition, for the sake of simplicity we denote  $E_0 = T_L$  and  $E_{N+1} = T_R$ .

2.3. *Main result.* To state our main result precisely, the following functions and measure classes are necessary. Let

$$W(\mathbf{E}) = \sum_{i=1}^N E_i.$$

For any  $0 < \eta \ll 1$ , let

$$V(\mathbf{E}) = V_\eta(\mathbf{E}) = \sum_{m=1}^N \sum_{i=1}^{N-m+1} \left( \sum_{j=0}^{m-1} E_{i+j} \right)^{a_m \eta - 1},$$

where  $a_m = 1 - (2^{m-1} - 1)/(2^N - 1)$  for  $m = 1, \dots, N$ . Note that  $0 < a_m < 1$  for each  $m$ , hence all powers  $a_m \eta - 1$  are negative. Let  $\mathcal{M}_\eta$  be the collection of probability measure  $\mu$  on  $\mathbb{R}_+^N$  such that

$$\int_{\mathbb{R}_+^N} (W(\mathbf{E}) + V(\mathbf{E})) \mu(d\mathbf{E}) < \infty.$$

$\mathcal{M}_\eta$  covers a large class of probability measures. For example, if  $\mathbf{X}$  is a random energy configuration that has finite expectation and its density at the neighborhood of the ‘‘boundary’’  $\{\mathbf{E} = (E_1, \dots, E_N) \mid E_i = 0 \text{ for some } i = 1 \sim N\}$  is uniformly bounded, then the probability measure induced by  $\mathbf{X}$  belongs to  $\mathcal{M}_\eta$  for any sufficiently small  $\eta > 0$ .

We have the following results regarding the stochastic stability of  $\mathbf{E}_t$ .

**THEOREM 1 (Polynomial contraction of the Markov operator).** *For any  $\gamma > 0$ , there exists  $\eta > 0$  such that*

$$\lim_{t \rightarrow \infty} t^{2-\gamma} \|\mu P^t - \nu P^t\|_{\text{TV}} = 0$$

for any  $\mu, \nu \in \mathcal{M}_\eta$ , where  $\|\cdot\|_{\text{TV}}$  is the total variation norm.

**THEOREM 2 (Properties of the invariant measure).** *There exists a unique invariant measure  $\pi$  that is absolutely continuous with respect to the Lebesgue measure. In addition, for any  $\gamma > 0$  there exists  $\eta > 0$  such that*

$$\lim_{t \rightarrow \infty} t^{1-\gamma} \|\mu P^t - \pi\|_{\text{TV}} = 0$$

for any  $\mu \in \mathcal{M}_\eta$ .

Note that  $\pi$  may not be in  $\mathcal{M}_\eta$ , which leads to a different rate in Theorem 2.

A corollary of polynomial contraction of Markov operator is the rate of correlation decay.

**THEOREM 3 (Polynomial correlation decay).** *Let functions  $\xi$  and  $\zeta$  be in  $L^\infty(\mathbb{R}_+^N)$ . For any  $\gamma > 0$ , there exists  $\eta > 0$  such that for any  $\mu \in \mathcal{M}_\eta$*

$$\begin{aligned} & \left| \int_{\mathbb{R}_+^N} (P^t \zeta)(\mathbf{E}) \xi(\mathbf{E}) \mu(d\mathbf{E}) - \int_{\mathbb{R}_+^N} (P^t \zeta)(\mathbf{E}) \mu(d\mathbf{E}) \int_{\mathbb{R}_+^N} \xi(\mathbf{E}) \mu(d\mathbf{E}) \right| \\ & \leq O(1) \cdot \|\xi\|_{L^\infty} \|\zeta\|_{L^\infty} \left( \frac{1}{t^{2-\gamma}} \right) \end{aligned}$$

as  $t \rightarrow \infty$ , where the  $O(1)$  term depends on  $\gamma, N$ , and  $\mu$ .

Finally, the following proposition gives the lower bound of convergence speed.

**PROPOSITION 4 (Lower bound of convergence).** *There exists a probability measure  $\nu$  satisfying  $d\nu \ll d\pi$  such that*

$$\|\nu P^t - \pi\|_{\text{TV}} \geq c(t + 1)^{-1-\gamma}$$

for any sufficiently small  $\gamma > 0$ .

**3. Approach toward polynomial ergodicity.** The purpose of this section is to introduce a general approach, called the induced chain method, toward the polynomial ergodicity of a Markov process. We introduce this method under the generic setting, as it can be potentially applied to other models. Our aim is to make this section self-contained. When citing results from references, we will explain how statements of those theorems are rephrased.

Throughout this section, we let  $\Psi_n$  be a discrete-time Markov chain on a measurable space  $(X, \mathcal{B})$ . The transition kernel of  $\Psi_n$  is  $\mathcal{P}(x, \cdot)$ .

For  $A \in \mathcal{B}$ , we let  $\tau_A$  be the first passage time to  $A$ :

$$\tau_A = \inf\{n > 0 \mid \Psi_n \in A\}.$$

A set  $A \in \mathcal{B}$  is said to be *accessible* if  $\mathbb{P}_x[\tau_A < \infty] = 1$  for every  $x \in X$ .

We say a Markov chain is *irreducible* with respect to a measure  $\phi$  on  $\mathcal{B}$  if for any  $A \in \mathcal{B}$  with  $\phi(A) > 0$  and any  $x \in X$ , there exists  $n$  such that  $\mathcal{P}^n(x, A) > 0$ . In other word, every set with positive  $\phi$ -measure is accessible. We refer readers to Chapter 4 of [29] for the well-definedness of irreducibility. In fact, if  $\Psi_n$  is irreducible with respect to some measure  $\phi$ , then there exists a “maximal irreducible measure”  $\psi$  that is unique up to equivalence class.

3.1. *Splitting, coupling and moments of return times.* By the polynomial ergodicity of  $\Psi_n$ , we mean the polynomial rate of contraction of the Markov operator  $\mathcal{P}$ , the polynomial rate of convergence to the invariant measure of  $\Psi_n$  and the polynomial rate of correlation decay of  $\Psi_n$ . To prove the polynomial ergodicity, the bound on return times to a certain *uniform reference set*  $\mathcal{C} \in \mathcal{B}$  is crucial.

DEFINITION 3.1. A set  $\mathcal{C} \in \mathcal{B}$  is said to be a uniform reference set if it satisfies

$$\inf_{x \in \mathcal{C}} \mathcal{P}(x, \cdot) \geq \delta \theta(\cdot),$$

where  $\theta$  is a probability measure on  $(X, \mathcal{B})$  and  $\delta$  is a strictly positive real number.

A uniform reference set is a special case of *small set* or *petite set* defined in [29].

We call  $\mathcal{C}$  a uniform reference set because processes starting from  $\mathcal{C}$  have some uniform “common future.” If such a uniform reference set  $\mathcal{C}$  exists,  $\Psi_n$  can be converted to a new process  $\tilde{\Psi}_n$  on a modified state space  $\tilde{X} = X \cup \mathcal{C}_1$ , where  $\mathcal{C}_1$  is an identical copy of  $\mathcal{C}_0 := \mathcal{C}$ .  $\mathcal{B}$  can be extended to  $\tilde{\mathcal{B}}$  on  $\tilde{X}$  accordingly. Then we can split a probability measure  $\mu$  on  $(X, \mathcal{B})$  to a probability measure  $\mu^*$  on  $(\tilde{X}, \tilde{\mathcal{B}})$ :

$$\begin{cases} \mu^*|_X = (1 - \delta)\mu|_{\mathcal{C}_0} + \mu|_{X \setminus \mathcal{C}_0}, \\ \mu^*|_{\mathcal{C}_1} = \delta\mu|_{\mathcal{C}_0}, \quad \mathcal{C}_0 \cong \mathcal{C}_1 \text{ via the natural identification.} \end{cases}$$

Then we can “lift”  $\Psi_n$  to a new Markov process  $\tilde{\Psi}_n$  on  $\tilde{X}$  with a transition kernel  $\tilde{\mathcal{P}}(x, \cdot)$ :

$$\begin{cases} \tilde{\mathcal{P}}(x, \cdot) = (\mathcal{P}(x, \cdot))^*, & x \in X \setminus \mathcal{C}_0, \\ \tilde{\mathcal{P}}(x, \cdot) = [(\mathcal{P}(x, \cdot))^* - \delta\theta^*(\cdot)]/(1 - \delta), & x \in \mathcal{C}_0, \\ \tilde{\mathcal{P}}(x, \cdot) = \theta^*(\cdot), & x \in \mathcal{C}_1. \end{cases}$$

It is straightforward to check that  $\tilde{\Psi}_n$  has an atom  $\alpha := \mathcal{C}_1$ , that is,  $\tilde{\mathcal{P}}(x, \cdot)$  is the same for all  $x \in \alpha$ . In addition, if the initial distribution of  $\tilde{\Psi}_n$  is split from some  $\mu$  on  $(X, \mathcal{B})$  as described above,  $\tilde{\Psi}_n$  projects to  $\Psi_n$  through the natural projection

from  $\tilde{X}$  to  $X$ . This transformation is called the Nummelin splitting [30]. We refer to [29, 30] for the details.

If  $\Psi_n$  admits an accessible uniform reference set, many results about the stochastic stability of  $\Psi_n$  can be implied by estimates about  $\tau_{\mathcal{C}}$ . To state the results, we need the aperiodicity of  $\Psi_n$ .

Let  $\mathcal{C}$  be a uniform reference set. Define  $E_{\mathcal{C}} \subset \mathbb{Z}^+$  be the set of integers such that  $M \in E_{\mathcal{C}}$  if and only if

$$\inf_{x \in \mathcal{C}} \mathcal{P}^M(x, \cdot) \geq \delta \theta(\cdot), \quad \theta(\mathcal{C}) > 0$$

for a probability measure  $\theta$  and a strictly positive number  $\delta$ . The greatest common divisor of  $E_{\mathcal{C}}$  is called the *period* of  $\Psi_n$  with respect to  $\mathcal{C}$ .

DEFINITION 3.2. An irreducible Markov process  $\Psi_n$  is said to be *aperiodic* if the period of  $\Psi_n$  with respect to any uniform reference set  $\mathcal{C}$  is 1.

DEFINITION 3.3. An irreducible Markov process  $\Psi_n$  is said to be *strongly aperiodic* if  $\Psi_n$  admits a uniform reference set  $\mathcal{C}$  such that  $\theta(\mathcal{C}) > 0$ .

If  $\Psi_n$  is strongly aperiodic,  $\Psi_n$  must be aperiodic such that no cyclic decomposition is possible. We refer to Theorem 5.4.4 of [29] for the precise result about the cyclic decomposition for Markov processes on measurable state spaces.

DEFINITION 3.4. A probability measure  $\pi$  is said to be *invariant* if

$$\pi(A) = \int_X \pi(dx) \mathcal{P}(x, A)$$

for any  $A \in \mathcal{B}$ .

THEOREM 3.5. Let  $\Psi_n$  be an irreducible Markov chain on  $(X, \mathcal{B})$  with transition kernel  $\mathcal{P}$ . If  $\mathcal{C} \in \mathcal{B}$  is an accessible uniform reference set such that

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x[\tau_{\mathcal{C}}] < \infty,$$

then there exists an invariant probability measure  $\pi$ .

PROOF. Define

$$L(x, A) = \mathbb{P}_x[\tau_A < \infty].$$

Then obviously  $L(x, \mathcal{C}) = 1$  for every  $x \in \mathcal{C}$ . By Theorem 8.3.6 of [29],  $\Psi_n$  is recurrent. The theorem then follows from Theorem 10.0.1 of [29].  $\square$

The polynomial ergodicity of  $\Psi_n$  follows from the finiteness of moments of  $\tau_{\mathcal{C}}$ .

**THEOREM 3.6** (Theorem 2.6 and 2.7 in [31]). *Let  $\Psi_n$  be an aperiodic Markov chain on  $(X, \mathcal{B})$  with transition kernel  $\mathcal{P}$ . Let  $\pi$  be an invariant probability measure of  $\Psi_n$ . If  $\mathcal{C} \in \mathcal{B}$  is an accessible uniform reference set and*

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x[\tau_{\mathcal{C}}^\beta] < \infty$$

for some  $\beta > 0$ , then for any probability measures  $\mu, \nu$  on  $X$  that satisfy

$$\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] < \infty \quad \text{and} \quad \mathbb{E}_\nu[\tau_{\mathcal{C}}^\beta] < \infty,$$

we have

$$\lim_{n \rightarrow \infty} n^\beta \|\mu \mathcal{P}^n - \nu \mathcal{P}^n\|_{\text{TV}} = 0.$$

In addition, if  $\beta > 1$ , then

$$\lim_{n \rightarrow \infty} n^{\beta-1} \|\mu \mathcal{P}^n - \pi\|_{\text{TV}} = 0$$

for any  $\mu$  which satisfies

$$\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] < \infty.$$

**REMARK 3.7.** We rephrased statements of Theorem 2.6 and 2.7 in [31] to make the notation consistent. Functions  $\psi(n)$  and  $\psi_0(n)$  defined in [31] correspond to  $n^{\beta-1}$  and  $n^\beta$ . Then by Theorem 2.7(i) of [31],

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x[\tau_{\mathcal{C}}^\beta] < \infty$$

implies that  $\Psi_n$  is ergodic of order  $\psi$ . By Theorem 2.6, we have

$$\mathbb{E}_\pi[\tau_{\mathcal{C}}^{\beta-1}] < \infty.$$

The rest of the results follows from Theorem 2.7(iii) of [31].

Below we give a short self-contained proof for Theorem 3.6 based on the discrete renewal theory in [26]. We refer to [25, 31] for the full details, and [16] for a modern treatment of continuous time Feller processes.

**PROOF OF THEOREM 3.6.** We first apply the Nummelin splitting to  $\Psi_n$  to obtain  $\tilde{\Psi}_n$  on  $\tilde{X}$ .  $\tilde{\Psi}_n$  possesses an accessible atom  $\alpha$ .

Let  $\tilde{\Psi}_n^1$  and  $\tilde{\Psi}_n^2$  be two independent copies of  $\tilde{\Psi}_n$  starting from  $\mu^*$  and  $\nu^*$ , respectively. Let  $Y_0^1, Y_1^1, Y_2^1, \dots$  and  $Y_0^2, Y_1^2, Y_2^2, \dots$  be the passage times to  $\alpha$  for the two independent processes, respectively. Since  $\alpha$  is an atom,  $\{Y_i^1\}_{i=1}^\infty$  and  $\{Y_i^2\}_{i=1}^\infty$  are i.i.d. random variables. Therefore,

$$S_n^1 := \sum_{i=0}^n Y_i^1 \quad \text{and} \quad S_n^2 := \sum_{i=0}^n Y_i^2$$

form two delayed renewal processes.  $Y_0^1$  and  $Y_0^2$  are called the *delay distributions* of the renewal processes. Let  $T$  be the first simultaneous renewal time

$$T = \inf\{m \geq 0 \mid S_{k_1}^1 = S_{k_2}^2 = m \text{ for some } k_1, k_2\}.$$

Since  $\Psi_n$  (and  $\tilde{\Psi}_n$ ) is aperiodic, the renewal processes  $S_n^1$  and  $S_n^2$  are aperiodic.

Then it is easy to see that after  $T$ ,  $\tilde{\Psi}_n^1$  and  $\tilde{\Psi}_n^2$  become indistinguishable.  $T$  is called the *coupling time* of  $\tilde{\Psi}_n^1$  and  $\tilde{\Psi}_n^2$ . It is well known that

$$\|\mu\mathcal{P}^n - \nu\mathcal{P}^n\|_{\text{TV}} \leq \|\mu^*\tilde{\mathcal{P}}^n - \nu^*\tilde{\mathcal{P}}^n\|_{\text{TV}} \leq 2\mathbb{P}_{\mu^*, \nu^*}[T > n].$$

Therefore, it is sufficient to show the polynomial tail of  $\mathbb{P}[T > n]$ . Instead of the polynomial tail, we first prove the finiteness of moments of  $T$  by using the delayed renewal processes constructed above and the following two lemmas.

LEMMA 3.8. *Let  $\Psi_n$  be an aperiodic Markov chain on  $(X, \mathcal{B})$  with transition kernel  $\mathcal{P}$ . If  $\mathcal{C} \in \mathcal{B}$  is an accessible uniform reference set and*

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x[\tau_{\mathcal{C}}^\beta] < \infty$$

for some  $\beta > 0$ , then for any probability measure  $\mu$  such that

$$\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] < \infty,$$

there exists a constant  $C < \infty$ , such that

$$\mathbb{E}_{\mu^*}[\tau_\alpha^\beta] \leq C\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] < \infty.$$

PROOF. Apply Nummelin splitting to  $\Psi_n$  with respect to  $\mathcal{C}$ . Define the stopping time  $\tau = \tau_{\mathcal{C}_0 \cup \alpha}$  for  $\tilde{\Psi}_n$ . Let  $\{\tau^n\}_{n=1}^\infty$  be the sequence of iterates of  $\tau$ , that is,

$$\tau^0 = 0, \quad \tau^1 = \tau, \quad \tau^{n+1} = \tau^n + \tau \circ \Theta^{\tau^n},$$

where  $\Theta$  is the usual shift operator. Further let  $Z_n$  be a sequence of  $\{0, 1\}$  random variables such that  $Z_n = 1$  if and only if  $\tilde{\Psi}_{\tau^n} \in \alpha$ . According to the definition of  $\tilde{\Psi}_n$ , the probability of  $\tilde{\Psi}_{\tau^n} \in \alpha$  is at least  $\delta$  whenever the split chain  $\tilde{\Psi}_n$  jumps to  $\mathcal{C}_0 \cup \alpha$  at the step  $\tau^n$ . Hence  $Z_n$  is  $\mathcal{F}_n = \sigma\{\tilde{\Psi}_0, \dots, \tilde{\Psi}_{\tau^n}\}$  measurable and

$$\mathbb{P}[Z_n = 1 \mid \mathcal{F}_{n-1}] \geq \delta > 0.$$

Let  $\zeta = \inf\{n > 0 \mid Z_n = 1\}$ . Then  $\tau_\alpha = \tau^\zeta$ . The lemma then follows from Lemma 3.1(iii) of [31]. From the proof of Lemma 3.1 of [31], we can see that there exists a constant  $C < \infty$ , such that

$$\mathbb{E}_{\mu^*}[\tau_\alpha^\beta] \leq C\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] < \infty. \quad \square$$

LEMMA 3.9. *Let  $S_n^1$  and  $S_n^2$  be the delayed renewal processes as above. If there exists  $\beta > 1$  such that  $\mathbb{E}[(Y_0^1)^\beta]$ ,  $\mathbb{E}[(Y_0^2)^\beta]$ , and  $\mathbb{E}[(Y_1^1)^\beta]$  are all finite, then there exists a constant  $C < \infty$  depending on  $\mathbb{E}[(Y_1^1)^\beta]$ , such that*

$$\mathbb{E}[T^\beta] < C(\mathbb{E}[(Y_0^1)^\beta] + \mathbb{E}[(Y_0^2)^\beta]) < \infty.$$

*In addition, there exists a delay distribution  $Y_0^c$  with  $\mathbb{E}[(Y_0^c)^{\beta-1}] < \infty$  such that*

$$S_n^c := Y_0^c + \sum_{i=1}^n Y_i^c, \quad n \geq 1$$

*is a stationary renewal process, where  $\{Y_i^c\}_{i=1}^n$  are independent random variables that have the same distribution as  $Y_1^1$ .*

PROOF. This lemma follows from Section II of [26]. The finiteness of  $\mathbb{E}[T^\beta]$  follows from Theorem 4.2 of [26]. Tracking the proof, we can see that  $\mathbb{E}[T^\beta]$  is actually bounded by a constant times  $\mathbb{E}[(Y_0^1)^\beta] + \mathbb{E}[(Y_0^2)^\beta]$ .

Let  $p_k = \mathbb{P}[Y_1^1 = k]$ . Choose a suitable normalizer  $\lambda$  such that

$$c_k = \lambda \sum_{i=k+1}^{\infty} p_i, \quad k \geq 0$$

is a probability distribution. Then the delay distribution  $Y_0^c$  with  $\mathbb{P}[Y_0^c = k] = c_k$  gives a stationary renewal process (Section II.2 of [26]). Further, it is easy to see that  $\mathbb{E}[(Y_0^c)^{\beta-1}] < \infty$  (Section II.5 of [26]).  $\square$

Bounds of  $\mathbb{E}[(Y_0^1)^\beta]$ ,  $\mathbb{E}[(Y_0^2)^\beta]$ , and  $\mathbb{E}[(Y_1^1)^\beta]$ , that is,  $\mathbb{E}_{\mu^*}[\tau_\alpha^\beta]$ ,  $\mathbb{E}_{\nu^*}[\tau_\alpha^\beta]$ , and  $\mathbb{E}_\alpha[\tau_\alpha^\beta]$ , are given in Lemma 3.8. Bounds of  $\mathbb{E}[T^\beta]$  when starting from  $\mu$  and  $\nu$  follow from Lemma 3.9. When starting from  $\pi$ , passage times to  $\alpha$  form a (delayed) stationary renewal process. The delay distribution of the corresponding renewal process must be  $Y_0^c$ , as the delay distribution that lead to a stationary renewal process is unique ([2], Chapter 2). Therefore, bounds of  $\mathbb{E}[T^\beta]$  when starting from  $\mu$  and  $\pi$  also follow from Lemma 3.9.

Theorem 3.6 is then implied by the following simple probability fact (see Lemma 3.12). Let  $Z$  be any nonnegative integer-valued random variable. For any  $\beta > 0$ ,

$$\mathbb{E}[Z^\beta] < \infty \quad \Rightarrow \quad \lim_{n \rightarrow \infty} n^\beta \mathbb{P}[Z > n] = 0. \quad \square$$

We remark that the bounds in Lemma 3.8 and Lemma 3.9 now depend on the corresponding initial conditions. One needs to track the proof of theorems in [26] and [31] to see such dependence. The dependence on initial conditions implies the following corollary, which is used in the proof of Theorem 3.

COROLLARY 3.10. *Let  $\Psi_n$  and  $\mathcal{C}$  be as in Theorem 3.6, then for any probability measures  $\mu, \nu$  on  $X$  that satisfy*

$$\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] < \infty \quad \text{and} \quad \mathbb{E}_\nu[\tau_{\mathcal{C}}^\beta] < \infty,$$

*there exists a constant  $C$  depending on  $\mu, \nu$ , and  $\mathcal{C}$  such that*

$$\sup_n n^\beta \|\mu\mathcal{P}^n - \nu\mathcal{P}^n\|_{\text{TV}} \leq C(\mathbb{E}_\mu[\tau_{\mathcal{C}}^\beta] + \mathbb{E}_\nu[\tau_{\mathcal{C}}^\beta]).$$

PROOF. We have

$$\sup_n n^\beta \|\mu\mathcal{P}^n - \nu\mathcal{P}^n\|_{\text{TV}} \leq 2 \sup_n n^\beta \mathbb{P}[T > n] \leq 2\mathbb{E}_{\mu^*, \nu^*}[T^\beta].$$

The corollary then follows immediately from Lemma 3.8 and Lemma 3.9.  $\square$

3.2. *Induced chain method.* As discussed above, a crucial step toward polynomial ergodicity is to estimate the moments of  $\tau_{\mathcal{C}}$ , that is, the first passage time to a certain uniform reference set. In some simple cases, such as the model in [25], this can be done by constructing a Lyapunov function. However, in our model and many other problems, it is extremely difficult to find a single Lyapunov function to complete this task. Here, we introduce a method, called the induced chain method, that can be used to estimate the moments of  $\tau_{\mathcal{C}}$  under more general settings.

Let  $X = G \cup B$  be a partition of the state space of  $\Psi_n$ , where  $G$  is the “good set” on which  $\Psi_n$  is sufficiently “active”, while  $B$  is the “bad set” on which  $\Psi_n$  may hover for a long time. Define  $0 = T_0 < T_1 < \dots < T_n < \dots$  to be return times to  $G$  such that

$$T_{n+1} = \inf\{k > T_n \mid \Psi_k \in G\}$$

and let  $\hat{\Psi}_n = \Psi_{T_n}$ . Then it is easy to check that  $\hat{\Psi}_n$  is a Markov chain induced by  $G$ .

Assume  $\mathcal{C} \subset G$ . The tail of  $\tau_{\mathcal{C}}$  can be estimated by the following two assumptions:

(i)  $T_{n+1} - T_n$  has a polynomial tail for each  $T_n$ . There exists a constant  $\alpha > 0$  such that

$$\mathbb{P}[T_{n+1} - T_n > k \mid \Psi_{T_n}] \leq \xi(\Psi_{T_n})k^{-\alpha},$$

where  $1 \leq \xi(\Psi_{T_n}) < \infty$  is a constant depending on  $\Psi_{T_n}$ . Furthermore,  $\xi(\Psi_{T_n})$  is uniformly bounded by  $\xi_1 < \infty$  for  $\Psi_{T_n} \in G$ .

(ii)  $\hat{\tau}_{\mathcal{C}} = \inf\{n > 0 \mid \hat{\Psi}_n \in \mathcal{C}\}$  has an exponential tail. There exist constants  $\omega, \eta = \eta(\Psi_0) > 0$  such that

$$\mathbb{P}_{\Psi_0}[\hat{\tau}_{\mathcal{C}} > k] \leq \eta(\Psi_0)e^{-\omega k},$$

where  $\eta(\Psi_0)$  depends on  $\Psi_0$ .

**THEOREM 3.11.** *Assuming (i) and (ii) above, for any  $\epsilon > 0$ , there exists a constant  $c$  such that*

$$\mathbb{P}_{\Psi_0}[\tau_{\mathcal{E}} > n] \leq c(\eta(\Psi_0) + \xi(\Psi_0))n^{-(\alpha-\epsilon)}$$

for any  $\Psi_0 \in X$ .

**PROOF.** For any small  $\delta > 0$ , we have

$$\{\tau_{\mathcal{E}} > k^{1+\delta}\} \subset \{\hat{\tau}_{\mathcal{E}} > k^\delta\} \bigcup_{n=0}^{\lfloor k^\delta \rfloor} \{T_{n+1} - T_n > k, \hat{\tau}_{\mathcal{E}} > n\}.$$

This implies

$$\begin{aligned} \mathbb{P}_{\Psi_0}[\tau_{\mathcal{E}} > k^{1+\delta}] &\leq \mathbb{P}_{\Psi_0}[\hat{\tau}_{\mathcal{E}} > k^\delta] \\ &\quad + \sum_{n=0}^{\lfloor k^\delta \rfloor} \mathbb{P}_{\Psi_0}[T_{n+1} - T_n > k \mid \hat{\tau}_{\mathcal{E}} > n]. \end{aligned}$$

By assumption (i) and the Markov property, we have

$$\mathbb{P}_{\Psi_0}[T_1 - T_0 > k \mid \hat{\tau}_{\mathcal{E}} > 0] \leq \xi(\Psi_0)k^{-\alpha}$$

for  $n = 0$  and

$$\begin{aligned} &\mathbb{P}_{\Psi_0}[T_{n+1} - T_n > k \mid \hat{\tau}_{\mathcal{E}} > n] \\ &= \int_G \mathbb{P}[T_{n+1} - T_n > k \mid \Psi_{T_n} = x, \hat{\tau}_{\mathcal{E}} > n] \mathbb{P}_{\Psi_0}[\Psi_{T_n} = dx \mid \hat{\tau}_{\mathcal{E}} > n] \\ &\leq \xi_1 k^{-\alpha} \end{aligned}$$

for  $n \geq 1$ .

By assumption (ii), we have

$$\mathbb{P}_{\Psi_0}[\hat{\tau}_{\mathcal{E}} > k^\delta] \leq \eta(\Psi_0)e^{-\omega k^\delta}.$$

Therefore,

$$\begin{aligned} \mathbb{P}_{\Psi_0}[\tau_{\mathcal{E}} > k^{1+\delta}] &\leq \eta(\Psi_0)e^{-\omega k^\delta} + \max\{\xi(\Psi_0), \xi_1\}k^\delta k^{-\alpha} \\ &\leq c(\delta)(\eta(\Psi_0) + \xi(\Psi_0))k^{-(\alpha-\delta)} \end{aligned}$$

for some  $c(\delta) > 0$  that depends on  $\delta$ , as  $e^{-\omega k^\delta}$  decays faster than  $k^{-\alpha}$  when  $k \rightarrow \infty$ .

Let  $k = n^{\frac{1}{1+\delta}}$ . The theorem then follows by making  $\delta > 0$  sufficiently small and letting  $c = c(\delta)$  for the  $\delta$  we choose.  $\square$

The finiteness of moments of a random variable is closely related to its polynomial tails. We finish this subsection by proving two simple probabilistic facts that will be used frequently in this paper.

Let  $Z$  be a random variable that takes nonnegative integer values, and let  $\beta > 1$ .

LEMMA 3.12. *For any  $\beta > 0$ ,  $\mathbb{E}[Z^\beta] < \infty \implies \lim_{n \rightarrow \infty} n^\beta \mathbb{P}[Z > n] = 0$ .*

PROOF. First,

$$\begin{aligned} \sum_{n=0}^{\infty} n^{\beta-1} \mathbb{P}[Z > n] &= \sum_{n=0}^{\infty} n^{\beta-1} \sum_{m=n+1}^{\infty} \mathbb{P}[Z = m] \\ &= \sum_{m=1}^{\infty} \left( \sum_{n=0}^{m-1} n^{\beta-1} \right) \mathbb{P}[Z = m] \\ &\leq \text{constant} \cdot \mathbb{E}[Z^\beta]. \end{aligned}$$

Then, letting  $a_n = \mathbb{P}[Z > n]$  so that  $a_0 \geq a_1 \geq \dots$ , we claim that

$$\text{if } \sum_{n=0}^{\infty} a_n n^{\beta-1} < \infty \quad \text{then } a_n n^\beta \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

To see that, write

$$\sum_{n=0}^{\infty} (n+1)^\beta (a_n - a_{n+1}) = \sum_{n=0}^{\infty} [(n+1)^\beta - n^\beta] a_n.$$

Since  $a_n$  is monotone, our hypothesis implies the sum on the left-hand side converges. Since

$$\sum_{n \geq N} (n+1)^\beta (a_n - a_{n+1}) = a_N (N+1)^\beta + \sum_{n \geq N+1} a_n ((n+1)^\beta - n^\beta),$$

it follows that both terms on the right tend to 0 as  $N \rightarrow \infty$ .  $\square$

LEMMA 3.13. *If  $\mathbb{P}[Z > n] \leq C(n+1)^{-\beta}$  for  $\beta > 1$ , then for any  $\beta - 1 > \varepsilon > 0$ , there exists a constant  $K$  that depends on  $\beta$  and  $\varepsilon$ , such that  $\mathbb{E}[Z^{\beta-\varepsilon}] < KC$ .*

PROOF. There exists a constant  $K_0$  that only depends on  $\varepsilon$  and  $\beta$ , such that

$$\begin{aligned} \mathbb{E}[Z^{\beta-\varepsilon}] &= \sum_{n=0}^{\infty} n^{\beta-\varepsilon} \mathbb{P}[Z = n] \\ &\leq K_0 \sum_{n=0}^{\infty} \sum_{m=0}^n m^{\beta-1-\varepsilon} \mathbb{P}[Z = n] \end{aligned}$$

$$\begin{aligned}
 &= K_0 \sum_{n=0}^{\infty} (n+1)^{\beta-1-\epsilon} \mathbb{P}[Z > n] \\
 &\leq K_0 \sum_{n=0}^{\infty} (n+1)^{\beta-1-\epsilon} C(n+1)^{-\beta}.
 \end{aligned}$$

Hence there exists a constant  $K$  that depends on  $\beta$  and  $\epsilon$ , such that the last quantity is less than  $K C$ .  $\square$

We remark that the induced chain method can be extended to a hierarchical setting. Let  $X = B \cup G$  be the same partition as above. If we have  $\mathcal{C} := G_m \subset G_{m-1} \subset \dots \subset G_0 := G$  for  $m \geq 1$ , and the first return time to  $G_{i+1}$  for the  $G_i$ -induced chain has exponential tail for each  $i = 0 \sim m - 1$ , then the similar argument in Theorem 3.11 still follows.

3.3. *Lyapunov function and moments of the return time.* In the induced chain argument above, it remains to find sufficient conditions to estimate tails of  $\hat{\tau}_{\mathcal{C}}$  and  $T_{n+1} - T_n$ . This can be done by constructing Lyapunov-type functions. We introduce the following two theorems that will be used later.

**THEOREM 3.14** (Theorem 15.2.5 of [29]). *Let  $\Psi_n$  be a Markov chain on  $(X, \mathcal{B})$  with transition kernel  $\mathcal{P}$ . We assume that there exist a function  $W : X \rightarrow [1, \infty]$ , a set  $A \in \mathcal{B}$ , constants  $b > 0$  and  $0 \leq \beta < 1$  such that*

$$\mathcal{P}W - W \leq -\beta W + b\mathbf{1}_A.$$

*Then for any  $r \in (1, (1 - \beta)^{-1})$ , there exists  $\epsilon > 0$  such that*

$$\mathbb{E}_x \left[ \sum_{k=0}^{\tau_A-1} W(\Psi_k) r^k \right] \leq \epsilon^{-1} r^{-1} W(x) + \epsilon^{-1} b\mathbf{1}_A.$$

**THEOREM 3.15** (Modified from Theorem 3.6 of [18]). *Let  $\Psi_n$  be a Markov chain on  $(X, \mathcal{B})$  with transition kernel  $\mathcal{P}$ . We assume that there exist a function  $W : X \rightarrow [1, \infty]$ , a set  $A \in \mathcal{B}$ , constants  $b, c > 0$  and  $0 \leq \beta < 1$  such that*

$$\mathcal{P}W - W \leq -cW^\beta + b\mathbf{1}_A.$$

*Then there exists a constant  $\hat{c}$  such that*

$$\mathbb{E}_x \left[ \sum_{k=0}^{\tau_A-1} (k+1)^{\hat{\beta}-1} \right] \leq \hat{c}W(x), \quad \hat{\beta} = (1 - \beta)^{-1}$$

*for any  $x \in X$*

REMARK 3.16. This theorem follows from equation (37) in the proof of Theorem 3.6 of [18]. It is actually a special case of equation (37) when  $B = C$ . In this special case, the quantity

$$\sum_{k=0}^{\tau_B-1} (k+1)^i \mathbf{1}_C(\Psi_k)$$

in the proof of Theorem 3.2 of [18] is  $\mathbf{1}_C(\Psi_0)$ . Therefore, we only need to estimate the first passage time to  $A$  by using Proposition 11.3.3 of [29]. (See the proof of Theorem 3.2 of [18] for details.) The original theorem in [18] estimates

$$\mathbb{E}_x \left[ \sum_{k=0}^{\tau_B-1} (k+1)^{\hat{\beta}-1} \right]$$

for any set  $B$ , hence more assumptions are needed than in our case.

**4. Excursion time on low energy set.** It is obvious that for  $\mathbf{E}_t$ , the “bad set”  $B$  (see Section 3.2) should consist of energy configurations at which at least one of  $E_i$  is sufficiently small. It remains to estimate the excursion time on this “bad set.” To do so, we define the following sequence and function. Let  $0 < \eta \ll 1$  be a parameter that will be determined later. Let  $a_1, \dots, a_N$  be the sequence as in Section 2.3:

$$a_i = 1 - \frac{2^{i-1} - 1}{2^N - 1}.$$

Then it is easy to see that  $1 = a_1 > a_2 > \dots > a_N > 0$  and  $2a_{i-1} - a_i > a_1$  for each  $2 \leq i \leq N$ .

Define functions

$$\begin{aligned} V_{n,k} &= \left( \sum_{j=0}^{n-1} E_{k+j} \right)^{a_n \eta^{-1}}, \\ V_1(\mathbf{E}) &= \sum_{i=1}^N V_{1,i} = \sum_{i=1}^N E_i^{a_1 \eta^{-1}}, \\ V_2(\mathbf{E}) &= \sum_{i=1}^{N-1} V_{2,i} = \sum_{i=1}^{N-1} (E_i + E_{i+1})^{a_2 \eta^{-1}}, \\ &\vdots \\ V_n(\mathbf{E}) &= \sum_{k=1}^{N-n+1} V_{n,k} = \sum_{k=1}^{N-n+1} \left( \sum_{j=0}^{n-1} E_{k+j} \right)^{a_n \eta^{-1}}, \\ &\vdots \end{aligned}$$

$$V_N(\mathbf{E}) = V_{N,1} = \left( \sum_{i=1}^N E_i \right)^{a_N \eta - 1},$$

and

$$V(\mathbf{E}) = \sum_{i=1}^N V_i(\mathbf{E}).$$

The motivation of constructing  $V$  is to control the entire chain through nearest neighbor interactions.  $V_{1,i}(E)$  is the “first level Lyapunov function” with respect to  $E_i$ , whose value decreases when  $E_i$  increases.  $V_{2,i}$  is the “second level,” which is dominantly larger than the “first level” functions  $V_{1,i}$  and  $V_{1,i+1}$ , such that its decrease can compensate the possible increase of  $V_{1,i}$  and  $V_{1,i+1}$ . Higher levels can be constructed analogously. This tower construction of Lyapunov functions stops at the  $N$ th level that covers the entire chain.

Our aim is to show that  $V(\mathbf{E})$  is a Lyapunov function when the value of  $V(\mathbf{E})$  is sufficiently large. The excursion time on low energy set then follows from Theorem 3.15.

The main theorem in this section is as follows.

**THEOREM 4.1.** *For any  $\eta > 0$  and  $h > 0$  small enough, there exist  $c_0 > 0$ ,  $M_0 > 1$  depending on  $\eta$ ,  $N$  and  $h$ , such that*

$$(P^h)V(\mathbf{E}) - V(\mathbf{E}) \leq -c_0 V^\alpha(\mathbf{E})$$

for every  $\mathbf{E} \in \{V > M_0\}$ , where  $\alpha = 1 - \frac{1}{2(1-\eta)}$ .

Let  $\mathbf{E} = (E_1, \dots, E_N)$  be the initial condition. For the sake of simplicity, we let  $R_i = R(E_{i-1}, E_i)$  for all  $i = 1, \dots, N + 1$ .  $\mathbf{E}_t$  is given by the following equivalent description: Starting from  $t = 0$ , a clock rings at an exponentially distributed random time  $\tau_1$  with rate  $\sum_{i=0}^N R_i$ . When the clock rings, one and only one energy exchange occurs between site  $i$  and  $i + 1$  with probability

$$\frac{R_i}{\sum_{i=0}^N R_i} := R_i / \mathcal{R}.$$

In other words,  $\mathbb{P}_{\mathbf{E}}[C_i(\tau_1)] = R_i / \mathcal{R}$ .

Let  $\mathbf{E}_{\tau_1^+}$  be the energy configuration immediately after the first energy exchange occurs at  $\tau_1$ . We use  $V(\mathbf{E}_{\tau_1^+})$  to estimate  $P^h V(\mathbf{E})$ . The strategy of proving Theorem 4.1 is as follows. In Lemma 4.2 and Lemma 4.3, we show that an energy exchange event can increase a  $V_{n,k}$  by at most multiplying with a constant. However, if  $V_{n,k}$  is sufficiently large and the energy stored at its “boundary site,” that is,  $E_{k-1}$  or  $E_{k+n}$ , is sufficiently large, then in Lemma 4.4 we prove that the corresponding energy exchange event will reduce the value of such a  $V_{n,k}$  by at least

one-half. Then in Lemma 4.5 we manage to prove that if  $V_{n,k}$  is sufficiently large, the “expected jump” of  $V_{n,k}$  at  $\tau_1$  can be compensated by the “expected drop” of  $V_{n',k'}/4N^2$  for *some*  $V_{n',k'}$ . This can always be achieved because  $V_{n',k'}$  is significantly greater than  $V_{n,k}$  when  $V_{n',k'}$  “covers”  $V_{n,k}$ . Finally, in Lemma 4.6 we prove that the “expected drop” of  $V$  at  $\tau_1$  is proportional to  $V^\alpha$  for  $\alpha = 1 - \frac{1}{2(1-\eta)}$ . The heuristics of Lemma 4.6 is that for every sufficiently large  $V_{n,k}$ , the “expected jump” is compensated by  $1/4N^2$  of the “expect drop” of *some*  $V_{n',k'}$  (Lemma 4.5). Therefore, at  $\tau_1$  the total “expected jump” of those large  $V_{n,k}$ s is compensated by  $1/4$  of the largest “expected drops” on the left and right side, denoted by  $V_{n_R,k_R}$  and  $V_{n_L,k_L}$ , respectively. Because the “expected drop” of at least one  $V_{n,k}$  is proportional to  $V^\alpha$ , we know that the “expected drop” contributed by  $V_{n_R,k_R}$  and  $V_{n_L,k_L}$  minus the total “expected jump” of all large  $V_{n,k}$ s, is also proportional to  $V^\alpha$ . On the other hand, Lemma 4.2 implies the “expected jump” of small  $V_{n,k}$ s at  $\tau_1$  can be controlled by a constant.

LEMMA 4.2. For any  $\mathbf{E} \in \mathbb{R}_+^N$ ,  $1 \leq n \leq N$ ,  $1 \leq k \leq N - n + 1$ , and  $0 < \eta < \frac{1}{2}$ , we have

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_k(\tau_1)}] \leq \frac{C}{\mathcal{R}} \left( \sum_{i=0}^{n-1} E_{k+i} \right)^{a_n \eta - \frac{1}{2}}$$

and

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k+n}(\tau_1)}] \leq \frac{C}{\mathcal{R}} \left( \sum_{i=0}^{n-1} E_{k+i} \right)^{a_n \eta - \frac{1}{2}}$$

for some  $C > 0$  depending on  $N$  and  $\eta$ .

PROOF. At the first energy exchange, if  $k \neq 1$  and  $k + n \neq N + 1$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_k(\tau_1)}] &= \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid C_k(\tau_1)] \mathbb{P}_{\mathbf{E}}[C_k(\tau_1)] \\ &= \frac{\min\{K, \sqrt{\min\{E_{k-1}, E_k\}}\}}{\mathcal{R}} \\ &\quad \times \int_0^1 [p(E_{k-1} + E_k) + E_{k+1} + \dots + E_{k+n-1}]^{a_n \eta - 1} dp \\ &:= I_1. \end{aligned}$$

Notice that  $a_n \eta$  is small so  $a_n \eta - \frac{1}{2} < 0$ . If  $E_k \geq \frac{1}{2} \sum_{i=1}^{n-1} E_{k+i}$ , then

$$I_1 \leq \frac{E_k^{1/2}}{\mathcal{R}} \cdot E_k^{a_n \eta - 1} \int_0^1 p^{a_n \eta - 1} dp \leq \frac{2}{a_n \eta \mathcal{R}} \left( \sum_{i=0}^{n-1} E_{k+i} \right)^{a_n \eta - \frac{1}{2}}.$$

Otherwise,

$$\begin{aligned} I_1 &\leq \frac{E_k^{1/2}}{\mathcal{R}} \cdot (E_{k+1} + \dots + E_{k+n-1})^{a_n \eta - 1} \\ &\leq 2 \frac{(E_k + \dots + E_{k+n-1})^{1/2}}{\mathcal{R}} \cdot (E_k + \dots + E_{k+n-1})^{a_n \eta - 1} \\ &\leq \frac{2}{\mathcal{R}} \left( \sum_{i=0}^{n-1} E_{k+i} \right)^{a_n \eta - \frac{1}{2}}. \end{aligned}$$

The same argument holds for

$$\begin{aligned} &\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] \\ &= \frac{\min\{K, \sqrt{\min\{E_{k+n-1}, E_{k+n}\}}\}}{\mathcal{R}} \int_0^1 [E_k + \dots + p(E_{k+n-1} + E_{k+n})]^{a_n \eta - 1} dp \end{aligned}$$

by discussing cases  $E_{k+n-1} \geq \frac{1}{2} \sum_{i=1}^{n-2} E_{k+i}$  and  $E_{k+n-1} < \frac{1}{2} \sum_{i=1}^{n-2} E_{k+i}$ .  
 If  $V_{k,n}$  involves the boundary, say  $k = 1$ , then

$$\begin{aligned} &\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_k(\tau_1)}] \\ &= \frac{\sqrt{\min\{T_L, E_1\}}}{\mathcal{R}} \int_0^\infty \int_0^1 \frac{1}{T_L} [p(x + E_1) + E_2 + \dots + E_n]^{a_n \eta - 1} e^{-x/T_L} dp dx. \end{aligned}$$

It is easy to check that the same argument above still holds. The case of  $k + n = N + 1$  can be estimated analogously.

The proof is completed by letting  $C = \frac{2}{a_N \eta} \geq \frac{2}{a_n \eta}$ .  $\square$

The calculation in the proof of Lemma 4.2 gives the following lemma.

LEMMA 4.3. For any  $\mathbf{E} \in \mathbb{R}_+^N$ ,  $\eta > 0$ ,  $1 \leq n \leq N$ , and  $1 \leq k \leq N - n + 1$ ,

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_k(\tau_1)] \leq C V_{n,k}(\mathbf{E})$$

and

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] \leq C V_{n,k}(\mathbf{E}),$$

where  $C > 0$  is the same as that in Lemma 4.2.

PROOF. We have

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_k(\tau_1)] = \int_0^1 [p(E_{k-1} + E_k) + E_{k+1} + \dots + E_{k+n-1}]^{a_n \eta - 1} dp.$$

It follows from the same argument as in the proof of Lemma 4.2 that

$$\begin{aligned} & \int_0^1 [p(E_{k-1} + E_k) + E_{k+1} + \dots + E_{k+n-1}]^{a_n \eta - 1} dp \\ & \leq \max\left\{\frac{2}{a_n \eta}, 2\right\} (E_k + \dots + E_{k+n-1})^{a_n \eta - 1} := C V_{n,k}(\mathbf{E}). \end{aligned}$$

The proof of

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] \leq C V_{n,k}(\mathbf{E})$$

is similar.  $\square$

LEMMA 4.4. *For any  $\mathbf{E} \in \mathbb{R}_+^N$  and any  $0 < \eta \ll 1$ , there exists a constant  $C'$  depending on  $N, T_L, T_R$  and  $\eta$ , such that whenever  $E_{k+n} > C'(E_k + \dots + E_{k+n-1})$  [resp.  $E_{k-1} > C'(E_k + \dots + E_{k+n-1})$ ], we have*

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] < \frac{1}{2} V_{n,k}(\mathbf{E}) \\ & \left(\text{resp. } \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_k(\tau_1)] < \frac{1}{2} V_{n,k}(\mathbf{E})\right). \end{aligned}$$

PROOF. First, assume  $k + n \neq N + 1$ , then we have

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] \\ & = \int_0^1 (E_k + \dots + E_{k+n-2} + p(E_{k+n-1} + E_{k+n}))^{a_n \eta - 1} dp \\ & < E_{k+n}^{a_n \eta - 1} \int_0^1 p^{a_n \eta - 1} dp. \end{aligned}$$

If  $E_{k+n} > C'(E_k + \dots + E_{k+n-1})$ , we have

$$E_{k+n}^{a_n \eta - 1} \int_0^1 p^{a_n \eta - 1} dp \leq \frac{C'^{a_n \eta - 1}}{a_n \eta} V_{n,k}(\mathbf{E}).$$

To make

$$\frac{C'^{a_n \eta - 1}}{a_n \eta} \leq \frac{1}{2},$$

one needs

$$C' \geq \left(\frac{1}{2} a_n \eta\right)^{\frac{1}{a_n \eta - 1}}.$$

If  $k + n = N + 1$ , then

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] \\ &= \int_0^1 \int_0^\infty \frac{1}{T_R} (E_k + \dots + E_{k+n-2} + p(E_{k+n-1} + x))^{a_n \eta - 1} e^{-x/T_R} dp dx \\ &< \int_0^\infty \frac{1}{T_R} x^{a_n \eta - 1} e^{-x/T_R} dx \int_0^1 p^{a_n \eta - 1} dp \\ &= T_R^{a_n \eta - 1} \frac{\Gamma(a_n \eta)}{a_n \eta}, \end{aligned}$$

where  $\Gamma(\cdot)$  is the Gamma function. Therefore, to make

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] < \frac{1}{2} V_{n,k}(\mathbf{E}),$$

we need

$$V_{n,k}(\mathbf{E}) = (E_k + \dots + E_{k+n-1})^{a_n \eta - 1} > 2 T_R^{a_n \eta - 1} \frac{\Gamma(a_n \eta)}{a_n \eta}.$$

Since  $E_{N+1} = T_R$ , this is equivalent to

$$\begin{aligned} E_{N+1} &> T_R \left[ 2 T_R^{a_n \eta - 1} \cdot \frac{\Gamma(a_n \eta)}{a_n \eta} \right]^{-\frac{1}{a_n \eta - 1}} (E_k + \dots + E_{k+n-1}) \\ &= \left( \frac{2 \Gamma(a_n \eta)}{a_n \eta} \right)^{-\frac{1}{a_n \eta - 1}} (E_k + \dots + E_{k+n-1}). \end{aligned}$$

The case for  $\mathcal{C}_k$  is symmetric and can be calculated in the same way. By combining all cases, it is easy to check that if

$$C' = \max_{1 \leq n \leq N} \left\{ \left( \frac{1}{2} a_n \eta \right)^{\frac{1}{a_n \eta - 1}}, \left( \frac{2 \Gamma(a_n \eta)}{a_n \eta} \right)^{-\frac{1}{a_n \eta - 1}} \right\},$$

we have the desired property for all  $n$  and  $k$ .  $\square$

LEMMA 4.5. For any  $\mathbf{E} \in \mathbb{R}_+^N$  and any  $0 < \eta \ll 1$ , there exists a  $M < \infty$  depending on  $\eta, T_L, T_R$  and  $N$ , such that for any  $V_{n,k}(\mathbf{E}) > M$ , if

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] \geq V_{n,k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)] \\ & \text{(resp. } \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_k(\tau_1)}] \geq V_{n,k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_k(\tau_1)]), \end{aligned}$$

then there exists  $k'$  and  $n'$ , such that

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] - V_{n,k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)] \\ & \leq \frac{1}{4N^2} \{ V_{n',k'}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k'+n'}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n',k'}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k'+n'}(\tau_1)}] \} \end{aligned}$$

$$\begin{aligned} & \left( \text{resp. } \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_k(\tau_1)}] - V_{n,k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_k(\tau_1)] \right. \\ & \quad \left. \leq \frac{1}{4N^2} \{V_{n',k'}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k'}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n',k'}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k'}(\tau_1)}] \} \right). \end{aligned}$$

PROOF. By symmetry, it is sufficient to consider the case of

$$\mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] \geq V_{n,k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)].$$

Let  $k' = k$  and  $n'$  be the first  $n' > n - 1$  such that

$$E_{k+n'} > C'(E_k + \dots + E_{k+n'-1}),$$

where  $C'$  is the constant defined in Lemma 4.4. When  $M$  is large, the sum  $E_k + \dots + E_{k+n'-1}$  is small. Note that  $E_{N+1} = T_R$ . Hence when  $M$  is sufficiently large, for any  $V_{n,k}(\mathbf{E}_0) > M$ , one can always find such an  $n'$ .

By Lemma 4.4, if  $n' = n$ , the energy exchange event  $\mathcal{C}_{k+n}(\tau_1)$  will only bring the expected value of  $V_{n,k}$  down. Therefore, it is sufficient to consider the case of  $n' \geq n + 1$ . The lemma follows if one can prove either

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] - V_{n,k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)] \\ & \leq \frac{1}{4N^2} \{V_{n',k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n'}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n',k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k+n'}(\tau_1)}] \} \\ & := \frac{1}{4N^2} I_1 \end{aligned}$$

or

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] - V_{n,k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)] \\ & \leq \frac{1}{4N^2} \{V_{1,k+n'-1}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n'}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{1,k+n'-1}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k+n'}(\tau_1)}] \} \\ & := \frac{1}{4N^2} I_2. \end{aligned}$$

Let  $x = E_k + \dots + E_{k+n'-2}$ ,  $y = E_{k+n'-1}$ , and  $z = E_{k+n'}$ . Since  $a_{n'} < a_n$ , if  $x$  is sufficiently small, by Lemma 4.2,

$$\begin{aligned} \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] & \leq \frac{C}{\mathcal{R}}(E_k + \dots + E_{k+n-1})^{a_n\eta - \frac{1}{2}} \\ & \leq \frac{C}{\mathcal{R}} \left( \frac{1}{NC'^{N-1}}(E_k + \dots + E_{k+n'-2}) \right)^{a_n\eta - \frac{1}{2}} \\ & \leq \frac{1}{\mathcal{R}} x^{a_{n'-1}\eta - \frac{1}{2}} \cdot (C \cdot C'^{\frac{1}{2} - a_n\eta}(N-1) \cdot N^{\frac{1}{2} - a_n\eta} \cdot x^{(a_n - a_{n'-1})\eta}) \\ & \leq \frac{1}{\mathcal{R}} x^{a_{n'-1}\eta - \frac{1}{2}}, \end{aligned}$$

where  $C$  is the constant in Lemma 4.2. The second inequality above follows from

$$E_{k+m} \leq C'(E_k + \dots + E_{k+m-1})$$

for each  $n \leq m < n'$ .

The requirement of small  $x$  can be satisfied by making  $M = M(\eta)$  sufficiently large, because

$$x \leq NC'^{N-1}(E_k + \dots + E_{k+n-1}) \leq NC'^{N-1}M^{\frac{1}{a_n\eta-1}}.$$

Similarly, we can make  $M = M(\eta)$  sufficiently large such that

$$y \leq NC'^{N-1}M^{\frac{1}{a_n\eta-1}} < K.$$

Notice that  $y < z$  because  $C'$  in Lemma 4.4 is greater than 1 [ $C' > (\frac{1}{2}a_n\eta)^{(a_n\eta-1)^{-1}}$ ,  $a_n < 1$ ,  $\eta < 1$ ]. Therefore, we have  $R(E_{k+n'-1}, E_{k+n'}) = \sqrt{y}$ . Hence we have

$$I_1 = \left\{ (x+y)^{a_{n'}\eta-1} - \int_0^1 [x+p(y+z)]^{a_{n'}\eta-1} dp \right\} \frac{\sqrt{y}}{\mathcal{R}}$$

and

$$I_2 = \left\{ y^{a_1\eta-1} - \int_0^1 [p(y+z)]^{a_1\eta-1} dp \right\} \frac{\sqrt{y}}{\mathcal{R}}.$$

Since  $z > C'(y+x) > C'y$  and  $y < C'x$ , by Lemma 4.4, we have

$$I_1 \geq \frac{1}{2}(x+y)^{a_{n'}\eta-1} \frac{\sqrt{y}}{\mathcal{R}} \geq \frac{1}{2(1+C')} \frac{x^{a_{n'}\eta-1} \sqrt{y}}{\mathcal{R}}$$

and

$$I_2 \geq \frac{1}{2}y^{a_1\eta-1} \frac{\sqrt{y}}{\mathcal{R}}.$$

We claim that

$$\max \left\{ \frac{1}{2}y^{a_1\eta-\frac{1}{2}}, \frac{1}{2(1+C')}x^{a_{n'}\eta-1}\sqrt{y} \right\} > 4N^2x^{a_{n'-1}\eta-\frac{1}{2}}.$$

It is easy to see that the lemma follows from this claim.

**PROOF OF THE CLAIM.** Let  $\epsilon = \eta^2$ . The proof is split to two cases.

*Case 1:*  $y < x^{1+2(a_{n'-1}-a_{n'})\eta-\epsilon}$ , then

$$\frac{1}{2}y^{a_1\eta-\frac{1}{2}} > \frac{1}{2}x^{(a_1\eta-\frac{1}{2})[1+2(a_{n'-1}-a_{n'})\eta-\epsilon]}.$$

Note that  $2a_{n'-1} - a_{n'} > a_1$  and  $\eta \ll 1$ , we have

$$\begin{aligned} & \left(a_1\eta - \frac{1}{2}\right) [1 + 2(a_{n'-1} - a_{n'})\eta - \epsilon] \\ &= -\frac{1}{2} + a_{n'-1}\eta - \epsilon + (a_1 + a_{n'} - 2a_{n'-1})\eta \\ & \quad + \epsilon + \epsilon\left(\frac{1}{2} - a_1\eta\right) + 2a_1(a_{n'-1} - a_{n'})\eta^2 \\ &< -\frac{1}{2} + a_{n'-1}\eta - \epsilon \end{aligned}$$

if

$$(2a_{n'-1} - a_{n'} - a_1)\eta > \frac{3}{2}\epsilon + 2a_1(a_{n'-1} - a_{n'})\eta^2,$$

which can be achieved by making  $\eta$  sufficiently small (because  $\epsilon = \eta^2$ ).

Therefore,

$$\frac{1}{2}y^{a_1\eta-\frac{1}{2}} > \frac{1}{2}x^{-\epsilon} \cdot x^{a_{n'-1}\eta-\frac{1}{2}} > 4N^2x^{a_{n'-1}\eta-\frac{1}{2}},$$

when  $x$  is sufficiently small, which can be made by letting  $M$  large enough.

Case 2:  $y \geq x^{1+2(a_{n'-1}-a_{n'})\eta-\epsilon}$ . Recall that we have  $\epsilon = \eta^2$ , we have

$$\begin{aligned} \frac{1}{2C'}x^{a_{n'}\eta-1}\sqrt{y} &\geq \frac{1}{2C'}x^{a_{n'}\eta-1} \cdot x^{\frac{1}{2}+a_{n'-1}\eta-a_{n'}\eta-\frac{\epsilon}{2}} \\ &= \frac{1}{2C'}x^{a_{n'-1}\eta-\frac{1}{2}} \cdot x^{-\frac{\epsilon}{2}} \\ &> 4N^2x^{a_{n'-1}\eta-\frac{1}{2}} \end{aligned}$$

if  $x$  is sufficiently small. Again, this can be achieved by letting  $M$  sufficiently large.  $\square \square$

LEMMA 4.6. For any  $\mathbf{E} \in \mathbb{R}_+^N$  and any  $0 < \eta \ll 1$ , there exist constants  $\alpha = 1 - \frac{1}{2(1-\eta)}$ ,  $C_1$  and  $M'$  depending on  $N$  and  $\eta$ , such that

$$\mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\tau_1^+})] \leq V(\mathbf{E}) - \frac{C_1}{\mathcal{R}}V^\alpha(\mathbf{E})$$

for all  $V(\mathbf{E}) > M'$ .

PROOF. Step 1. First, we show that there exists  $\tilde{n}, \tilde{k}$  such that the ‘‘expected drop’’ of  $V_{\tilde{n}, \tilde{k}}$  at  $\tau_1$  is proportional to  $V^\alpha(\mathbf{E})$ . Let  $V_{\tilde{n}, \tilde{k}}(\mathbf{E})$  be the maximum of  $\{V_{i,j}(\mathbf{E})\}$ . Therefore,  $V_{\tilde{n}, \tilde{k}}(\mathbf{E}) \geq \frac{1}{N(N+1)}V(\mathbf{E})$ . Since  $V_{\tilde{n}, \tilde{k}}(\mathbf{E})$  is the maximum, we have

$$(E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}}\eta-1} > (E_{\tilde{k}-1} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}+1}\eta-1}$$

and

$$(E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}}\eta-1} > (E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}})^{a_{\tilde{n}+1}\eta-1}.$$

Since  $a_{\tilde{n}+1} < a_{\tilde{n}}$ , when  $M'$  is sufficiently large, we have

$$E_{\tilde{k}-1} > C'(E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})$$

and

$$E_{\tilde{k}+\tilde{n}} > C'(E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1}),$$

where  $C'$  is as in Lemma 4.4. In addition, we have

$$E_{\tilde{k}}^{a_1\eta-1} \leq (E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}}\eta-1}$$

and

$$E_{\tilde{k}+\tilde{n}-1}^{a_1\eta-1} \leq (E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}}\eta-1}.$$

Since  $M'$  is assumed to be sufficiently large, we have  $\min\{K, \min\{E_{\tilde{k}-1}, E_{\tilde{k}}\}\} = E_{\tilde{k}}$  and  $\min\{K, \min\{E_{\tilde{k}+\tilde{n}-1}, E_{\tilde{k}+\tilde{n}}\}\} = E_{\tilde{k}+\tilde{n}-1}$ . Therefore,

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V_{\tilde{n},\tilde{k}}(\mathbf{E}_{\tau_1^+})] - V_{\tilde{n},\tilde{k}}(\mathbf{E}) \\ &= \frac{\sqrt{E_{\tilde{k}}}}{\mathcal{R}} \left\{ \int_0^1 [p(E_{\tilde{k}-1} + E_{\tilde{k}}) + \dots + E_{\tilde{n}+\tilde{k}-1}]^{a_{\tilde{n}}\eta-1} dp \right. \\ & \quad \left. - (E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}}\eta-1} \right\} \\ & \quad + \frac{\sqrt{E_{\tilde{k}+\tilde{n}-1}}}{\mathcal{R}} \left\{ \int_0^1 [E_{\tilde{k}} + \dots + p(E_{\tilde{n}+\tilde{k}-1} + E_{\tilde{n}+\tilde{k}})]^{a_{\tilde{n}}\eta-1} dp \right. \\ & \quad \left. - (E_{\tilde{k}} + \dots + E_{\tilde{k}+\tilde{n}-1})^{a_{\tilde{n}}\eta-1} \right\} \\ & := I_1 + I_2. \end{aligned}$$

It follows from the same calculation as in Lemma 4.4 that

$$\begin{aligned} (4.1) \quad I_i &\leq -\frac{1}{2}(E_{\tilde{k}} + \dots + E_{\tilde{n}+\tilde{k}-1})^{a_{\tilde{n}}\eta-1} \cdot (E_{\tilde{k}} + \dots + E_{\tilde{n}+\tilde{k}-1})^{\frac{a_{\tilde{n}}\eta-1}{2(a_1\eta-1)}} \cdot \frac{1}{\mathcal{R}} \\ &= -\frac{1}{2\mathcal{R}} V_{\tilde{n},\tilde{k}}^\alpha(\mathbf{E}) \end{aligned}$$

for  $i = 1, 2$ , where

$$\alpha = 1 - \frac{1}{2(1 - a_1\eta)}.$$

Step 2. Let  $M$  be as in Lemma 4.5. Let  $\{(n_i, k_i)\}_{i=1}^m$  be indices for which  $V_{n_i, k_i}(\mathbf{E}) \geq M$ . The aim of this step is to prove that the total “expected jump” of these  $V_{n_i, k_i}$  at  $\tau_1$  is dominated by the “expected drop” of some  $V_{n_R, k_R}$  and  $V_{n_L, k_L}$  from left-hand and right-hand side, respectively.

By Lemma 4.5, we can construct sequences  $\{(\hat{n}'_i, \hat{k}'_i)\}_{i=1}^m$  and  $\{(\hat{n}''_i, \hat{k}''_i)\}_{i=1}^m$  such that if

$$\mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_i+n_i}(\tau_1)}] \geq V_{n_i, k_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_i+n_i}(\tau_1)],$$

then

$$\begin{aligned} &\mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_i+n_i}(\tau_1)}] - V_{n_i, k_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_i+n_i}(\tau_1)] \\ &\leq \frac{1}{4N^2} \{V_{\hat{n}'_i, \hat{k}'_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{\hat{k}'_i+\hat{n}'_i}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{\hat{n}'_i, \hat{k}'_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{\hat{k}'_i+\hat{n}'_i}(\tau_1)}]\} \end{aligned}$$

and if

$$\mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_i}(\tau_1)}] \geq V_{n_i, k_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_i}(\tau_1)],$$

then

$$\begin{aligned} &\mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_i}(\tau_1)}] - V_{n_i, k_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_i}(\tau_1)] \\ &\leq \frac{1}{4N^2} \{V_{\hat{n}''_i, \hat{k}''_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{\hat{k}''_i}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{\hat{n}''_i, \hat{k}''_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{\hat{k}''_i}(\tau_1)}]\}. \end{aligned}$$

When condition

$$\begin{aligned} &\mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_i+n_i}(\tau_1)}] \geq V_{n_i, k_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_i+n_i}(\tau_1)] \\ &\text{(resp. } \mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_i}(\tau_1)}] \geq V_{n_i, k_i}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_i}(\tau_1)]) \end{aligned}$$

is not satisfied, we simply let  $\hat{n}'_i = n_i, \hat{k}'_i = k_i$  (resp.,  $\hat{n}''_i = n_i, \hat{k}''_i = k_i$ ). Therefore, the “expected jump” of  $V_{n_i, k_i}$  at  $\tau_1$  is dominated by the right and left “expected drop” of  $V_{\hat{n}'_i, \hat{k}'_i}$  and  $V_{\hat{n}''_i, \hat{k}''_i}$ , respectively.

Let  $(n_R, k_R)$  be the index for which  $V_{n_R, k_R}(\mathbf{E})$  has biggest “right drop” at  $\tau_1$ , that is,

$$\begin{aligned} &V_{n, k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k+n}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k+n}(\tau_1)}] \\ &\leq V_{n_R, k_R}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_R+n_R}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n_R, k_R}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_R+n_R}(\tau_1)}] \end{aligned}$$

for all  $n, k$ , and  $(n_L, k_L)$  be the index for which  $V_{n_L, k_L}(\mathbf{E})$  has the biggest “left drop,” that is,

$$\begin{aligned} &V_{n, k}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_k(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_k(\tau_1)}] \\ &\leq V_{n_L, k_L}(\mathbf{E})\mathbb{P}_{\mathbf{E}}[C_{k_L}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n_L, k_L}(\mathbf{E}_{\tau_1^+})\mathbf{1}_{C_{k_L}(\tau_1)}] \end{aligned}$$

for all  $n, k$ . Both terms should be positive when  $M'$  is sufficiently large because of the argument about  $V_{\tilde{n}, \tilde{k}}$  in Step 1.

Since there is only  $\frac{1}{2}N(N + 1)$   $V_{n,k}$ 's, we have  $m \leq \frac{1}{2}N(N + 1)$ . Hence

$$\begin{aligned} & \sum_{i=1}^m \{ \mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+})] - V_{n_i, k_i}(\mathbf{E}) \} \\ &= \sum_{i=1}^m \{ \mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_i+n_i}(\tau_1)}] - V_{n_i, k_i}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_i+n_i}(\tau_1)] \\ & \quad + \mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_i}(\tau_1)}] - V_{n_i, k_i}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_i}(\tau_1)] \} \\ &\leq \frac{1}{4N^2} \sum_{i=1}^m \{ \max\{ V_{\hat{n}'_i, \hat{k}'_i}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{\hat{k}'_i+\hat{n}'_i}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{\hat{n}'_i, \hat{k}'_i}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{\hat{k}'_i+\hat{n}'_i}(\tau_1)}], 0 \} \\ & \quad + \max\{ V_{\hat{n}''_i, \hat{k}''_i}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{\hat{k}''_i}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{\hat{n}''_i, \hat{k}''_i}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{\hat{k}''_i}(\tau_1)}], 0 \} \} \\ &\leq \frac{1}{4} \{ V_{n_R, k_R}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_R+n_R}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n_R, k_R}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_R+n_R}(\tau_1)}] \\ & \quad + V_{n_L, k_L}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_L}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n_L, k_L}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_L}(\tau_1)}] \}. \end{aligned}$$

Step 3. Then we can finalize the entire proof. We have

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}(V(\mathbf{E}_{\tau_1^+})) - V(\mathbf{E}) \\ &= \sum_{n,k} \{ \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k+n}(\tau_1)}] - V_{n,k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k+n}(\tau_1)] \\ & \quad + \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_k(\tau_1)}] - V_{n,k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_k(\tau_1)] \} \\ & \quad \text{(Law of total expectation)} \\ &= \sum_{i=1}^m \{ \mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_i+n_i}(\tau_1)}] - V_{n_i, k_i}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_i+n_i}(\tau_1)] \\ & \quad + \mathbb{E}_{\mathbf{E}}[V_{n_i, k_i}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_i}(\tau_1)}] - V_{n_i, k_i}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_i}(\tau_1)] \} \\ & \quad + \sum_{k,n} \{ \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})] - V_{n,k}(\mathbf{E}) \} \quad \text{(by Lemma 4.5)} \\ &\leq \frac{3}{4} \{ \mathbb{E}_{\mathbf{E}}[V_{n_R, k_R}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{n_R+k_R}(\tau_1)}] - V_{n_R, k_R}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{n_R+k_R}(\tau_1)] \\ & \quad + \mathbb{E}_{\mathbf{E}}[V_{n_L, k_L}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{C_{k_L}(\tau_1)}] - V_{n_L, k_L}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[C_{k_L}(\tau_1)] \} \\ & \quad + \sum_{k,n} \{ \mathbb{E}_{\mathbf{E}}[V_{n,k}(\mathbf{E}_{\tau_1^+})] - V_{n,k}(\mathbf{E}) \} \quad \text{(by Step 2)} \\ &\leq \frac{3}{4}(I_3 + I_4) + I_5, \end{aligned}$$

where

$$I_3 = \mathbb{E}_{\mathbf{E}}[V_{n_R, k_R}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{n_R+k_R}(\tau_1)}] - V_{n_R, k_R}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{n_R+k_R}(\tau_1)],$$

$$I_4 = \mathbb{E}_{\mathbf{E}}[V_{n_L, k_L}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k_L}(\tau_1)}] - V_{n_L, k_L}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k_L}(\tau_1)],$$

and

$$I_5 = \sum_{k, n, V_{n, k} < M} \{ \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+})] - V_{n, k}(\mathbf{E}) \}.$$

It follows from Lemma 4.3 that

$$\begin{aligned} I_5 &= \sum_{k, n, V_{n, k} < M} \{ \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k+n}(\tau_1)}] - V_{n, k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)] \\ &\quad + \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+}), \mathbf{1}_{\mathcal{C}_k(\tau_1)}] - V_{n, k}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_k(\tau_1)] \} \\ &\leq \sum_{k, n, V_{n, k} < M} \{ \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k+n}(\tau_1)] \\ &\quad + \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_k(\tau_1)] \mathbb{P}_{\mathbf{E}}[\mathcal{C}_k(\tau_1)] \} \\ &\leq \sum_{k, n, V_{n, k} < M} \{ \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+n}(\tau_1)] + \mathbb{E}_{\mathbf{E}}[V_{n, k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_k(\tau_1)] \} \cdot \frac{2K}{\mathcal{R}} \\ &\leq N(N+1)CMK \cdot \frac{1}{\mathcal{R}}. \end{aligned}$$

In addition, by the definition of  $n_R, k_R$  and  $n_L, k_L$ , we have

$$\begin{aligned} &V_{\tilde{n}, \tilde{k}}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{\tilde{k}+\tilde{n}}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{\tilde{n}, \tilde{k}}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{\tilde{k}+\tilde{n}}(\tau_1)}] \\ &\leq V_{n_R, k_R}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k_R+n_R}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n_R, k_R}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k_R+n_R}(\tau_1)}] \end{aligned}$$

and

$$\begin{aligned} &V_{\tilde{n}, \tilde{k}}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{\tilde{k}}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{\tilde{n}, \tilde{k}}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{\tilde{k}}(\tau_1)}] \\ &\leq V_{n_L, k_L}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{k_L}(\tau_1)] - \mathbb{E}_{\mathbf{E}}[V_{n_L, k_L}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{k_L}(\tau_1)}], \end{aligned}$$

where  $\tilde{n}$  and  $\tilde{k}$  are from Step 1.

By inequality (4.1) in Step 1, we have

$$\begin{aligned} I_3 + I_4 &\leq (\mathbb{E}_{\mathbf{E}}[V_{\tilde{n}, \tilde{k}}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{\tilde{k}+\tilde{n}}(\tau_1)}] - V_{\tilde{n}, \tilde{k}}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{\tilde{k}+\tilde{n}}(\tau_1)]) \\ &\quad + (\mathbb{E}_{\mathbf{E}}[V_{\tilde{n}, \tilde{k}}(\mathbf{E}_{\tau_1^+}) \mathbf{1}_{\mathcal{C}_{\tilde{k}}(\tau_1)}] - V_{\tilde{n}, \tilde{k}}(\mathbf{E}) \mathbb{P}_{\mathbf{E}}[\mathcal{C}_{\tilde{k}}(\tau_1)]) \\ &= \mathbb{E}_{\mathbf{E}}[V_{\tilde{n}, \tilde{k}}(\mathbf{E}_{\tau_1^+})] - V_{\tilde{n}, \tilde{k}}(\mathbf{E}) \\ &\leq -\frac{1}{\mathcal{R}} V_{\tilde{n}, \tilde{k}}^\alpha(\mathbf{E}) \leq -\frac{1}{\mathcal{R}} \cdot \left( \frac{1}{N(N+1)} \right)^\alpha V^\alpha(\mathbf{E}). \end{aligned}$$

One can then further choose  $M' > M$  such that

$$\frac{1}{4\mathcal{R}} \cdot \left(\frac{1}{N(N+1)}\right)^\alpha M'^\alpha > I_5.$$

Therefore, for every  $\mathbf{E}$  such that  $V(\mathbf{E}) > M'$ , we have

$$\mathbb{E}_{\mathbf{E}}(V(\mathbf{E}_{\tau_1^+})) - V(\mathbf{E}) \leq -\frac{1}{\mathcal{R}} \cdot \frac{1}{2} \left(\frac{1}{N(N+1)}\right)^\alpha V^\alpha(\mathbf{E}) := -\frac{C_1}{\mathcal{R}} V^\alpha(\mathbf{E}).$$

This completes the proof.  $\square$

**PROOF OF THEOREM 4.1.** Let  $0 = \tau_0 < \tau_1 < \tau_2 < \dots$  be the times of clock rings. Let  $B_0 = \{\mathbf{E} \mid V(\mathbf{E}) > M'\}$  where  $M'$  is the constant in Lemma 4.6. By the Markov property and Lemma 4.6, for any  $\mathbf{E} \in \mathbb{R}_+^N$  we have

$$(4.2) \quad \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\tau_{n+1}^+}) \mid \mathbf{E}_{\tau_n^+}] = \mathbb{E}_{\mathbf{E}_{\tau_n^+}}[V(\mathbf{E}_{\tau_1^+})] \leq V(\mathbf{E}_{\tau_n^+}) - \frac{C_1}{\mathcal{R}} V(\mathbf{E}_{\tau_n^+})^\alpha$$

if  $\mathbf{E}_{\tau_n^+} \in B_0$ . Otherwise, for each  $\mathbf{E}_{\tau_n^+} \notin B_0$ , by the Markov property we have the uniform bound

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\tau_{n+1}^+}) \mid \mathbf{E}_{\tau_n^+}] \\ &= \sum_{m=1}^N \sum_{k=1}^{N-m+1} \{ \mathbb{E}_{\mathbf{E}_{\tau_n^+}}[V_{m,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_k(\tau_1)] \mathbb{P}_{\mathbf{E}_{\tau_n^+}}[\mathcal{C}_k(\tau_1)] \\ & \quad + \mathbb{E}_{\mathbf{E}_{\tau_n^+}}[V_{m,k}(\mathbf{E}_{\tau_1^+}) \mid \mathcal{C}_{k+m}(\tau_1)] \mathbb{P}_{\mathbf{E}_{\tau_n^+}}[\mathcal{C}_{k+m}(\tau_1)] \} \\ (4.3) \quad & \leq \sum_{m=1}^N \sum_{k=1}^{N-m+1} C V_{m,k}(\mathbf{E}_{\tau_n^+}) (\mathbb{P}_{\mathbf{E}_{\tau_n^+}}[\mathcal{C}_k(\tau_1)] + \mathbb{P}_{\mathbf{E}_{\tau_n^+}}[\mathcal{C}_{k+m}(\tau_1)]) \\ & \quad \text{(by Lemma 4.3)} \\ & \leq C \sum_{m=1}^N \sum_{k=1}^{N-m+1} V_{m,k}(\mathbf{E}_{\tau_n^+}) \leq CM', \end{aligned}$$

where  $C$  is the constant in Lemma 4.2.

Let  $S = \inf\{n \mid \tau_n > h\}$ , and define  $\hat{\tau}_n = \min\{\tau_n, \tau_{S-1}\}$ . Then

$$P^h V(\mathbf{E}) = \lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+}) \mathbf{1}_{\{S \leq n+1\}}] \leq \limsup_{n \rightarrow \infty} \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+})].$$

We will prove a uniform bound for  $\mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+})]$ . Equation (4.2) implies the expectation of  $V$  drops when starting from  $B_0$ . Equation (4.3) means the expectation of  $V$  can grow with  $CM'$  at most. By assuming the worse of (4.2) and (4.3), we have

$$\mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\tau_{n+1}^+}) \mid \tau_{n+1} \leq h] \leq \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+}) \mid \tau_{n+1} \leq h] + CM'$$

for every  $n \geq 0$ . Notice that for given  $\mathbf{E}_{\tau_n^+}$ ,  $\mathbf{E}_{\tau_{n+1}^+}$  is independent of  $\tau_{n+1} - \tau_n$ . Therefore conditioning on  $\tau_{n+1} \leq h$  does not affect the bounds in (4.2) and (4.3). Since  $\mathcal{R} \leq (N + 1)K$ , for each  $\mathbf{E}_{\tau_n^+}$ , we have

$$\mathbb{P}[\tau_{n+1} \leq h | \mathbf{E}_{\tau_n^+}, \tau_n \leq h] \leq (1 - e^{-hNK}).$$

Therefore, inductively we have

$$\mathbb{P}_{\mathbf{E}}[\tau_{n+1} \leq h] \leq (1 - e^{-hNK})^{n+1}$$

for all  $n \geq 0$ . This implies

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_{n+1}^+})] \\ &= \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_{n+1}^+}) | \tau_{n+1} > h] \cdot \mathbb{P}_{\mathbf{E}}[\tau_{n+1} > h] \\ & \quad + \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_{n+1}^+}) | \tau_{n+1} \leq h] \cdot \mathbb{P}_{\mathbf{E}}[\tau_{n+1} \leq h] \\ (4.4) \quad & \leq \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+}) | \tau_{n+1} > h] \cdot \mathbb{P}_{\mathbf{E}}[\tau_{n+1} > h] \\ & \quad + (\mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+}) | \tau_{n+1} \leq h] + CM') \cdot \mathbb{P}_{\mathbf{E}}[\tau_{n+1} \leq h] \\ & \leq \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_n^+})] + CM'(1 - e^{-hNK})^{n+1}. \end{aligned}$$

Adding up from  $n = 1$  to  $\infty$ , this gives

$$P^h V(\mathbf{E}) \leq \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_1^+})] + \frac{CM'}{e^{-hNK}}.$$

Let  $h > 0$  be small enough so that

$$\mathbb{P}_{\mathbf{E}}[\tau_1 \leq h] = 1 - e^{-h\mathcal{R}} > \frac{h}{2}\mathcal{R}.$$

This is the only condition we impose on  $h$ . There exists such an  $h$  independently of  $\mathbf{E}$  because of the bound  $\mathcal{R} \leq NK$ .

Notice that the energy exchange at  $\tau_1$  is independent of  $\tau_1$ . We have, by Lemma 4.6,

$$\begin{aligned} \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{\tau}_1^+})] &= \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\tau_1^+}) | \tau_1 \leq h] \cdot \mathbb{P}_{\mathbf{E}}[\tau_1 \leq h] + V(\mathbf{E}) \cdot \mathbb{P}_{\mathbf{E}}[\tau_1 > h] \\ &\leq \left( V(\mathbf{E}) - \frac{C_1}{\mathcal{R}} V(\mathbf{E})^\alpha \right) \cdot \mathbb{P}_{\mathbf{E}}[\tau_1 \leq h] + V(\mathbf{E}) \cdot \mathbb{P}_{\mathbf{E}}[\tau_1 > h] \\ &\leq V(\mathbf{E}) - C_1 \frac{h}{2} V(\mathbf{E})^\alpha. \end{aligned}$$

This gives

$$P^h V(\mathbf{E}) \leq V(\mathbf{E}) - C_1 \frac{h}{2} V(\mathbf{E})^\alpha + CM' e^{hNK}$$

for any  $\mathbf{E} \in B_0$ .

To complete the proof of Theorem 4.1, it suffices to replace  $M'$  by a large enough number  $M_0 > 1$  so that for  $\mathbf{E} \in \{V(\mathbf{E}) > M_0\}$ , the constant  $CM'e^{hNK}$  is absorbed into  $c_0V(\mathbf{E})^\alpha$  for  $c_0 := C_1^{\frac{h}{4}}$ .  $\square$

The same calculation in the proof of Theorem 4.1 also yields the following.

LEMMA 4.7.

$$\sup_{\{\mathbf{E}:V(\mathbf{E})\leq\hat{M}\}} P^h V(\mathbf{E}) \leq \hat{M} + CM'e^{hNK} < \infty$$

for any  $\hat{M} > 0$ .

PROOF. The calculation in (4.4) gives

$$\mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{t}_{n+1}^+})] \leq \mathbb{E}_{\mathbf{E}}[V(\mathbf{E}_{\hat{t}_n^+})] + CM'(1 - e^{-hNK})^{n+1}.$$

Adding up from  $n = 0$  to  $\infty$ , this gives

$$P^h V(\mathbf{E}) \leq V(\mathbf{E}) + CM'e^{hNK} \leq \hat{M} + CM'e^{hNK}. \quad \square$$

**5. Excursion time for the induced chain.** For sufficiently small given parameters  $h > 0$  and  $\eta > 0$ , let  $M_0$  be the constant defined in Theorem 4.1 and  $B := \{\mathbf{E} \mid V(\mathbf{E}) > M_0\}$ . Define  $G = \mathbb{R}_+^N \setminus B$ . The previous section together with Theorem 3.11 gives bounds on the excursion time in  $B$ . As introduced in Section 3.2, now one should work on the  $G$ -induced chain.  $G$  is not a compact set as the energy at each site can be arbitrarily high. A common way to show the tightness of a Markov process on noncompact state space is to construct a Lyapunov function, as we do in this section for the  $G$ -induced chain.

Let  $h > 0$  be the given size of a time step (defined in Theorem 4.1). We consider the time- $h$  sampling chain  $\{\mathbf{E}_{nh}\}_{n=0}^\infty$  of  $\mathbf{E}_t$ . For the sake of simplicity, we use the notation  $\mathbf{E}_n$  to represent  $\mathbf{E}_{nh}$  when it does not lead to a confusion.

Let  $0 = T_0 < T_1 < T_2 < \dots$  be discrete stopping times such that

$$T_1 = \inf\{k > 0 \mid \mathbf{E}_k \in G\}$$

and

$$T_{n+1} = \inf\{k > T_n \mid \mathbf{E}_k \in G\}$$

for  $n = 1, 2, \dots$ . We define

$$\hat{\mathbf{E}}_n = \mathbf{E}_{T_n}$$

for  $n = 1, 2, \dots$  as the  $G$ -induced chain. It is easy to see that  $\hat{\mathbf{E}}_n$  is also a Markov chain. We denote the transition kernel of  $\hat{\mathbf{E}}_n$  by  $\hat{P}$ .

As in [25], a natural Lyapunov function is the total energy in the system. Let

$$W(\mathbf{E}) = \sum_{n=1}^N E_n.$$

The main theorem of this section reads the following.

**THEOREM 5.1.** *There exist constants  $M_1 > 1$  and  $\delta > 0$ , such that*

$$\hat{P}W(\mathbf{E}) \leq (1 - \delta)W(\mathbf{E})$$

for every  $\mathbf{E} \in G$  with  $W(\mathbf{E}) > M_1$ .

To prove Theorem 5.1, the first task is to bound the length of the time step for  $\hat{\mathbf{E}}_n$ . Such estimate follows from Theorem 4.1 and Theorem 3.15 immediately.

**PROPOSITION 5.2.** *There exists a constant  $C_5$  such that*

$$\mathbb{E}[(T_{n+1} - T_n)^{\hat{\alpha}} \mid \hat{\mathbf{E}}_n] \leq C_5 \hat{V}(\hat{\mathbf{E}}_n)$$

for any  $\hat{\mathbf{E}}_n$  and any  $n \geq 0$ , where  $\hat{\alpha} = (1 - \alpha)^{-1} = 2 - 2\eta$  for the constant  $\alpha$  given in Theorem 4.1, and  $\hat{V}(\mathbf{E}) = \max\{V(\mathbf{E}), 1\}$ . In particular, if  $n \geq 1$ , then

$$\mathbb{E}[(T_{n+1} - T_n)^{\hat{\alpha}} \mid \hat{\mathbf{E}}_n] \leq C_5 M_0,$$

where  $M_0$  is as in Theorem 4.1.

**PROOF.** By the definition of  $\hat{\mathbf{E}}_n$ ,  $\hat{\mathbf{E}}_n = \mathbf{E}_{T_n}$  is the energy configuration at the stopping time  $T_n$ . Notice that

$$\mathbb{E}[(T_{n+1} - T_n)^{\hat{\alpha}} \mid \hat{\mathbf{E}}_n] \leq 2 \cdot \mathbb{E}_{\mathbf{E}_{T_n}} \left[ \sum_{k=0}^{\tau_G-1} (k+1)^{\hat{\alpha}-1} \right]$$

for the constant  $\alpha$  we use. The proposition follows immediately by applying Theorem 3.15 to  $\mathbf{E}_n$ . Since  $M_0 > 1$ , let

$$\hat{V}(\mathbf{E}) = \max\{1, V(\mathbf{E})\}.$$

Then since  $\hat{V} \leq V + 1$ , it follows from Theorem 4.1 and Lemma 4.7 that

$$\hat{P}\hat{V}(\mathbf{E}) - \hat{V}(\mathbf{E}) \leq -c_0 \hat{V}^\alpha(\mathbf{E}) + (1 + M_0 + CM'e^{hNK})\mathbf{1}_G.$$

The proposition then follows from Theorem 3.15.  $\square$

The following definitions regarding the energy flux in the system are necessary for the proof of Theorem 5.1. Let  $t_1, t_2, \dots$  be the times at which either clock 1 or clock  $N + 1$  rings. The energy in-flow and out-flow on  $[0, T)$  are denoted by

$$F_I([0, T)) = \sum_{0 \leq t_i < T} (W(\mathbf{E}_{t_i^+}) - W(\mathbf{E}_{t_i}))^+$$

and

$$F_O([0, T]) = \sum_{0 \leq t_i < T} (W(\mathbf{E}_{t_i}) - W(\mathbf{E}_{t_i^+}))^+,$$

respectively. Next, we need to estimate the energy flux with respect to  $\hat{\mathbf{E}}$ .

LEMMA 5.3. *There exists a constant  $C_2$  such that*

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T])] \leq C_2 T$$

for any  $\mathbf{E} \in \mathbb{R}_+^N$  and any  $T > 0$ . In addition,

$$X_n := F_I([0, nh]) - C_2 hn$$

is a supermartingale relative to  $\mathcal{F}_n$ , where  $\mathcal{F}_n$  is the  $\sigma$  field generated by  $\{\mathbf{E}_0, \dots, \mathbf{E}_n\}$ .

PROOF. It is easy to see that for any  $\mathbf{E}(t) \in \mathbb{R}_+^N$  we have

$$\begin{aligned} & \mathbb{E}_{\mathbf{E}(t)}[F_I([t, t + dt])] \\ & \leq \sqrt{T_L} \left\{ \int_0^\infty \int_0^1 \frac{1}{T_L} p(x + E_1(t)) e^{-x/T_L} dx dp - E_1 \right\}^+ dt \\ & \quad + \sqrt{T_R} \left\{ \int_0^\infty \int_0^1 \frac{1}{T_R} p(x + E_N(t)) e^{-x/T_R} dx dp - E_N \right\}^+ dt \\ & \leq \sqrt{T_L} \left\{ \int_0^\infty \int_0^1 \frac{1}{T_L} px e^{-x/T_L} dx dp \right\} dt \\ & \quad + \sqrt{T_R} \left\{ \int_0^\infty \int_0^1 \frac{1}{T_R} px e^{-x/T_R} dx dp \right\} dt \\ & = \frac{1}{2} (T_L^{3/2} + T_R^{3/2}) dt := C_2 dt. \end{aligned}$$

This estimate is independent of  $\mathbf{E}(t)$ . Therefore,

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T])] \leq C_2 T$$

for any  $\mathbf{E}$  and  $T$ . In addition,

$$\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = \mathbb{E}[F_I([nh, (n + 1)h]) \mid \mathcal{F}_n] - C_2 h \leq 0$$

for any  $\mathbf{E}_n$ . This completes the proof.  $\square$

By the definition of  $T_1$ ,  $T_1$  is a stopping time relative to  $\mathcal{F}_n$ . We have the following.

PROPOSITION 5.4.

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C_2 h \mathbb{E}_{\mathbf{E}}[T_1],$$

where  $C_2$  is the constant in Lemma 5.3.

PROOF. By Proposition 5.2,  $\mathbb{E}_{\mathbf{E}}[T_1] < \infty$  for any  $\mathbf{E} \in \mathbb{R}_+^N$ . In addition, by Lemma 5.3,

$$\mathbb{E}_{\mathbf{E}}[|X_{n+1} - X_n| \mid \mathcal{F}_n] \leq 2C_2 h < \infty,$$

where the supermartingale  $X_n$  is defined in Lemma 5.3. It then follows from the optional stopping theorem that

$$\mathbb{E}_{\mathbf{E}}[X_{T_1}] \leq \mathbb{E}_{\mathbf{E}}[X_0] = 0.$$

Therefore,

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C_2 h \mathbb{E}_{\mathbf{E}}[T_1]. \quad \square$$

The following estimate about the energy influx with respect to the  $G$ -induced chain follows easily.

LEMMA 5.5. *There exist constants  $C_3 < \infty$  and  $C'_3 < \infty$  such that*

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C_3$$

if  $\mathbf{E} \in G$  and

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C'_3 \hat{V}(\mathbf{E})$$

if  $\mathbf{E} \notin G$ .

PROOF. By Proposition 5.2,

$$\mathbb{E}_{\mathbf{E}}[T_1] \leq \mathbb{E}_{\mathbf{E}}[(T_1)^{\hat{\alpha}}] \leq C_5 \hat{V}(\mathbf{E}).$$

It then follows from Proposition 5.4 that

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C_2 h \cdot C_5 \hat{V}(\mathbf{E}).$$

Let

$$C_3 := C_2 h \cdot C_5 M_0.$$

If  $\mathbf{E}_0 \in G$ , the lemma follows immediately. If  $\mathbf{E}_0 \notin G$ , by letting

$$C'_3 = C_2 h C_5,$$

we will have

$$\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C'_3 \hat{V}(\mathbf{E}).$$

This completes the proof.  $\square$

The following lemma controls the out flow of the energy.

LEMMA 5.6. Assume  $\mathbf{E} \in G$ . There exist constants  $\sigma, M'' > 0$ , such that

$$\mathbb{E}_{\mathbf{E}}[F_O([0, T_1])] \geq \sigma W(\mathbf{E})$$

whenever  $W(\mathbf{E}) > M''$ .

PROOF. Since

$$\mathbb{E}_{\mathbf{E}}[F_O([0, T_1])] \geq \mathbb{E}_{\mathbf{E}}[F_O([0, h])],$$

it is sufficient to construct an event within the time interval  $[0, h)$  such that whenever  $\mathbf{E} \in G$ , a certain proportion of energy can be dumped out of the system.

Let  $E_{n_1}$  be the site that holds the largest amount of energy. Let  $\mathcal{E}$  denote the following event:

- Clocks  $n_1, n_1 - 1, \dots, 2$ , and 1 ring in the time interval  $[0, \frac{h}{n_1}), [\frac{h}{n_1}, \frac{2h}{n_1}), \dots, [\frac{(n_1-1)h}{n_1}, h)$ , respectively.
- At the  $i$ th ring,  $E_{n_1-i+1}$  gives at least half of its energy to  $E_{n_1-i}$  for  $i = 1, \dots, n_1 - 1$ .
- At the  $n_1$ th ring,  $E_1$  dumps  $1/3$  of its energy to the left heat bath.
- Besides what described above, all other clocks do not ring during the time period  $[0, h)$ .

If  $W(\mathbf{E}) \geq 3N \cdot 2^{N-1} T_L$ , we have  $E_{n_1}(0) \geq 3 \cdot 2^{N-1} \cdot T_L$  and  $E_1(\frac{(n_1-1)h}{n_1}) > 3T_L$  conditioning with event  $\mathcal{E}$ . Note that all clock rates are bounded above by  $K$ . In addition, since  $E_{n_1}$  holds the largest amount of energy, right before the  $i$ th ring (for  $i < n_1$ ) we have  $E_{n_1-i+1} \geq 2^{-(i-1)} E_{n_1-i}$ . Hence the probability that  $E_{n_1-i+1}$  gives at least half of its energy to  $E_{n_1-i}$  is at least  $2^{-i}/(1 + 2^{-(i-1)})$ . Therefore, it is a straightforward exercise to check that there exists a constant  $c_0 > 0$  such that

$$\mathbb{P}[\mathcal{E}] \geq c_0 > 0$$

for every  $\mathbf{E} \in G$ . The proof is completed by letting

$$\sigma = \frac{c_0}{3} \cdot 2^{-(N-1)}$$

and

$$M'' = 3N \cdot 2^{N-1} T_L. \quad \square$$

PROOF OF THEOREM 5.1. By the definition of  $W(\mathbf{E})$ , we have

$$\hat{P}W(\mathbf{E}) - W(\mathbf{E}) = \mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] - \mathbb{E}_{\mathbf{E}}[F_O([0, T_1])].$$

Since  $\mathbf{E} \in G$ , by Lemma 5.5, we have  $\mathbb{E}_{\mathbf{E}}[F_I([0, T_1])] \leq C_3$ . On the other hand, by Lemma 5.6, if  $W(\mathbf{E}) > M''$ , we have

$$\mathbb{E}_{\mathbf{E}}[F_O([0, T_1])] \geq \sigma W(\mathbf{E}).$$

Therefore, let  $\delta = \frac{1}{2}\sigma$  and  $M_1 = \max\{2C_3/\sigma, M''\}$ , we have

$$\hat{P}W(\mathbf{E}) \leq (1 - \delta)W(\mathbf{E}).$$

This completes the proof.  $\square$

**6. Proof of the theorems.**

6.1. *Existence of a uniform reference set  $\mathcal{C}$ .* Define

$$\mathcal{C} = \{\mathbf{E} \mid V(\mathbf{E}) \leq M_0, W(\mathbf{E}) \leq M_1\}.$$

The aim of this subsection is to prove that  $\mathcal{C}$  is a uniform reference set. This follows immediate from the theorem below.

**THEOREM 6.1.** *For any  $t > 0$ , there exists a constant  $\eta > 0$  such that*

$$P^t(\mathbf{E}, \cdot) > \eta U_{\mathcal{C}}(\cdot) \quad \text{for all } \mathbf{E} \in \mathcal{C},$$

where  $U_{\mathcal{C}}$  is the uniform probability measure on  $\mathcal{C}$ .

**PROOF.** Let  $e = \inf\{\min(E_1, \dots, E_N) \mid \mathbf{E} = (E_1, \dots, E_N) \in \mathcal{C}\}$ . Clearly  $e > 0$ .

Note that  $\mathcal{C}$  is compact due to the condition involving  $W$ . Then we cover  $\mathcal{C}$  by finitely many disks  $D = D(\bar{\mathbf{E}}, \xi) = \{\mathbf{E} \mid |\mathbf{E} - \bar{\mathbf{E}}| \leq \xi\}$  for  $\xi < e/2$ . It is sufficient to show that for any  $t > 0$ , there exists  $\eta > 0$  independent of  $\bar{\mathbf{E}} \in \mathcal{C}$ , such that for any  $\mathbf{E} \in \mathcal{C}$ ,  $P^t(\mathbf{E}, \cdot) \geq \eta U_D(\cdot)$  for all  $D$  in this cover.

Let  $\bar{\mathbf{E}}$  and  $D$  be fixed. We prescribe the following sequence of events:

(i) On the time intervals  $[\frac{(i-1)t}{2N}, \frac{it}{2N})$  for each  $i = 1, \dots, N - 1$ , site  $i$  exchanges energy with site  $i + 1$ . After the energy exchange, the remaining energy at site  $i$  is between  $e/2$  and  $e$ . Other clocks do not ring during this time period.

(ii) On the time interval  $[\frac{(N-1)t}{2N}, \frac{1}{2})$ , site  $N$  exchanges energy with the right heat bath. After the energy exchange,  $E_N$  is greater than  $\sup_{\bar{\mathbf{E}} \in D} \sum_{i=1}^N (\bar{E}_i + \xi)$ . Other clocks do not ring during this time period.

(iii) On the time intervals  $[\frac{(N+i-1)t}{2N}, \frac{(N+i)t}{2N})$  for each  $i = 1, \dots, N$ , site  $N - i$  exchanges energy with site  $N + 1 - i$ . (Note that site 0 is the left heat bath.) After each energy exchange, the energy left at site  $N + 1 - i$  is uniformly distributed in  $E_{N+1-i} \in [\bar{E}_{N+1-i} - \xi, \bar{E}_{N+1-i} + \xi]$ . Other clocks do not ring during this time period.

It is then easy to check that the event above occurs with probability at least  $\eta$ , where  $\eta > 0$  is independent of  $\mathbf{E}$  provided  $\mathbf{E} \in \mathcal{C}$ .  $\square$

Since  $P^t(\mathbf{E}, \cdot)$  now has positive density everywhere in  $\mathbb{R}_+^N$ , the strong aperiodicity and irreducibility of  $\mathbf{E}_n$  follows immediately.

**COROLLARY 6.2.**  *$\mathbf{E}_n$  is a strongly aperiodic Markov chain.*

**PROOF.** By Theorem 6.1,  $\mathcal{C}$  is a uniform reference set. In addition,  $U_{\mathcal{C}}(\mathcal{C}) > 0$ . The strong aperiodicity follows from its definition.  $\square$

Therefore,  $\mathbf{E}_n$  is aperiodic.

COROLLARY 6.3.  $\mathbf{E}_n$  is  $\lambda$ -irreducible, where  $\lambda$  is the Lebesgue measure on  $\mathbb{R}_+^N$ .

PROOF. Let  $A \subset \mathbb{R}_+^N$  be a set with strictly positive Lebesgue measure. Then there exists a set  $O$  that has the form

$$O = \{\mathbf{E} \mid 0 < c \leq E_i \leq C < \infty, i = 1, \dots, N\}$$

such that  $\lambda(O \cap A) > 0$ .

For any  $\mathbf{E}_0 \in \mathbb{R}_+^N$  and the time step  $h$  as in Theorem 4.1, the same construction as in Theorem 6.1 implies that  $P^h(\mathbf{E}_0, \cdot) > \eta U_O(\cdot)$  for some  $\eta > 0$ . Hence  $P^h(\mathbf{E}_0, A) > \eta U_O(A) > 0$ .  $\square$

6.2. *Absolute continuity of the invariant measure.* This subsection aims to prove the absolute continuity of the invariant probability measure with respect to the Lebesgue measure. For the sake of simplicity, we denote the Lebesgue measure on  $\mathbb{R}_+^N$  by  $\lambda$ .

PROPOSITION 6.4. *If  $\pi$  is an invariant measure of  $\mathbf{E}_t$ , then  $\pi$  is absolutely continuous with respect to  $\lambda$  with a strictly positive density.*

The proof is similar to that of Proposition 6.1 of [24]. For  $\mathbf{E} \in \mathbb{R}_+^N$  and  $t > 0$ , we have decomposition

$$P^t(\mathbf{E}, \cdot) = \nu_\perp + \nu_{abs},$$

where  $\nu_{abs}$  and  $\nu_\perp$  are absolutely continuous and singular component with respect to  $\lambda$ , respectively. We need to show that an absolutely continuous component cannot revert back to singularity as time evolves.

LEMMA 6.5. *For any probability measure  $\mu \ll \lambda$ ,  $\mu P^t \ll \lambda$  for any  $t > 0$ .*

PROOF. This proof is similar to that of Lemma 6.3 of [24]. We include it here for the sake of completeness of this paper.

Let  $t > 0$  be fixed. We define  $l = (c_1, \dots, c_n)$  be the sequence of energy exchanges, where  $c_i = k$  means site  $k$  exchanges energy with site  $k + 1$ . (As before, heat baths are sites 0 and  $N + 1$ .) Let  $S(l)$  be the event that energy exchanges  $(c_1, \dots, c_n)$  occur during the time period  $[0, t)$  in the order specified, and no other energy exchanges occur. If zero (resp., infinite many) energy exchange occurs on  $[0, t)$ , we denote the corresponding event by  $S(\emptyset)$  [resp.,  $S(\infty)$ ]. Obviously,  $\mathbb{P}_{\mathbf{E}}[S(\infty)] = 0$ .

For  $S = S(l)$  or  $S(\emptyset)$ , we define the conditional Markov operator

$$(\mu P_S)(A) := \int_{\mathbb{R}_+^N} \mathbb{P}[\mathbf{E}_t \in A \mid \mathbf{E}_0 = \mathbf{E} \mid S] \mu(d\mathbf{E})$$

and the measure

$$\frac{d\mu_Q}{d\lambda}(\mathbf{E}) := \mathbb{P}_{\mathbf{E}}[S(Q)] \frac{d\mu}{d\lambda}(\mathbf{E})$$

for  $Q = l$  or  $\emptyset$ . Then by the law of total probability,

$$\mu P^t = \mu_{\emptyset} + \sum_l \mu_l P_{S(l)}.$$

Therefore, it is sufficient to show that each term above is absolutely continuous with respect to  $\lambda$ . Since for each  $l = (c_1, \dots, c_n)$ , we have the decomposition

$$P_{S(l)} = P_{S(c_n)} \cdots P_{S(c_1)}.$$

Hence the proof is reduced to proving the absolute continuity of  $\mu P_{c_i}$  for each  $i = 0 \sim N + 1$ , which is a straightforward exercise. Let  $\xi$  and  $\hat{\xi}_k$  be the density of  $\mu$  and  $\mu P_{S(c_k)}$ , respectively. Then if  $0 < k < N$ , we have

$$\begin{aligned} &\hat{\xi}_k(E_1, E_2, \dots, E_N) \\ &= \int_0^1 \xi(E_1, \dots, E_{k-1}, p(E_k + E_{k+1}), (1-p)(E_k + E_{k+1}), E_{k+2}, \dots, E_N) dp. \end{aligned}$$

For  $k = 0$  and  $N$ , we have

$$\begin{aligned} &\hat{\xi}_0(E_1, E_2, \dots, E_N) \\ &= \int_0^\infty \int_{(E_1 - \hat{E}_1)^+}^\infty \xi(\hat{E}_1, E_2, \dots, E_N) \frac{1}{\hat{E}_1 + E} \frac{1}{T_L} e^{-E/T_L} dE d\hat{E}_1 \end{aligned}$$

and

$$\begin{aligned} &\hat{\xi}_N(E_1, E_2, \dots, E_N) \\ &= \int_0^\infty \int_{(E_N - \hat{E}_N)^+}^\infty \xi(E_1, E_2, \dots, \hat{E}_N) \frac{1}{\hat{E}_N + E} \frac{1}{T_R} e^{-E/T_R} dE d\hat{E}_N. \quad \square \end{aligned}$$

**PROOF OF PROPOSITION 6.4.** Let  $\pi = \pi_{abs} + \pi_{\perp}$  be an invariant measure. Assume  $\pi_{\perp} \neq 0$ . For  $t > 0$ ,  $\pi_{abs} P^t \ll \lambda$  by Lemma 6.5. By Theorem 6.1, for any  $\mathbf{E} \in \mathcal{C}$ ,  $P^{t/2}(\mathbf{E}, \cdot)$  has a nonzero absolutely continuous component with respect to the Lebesgue measure, which has a strictly positive density on  $\mathcal{C}$ . Since  $\mathcal{C}$  is accessible within finitely many energy exchanges,  $P^{t/2}(\mathbf{E}, \mathcal{C}) > 0$  for all  $\mathbf{E} \in \mathbb{R}_+^N$ . Hence for all  $\mathbf{E} \in \mathbb{R}_+^N$ ,  $P^t(\mathbf{E}, \cdot)$  has a nonzero absolutely continuous component with respect to the Lebesgue measure, which has a strictly positive density on  $\mathcal{C}$ .

If  $\pi_{\perp} \neq 0$ , there must exist  $M_2, M_3 < \infty$  such that

$$\pi_{\perp}(\{\mathbf{E} \mid V(\mathbf{E}) \leq M_2, W(\mathbf{E}) \leq M_3\}) > 0.$$

Therefore, Theorem 6.1 implies that  $\pi_{\perp} P^t$  must have an absolutely continuous component. The absolutely continuous component of  $\pi P^t$  is strictly larger than that of  $\pi$ . This contradicts to the invariance of  $\pi$ .  $\square$

6.3. *Excursion time before entering  $\mathcal{C}$ .* Now we are ready to estimate the tail of  $\tau_{\mathcal{C}}$ . Let  $\hat{V} = \max\{1, V\}$  and  $\hat{W} = \max\{1, W\}$ .

**THEOREM 6.6.** *For any  $\epsilon > 0$ , there exists a constant  $C_6 < \infty$  such that*

$$\mathbb{P}_{\mathbf{E}}[\tau_{\mathcal{C}} > n] \leq C_6(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E}))n^{-(\hat{\alpha}-\epsilon)}$$

for any  $\mathbf{E} \in \mathbb{R}_+^N$ , where  $\hat{\alpha} = 2 - 2\eta$ .

**PROOF.** Note that  $\mathcal{C} \subset G$ , therefore, one can define  $\hat{\tau}_{\mathcal{C}}$  as the first passage time to  $\mathcal{C}$  for the induced chain  $\hat{\mathbf{E}}_n$ .

Note that  $M_0, M_1 > 1$ . Apply Theorem 3.14 to  $\hat{W} = \max\{1, W(\mathbf{E})\}$ . It follows from Theorem 3.14 and Theorem 5.1 that there exist constants  $C_7 > 0$  and  $r > 1$  such that, for every  $\mathbf{E} \in G$ ,

$$\mathbb{E}_{\mathbf{E}}[r^{\hat{\tau}_{\mathcal{C}}}] \leq \mathbb{E}_{\mathbf{E}}\left[\sum_{k=0}^{\tau_{\mathcal{C}}-1} W(\hat{\mathbf{E}}_k)r^k\right] < \hat{W}(\mathbf{E})C_7.$$

Applying Markov’s inequality to  $r^{\hat{\tau}_{\mathcal{C}}}$ , we have

$$(6.1) \quad \mathbb{P}_{\mathbf{E}}[\hat{\tau}_{\mathcal{C}} > n] < C_7\hat{W}(\mathbf{E})e^{-cn},$$

where  $c = \log r$ . For any given initial condition  $\mathbf{E} \in \mathbb{R}_+^N$ , we have

$$\begin{aligned} \mathbb{P}_{\mathbf{E}}[\hat{\tau}_{\mathcal{C}} > n] &= \int_{\mathbb{R}_+^N} \mathbb{P}[\hat{\tau}_{\mathcal{C}} > n \mid \mathbf{E}_{T_1} = \tilde{E}] \mathbb{P}_{\mathbf{E}}[\mathbf{E}_{T_1} = d\tilde{E}] \\ &\leq \int_{\mathbb{R}_+^N} C_7\hat{W}(\tilde{E})e^{-cn} \mathbb{P}_{\mathbf{E}}[\mathbf{E}_{T_1} = \tilde{E}] \\ &= C_7e^{-cn} \mathbb{E}_{\mathbf{E}}[\hat{W}(\mathbf{E}_{T_1})] \\ &\leq C_7e^{-cn} \cdot (\hat{W}(\mathbf{E}) + \mathbb{E}_{\mathbf{E}}[F_I([0, T_1])]) \\ &\leq C_7e^{-cn} \cdot (\hat{W}(\mathbf{E}) + \max\{C_3, C'_3\hat{V}(\mathbf{E})\}) \\ &\leq C_7e^{-cn} \cdot (\hat{W}(\mathbf{E}) + \max\{C_3, C'_3\}\hat{V}(\mathbf{E})), \end{aligned}$$

where the third line follows from equation (6.1) and the second to last line follows from Lemma 5.5, constants  $C_3$  and  $C'_3$  are as in Lemma 5.5.

By Proposition 5.2, we have

$$\mathbb{E}[(T_{n+1} - T_n)^{\hat{\alpha}} \mid \hat{\mathbf{E}}_n] \leq C_5\hat{V}(\hat{\mathbf{E}}_n),$$

where  $C_5$  is from Proposition 5.2. Without loss of generality, we let  $C_5 \geq 1$ .

Applying Markov’s inequality to  $(T_{n+1} - T_n)^{\hat{\alpha}}$ , we have

$$\mathbb{P}[T_{n+1} - T_n > k \mid \mathbf{E}_{T_n}] \leq C_5\hat{V}(\mathbf{E}_{T_n})k^{-\hat{\alpha}}.$$

If  $n \geq 1$ , since  $\mathbf{E}_{T_n} = \hat{\mathbf{E}}_n \in G$ , we have a uniform bound  $\hat{V}(\hat{\mathbf{E}}_n) \leq M_0$ .

Therefore, assumptions of Theorem 3.11 are satisfied for  $\xi = \hat{V}$  and  $\eta = \hat{W}$ . For each given  $\epsilon > 0$ , notice that  $\hat{V} \geq 1$  and  $\hat{W} \geq 1$ , we have

$$\begin{aligned} \mathbb{P}_{\mathbf{E}}[\tau_{\mathcal{C}} > n] &\leq c(C_5 \hat{V}(\mathbf{E}) + \hat{W}(\mathbf{E}) + \max\{C_3, C'_3\} \hat{V}(\mathbf{E}))n^{-(\hat{\alpha}-\epsilon)} \\ &\leq C_6(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E}))n^{-(\hat{\alpha}-\epsilon)}, \end{aligned}$$

where  $C_6$  is a constant depending on  $\epsilon$  and  $\eta$ . This completes the proof.  $\square$

**PROOFS OF THEOREMS 1 AND 2.** We first prove Theorems 1 and 2 for  $\mathbf{E}_n$ . It follows from Theorem 6.6 that

$$\mathbb{P}_{\mathbf{E}}[\tau_{\mathcal{C}} > n] \leq C_6(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E}))n^{-(\hat{\alpha}-\epsilon)},$$

where  $C_6$  is a constant that depends on both  $\eta$  and  $\epsilon$ . Since  $\mathbb{P}_{\mathbf{E}}[\tau_{\mathcal{C}} > n]$  is at most 1, we have

$$\mathbb{P}_{\mathbf{E}}[\tau_{\mathcal{C}} > n] \leq \max\{1, 2C_6\}(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E}))(n + 1)^{-(\hat{\alpha}-\epsilon)}.$$

Therefore, by Lemma 3.13, there exists a constant  $C_8$  that depends on  $\eta$  and  $\epsilon$ , such that

$$\mathbb{E}_{\mathbf{E}}[\tau_{\mathcal{C}}^{\hat{\alpha}-2\epsilon}] \leq C_8(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E})).$$

Note that  $\hat{\alpha} = 2 - 2\eta$ . Let  $\epsilon = \eta$  and  $\eta = \frac{1}{4}\gamma$ , we have

$$(6.2) \quad \mathbb{E}_{\mathbf{E}_0}[\tau_{\mathcal{C}}^{2-\gamma}] \leq C_8(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E})).$$

By the compactness of  $\mathcal{C}$ , it is easy to see that

$$\sup_{\mathbf{E} \in \mathcal{C}} \mathbb{E}_{\mathbf{E}}[\tau_{\mathcal{C}}^{2-\gamma}] < \infty.$$

Note that  $\hat{W} \leq W + 1$  and  $\hat{V} \leq V + 1$ . Therefore, for any probability measure  $\mu \in \mathcal{M}_{\gamma}$ ,  $C_8(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E}))$  is integrable and

$$(6.3) \quad \mathbb{E}_{\mu}[\tau_{\mathcal{C}}^{2-\gamma}] < \infty.$$

Theorem 1 for  $\mathbf{E}_n$  is then proved by applying Theorem 3.6 to  $\mu, \nu \in \mathcal{M}_{\gamma}$ .

It follows from Corollaries 6.2 and 6.3 that  $\mathbf{E}_n$  is a strongly aperiodic irreducible Markov chain. The existence of an invariant measure  $\pi$  then follows from

$$\sup_{\mathbf{E} \in \mathcal{C}} \mathbb{E}_{\mathbf{E}}[\tau_{\mathcal{C}}] < \sup_{\mathbf{E} \in \mathcal{C}} \mathbb{E}_{\mathbf{E}}[\tau_{\mathcal{C}}^{2-\gamma}] < \infty$$

and Theorem 3.5. The absolute continuity of  $\pi$  comes from Proposition 6.4. The  $n^{-(1-\gamma)}$  speed of convergence to  $\pi$  is given by the existence of  $\pi$  and Theorem 3.6.

It remains to prove the uniqueness of  $\pi$ . We prove the uniqueness of  $\pi$  for any  $\mathbf{E}_t$  instead of  $\mathbf{E}_n$ .

Recall the proof of Proposition 6.4, for any  $t > 0$  and  $\mathbf{E} \in \mathbb{R}_+^N$ ,  $P^t(\mathbf{E}, \cdot)$  has a strictly positive density on  $\mathcal{C}$ . This implies  $\mathcal{C}$  belongs to the support of any invariant probability measure. However, any two distinct ergodic invariant probability measures must be mutually singular. In addition, every invariant probability measure must be a convex combination of ergodic invariant measures. Hence  $\mathbf{E}_t$  has at most one invariant probability measure, which must be  $\pi$  (see, e.g., Theorem 1.7 of [16]).

This completes the proof of Theorem 2 for  $\mathbf{E}_n$ .  $\square$

The return time argument for  $\mathbf{E}_n$  also help us to obtain the tail of a marginal distribution of  $\pi$ . Since  $\pi$  is absolutely continuous with respect to the Lebesgue measure, any marginal distribution of  $\pi$  also has absolute continuity. Let  $\rho_i(E)$  be the density of the marginal distribution of  $\pi$  with respect to site  $i$ . Let

$$q_i(E) = \int_0^E \rho_i(\hat{E}) d\hat{E} = \mathbb{P}_\pi[E_i < E]$$

be the marginal distribution function. The following lemma holds.

LEMMA 6.7. *For any  $i = 1, \dots, N$  and any sufficiently small  $\gamma > 0$ , there exists  $0 < \delta < 1$  such that*

$$q_i(E) \geq E^{1/2+\gamma}$$

if  $0 < E < \delta$ .

PROOF. Define sets  $A = \{(E_1, \dots, E_N) \mid 1 \leq E_i \leq 2 \text{ for all } i\}$  and  $B_i(E) = \{(E_1, \dots, E_N) \mid E_i \leq E\}$ . It is then well known that  $q_i(E) = \pi(B_i(E))$  is equal to the expected occupation time for  $\mathbf{E}_n$  on  $B_i(E)$ , that is,

$$\pi(B_i(E)) = \int_A \pi(dy) \mathbb{E}_y \left[ \sum_{k=0}^{\tau_A-1} \mathbf{1}_{\{\mathbf{E}_k \in B_i(E)\}} \right]$$

(Theorem 10.4.9 of [29]). Let  $h$  be the time step size when defining  $\mathbf{E}_n$  in Theorem 4.1. We have

$$\pi(B_i(E)) \geq \int_A \pi(dy) \mathbb{P}_y[S_1] \mathbb{P}[S_2 \mid S_1] \mathbb{P}[S_3 \mid S_1, S_2] E^{-1/2},$$

where  $S_1, S_2, S_3$  are the following three events:

$$S_1 = \{\text{clock } i \text{ rings exactly once at } t_0 < h, \text{ all other clocks are silent}\},$$

$$S_2 = \{E_i(t_0^+) \in (0, E) \text{ after } S_1 \text{ occurs}\},$$

and

$$S_3 = \{\text{after } t = h, \text{ no energy exchange involves } E_i \text{ before } t = h + \lfloor E^{-1/2} \rfloor\}.$$

Then it is easy to see that  $P_y[S_1]$  is uniformly positive for  $y \in A$ ,  $\mathbb{P}[S_2 | S_1] > E/4$  is independent of the initial condition, and  $\mathbb{P}[S_3 | S_1, S_2] \geq e^{-1}$  for any initial condition because  $E_i(h) < E$ . In addition, we have  $\pi(A) > 0$ . Hence for all sufficiently small  $E > 0$ , we have

$$q_i(E) \geq cE^{1/2}$$

for some constant  $c > 0$  that is independent of  $E$  and  $i$ . This completes the proof. □

Theorems 1 and 2 for  $\mathbf{E}_t$ . The last step is to pass results from  $\mathbf{E}_n$  to  $\mathbf{E}_t$ . The contraction of the Markov operator is easy to pass because we have

$$\|\mu P^t - \nu P^t\|_{TV} = \|(\mu P^{\lfloor \frac{t}{h} \rfloor h} - \nu P^{\lfloor \frac{t}{h} \rfloor h}) P^{(t - \lfloor \frac{t}{h} \rfloor h)}\|_{TV} \leq \|\mu P^{\lfloor \frac{t}{h} \rfloor h} - \nu P^{\lfloor \frac{t}{h} \rfloor h}\|_{TV}.$$

It remains to show that  $\pi$ , the invariant probability measure of  $\mathbf{E}_n$ , is invariant for any  $\mathbf{E}_t$ ,  $t > 0$ .

LEMMA 6.8.  $\pi P^t = \pi$  for any  $t > 0$ .

PROOF. Note that the argument in Sections 4 and 5 works for all sufficiently small time steps. Let  $h$  be the time step we have chosen for  $\mathbf{E}_n$ . For any  $r < h$ ,  $P^r$  also admits an invariant measure  $\pi_r$ . It is then sufficient to show that  $\pi_r = \pi$  because any  $t$  can be written as

$$t = \left\lfloor \frac{t}{h} \right\rfloor \cdot h + r$$

for some  $r < h$ .

In addition, we have the “continuity at zero,”

$$\|\pi_r P^\delta - \pi_r\|_{TV} \rightarrow 0 \quad \text{as } \delta \rightarrow 0$$

because all clock rates are less than  $K$ .

Without loss of generality, assume  $r/h \notin \mathbb{Q}$ . By the density of orbits in irrational rotations, there exist sequences  $i_n, j_n \in \mathbb{Z}^+$ , such that  $\delta_n := h - \frac{i_n}{j_n}r \rightarrow 0$  from right. Then

$$\pi_r P^h = \pi_r P^{\frac{i_n}{j_n}r} P^{\delta_n} = \pi_r P^{\delta_n} \rightarrow \pi_r$$

by the “continuity at zero.”

Hence  $\pi_r$  is invariant with respect to  $P^h$ . By uniqueness,  $\pi_r = \pi$ . □

Therefore, Theorems 1 and 2 also hold for  $\mathbf{E}_t$ .

REMARK. We expect Theorem 1 to be close to optimal. Let  $\mathcal{C}$  be a uniform reference set. Then it is easy to see that there exists an  $\epsilon > 0$  such that  $B_\epsilon = \{\mathbf{E} \mid E_i < \epsilon \text{ for some } i = 1, \dots, N\}$  is disjoint with  $\mathcal{C}$ . Let  $D = \{\mathbf{E} \mid E_i < L\}$  for some large  $L$ . Then it is easy to see that for any  $\mu$  and  $\nu$  that have uniformly positive density on  $B_\epsilon \cap D$ , we have  $\mu(B_{t^{-2}}) \sim O(t^{-2})$  and  $\nu(B_{t^{-2}}) \sim O(t^{-2})$  for  $t \gg 1$ . When  $E_i < t^{-2}$ , the probability that no energy exchange occurs between site  $i - 1$  and  $i$ , or between site  $i$  and  $i + 1$ , before time  $t$  is  $O(1)$ . Therefore, we have the lower bound on tails  $\mathbb{P}_\mu[\tau_{\mathcal{C}} > t] \geq O(1) \cdot t^{-2}$  (and  $\mathbb{P}_\nu[\tau_{\mathcal{C}} > t] \geq O(1) \cdot t^{-2}$ ). This implies the coupling time  $T$  has the tail

$$\mathbb{P}_{\mu, \nu}[T > t] \geq O(1) \cdot t^{-2}.$$

This is consistent with our numerical result in [22] that the tail of  $\mathbb{P}[\tau_{\mathcal{C}} > t]$  is  $\sim t^{-2}$ . Similar argument leads to the proof of Proposition 4.

PROOF OF THEOREM 3. The following calculation is straightforward:

$$\begin{aligned} & \left| \int (P^t \zeta)(\mathbf{E}) \xi(\mathbf{E}) \mu(d\mathbf{E}) - \int (P^t \zeta)(\mathbf{E}) \mu(d\mathbf{E}) \int \xi(\mathbf{E}) \mu(d\mathbf{E}) \right| \\ &= \left| \int \xi(\mathbf{E}) \left( (P^t \zeta)(\mathbf{E}) - \int (P^t \zeta)(\mathbf{Z}) \mu(d\mathbf{Z}) \right) \mu(d\mathbf{E}) \right| \\ &\leq \|\xi\|_{L^\infty} \|\zeta\|_{L^\infty} \int \|\delta_{\mathbf{E}} P^t - \mu P^t\|_{TV} \mu(d\mathbf{E}). \end{aligned}$$

It then follows from Corollary 3.10 and equation (6.2) that

$$\|\delta_{\mathbf{E}} P^t - \mu P^t\|_{TV} \leq C(\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E}) + C_\mu)([t])^{\gamma-2}$$

for some  $C, C_\mu < \infty$  that is independent of  $\mathbf{E}$ . Since  $\hat{W}(\mathbf{E}) + \hat{V}(\mathbf{E})$  is  $\mu$ -integrable, we have

$$\begin{aligned} & \left| \int (P^t \zeta)(\mathbf{E}) \xi(\mathbf{E}) \mu(d\mathbf{E}) - \int (P^t \zeta)(\mathbf{E}) \mu(d\mathbf{E}) \int \xi(\mathbf{E}) \mu(d\mathbf{E}) \right| \\ &\leq O(1) \cdot \|\xi\|_{L^\infty} \|\zeta\|_{L^\infty} t^{\gamma-2}, \end{aligned}$$

where the  $O(1)$  term depends on  $\gamma, N$  and  $\mu$ .  $\square$

PROOF OF PROPOSITION 4. Let  $\nu$  be a probability measure that satisfies the following properties:

- $\nu$  is absolutely continuous with respect to  $\pi$ .
- Let  $B_\epsilon = \{\mathbf{E} = (E_1, \dots, E_N) \in \mathbb{R}_+^N \mid E_1 < \epsilon\}$  for some fixed small  $\epsilon > 0$ .  $\nu$  satisfies  $d\nu/d\pi = 4$  on  $B_\epsilon$ .

Such  $\nu$  must exist because  $\pi$  is absolutely continuous with respect the Lebesgue measure. Hence we can always find a small  $\epsilon > 0$  such that  $\pi(B_\epsilon) < \frac{1}{4}$ .

Then for  $t > 0$ , we have

$$\begin{aligned} \|\nu P^t - \pi\|_{\text{TV}} &\geq \|(\nu P^t)(B_{t-2}) - \pi(B_{t-2})\|_{\text{TV}} \geq (\nu P^t)(B_{t-2}) - \pi(B_{t-2}) \\ &\geq \mathbb{P}_\nu[\text{clock 1 does not ring before } t, \mathbf{E}_0 \in B_{t-2}] - \pi(B_{t-2}) \\ &= \nu(B_{t-2}) \mathbb{P}_{\nu|_{B_{t-2}}}[\text{clock 0 and 1 does not ring before } t] - \pi(B_{t-2}) \\ &\geq \nu(B_{t-2})e^{-1} - \pi(B_{t-2}) = (4e^{-1} - 1)\pi(B_{t-2}), \end{aligned}$$

where  $\nu|_{B_{t-2}}$  is the restricted probability measure  $\nu$  on  $B_{t-2}$ . Let  $\gamma > 0$  be a sufficiently small number. Apply Lemma 6.7 to the marginal distribution function  $q_1(E)$  and small parameter  $\gamma/2$ . There should exist a  $T < \infty$  such that

$$\pi(B_{t-2}) = q_1(t^{-2}) \geq t^{-1-\gamma}$$

for any  $t > T$ . Hence

$$\|\nu P^t - \pi\|_{\text{TV}} \geq (4e^{-1} - 1)(1+t)^{-1-\gamma}, \quad t > T.$$

In addition, we can always find  $c > 0$  such that

$$\|\nu P^t - \pi\|_{\text{TV}} \geq c(1+t)^{-1-\gamma}$$

for all  $0 \leq t \leq T$ . The proof is completed by combining the two estimates.  $\square$

**REMARK.** The result of Lemma 6.7 implies that the invariant probability measure  $\pi$  may not belong to  $\mathcal{M}_\eta$  for sufficiently small  $\eta > 0$ . As a result, the initial probability distribution  $\nu$  constructed in the proof of Proposition 4 may not be in  $\mathcal{M}_\eta$  either. Our numerical simulation shows that the lower bound of convergence holds for many initial probability distributions within the measure class  $\mathcal{M}_\eta$  as well. But a rigorous proof requires many detailed properties of  $\pi$ , which turns out to be very difficult due to the nonequilibrium nature of the system.

**Acknowledgments.** The author would like to thank Lai-Sang Young, Jonathan Mattingly and Martin Hairer for many enlightening discussions.

## REFERENCES

- [1] BALADI, V., DEMERS, M. F. and LIVERANI, C. (2018). Exponential decay of correlations for finite horizon Sinai billiard flows. *Invent. Math.* **211** 39–177. [MR3742756](#)
- [2] BARBU, V. S. and LIMNIOS, N. (2008). *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: Their Use in Reliability and DNA Analysis. Lecture Notes in Statistics* **191**. Springer, New York. [MR2452304](#)
- [3] BERNARDIN, C. and OLLA, S. (2005). Fourier’s law for a microscopic model of heat conduction. *J. Stat. Phys.* **121** 271–289. [MR2185330](#)
- [4] BUNIMOVICH, L., LIVERANI, C., PELLEGRINOTTI, A. and SUHOV, Y. (1992). Ergodic systems of  $n$  balls in a billiard table. *Comm. Math. Phys.* **146** 357–396. [MR1165188](#)
- [5] CHERNOV, N. (1999). Decay of correlations and dispersing billiards. *J. Stat. Phys.* **94** 513–556. [MR1675363](#)

- [6] CHERNOV, N. and MARKARIAN, R. (2006). *Chaotic Billiards. Mathematical Surveys and Monographs* **127**. Amer. Math. Soc., Providence, RI. [MR2229799](#)
- [7] CHERNOV, N. and YOUNG, L. S. (2000). Decay of correlations for Lorentz gases and hard balls. In *Hard Ball Systems and the Lorentz Gas. Encyclopaedia Math. Sci.* **101** 89–120. Springer, Berlin. [MR1805327](#)
- [8] CUNEO, N. and ECKMANN, J.-P. (2016). Non-equilibrium steady states for chains of four rotors. *Comm. Math. Phys.* **345** 185–221. [MR3509013](#)
- [9] CUNEO, N., ECKMANN, J.-P. and POQUET, C. (2015). Non-equilibrium steady state and subgeometric ergodicity for a chain of three coupled rotors. *Nonlinearity* **28** 2397. [MR3366649](#)
- [10] DOLGOPYAT, D. and LIVERANI, C. (2011). Energy transfer in a fast-slow Hamiltonian system. *Comm. Math. Phys.* **308** 201–225. [MR2842975](#)
- [11] ECKMANN, J.-P., PILLET, C.-A. and REY-BELLET, L. (1999). Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.* **201** 657–697. [MR1685893](#)
- [12] ECKMANN, J.-P. and YOUNG, L.-S. (2006). Nonequilibrium energy profiles for a class of 1-D models. *Comm. Math. Phys.* **262** 237–267. [MR2200889](#)
- [13] GASPARD, P. and GILBERT, T. (2008). Heat conduction and Fourier’s law in a class of many particle dispersing billiards. *New J. Phys.* **10** Art. ID 103004.
- [14] GASPARD, P. and GILBERT, T. (2008). Heat conduction and Fourier’s law by consecutive local mixing and thermalization. *Phys. Rev. Lett.* **101** Art. ID 020601.
- [15] GRIGO, A., KHANIN, K. and SZÁSZ, D. (2012). Mixing rates of particle systems with energy exchange. *Nonlinearity* **25** 2349. [MR2956578](#)
- [16] HAIRER, M. (2010). Convergence of Markov processes. Lecture notes.
- [17] HAYDN, N., LACROIX, Y. and VAIENTI, S. (2005). Hitting and return times in ergodic dynamical systems. *Ann. Probab.* **33** 2043–2050. [MR2165587](#)
- [18] JARNER, S. F. and ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. [MR1890063](#)
- [19] KIPNIS, C., MARCHIORO, C. and PRESUTTI, E. (1982). Heat flow in an exactly solvable model. *J. Stat. Phys.* **27** 65–74. [MR0656869](#)
- [20] LEFEVERE, R. and ZAMBOTTI, L. (2010). Hot scatterers and tracers for the transfer of heat in collisional dynamics. *J. Stat. Phys.* **139** 686–713. [MR2638933](#)
- [21] LI, Y. (2015). On the stochastic behaviors of locally confined particle systems. *Chaos* **25** Art. ID 073121. [MR3405854](#)
- [22] LI, Y. and XU, H. (2017). Numerical simulation of polynomial-speed convergence phenomenon. *J. Stat. Phys.* **169** 697–729. [MR3715860](#)
- [23] LI, Y. and YOUNG, L.-S. (2013). Existence of nonequilibrium steady state for a simple model of heat conduction. *J. Stat. Phys.* **152** 1170–1193. [MR3101481](#)
- [24] LI, Y. and YOUNG, L.-S. (2014). Nonequilibrium steady states for a class of particle systems. *Nonlinearity* **27** 607.
- [25] LI, Y. and YOUNG, L.-S. (2017). Polynomial convergence to equilibrium for a system of interacting particles. *Ann. Appl. Probab.* **27** 65–90.
- [26] LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover Publications, Inc., Mineola, NY. Corrected reprint of the 1992 original. [MR1924231](#)
- [27] LIVERANI, C. and OLLA, S. (2012). Toward the Fourier law for a weakly interacting anharmonic crystal. *J. Amer. Math. Soc.* **25** 555–583. [MR2869027](#)
- [28] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. [MR1234295](#)
- [29] MEYN, S. P. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge Univ. Press, Cambridge.

- [30] NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43** 309–318.
- [31] NUMMELIN, E. and TUOMINEN, P. (1983). The rate of convergence in Orey’s theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process. Appl.* **15** 295–311. [MR0711187](#)
- [32] REY-BELLET, L. (2005). Nonequilibrium statistical mechanics of open classical systems. In *XIVth International Congress on Mathematical Physics* 447–454. World Sci. Publ., Hackensack, NJ. [MR2227858](#)
- [33] REY-BELLET, L. and THOMAS, L. E. (2000). Asymptotic behavior of thermal nonequilibrium steady states for a driven chain of anharmonic oscillators. *Comm. Math. Phys.* **215** 1–24. [MR1799873](#)
- [34] REY-BELLET, L. and THOMAS, L. E. (2002). Fluctuations of the entropy production in anharmonic chains. *Ann. Henri Poincaré* **3** 483–502. [MR1915300](#)
- [35] RUELLE, D. (2012). A mechanical model for Fourier’s law of heat conduction. *Comm. Math. Phys.* **311** 755–768. [MR2909762](#)
- [36] SASADA, M. (2015). Spectral gap for stochastic energy exchange model with nonuniformly positive rate function. *Ann. Probab.* **43** 1663–1711. [MR3353812](#)
- [37] YARMOLA, T. (2013). Sub-exponential mixing of random billiards driven by thermostats. *Nonlinearity* **26** 1825.
- [38] YARMOLA, T. (2014). Sub-exponential mixing of open systems with particle-disk interactions. *J. Stat. Phys.* **156** 473–492. [MR3217533](#)

DEPARTMENT OF MATHEMATICS AND STATISTICS  
UNIVERSITY OF MASSACHUSETTS AMHERST  
AMHERST, MASSACHUSETTS 01003  
USA  
E-MAIL: [yaoli@math.umass.edu](mailto:yaoli@math.umass.edu)