Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Quenched LDP for words in a letter sequence

Matthias Birkner[1 2], Andreas Greven[3], Frank den Hollander[4 5]

submitted: 24 July 2008

[1] Weierstraß-Institut für Angewandte Analysis und Stochastik,
Mohrenstr. 39, 10117 Berlin, Germany

[2] Institut für Angewandte Mathematik, Universität Bonn,
Wegelerstr. 6, 53115 Bonn, Germany

[3] Mathematisches Institut, Universität Erlangen-Nürnberg,
Bismarckstrasse $1\frac{1}{2}$, 91054 Erlangen, Germany

[4] Mathematical Institute, Leiden University, P.O. Box 9512,
2300 RA Leiden, The Netherlands

[5] EURANDOM, P.O. Box 513,
5600 MB Eindhoven, The Netherlands

No. 1349
Berlin 2008

2000 *Mathematics Subject Classification.* 60F10, 60G10, 82D60.

Key words and phrases. Letters and words, renewal process, empirical process, annealed vs. quenched, large deviation principle, rate function, specific relative entropy, collision local time.

This work was supported in part by DFG and NWO through the Dutch-German Bilateral Research Group "Mathematics of Random Spatial Models from Physics and Biology". MB and AG are grateful for hospitality at EURANDOM.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

When we cut an i.i.d. sequence of letters into words according to an independent renewal process, we obtain an i.i.d. sequence of words. In the *annealed* large deviation principle (LDP) for the empirical process of words, the rate function is the specific relative entropy of the observed law of words w.r.t. the reference law of words. In the present paper we consider the *quenched* LDP, i.e., we condition on a typical letter sequence. We focus on the case where the renewal process has an *algebraic* tail. The rate function turns out to be a sum of two terms, one being the annealed rate function, the other being proportional to the specific relative entropy of the observed law of letters w.r.t. the reference law of letters, with the former being obtained by concatenating the words and randomising the location of the origin. The proportionality constant equals the tail exponent of the renewal process. Earlier work by Birkner considered the case where the renewal process has an exponential tail, in which case the rate function turns out to be the first term on the set where the second term vanishes and to be infinite elsewhere.

We apply our LDP to prove that the radius of convergence of the moment generating function of the collision local time of two strongly transient random walks on $\mathbb{Z}^d$, $d \geq 1$, strictly increases when we condition on one of the random walks, both in discrete time and in continuous time. The presence of these gaps implies the existence of an *intermediate phase* for the long-time behaviour of a class of coupled branching processes, interacting diffusions, respectively, directed polymers in random environments.

1 Introduction and main results

1.1 Problem setting

Let E be a finite set of *letters*. Let $\widetilde{E} = \cup_{n\in\mathbb{N}} E^n$ be the set of finite *words* drawn from E. Both E and $\widetilde{E}$ are Polish spaces under the discrete topology. Let $\mathcal{P}(E^{\mathbb{N}})$ and $\mathcal{P}(\widetilde{E}^{\mathbb{N}})$ denote the set of probability measures on sequences drawn from E, respectively, $\widetilde{E}$, equipped with the topology of weak convergence. Write θ and $\widetilde{\theta}$ for the left-shift acting on $E^{\mathbb{N}}$, respectively, $\widetilde{E}^{\mathbb{N}}$. Write $\mathcal{P}^{\mathrm{inv}}(E^{\mathbb{N}}), \mathcal{P}^{\mathrm{erg}}(E^{\mathbb{N}})$ and $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}), \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ for the set of probability measures that are invariant and ergodic under θ, respectively, $\widetilde{\theta}$.

For $\nu \in \mathcal{P}(E)$, let $X = (X_i)_{i\in\mathbb{N}}$ be i.i.d. with law ν. Without loss of generality we will assume that $\mathrm{supp}(\nu) = E$ (otherwise we replace E by $\mathrm{supp}(\nu)$). For $\rho \in \mathcal{P}(\mathbb{N})$, let $\tau = (\tau_i)_{i\in\mathbb{N}}$ be i.i.d. with law ρ having infinite support and satisfying the *algebraic tail property*

$$\lim_{\substack{n\to\infty \\ \rho(n)>0}} \frac{\log \rho(n)}{\log n} =: -\alpha, \quad \alpha \in (1,\infty). \tag{1.1}$$

(No regularity assumption will be necessary for $\mathrm{supp}(\rho)$.) Assume that X and τ are independent and write $\mathbb{P}$ to denote their joint law. Cut words out of X according to τ, i.e., put (see Figure 1)

$$T_0 := 0 \quad \text{and} \quad T_i := T_{i-1} + \tau_i, \quad i \in \mathbb{N}, \tag{1.2}$$

and let

$$Y^{(i)} := \big(X_{T_{i-1}+1}, X_{T_{i-1}+2}, \dots, X_{T_i}\big), \quad i \in \mathbb{N}. \tag{1.3}$$

Then, under the law $\mathbb{P}$, $Y = (Y^{(i)})_{i\in\mathbb{N}}$ is an i.i.d. sequence of words with marginal law $q_{\rho,\nu}$ on $\widetilde{E}$ given by

$$\begin{aligned} &q_{\rho,\nu}\big((x_1,\dots,x_n)\big) := \mathbb{P}\big(Y^{(1)} = (x_1,\dots,x_n)\big) = \rho(n)\,\nu(x_1)\cdots\nu(x_n), \\ &n \in \mathbb{N},\, x_1,\dots,x_n \in E. \end{aligned} \tag{1.4}$$

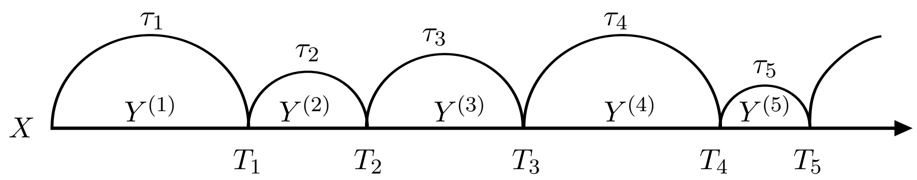

Figure 1: Cutting words from a letter sequence according to a renewal process.

For $N \in \mathbb{N}$, let $(Y^{(1)},\dots,Y^{(N)})^{\mathrm{per}}$ stand for the periodic extension of $(Y^{(1)},\dots,Y^{(N)})$ to an element of $\widetilde{E}^{\mathbb{N}}$, and define

$$R_N := \frac{1}{N}\sum_{i=0}^{N-1} \delta_{\widetilde{\theta}^i(Y^{(1)},\dots,Y^{(N)})^{\mathrm{per}}} \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}), \tag{1.5}$$

the *empirical process of N-tuples of words*. By the ergodic theorem, we have

$$\mathrm{w}-\lim_{N\to\infty} R_N = q_{\rho,\nu}^{\otimes\mathbb{N}} \quad \mathbb{P}\text{–a.s.}, \tag{1.6}$$

with $\mathrm{w}-\lim$ denoting the weak limit. The following large deviation principle (LDP) is standard (see e.g. Dembo and Zeitouni [10], Corollaries 6.5.15 and 6.5.17). Let

$$H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) := \lim_{N\to\infty} \frac{1}{N}\, h\left(Q_{|\mathscr{F}_N} \mid (q_{\rho,\nu}^{\otimes\mathbb{N}})_{|\mathscr{F}_N}\right) \in [0,\infty] \tag{1.7}$$

be the *specific relative entropy of Q w.r.t. $q_{\rho,\nu}^{\otimes\mathbb{N}}$*, where $\mathscr{F}_N = \sigma(Y^{(1)},\dots,Y^{(N)})$ is the sigma-algebra generated by the first N words, $Q_{|\mathscr{F}_N}$ is the restriction of Q to $\mathscr{F}_N$, and $h(\,\cdot \mid \cdot\,)$ denotes relative entropy. (For general properties of entropy, see Walters [25], Chapter 4.)

Theorem 1.1. [Annealed LDP] *The family of probability distributions $\mathbb{P}(R_N \in \cdot)$, $N \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with rate N and with rate function $I^{\mathrm{ann}}\colon \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \to [0,\infty]$ given by*

$$I^{\mathrm{ann}}(Q) = H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}). \tag{1.8}$$

This rate function is lower semi-continuous, has compact level sets, has a unique zero at $Q = q_{\rho,\nu}^{\otimes\mathbb{N}}$, and is affine.

The LDP for R_N arises from the LDP for N-tuples via a projective limit theorem. The ratio under the limit in (1.7) is the rate function for N-tuples according to Sanov's theorem (see e.g. den Hollander [17], Section II.5), and is non-decreasing in N.

1.2 Main theorems

Our aim in the present paper is to derive the LDP for $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$. To state our result, we need some more notation.

Let $\kappa\colon \widetilde{E}^{\mathbb{N}} \to E^{\mathbb{N}}$ denote the *concatenation map* that glues a sequence of words into a sequence of letters. For $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ such that

$$m_Q := \mathbb{E}_Q[\tau_1] < \infty, \tag{1.9}$$

define $\Psi_Q \in \mathcal{P}^{\mathrm{inv}}(E^{\mathbb{N}})$ as

$$\Psi_Q(\cdot) := \frac{1}{m_Q}\mathbb{E}_Q\left[\sum_{k=0}^{\tau_1-1}\delta_{\theta^k\kappa(Y)}(\cdot)\right]. \tag{1.10}$$

Think of Ψ_Q as *the shift-invariant version of the concatenation of* Y *under the law* Q *obtained after randomising the location of the origin.*

For $\mathrm{tr} \in \mathbb{N}$, let $[\cdot]_{\mathrm{tr}}\colon\ \widetilde{E} \to [\widetilde{E}]_{\mathrm{tr}} := \cup_{n=1}^{\mathrm{tr}} E^n$ denote the *word length truncation* map defined by

$$y = (x_1, \ldots, x_n) \mapsto [y]_{\mathrm{tr}} := (x_1, \ldots, x_{n\wedge \mathrm{tr}}), \qquad n \in \mathbb{N},\ x_1, \ldots, x_n \in E. \tag{1.11}$$

Extend this to a map from $\widetilde{E}^{\mathbb{N}}$ to $[\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}}$ via

$$\big[(y^{(1)}, y^{(2)}, \ldots)\big]_{\mathrm{tr}} := \big([y^{(1)}]_{\mathrm{tr}}, [y^{(2)}]_{\mathrm{tr}}, \ldots\big) \tag{1.12}$$

and to a map from $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}([\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}})$ via

$$[Q]_{\mathrm{tr}}(A) := Q(\{z \in \widetilde{E}^{\mathbb{N}}\colon\ [z]_{\mathrm{tr}} \in A\}), \qquad A \subset [\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}} \text{ measurable.} \tag{1.13}$$

Note that if $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, then $[Q]_{\mathrm{tr}}$ is an element of the set

$$\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}) = \{Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon\ m_Q < \infty\}. \tag{1.14}$$

Theorem 1.2. [Quenched LDP] *Assume* (1.1). *Then, for* $\nu^{\otimes\mathbb{N}}$*–a.s. all* X*, the family of (regular) conditional probability distributions* $\mathbb{P}(R_N \in \cdot \mid X)$*,* $N \in \mathbb{N}$*, satisfies the LDP on* $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ *with rate* N *and with deterministic rate function* $I^{\mathrm{que}}\colon\ \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \to [0,\infty]$ *given by*

$$I^{\mathrm{que}}(Q) := \begin{cases} I^{\mathrm{fin}}(Q), & \text{if } Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}),\\ \lim\limits_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}\big([Q]_{\mathrm{tr}}\big), & \text{otherwise,} \end{cases} \tag{1.15}$$

where

$$I^{\mathrm{fin}}(Q) := H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1)\, m_Q\, H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}). \tag{1.16}$$

Theorem 1.3. *The rate function* I^{que} *is lower semi-continuous, has compact level sets, has a unique zero at* $Q = q_{\rho,\nu}^{\otimes\mathbb{N}}$*, and is affine. Moreover, it is equal to the lower semi-continuous extension of* I^{fin} *from* $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ *to* $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$*.*

Theorem 1.2 will be proved in Sections 3–5, Theorem 1.3 in Section 6.

A remarkable aspect of (1.16) in relation to (1.8) is that it *quantifies* the difference between the quenched and the annealed rate function. Note the appearance of the tail exponent α. This will be important in the applications described in Sections 1.4–1.5. We have not been able to find a simple formula for $I^{\mathrm{que}}(Q)$ when $m_Q = \infty$. In Appendix A we will show that the truncation map is continuous on all of $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, i.e.,

$$I^{\mathrm{ann}}(Q) = \lim_{\mathrm{tr}\to\infty} I^{\mathrm{ann}}([Q]_{\mathrm{tr}}), \qquad I^{\mathrm{que}}(Q) = \lim_{\mathrm{tr}\to\infty} I^{\mathrm{que}}([Q]_{\mathrm{tr}}), \qquad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{1.17}$$

Theorem 1.2 is an extension of Birkner [3], Theorem 1. In that paper, the quenched LDP is derived under the assumption that the law ρ satisfies the *exponential tail property*

$$\exists\, C < \infty,\, \lambda > 0\colon \quad \rho(n) \leq C e^{-\lambda n} \quad \forall\, n \in \mathbb{N} \tag{1.18}$$

(which includes the case where $\mathrm{supp}(\rho)$ is finite). The rate function governing the LDP is given by

$$I^{\mathrm{que}}(Q) := \begin{cases} H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}), & \text{if } Q \in \mathscr{R}_\nu, \\ \infty, & \text{if } Q \notin \mathscr{R}_\nu, \end{cases} \tag{1.19}$$

where

$$\mathscr{R}_\nu := \left\{ Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon \ \mathrm{w-}\lim_{L\to\infty} \frac{1}{L} \sum_{k=0}^{L-1} \delta_{\theta^k \kappa(Y)} = \nu^{\otimes\mathbb{N}} \ Q-a.s. \right\}. \tag{1.20}$$

Think of $\mathscr{R}_\nu$ as the set of those Q's for which *the concatenation of words has the same statistical properties as the letter sequence* X. This set is not closed in the weak topology: its closure is $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$.

We can include the cases where ρ satisfies (1.1) with $\alpha = 1$ or $\alpha = \infty$.

Theorem 1.4. (a) *If $\alpha = 1$, then the quenched LDP holds with $I^{\mathrm{que}} = I^{\mathrm{ann}}$ given by* (1.8).
(b) *If $\alpha = \infty$, then the quenched LDP holds with rate function*

$$I^{\mathrm{que}}(Q) = \begin{cases} H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) & \textit{if } \lim_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}}) = 0, \\ \infty & \textit{otherwise.} \end{cases} \tag{1.21}$$

Theorem 1.4 will be proved in Section 7. Part (a) says that the quenched and the annealed rate function are identical when $\alpha = 1$. Part (b) says that (1.19) can be viewed as the limiting case of (1.16) as $\alpha \to \infty$. Indeed, it was shown in Birkner [3], Lemma 2, that on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$:

$$\Psi_Q = \nu^{\otimes\mathbb{N}} \text{ if and only if } Q \in \mathscr{R}_\nu. \tag{1.22}$$

Hence, (1.21) and (1.19) agree on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, and the rate function (1.21) is the lower semicontinuous extension of (1.19) to $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$. Note that by Lemma 7 in Birkner [3], the expressions in (1.21) and (1.19) are identical if ρ has exponentially decaying tails. In this sense, Part (b) generalises the result in Birkner [3], Theorem 1, to arbitrary ρ with a tail that decays faster than algebraic.

Let $\pi_1\colon \widetilde{E}^{\mathbb{N}} \to \widetilde{E}$ be the projection onto the first word, and let $\mathcal{P}(\widetilde{E})$ be the set of probability measures on $\widetilde{E}$. An application of the contraction principle to Theorem 1.2 yields the following.

Corollary 1.5. *Under the assumptions of Theorem* 1.2*, for* $\nu^{\otimes\mathbb{N}}$*–a.s. all* X*, the family of (regular) conditional probability distributions* $\mathbb{P}(\pi_1 R_N \in \cdot \mid X)$*,* $N \in \mathbb{N}$*, satisfies the LDP on* $\mathcal{P}(\widetilde{E})$ *with rate* N *and with deterministic rate function* $I_1^{\mathrm{que}}\colon \mathcal{P}(\widetilde{E}) \to [0,\infty]$ *given by*

$$I_1^{\mathrm{que}}(q) := \inf\left\{I^{\mathrm{que}}(Q)\colon\ Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}),\ \pi_1 Q = q\right\}. \tag{1.23}$$

This rate function is lower semi-continuous, has compact levels sets, has a unique zero at $q = q_{\rho,\nu}$*, and is convex.*

By taking projective limits, it is possible to extend Theorems 1.2–1.3 to more general letter spaces. The following corollary will be proved in Section 8.

Corollary 1.6. *The quenched LDP also holds when* E *is countable, with the same rate function as in* (1.15–1.16).

One can push further and obtain an LDP for $E = \mathbb{R}$, by picking $E = 2^{-n}\mathbb{Z}$, $n \in \mathbb{Z}$, and taking the limit as $n \to \infty$. We will, however, not pursue this extension here, since control of the limiting relative entropies adds on an extra technical layer.

1.3 Heuristic explanation of main theorems

To explain the background of Theorem 1.2, we begin by recalling a few properties of entropy. Let $H(Q)$ denote the *specific entropy of* $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ defined by

$$H(Q) := \lim_{N\to\infty} \frac{1}{N}\, h\big(Q_{|\mathscr{F}_N}\big) \in [0,\infty], \tag{1.24}$$

where $h(\cdot)$ denotes entropy. The sequence under the limit in (1.24) is non-increasing in N. Since $q_{\rho,\nu}^{\otimes\mathbb{N}}$ is a product measure, we have the identity (recall (1.2–1.4))

$$\begin{aligned} H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) &= -H(Q) - \mathbb{E}_Q[\log q_{\rho,\nu}(Y_1)] \\ &= -H(Q) - \mathbb{E}_Q[\log \rho(\tau_1)] - m_Q\, \mathbb{E}_{\Psi_Q}[\log \nu(X_1)]. \end{aligned} \tag{1.25}$$

Similarly,

$$H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) = -H(\Psi_Q) - \mathbb{E}_{\Psi_Q}[\log \nu(X_1)]. \tag{1.26}$$

Below, for a discrete random variable Z with a law Q on a state space $\mathcal{Z}$ we will write $Q(Z)$ for the random variable $f(Z)$ with $f(z) = Q(Z = z)$, $z \in \mathcal{Z}$. Abbreviate

$$K^{(N)} := \kappa(Y^{(1)}, \dots, Y^{(N)}) \quad \text{and} \quad K^{(\infty)} := \kappa(Y). \tag{1.27}$$

In analogy with (1.14), define

$$\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}}) := \Big\{Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})\colon\ m_Q < \infty\Big\}. \tag{1.28}$$

Lemma 1.7. [Birkner [3], Lemmas 3 and 4]
Suppose that $Q \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ *and* $H(Q) < \infty$. *Then,* Q*-a.s.,*

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \log Q(K^{(N)}) &= -m_Q H(\Psi_Q), \\ \lim_{N\to\infty} \frac{1}{N} \log Q\big(\tau_1, \dots, \tau_N \mid K^{(N)}\big) &=: -H_{\tau|K}(Q), \\ \lim_{N\to\infty} \frac{1}{N} \log Q\big(Y^{(1)}, \dots, Y^{(N)}\big) &= -H(Q), \end{aligned} \tag{1.29}$$

with

$$m_Q H(\Psi_Q) + H_{\tau|K}(Q) = H(Q). \tag{1.30}$$

Equation (1.30), which follows from (1.29) and the identity

$$Q(K^{(N)}) Q(\tau_1, \dots, \tau_N \mid K^{(N)}) = Q(Y^{(1)}, \dots, Y^{(N)}), \tag{1.31}$$

identifies $H_{\tau|K}(Q)$. Think of $H_{\tau|K}(Q)$ as the *conditional specific entropy of word lengths under the law Q given the concatenation.* Combining (1.25–1.26) and (1.30), we have

$$H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) = m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - H_{\tau|K}(Q) - \mathbb{E}_Q[\log \rho(\tau_1)]. \tag{1.32}$$

The term $-H_{\tau|K}(Q) - \mathbb{E}_Q[\log \rho(\tau_1)]$ in (1.32) can be interpreted as the *conditional specific relative entropy of word lengths under the law Q w.r.t. $\rho^{\otimes\mathbb{N}}$ given the concatenation.*

Note that $m_Q < \infty$ and $H(Q) < \infty$ imply that $H(\Psi_Q) < \infty$, as can be seen from (1.30). Also note that $-\mathbb{E}_{\Psi_Q}[\log \nu(X_1)] < \infty$ because E is finite, and $-\mathbb{E}_Q[\log \rho(\tau_1)] < \infty$ because of (1.1) and $m_Q < \infty$, implying that (1.25–1.26) are proper.

We are now ready to give a heuristic explanation of Theorem 1.2. Let

$$R^N_{j_1,\dots,j_N}(X), \qquad 0 < j_1 < \cdots < j_N < \infty, \tag{1.33}$$

denote the empirical process of N-tuples of words when X is cut at the points $j_1, \dots, j_N$ (i.e., when $T_i = j_i$ for $i = 1, \dots, N$; see (3.16–3.17) for a precise definition). Fix $Q \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$. The probability $\mathbb{P}(R_N \approx Q \mid X)$ is a sum over all N-tuples $j_1, \dots, j_N$ such that $R^N_{j_1,\dots,j_N}(X) \approx Q$, weighted by $\prod_{i=1}^N \rho(j_i - j_{i-1})$ (with $j_0 = 0$). The fact that $R^N_{j_1,\dots,j_N}(X) \approx Q$ has three consequences:

(1) The $j_1, \dots, j_N$ must cut $\approx N$ substrings out of X of total length $\approx N m_Q$ that look like the concatenation of words that are Q-typical, i.e., that look as if generated by Ψ_Q (possibly with gaps in between). This means that most of the cut-points must hit atypical pieces of X. We expect to have to shift X by $\approx \exp[N m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}})]$ in order to find the first contiguous substring of length $N m_Q$ whose empirical shifts lie in a small neighbourhood of Ψ_Q. By (1.1), the probability for the single increment $j_1 - j_0$ to have the size of this shift is $\approx \exp[-N\alpha\, m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}})]$.

(2) The combinatorial factor $\exp[N H_{\tau|K}(Q)]$ counts how many "local perturbations" of $j_1, \dots, j_N$ preserve the property that $R^N_{j_1,\dots,j_N}(X) \approx Q$.

(3) The statistics of the increments $j_1 - j_0, \dots, j_N - j_{N-1}$ must be close to the distribution of word lengths under Q. Hence, the weight factor $\prod_{i=1}^N \rho(j_i - j_{i-1})$ must be $\approx \exp[N\mathbb{E}_Q[\log \rho(\tau_1)]]$ (at least, for Q-typical pieces).

The contributions from (1)–(3), together with the identity in (1.32), explain the formula in (1.16) on $\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$. Considerable work is needed to extend (1)–(3) from $\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$. This is explained in Section 3.5.

In (1), instead of having a single large increment preceding a single contiguous substring of length Nm_Q, it is possible to have several large increments preceding several contiguous substrings, which together have length Nm_Q. The latter gives rise to the same contribution, and so there is some entropy associated with the choice of the large increments. Lemma 2.1 in Section 2.1 is needed to control this entropy, and shows that it is negligible.

1.4 Application of LDP to collision local time of random walks

In this section we apply Theorems 1.1–1.2 to derive two results about the collision local time of random walks, which will in turn be used in Section 1.5.

1.4.1 Discrete time

Let $S = (S_k)_{k=0}^\infty$ and $S' = (S'_k)_{k=0}^\infty$ be two independent random walks on $\mathbb{Z}^d$, $d \geq 1$, both starting at the origin, with a *symmetric* and irreducible transition kernel $p(\cdot,\cdot)$. Suppose that

$$\lim_{n\to\infty} \frac{\log p^{2n}(0,0)}{\log(2n)} =: -\alpha, \quad \alpha \in (1,\infty). \tag{1.34}$$

Write $\mathbb{P}$ to denote the joint law of S, S'. Let

$$V := \sum_{k=0}^{\infty} 1_{\{S_k = S'_k\}} \tag{1.35}$$

be the *collision local time* of S, S', and define

$$z_1 := \sup\left\{z \geq 0\colon\ \mathbb{E}\left[\, z^V \mid S\,\right] < \infty \ \ S\text{-a.s.}\right\}, \quad z_2 := \sup\left\{z \geq 0\colon\ \mathbb{E}\left[\, z^V\,\right] < \infty\right\}. \tag{1.36}$$

(The lower indices indicate the number of random walks being averaged over.) Note that, by the tail triviality of S, the range of z's for which $\mathbb{E}[\, z^V \mid S\,]$ converges is S-a.s. constant. Also note that (1.34) implies that $p(\cdot,\cdot)$ is transient, so that $\mathbb{P}(V < \infty) = 1$. The following theorem holds when $p(\cdot,\cdot)$ is strongly transient, i.e., when $\sum_{n=1}^\infty np^n(0,0) < \infty$.

Theorem 1.8. *Assume* (1.34). *If $p(\cdot,\cdot)$ is strongly transient, then $1 < z_2 < z_1 < \infty$.*

Since $\mathbb{P}(V = k) = (1 - F^{(2)})[F^{(2)}]^{k-1}$, $k \in \mathbb{N}$, with

$$F^{(2)} := \mathbb{P}\big(\exists\, k \in \mathbb{N}\colon\ S_k = S'_k\big), \tag{1.37}$$

an easy computation gives

$$z_2 = 1/F^{(2)}. \tag{1.38}$$

Note that $F^{(2)} = 1 - [1/G^{(2)}(0,0)]$ with $G^{(2)}(0,0) = \sum_{n=0}^\infty p^{2n}(0,0)$ (see Spitzer [22], Section 1). There is no simple expression for z_1. In Section 9.1 we will give an upper bound.

1.4.2 Continuous time

Next we turn the discrete-time random walks S, S' into continuous-time random walks $\widetilde{S} = (S_t)_{t\geq 0}$ and $\widetilde{S}' = (\widetilde{S}'_t)_{t\geq 0}$ by allowing them to make steps at rate 1, keeping the same $p(\cdot,\cdot)$. Then the collision local time becomes

$$\widetilde{V} := \int_0^\infty 1_{\{\widetilde{S}_t = \widetilde{S}'_t\}}\, dt. \tag{1.39}$$

For the analogous quantities $\widetilde{z}_1$ and $\widetilde{z}_2$, we have the following.

Theorem 1.9. *Assume* (1.34)*. If $p(\cdot,\cdot)$ is strongly transient, then $0 < \widetilde{z}_2 < \widetilde{z}_1 < \infty$.*

An easy computation gives $\log \widetilde{z}_2 = 2/G(0,0)$ with $G(0,0) = \sum_{n=0}^\infty p^n(0,0)$. There is again no simple expression for $\widetilde{z}_1$.

1.4.3 Conjecture

We close with the following conjecture.

Conjecture 1.10. *The gaps in Theorems* 1.8–1.9 *are present also when $p(\cdot,\cdot)$ is transient but not strongly transient.*

Random walks with zero mean and finite variance are transient for $d \geq 3$ and strongly transient for $d \geq 5$ (Spitzer [22], Section 1). In a forthcoming paper by Birkner and Sun [4], the gap in Theorem 1.8 is proved for simple random walk on $\mathbb{Z}^d$, $d \geq 4$, and the proof is in principle extendable to more general random walks. It is an adaptation of the fractional moment technique developed by Derrida, Giacomin, Lacoin and Toninelli [11] in the context of pinning models. Note that simple random walk on $\mathbb{Z}^4$ is just on the border of not being strongly transient. Thus, part of the above conjecture is already giving way.

1.5 The gaps settle three conjectures

In this section we use Theorems 1.8–1.9 to prove the existence of an *intermediate phase* for three classes of interacting particle systems.

1.5.1 Coupled branching processes

Theorem 1.9 *proves* a conjecture put forward in Greven [14], [15]. Consider a spatial population model, defined as the Markov process $(\eta_t)_{t\geq 0}$ taking values in $(\mathbb{N} \cup \{0\})^{\mathbb{Z}^d}$ (counting the number of individuals at the different sites of $\mathbb{Z}^d$) evolving as follows:

(1) Individuals migrate at rate 1 according to $a(\cdot,\cdot)$.

(2) A new individual is born at site x at rate $b\eta(x)$.

(3) One individual at site x dies at rate $(1-p)b\eta(x)$.

(4) All individuals at site x die simultaneously at rate pb.

Here, $a(\cdot,\cdot)$ is an irreducible random walk transition kernel on $\mathbb{Z}^d \times \mathbb{Z}^d$, $b \in (0,\infty)$ is a birth-death rate, $p \in [0,1]$ is a coupling parameter, while (1)–(4) occur independently at every $x \in \mathbb{Z}^d$. The case $p = 0$ corresponds to a critical branching random walk, for which the average number of individuals per site is preserved. The case $p > 0$ is interesting because the individuals descending from different ancestors are no longer independent.

A critical branching random walk satisfies the following *dichotomy* (where for simplicity we restrict to the case where $a(\cdot,\cdot)$ is symmetric): if the initial configuration η_0 is drawn from a shift-invariant probability distribution with finite mean, then η_t as $t \to \infty$ locally dies out ("extinction") when $a(\cdot,\cdot)$ is recurrent, but converges to a non-trivial equilibrium ("survival") when $a(\cdot,\cdot)$ is transient, both irrespective of the value of b. In the latter case, the equilibrium has the same mean as the initial distribution and has all moments finite.

For the coupled branching process with $p > 0$ there is a dichotomy too, but it is controlled by a subtle interplay of $a(\cdot,\cdot)$, b and p: extinction holds when $a(\cdot,\cdot)$ is recurrent, but also when $a(\cdot,\cdot)$ is transient and p is sufficiently large. Indeed, it is shown in Greven [14] that if $a(\cdot,\cdot)$ is transient, then there is a unique $p_* \in (0,1)$ such that survival holds for $p < p_*$ and extinction holds for $p > p_*$.

Recall the critical values $\widetilde{z}_1, \widetilde{z}_2$ introduced in Section 1.4.2. Survival holds if $\mathbb{E}(\exp[bp\widetilde{V}] \mid \widetilde{S}) < \infty$ $\widetilde{S}$-a.s., i.e., if $p < p_1$ with

$$p_1 = \frac{1}{b} \log \widetilde{z}_1. \tag{1.40}$$

This is shown by a size-biasing of the population in the spirit of Kallenberg [19]. On the other hand, survival with a *finite second moment* holds if and only if $\mathbb{E}(\exp[bp\widetilde{V}]) < \infty$, i.e., if and only if $p < p_2$ with

$$p_2 = \frac{1}{b} \log \widetilde{z}_2. \tag{1.41}$$

Clearly, $p_* \geq p_1 \geq p_2$. Theorem 1.9 shows that if $a(\cdot,\cdot)$ satisfies (1.34) and is strongly transient, then $p_1 > p_2$, implying that there is an intermediate phase of survival with an *infinite second moment.*

Theorem 1.8 *corrects* an error in Birkner [1], Theorem 6. Here, a system of individuals living on $\mathbb{Z}^d$ is considered subject to migration and branching. Each individual independently migrates at rate 1 according to a random walk transition kernel $a(\cdot,\cdot)$, and branches at a rate that depends on the number of individuals present at the same location. It is argued that this system has an intermediate phase in which the numbers of individuals at different sites tend to an equilibrium with a finite first moment but an infinite second moment. The proof is, however, based on a wrong rate function. Corollary 1.5 shows that the rate function claimed in Birkner [1], Theorem 6, must be replaced by that in (1.23), after which the intermediate phase persists. This also affects [1], Theorem 5, which uses [1], Theorem 6, to compute z_1 in Section 1.4 and finds an incorrect formula. As we will see in Section 9.1, this formula actually is an upper bound for z_1.

1.5.2 Interacting diffusions

Theorem 1.9 *proves* a conjecture put forward in Greven and den Hollander [16]. Consider the system of interacting diffusions on $[0,\infty)$ defined by the following collection of coupled stochastic differential equations:

$$dX_x(t) = \sum_{y\in\mathbb{Z}^d} a(x,y)[X_y(t) - X_x(t)]\ dt + \sqrt{bX_x(t)^2}\ dW_x(t), \qquad x \in \mathbb{Z}^d, t \geq 0. \tag{1.42}$$

Here, $a(\cdot,\cdot)$ is an irreducible random walk transition kernel on $\mathbb{Z}^d \times \mathbb{Z}^d$, $b \in (0,\infty)$ is a diffusion parameter, and $(\{W_x(t)\}_{x\in\mathbb{Z}^d})_{t\geq 0}$ is a collection of independent standard Brownian motions on $\mathbb{R}$. The initial condition is chosen such that $\{X_x(0)\}_{x\in\mathbb{Z}^d}$ is a shift-invariant and shift-ergodic random field on $[0,\infty)$ with mean $\Theta \in (0,\infty)$ (the evolution preserves the mean).

It was shown in [16], Theorems 1.4–1.6, that if $a(\cdot,\cdot)$ is symmetric and transient, then there exist $0 < b_2 \leq b_*$ such that the system in (1.42) converges to an equilibrium when $0 < b < b_*$, and this equilibrium has a *finite second moment* when $0 < b < b_2$ and an *infinite second moment* when $b_2 \leq b < b_*$. It was conjectured in [16], Conjecture 1.8, that $b_* > b_2$. As explained in [16], Section 4.2, the gap in Theorem 1.9 settles this conjecture (at least when $a(\cdot,\cdot)$ is strongly transient), with

$$b_2 = \log \widetilde{z}_2 \quad \text{and} \quad b_* = \log \widetilde{z}_1.$$

1.5.3 Directed polymers in random environments

Theorem 1.8 *disproves* a conjecture put forward in Monthus and Garel [20]. Let $a(\cdot,\cdot)$ be a symmetric and irreducible random walk transition kernel on $\mathbb{Z}^d \times \mathbb{Z}^d$, let $S = (S_k)_{k=0}^\infty$ be the corresponding random walk, and let $\xi = \{\xi(x,n)\colon\ x \in \mathbb{Z}^d, n \in \mathbb{N}\}$ be i.i.d. $\mathbb{R}$-valued non-degenerate random variables satisfying

$$\lambda(\beta) := \log \mathbb{E}\big(\exp[\beta\xi(x,n)]\big) \in \mathbb{R} \qquad \forall\, \beta \in \mathbb{R}. \tag{1.43}$$

Put

$$e_n(\xi,S) := \exp\left[\sum_{k=1}^n \{\beta\xi(S_k,k) - \lambda(\beta)\}\right], \tag{1.44}$$

and set

$$Z_n(\xi) := \mathbb{E}[e_n(\xi,S)] = \sum_{s_1,\dots,s_n\in\mathbb{Z}^d} \left[\prod_{k=1}^n p(s_{k-1},s_k)\right]\ e_n(\xi,s), \qquad s = (s_k)_{k=0}^\infty,\ s_0 = 0, \tag{1.45}$$

i.e., $Z_n(\xi)$ is the normalizing constant in the probability distribution of the random walk S whose paths are reweighted by $e_n(\xi,S)$, which is referred to as the "polymer measure". The $\xi(x,n)$'s describe a random space-time medium with which S is interacting, with β playing the role of the interaction strength.

It is well known that $(Z_n)_{n\in\mathbb{N}}$ is a non-negative martingale with respect to the family of sigma-algebras $\mathcal{F}_n := \sigma(\xi(x,k), x \in \mathbb{Z}^d, 1 \leq k \leq n)$, $n \in \mathbb{N}$. Hence

$$\lim_{n\to\infty} Z_n = Z_\infty \geq 0 \qquad \xi - a.s., \tag{1.46}$$

with the event $\{Z_\infty = 0\}$ being ξ-trivial. One speaks of *weak disorder* if $Z_\infty > 0$ ξ-a.s. and of *strong disorder* otherwise. As shown in Comets and Yoshida [9], there is a unique critical value β_* such that weak disorder holds for $\beta < \beta_*$ and strong disorder holds for $\beta > \beta_*$. Moreover, in the weak disorder region the paths have a Gaussian scaling limit under the polymer measure, while this is not the case in the strong disorder region.

Recall the critical values z_1, z_2 defined in Section 1.4. Bolthausen [5] observed that

$$\mathbb{E}\left[Z_n^2\right] = \mathbb{E}\Big[\exp\left[\{\lambda(2\beta) - 2\lambda(\beta)\}\,|\{1 \leq k \leq n\colon\, S_k = S'_k\}|\right]\Big], \tag{1.47}$$

where S and S' are two independent random walks with transition kernel $p(\cdot,\cdot)$, and concluded that $(Z_n)_{n\in\mathbb{N}}$ is L^2-bounded if and only if $\beta < \beta_2$ with $\beta_2 \in (0,\infty]$ the unique solution of

$$\lambda(2\beta_2) - 2\lambda(\beta_2) = z_2. \tag{1.48}$$

Since $\mathbb{P}(Z_\infty > 0) \leq \mathbb{E}[Z_\infty]^2/\mathbb{E}[Z_\infty^2]$ and $\mathbb{E}[Z_\infty] = Z_0 = 1$, it follows that $\beta < \beta_2$ implies weak disorder, i.e., $\beta_* \geq \beta_2$. By a stochastic representation of the size-biased law of Z_n, it was shown in Birkner [2], Proposition 1, that in fact weak disorder holds if $\beta < \beta_1$ with $\beta_1 \in (0,\infty]$ the unique solution of

$$\lambda(2\beta_1) - 2\lambda(\beta_1) = z_1, \tag{1.49}$$

i.e., $\beta_* \geq \beta_1$. Since $\beta \mapsto \lambda(2\beta) - 2\lambda(\beta)$ is strictly increasing, it follows from (1.48–1.49) and Theorem 1.8 that $\beta_1 > \beta_2$ when $a(\cdot,\cdot)$ satisfies (1.34) and is strongly transient and when ξ is such that $\beta_2 < \infty$. In that case the weak disorder region contains a subregion for which $(Z_n)_{n\in\mathbb{N}}$ is *not* L^2-bounded. This disproves a conjecture of Monthus and Garel [20], who argued that $\beta_2 = \beta_*$. A similar conclusion is reached in a recent paper by Camanes and Carmona [6] with different techniques. The latter paper considers only simple random walk, but includes examples of ξ for which the gap is present also in $d = 3$ and $d = 4$.

1.6 Outline

Section 2 collects some preparatory facts that are needed for the proofs of the main theorems, including a lemma that controls the entropy associated with the locations of the large increments in the renewal process. In Section 3 and 4 we prove the large deviation upper, respectively, lower bound. The proof of the former is long (taking up more than half of the paper) and requires a somewhat lengthy construction with combinatorial, functional analytic and ergodic theoretic ingredients. In particular, extending the lower bound from ergodic to non-ergodic probability measures is technically involved. The proofs of Theorems 1.2–1.4 are in Sections 5–7, of Theorem 1.6 in Section 8, and of Theorems 1.8–1.9 in Section 9. Appendix A contains a proof that the annealed and the quenched rate function are continuous under the truncation of the word length approximation.

2 Preparatory facts

Section 2.1 proves a core lemma that is needed to control the entropy of large increments in the renewal process. Section 2.2 shows that the tail property of ρ is preserved under convolutions.

2.1 A core lemma

As announced at the end of Section 1.3, we need to account for the entropy that is associated with the locations of the large increments in the renewal process. This requires the following combinatorial lemma.

Lemma 2.1. *Let $\omega = (\omega_l)_{l\in\mathbb{N}}$ be i.i.d. with $\mathbb{P}(\omega_1 = 1) = 1 - \mathbb{P}(\omega_1 = 0) = p \in (0,1)$, and let $\alpha \in (1,\infty)$. For $N \in \mathbb{N}$, let*

$$S_N(\omega) := \sum_{\substack{0<j_1<\cdots<j_N<\infty \\ \omega_{j_1}=\cdots=\omega_{j_N}=1}} \prod_{i=1}^{N} (j_i - j_{i-1})^{-\alpha} \qquad (j_0 = 0) \tag{2.1}$$

and put

$$\limsup_{N\to\infty} \frac{1}{N} \log S_N(\omega) =: -\phi(\alpha,p) \qquad \omega - a.s. \tag{2.2}$$

(the limit being ω-a.s. constant by tail triviality). Then

$$\lim_{p\downarrow 0} \frac{\phi(\alpha,p)}{\alpha\log(1/p)} = 1. \tag{2.3}$$

Proof. Let $\tau_N := \min\{l \in \mathbb{N}\colon \omega_l = \omega_{l+1} = \cdots = \omega_{l+N-1}\}$. In (2.1), choosing $j_1 = \tau_N$ and $j_i = j_{i-1} + 1$ for $i = 2,\ldots,N$, we see that $S_N(\omega) \geq \tau_N^{-\alpha}$. Since

$$\lim_{N\to\infty} \frac{1}{N} \log \tau_N \to \log(1/p) \qquad \omega - a.s., \tag{2.4}$$

we have

$$\phi(\alpha,p) \leq \alpha \log(1/p) \qquad \forall\, p \in (0,1). \tag{2.5}$$

To show that this bound is sharp in the limit as $p \downarrow 0$, we estimate fractional moments of $S_N(\omega)$. For any $\beta \in (1/\alpha, 1]$, using that $(u+v)^\beta \leq u^\beta + v^\beta$, $u,v \geq 0$, we get

$$\begin{aligned} \mathbb{E}\Big[S_N(\omega)^\beta\Big] &\leq \sum_{0<j_1<\cdots<j_N<\infty} \mathbb{E}\Big[1_{\{\omega_{j_1}=\cdots=\omega_{j_N}=1\}} \prod_{i=1}^{N} (j_i - j_{i-1})^{-\alpha\beta}\Big] \\ &= \sum_{0<j_1<\cdots<j_N<\infty} p^N \prod_{i=1}^{N} (j_i - j_{i-1})^{-\alpha\beta} \\ &= \big[p\,\zeta(\alpha\beta)\big]^N, \end{aligned} \tag{2.6}$$

where $\zeta(s) = \sum_{n\in\mathbb{N}} n^{-s}$, $s > 1$, is Riemann's ζ-function. Hence

$$\mathbb{E}\Big[\frac{1}{N}\log S_N(\omega)\Big] \leq \frac{1}{N\beta} \log \mathbb{E}\Big[S_N(\omega)^\beta\Big] \leq \frac{1}{\beta}\big[\log p + \log\zeta(\alpha\beta)\big]. \tag{2.7}$$

Letting $N \to \infty$, and using (2.2) together with Fatou's lemma, we obtain that

$$\phi(\alpha,p) \geq \frac{1}{\beta}[\log(1/p) - \log\zeta(\alpha\beta)] \qquad \forall\, \beta \in (1/\alpha, 1]. \tag{2.8}$$

Now let $p \downarrow 0$, followed by $\beta \downarrow 1/\alpha$ to obtain the claim. □

Remark 2.2. Note that $\mathbb{E}[S_N(\omega)] = (p\zeta(\alpha))^N$, but we expect that typically $S_N(\omega) \approx p^{\alpha N}$. This is verified by bounding suitable non-integer moments of $S_N(\omega)/p^{\alpha N}$. Estimating non-integer moments in situations when the mean is inconclusive is a useful technique in various fields of probability, see, e.g., Holley and Liggett [18] and Toninelli [24] and the discussion and references there. It has recently been fruitfully applied by Toninelli [24] to pinning and copolymer models, and the proof above is similar to that of Theorem 2.1 there.

2.2 Convolution preserves polynomial tail

The following lemma will be needed in Section 3.6. For $m \in \mathbb{N}$, let ρ^{*m} denote the m-fold convolution of ρ.

Lemma 2.3. *Suppose that ρ satisfies $\rho(n) \le C_\rho n^{-\alpha}$, $n \in \mathbb{N}$, for some $C_\rho < \infty$. Then*

$$\rho^{*m}(n) \le (2^\alpha C_\rho \vee 1)\, m^{\alpha+1} n^{-\alpha} \qquad \forall\, m, n \in \mathbb{N}. \tag{2.9}$$

Proof. If $n \le m$, then the right-hand side of (2.9) is ≥ 1. So, let us assume that $n > m$. Then

$$\begin{aligned}
\rho^{*m}(n) &= \sum_{\substack{x_1,\dots,x_m \ge 1\\ x_1+\dots+x_m = n}} \prod_{i=1}^m \rho(x_i) \le \sum_{j=1}^n \sum_{\substack{x_1,\dots,x_m \ge 1\\ x_1+\dots+x_m = n\\ x_j = x_1 \vee \dots \vee x_m}} \rho(x_j) \prod_{i \ne j}^m \rho(x_i) \\
&\le m\, C_\rho \lfloor n/m \rfloor^{-\alpha} \sum_{x_1,\dots,x_{m-1} \ge 1} \prod_{i=1}^{m-1} \rho(x_i) \\
&= m\, C_\rho \lfloor n/m \rfloor^{-\alpha} \le 2^\alpha C_\rho\, m^{\alpha+1}\, n^{-\alpha}.
\end{aligned} \tag{2.10}$$

□

3 Upper bound

The following upper bound will be used in Section 5 to derive the upper bound in the definition of the LDP.

Proposition 3.1. *For any $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ and any $\varepsilon > 0$, there is an open neighbourhood $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ of Q such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in \mathcal{O}(Q) \mid X\big) \le -I^{\mathrm{fin}}(Q) + \varepsilon \quad X - a.s. \tag{3.1}$$

Proof. It suffices to consider the case $\Psi_Q \ne \nu^{\otimes\mathbb{N}}$. The case $\Psi_Q = \nu^{\otimes\mathbb{N}}$, for which $I^{\mathrm{fin}}(Q) = H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$ as is seen from (1.16), is contained in the upper bound in Birkner [3], Lemma 8. Alternatively, by lower semicontinuity of $Q' \mapsto H(Q' \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$, there is a neighbourhood $\mathcal{O}(Q)$ such that

$$\inf_{Q' \in \overline{\mathcal{O}(Q)}} H(Q' \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) \ge H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - \varepsilon = I^{\mathrm{fin}}(Q) - \varepsilon, \tag{3.2}$$

where $\overline{\mathcal{O}(Q)}$ denotes the closure of $\mathcal{O}(Q)$ (in the weak topology), and we can use the annealed bound.

In Sections 3.1–3.5 we first prove Proposition 3.1 under the assumption that there exist $\alpha \in (1,\infty)$, $C_\rho < \infty$ such that

$$\rho(n) \leq C_\rho \, n^{-\alpha}, \quad n \in \mathbb{N}, \tag{3.3}$$

which is needed in Lemma 2.3. In Section 3.6 we show that this can be replaced by (1.1). In Sections 3.1–3.4, we first consider $Q \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ (recall (1.28)). Here, we turn the heuristics from Section 1.3 into a rigorous proof. In Section 3.5 we remove the ergodicity restriction. The proof is long and technical (taking up more than half of the paper).

3.1 Step 1: Consequences of ergodicity

We will use the ergodic theorem to construct specific neighborhoods of $Q \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ that are well adapted to formalize the strategy of proof outlined in our heuristic explanation of the main theorem in Section 1.3.

Fix $\varepsilon_1, \delta_1 > 0$. By the ergodicity of Q and Lemma 1.7, the event (recall (1.9) and (1.27))

$$\begin{aligned}
&\left\{\frac{1}{M}|K^{(M)}| \in m_Q + [-\varepsilon_1,\varepsilon_1]\right\}\\
&\cap \left\{-\frac{1}{M}\log Q(K^{(M)}) \in m_Q H(\Psi_Q) + [-\varepsilon_1,\varepsilon_1]\right\}\\
&\cap \left\{-\frac{1}{M}\log Q(Y^{(1)},\dots,Y^{(M)}) \in H(Q) + [-\varepsilon_1,\varepsilon_1]\right\}\\
&\cap \left\{\frac{1}{M}\sum_{k=1}^{|K^{(M)}|}\log \nu((K^{(M)})_k) \in m_Q \mathbb{E}_{\Psi_Q}\big[\log \nu(X_1)\big] + [-\varepsilon_1,\varepsilon_1]\right\}\\
&\cap \left\{\frac{1}{M}\sum_{i=1}^{M}\log \rho(\tau_i) \in \mathbb{E}_{Q}\big[\log \rho(\tau_1)\big] + [-\varepsilon_1,\varepsilon_1]\right\}
\end{aligned} \tag{3.4}$$

has Q-probability at least $1-\delta_1/4$ for M large enough (depending on Q), where $|K^{(M)}|$ is the length of the string of letters $K^{(M)}$. Hence, there is a finite number A of sentences of length M, denoted by

$$(z_a)_{a=1,\dots,A} \text{ with } z_a := (y^{(a,1)},\dots,y^{(a,M)}) \in \widetilde{E}^M, \tag{3.5}$$

such that for $a = 1, \dots, A$,

$$
\begin{aligned}
|\kappa(z_a)| &\in \Big[M(m_Q - \varepsilon_1), M(m_Q + \varepsilon_1)\Big], \\
Q(K^{(M)} = \kappa(z_a)) &\in \Big[\exp[-M(m_Q H(\Psi_Q) + \varepsilon_1)], \exp[-M(m_Q H(\Psi_Q) - \varepsilon_1)]\Big], \\
Q\big((Y^{(1)}, \dots, Y^{(M)}) = z_a\big) &\in \Big[\exp[-M(H(Q) + \varepsilon_1)], \exp[-M(H(Q) - \varepsilon_1)]\Big], \\
\sum_{k=1}^{|\kappa(z_a)|} \log \nu((\kappa(z_a))_k) &\in \Big[M(m_Q \mathbb{E}_{\Psi_Q}[\log \nu(X_1)] - \varepsilon_1), M(m_Q \mathbb{E}_{\Psi_Q}[\log \nu(X_1)] + \varepsilon_1)\Big], \\
\sum_{i=1}^{M} \log \rho(|y^{(a,i)}|) &\in \Big[M(\mathbb{E}_Q[\log \rho(\tau_1)] - \varepsilon_1), M(\mathbb{E}_Q[\log \rho(\tau_1)] + \varepsilon_1)\Big],
\end{aligned}
\tag{3.6}
$$

and

$$
\sum_{a=1}^{A} Q\Big((Y^{(1)}, \dots, Y^{(M)}) = z_a\Big) \geq 1 - \frac{\delta_1}{2}. \tag{3.7}
$$

Note that (3.7) and the third line of (3.6) imply that

$$
A \in \Big[(1 - \delta_1) \exp\big[M(H(Q) - \varepsilon_1)\big], \exp\big[M(H(Q) + \varepsilon_1)\big]\Big]. \tag{3.8}
$$

Abbreviate

$$
\mathscr{A} := \{z_a,\, a = 1, \dots, A\}. \tag{3.9}
$$

Let

$$
\mathscr{B} := \big\{\zeta^{(b)}, b = 1, \dots, B\big\} = \big\{\kappa(z_a),\, a = 1, \dots, A\big\} \tag{3.10}
$$

be the set of strings of letters arising from concatenations of the individual z_a's, and let

$$
I_b := \big\{1 \leq a \leq A\colon\ \kappa(z_a) = \zeta^{(b)}\big\}, \quad b = 1, \dots, B, \tag{3.11}
$$

so that $|I_b|$ is the number of sentences in $\mathscr{A}$ giving a particular string in $\mathscr{B}$. By the second line of (3.6), we can bound B as

$$
B \leq \exp\big[M(m_Q H(\Psi_Q) + \varepsilon_1)\big], \tag{3.12}
$$

because $\sum_{b=1}^{B} Q(K^{(M)} = \zeta^{(b)}) \leq 1$ and each summand is at least $\exp[-M(m_Q H(\Psi_Q) + \varepsilon_1)]$. Furthermore, we have

$$
|I_b| \leq \exp\big[M(H_{\tau|K}(Q) + 2\varepsilon_1)\big], \quad b = 1, \dots, B, \tag{3.13}
$$

since

$$
\begin{aligned}
\exp\big[-M(m_Q H(\Psi_Q) - \varepsilon_1)\big] &\geq Q\big(\kappa(Y^{(1)}, \dots, Y^{(M)}) = \zeta^{(b)}\big) \\
&\geq \sum_{a \in I_b} Q\big((Y^{(1)}, \dots, Y^{(M)}) = z_a\big) \geq |I_b| \exp\big[-M(H(Q) + \varepsilon_1)\big],
\end{aligned}
\tag{3.14}
$$

and $H(Q) - m_Q H(\Psi_Q) = H_{\tau|K}(Q)$ by (1.32).

3.2 Step 2: Good sentences in open neighbourhoods

Define the following open neighbourhood of Q (recall (3.9))

$$\mathcal{O} := \Big\{ Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon\ Q'_{|\mathscr{F}_M}(\mathscr{A}) > 1 - \delta_1 \Big\}. \tag{3.15}$$

Here, $Q(z)$ is shorthand for $Q((Y^{(1)}, \dots, Y^{(M)}) = z)$. For $x \in E^{\mathbb{N}}$ and for a vector of cut-points $(j_1, \dots, j_N) \in \mathbb{N}^N$ with $0 < j_1 < \dots < j_N < \infty$ and $N > M$, let

$$\xi_N := (\xi^{(i)})_{i=1,\dots,N} = \big(x|_{(0,j_1]}, x|_{(j_1,j_2]}, \dots, x|_{(j_{N-1},j_N]} \big) \in \widetilde{E}^N \tag{3.16}$$

(with $(0, j_1]$ short-hand notation for $(0, j_1] \cap \mathbb{N}$, etc.) be the sequence of words obtained by cutting x at the positions j_i, and let

$$R^N_{j_1,\dots,j_N}(x) := \frac{1}{N} \sum_{i=0}^{N-1} \delta_{\widetilde{\theta}^i(\xi_N)^{\mathrm{per}}} \tag{3.17}$$

be the corresponding empirical process. By (3.15),

$$\begin{aligned} &R^N_{j_1,\dots,j_N}(x) \in \mathcal{O} \qquad \Longrightarrow \\ &\quad \#\Big\{ 1 \le i \le N - M\colon\ \big(x|_{(j_{i-1},j_i]}, \dots, x|_{(j_{i+M-1},j_{i+M}]} \big) \in \mathscr{A} \Big\} \ge N(1 - \delta_1) - M. \end{aligned} \tag{3.18}$$

Note that (3.18) implies that the sentence ξ_N contains at least

$$C := \lfloor (1 - \delta_1) N/M \rfloor - 1 \tag{3.19}$$

disjoint subsentences from the set $\mathscr{A}$, i.e., there are $1 \le i_1, \dots, i_C \le N - M$ with $i_c - i_{c-1} \ge M$ for $c = 1, \dots, C$ such that

$$\big(\xi^{(i_c)}, \xi^{(i_c+1)}, \dots, \xi^{(i_c+M-1)} \big) \in \mathscr{A} \tag{3.20}$$

(we implicitly assume that N is large enough so that $C > 1$). Indeed, we can e.g. construct the i_c's iteratively as

$$\begin{aligned} &i_0 = -M, \\ &i_c = \min\Big\{ k \ge i_{c-1} + M\colon\ \text{a sentence from } \mathscr{A} \text{ starts at position } k \text{ in } \xi_N \Big\}, \\ &\qquad c = 1, \dots, C, \end{aligned} \tag{3.21}$$

and we can continue the iteration as long as $cM + \delta_1 N \le N$. But (3.20) in turn implies that the j_{i_c}'s cut out of x at least C disjoint subwords from $\mathscr{B}$, i.e.,

$$x|_{(j_{i_c}, j_{i_c+M}]} \in \mathscr{B}, \quad c = 1, \dots, C. \tag{3.22}$$

3.3 Step 3: Estimate of the large deviation probability

Using Steps 1 and 2, we estimate (recall (3.15))

$$\mathbb{P}\big(R_N \in \mathcal{O} \mid X\big) = \sum_{0 < j_1 < \dots < j_N < \infty} \mathbb{1}_{\mathcal{O}}\left(R^N_{j_1,\dots,j_N}(X) \right) \prod_{i=1}^N \rho(j_i - j_{i-1}) \tag{3.23}$$

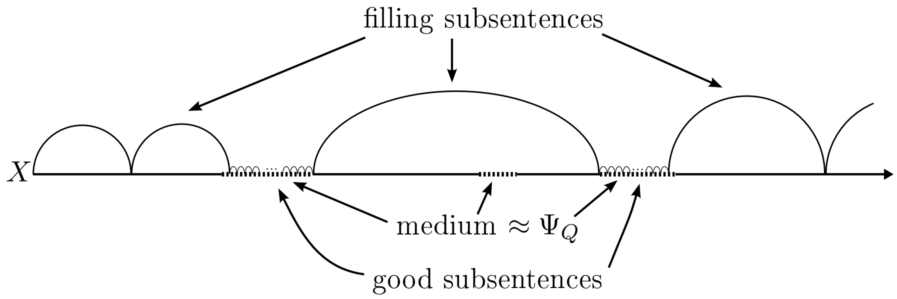

Figure 2: Looking for good subsentences and filling subsentences (see below (3.25)).

from above as follows. Fix a vector of cut-points $(j_1, \dots, j_N)$ giving rise to a non-zero contribution in the right-hand side of (3.23). We think of this vector as describing a particular way of cutting X into a sentence of N words. By (3.22), at least C (recall 3.19) of the j_c's must be cut-points where a word from $\mathscr{B}$ is written on X, and these C subwords must be disjoint. As words in $\mathscr{B}$ arise from concatenations of sentences from $\mathscr{A}$, this means we can find

$$\ell_1 < \cdots < \ell_C, \quad \{\ell_1, \dots, \ell_C\} \subset \{0, j_1, \dots, j_N\} \quad \text{and} \quad \zeta_1, \dots, \zeta_C \in \mathscr{A} \tag{3.24}$$

such that

$$X|_{(\ell_c, \ell_c + |\kappa(\zeta_c)|]} = \kappa(\zeta_c) =: \eta^{(c)} \in \mathscr{B} \text{ and } \ell_c \geq \ell_{c-1} + |\kappa(\zeta_{c-1})|, \quad c = 1, \dots, C-1. \tag{3.25}$$

We call $\zeta_1, \dots, \zeta_C$ the *good* subsentences.

Note that once we fix the ℓ_c's and the ζ_c's, this determines $C+1$ *filling* subsentences (some of which may be empty) consisting of the words between the good subsentences. See Figure 2 for an illustration. In particular, this determines numbers $m_1, \dots, m_{C+1} \in \mathbb{N}$ such that $m_1 + \cdots + m_{C+1} = N - CM$, where m_c is the number of words we cut between the $(c-1)$-st and the c-th good subsentence (and m_{C+1} is the number of words after the C-th good subsentence).

Next, let us fix good $\ell_1 < \cdots < \ell_C$ and $\eta^{(1)}, \dots, \eta^{(C)} \in \mathscr{B}$, satisfying

$$X|_{(\ell_c, \ell_c + |\eta^{(c)}|]} = \eta^{(c)}, \quad \ell_c \geq \ell_{c-1} + |\eta^{(c-1)}|, \qquad c = 1, \dots, C. \tag{3.26}$$

To estimate how many different choices of $(j_1, \dots, j_N)$ may lead to this particular $((\ell_c), (\eta^{(c)}))$, we proceed as follows. There are at most

$$\big(2M\varepsilon_1\big)^C \exp\big[M\big(H_{\tau|K}(Q) + 2\varepsilon_1\big)\big]^C \leq \exp\big[N\big(H_{\tau|K}(Q) + \delta_2\big)\big] \tag{3.27}$$

possible choices for the word lengths inside these good subsentences. Indeed, by the first line of (3.6), at most $2M\varepsilon_1$ different elements of $\mathscr{B}$ can start at any given position ℓ_c and, by (3.13), each of them can be cut in at most $\exp\big[M(H_{\tau|K}(Q) + 2\varepsilon_1)\big]$ different ways to obtain an element of $\mathscr{A}$. In (3.27), $\delta_2 = \delta_2(\varepsilon_1, \delta_1, M)$ can be made arbitrarily small by choosing M large and ε_1, δ_1 small. Furthermore, there are at most

$$\binom{N - C(M-1)}{C} \leq \exp[\delta_3 N] \tag{3.28}$$

possible choices of the m_c's, where $\delta_3 = \delta_3(\delta_1, M)$ can be made arbitrarily small by choosing M large and δ_1 small.

Next, we estimate the value of $\prod_{i=1}^N \rho(j_i - j_{i-1})$ for any $(j_1, \dots, j_N)$ leading to the given $((\ell_c), (\eta^{(c)}))$. In view of the fifth line of (3.6), we have

$$\begin{aligned}
&\prod_{i=1}^N \mathbb{1}_{\{\text{the } i\text{-th word falls inside the } C \text{ good subsentences}\}} \rho(j_i - j_{i-1}) \\
&\qquad \le \exp\big[CM\big(\mathbb{E}_Q[\log \rho(\tau_1)] + \varepsilon_1\big)\big] \\
&\qquad \le \exp\big[N\big(\mathbb{E}_Q[\log \rho(\tau_1)] + \delta_4\big)\big],
\end{aligned} \tag{3.29}$$

where $\delta_4 = \delta_4(\varepsilon_1, \delta_1, M)$ can be made arbitrarily small by choosing M large and ε_1, δ_1 small. The filling subsentences have to exactly fill up the gaps between the good subsentences and so, for a given choice of (ℓ_c), $(\eta^{(c)})$ and (m_c), the contribution to $\prod_{i=1}^N \rho(j_i - j_{i-1})$ from the filling subsentences is $\prod_{c=1}^C \rho^{*m_c}(\ell_c - \ell_{c-1} - |\eta^{(c-1)}|)$ (the term for $c = 1$ is to be interpreted as $\rho^{*m_1}(\ell_1)$, and ρ^{*0} as δ_0). By Lemma 2.3,

$$\begin{aligned}
&\prod_{c=1}^C \rho^{*m_c}\big(\ell_c - \ell_{c-1} - |\eta^{(c-1)}|\big) \\
&\qquad \le (2^\alpha C_\rho \vee 1)^C \left(\prod_{c=1}^C m_c^{\alpha+1}\right) \prod_{c=1}^C \big((\ell_c - \ell_{c-1} - |\eta^{(c-1)}|) \vee 1\big)^{-\alpha} \\
&\qquad \le (2^\alpha C_\rho \vee 1)^C \Big(\frac{N - CM}{G}\Big)^{(\alpha+1)C} \prod_{c=1}^C \big((\ell_c - \ell_{c-1} - |\eta^{(c-1)}|) \vee 1\big)^{-\alpha} \\
&\qquad \le \exp[N\delta_5] \prod_{c=1}^C \big((\ell_c - \ell_{c-1} - |\eta^{(c-1)}|) \vee 1\big)^{-\alpha},
\end{aligned} \tag{3.30}$$

where $\delta_5 = \delta(\delta_1, M)$ can be made arbitrarily small by choosing M large and δ_1 small. For the second inequality, we have used the fact that the product $\prod_{c=1}^C m_c^{\alpha+1}$ is maximal when all factors are equal.

Combining (3.23–3.30), we obtain

$$\begin{aligned}
\mathbb{P}\big(R_N \in \mathcal{O} \mid X\big) &\le \exp\Big[N\Big(H_{\tau|K}(Q) + \mathbb{E}_Q[\log \rho(\tau_1)] + \delta_2 + \delta_3 + \delta_4 + \delta_5\Big)\Big] \\
&\qquad \times \sum_{(\ell_c),\, (\eta^{(c)}) \text{ good}} \prod_{c=1}^C \big((\ell_c - \ell_{c-1} - |\eta^{(c-1)}|) \vee 1\big)^{-\alpha}.
\end{aligned} \tag{3.31}$$

Combining (3.31) with Lemma 3.2 below, and recalling the identity in (1.32), we obtain the result in Proposition 3.1 for ρ satisfying (3.3), with $\mathcal{O}$ defined in (3.15) and $\varepsilon = \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6$. Note that ε can be made arbitrarily small by choosing ε_1, δ_1 small and M large.

3.4 Step 4: Cost of finding good sentences

Lemma 3.2. *For $\varepsilon_1, \delta_1 > 0$ and $M \in \mathbb{N}$,*

$$\begin{aligned} &\limsup_{N\to\infty} \frac{1}{N} \log \left[\sum_{(\ell_c),\ (\eta^{(c)})\ \text{good}} \prod_{c=1}^{C} \big((\ell_c - \ell_{c-1} - |\eta^{(c-1)}|) \vee 1\big)^{-\alpha} \right] \\ &\qquad \leq -\alpha\, m_Q H(\Psi_Q \mid \nu^{\otimes \mathbb{N}}) + \delta_6 \quad a.s., \end{aligned} \tag{3.32}$$

where $\delta_6 = \delta(\varepsilon_1, \delta_1, M)$ can be made arbitrarily small by choosing M large and ε_1, δ_1 small.

Proof. Note that, by the fourth line of (3.6), for any $\eta \in \mathscr{B}$ (recall (3.10)) and $k \in \mathbb{N}$,

$$\mathbb{P}\big(\eta \text{ starts at position } k \text{ in } X\big) \leq \exp\big[M\big(m_Q \mathbb{E}_{\Psi_Q}[\log \nu(X_1)] + \varepsilon_1\big)\big]. \tag{3.33}$$

Combining this with (3.12), we get

$$\begin{aligned} &\mathbb{P}\big(\text{some element of } \mathscr{B} \text{ starts at position } k \text{ in } X\big) \\ &\qquad \leq \exp\big[M\big(m_Q \mathbb{E}_{\Psi_Q}[\log \nu(X_1)] + \varepsilon_1\big)\big] \times \exp\big[M\big(m_Q H(\Psi_Q) + \varepsilon_1\big)\big] \\ &\qquad = \exp\big[-M\big(m_Q H(\Psi_Q \mid \nu^{\otimes \mathbb{N}}) - 2\varepsilon_1\big)\big], \end{aligned} \tag{3.34}$$

where we use (1.26).

Next, we *coarse-grain* the sequence X into blocks of length

$$L := \lfloor M(m_Q - \varepsilon_1) \rfloor, \tag{3.35}$$

and compare the coarse-grained sequence with a *low-density Bernoulli sequence.* To this end, define a $\{0,1\}$-valued sequence $(A_l)_{l\in\mathbb{N}}$ inductively as follows. Put $A_0 := 0$, and, for $l \in \mathbb{N}$ given that $A_0, A_1, \dots, A_{l-1}$ have been assigned values, define A_l by distinguishing the following two cases:

(1) If $A_{l-1} = 0$, then

$$A_l := \begin{cases} 1, & \text{if in } X \text{ there is a word } \eta \in \mathscr{B} \text{ starting in } ((l-1)L, lL], \\ 0, & \text{otherwise.} \end{cases} \tag{3.36}$$

(2) If $A_{l-1} = 1$, then

$$A_l := \begin{cases} 1, & \begin{array}{l} \text{if in } X \text{ there are words } \eta, \eta' \in \mathscr{B} \text{ starting in } ((l-2)L, (l-1)L], \\ \text{respectively, } ((l-1)L, lL] \text{ and occurring disjointly,} \end{array} \\ 0, & \text{otherwise.} \end{cases} \tag{3.37}$$

Put

$$p := L \exp\big[-M\big(m_Q H(\Psi_Q \mid \nu^{\otimes \mathbb{N}}) - 2\varepsilon_1\big)\big]. \tag{3.38}$$

Then we claim

$$\mathbb{P}(A_1 = a_1, \dots, A_n = a_n) \le p^{a_1+\cdots+a_n}, \quad n \in \mathbb{N},\ a_1, \dots, a_n \in \{0,1\}. \tag{3.39}$$

In order to verify (3.39), fix $a_1, \dots, a_n \in \{0,1\}$ with $a_1 + \cdots + a_n = m$. By construction, for the event in the left-hand side of (3.39) to occur there must be m non-overlapping elements of $\mathscr{B}$ at certain positions in X. By (3.34), the occurrence of any m fixed starting positions has probability at most

$$\exp\big[-mM\big(m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - 2\varepsilon_1\big)\big], \tag{3.40}$$

while the choice of the a_l's dictates that there are at most L^m possibilities for the starting points of the m words.

By (3.39), we can couple the sequence $(A_l)_{l\in\mathbb{N}}$ with an i.i.d. Bernoulli(p)-sequence $(\omega_l)_{l\in\mathbb{N}}$ such that

$$A_l \le \omega_l \qquad \forall\, l \in \mathbb{N} \quad \text{a.s.} \tag{3.41}$$

(Note that (3.39) guarantees the existence of such a coupling for any fixed n. In order to extend this existence to the infinite sequence, observe that the set of functions depending on finitely many coordinates is dense in the set of continuous increasing functions on $\{0,1\}^{\mathbb{N}}$, and use the results in Strassen [23].)

Each admissible choice of $\ell_1, \dots, \ell_C$ in (3.32) leads to a C-tuple $i_1 < \cdots < i_C$ such that $A_{i_1} = \cdots = A_{i_C} = 1$ (since it cuts out non-overlapping words, which is compatible with (3.36–3.37)), and for any such $(i_1, \dots, i_C)$ there are at most L^C different admissible choices of the ℓ_c's. Thus, we have

$$\sum_{(\ell_c),\ (\eta^{(c)})\ \text{good}} \prod_{c=1}^{C} \big((\ell_c - \ell_{c-1} - |\eta^{(c-1)}|) \vee 1\big)^{-\alpha} \le L^C L^{-\alpha} \sum_{\substack{0<i_1<\cdots<i_C<\infty\\ A_{i_1}=\cdots=A_{i_C}=1}} \prod_{c=1}^{C} (i_c - i_{c-1})^{-\alpha}. \tag{3.42}$$

Using (3.3) and (3.19), and recalling the definition of $\phi(\alpha, p)$ in (2.2), we have

$$\limsup_{N\to\infty} \frac{1}{N} \log\big[\text{r.h.s. } (3.42)\big] \le \frac{1-\delta_1}{M} \Big(\log\big(Mm_Q\big) - \log C_\rho - \phi(\alpha,p)\Big) \qquad \omega - a.s. \tag{3.43}$$

From (3.38) we know that $\log(1/p) \sim M(m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - 2\varepsilon_1)$ as $M \to \infty$ and so, by Lemma 2.1, we have

$$\text{r.h.s. } (3.43) \le -(1-\varepsilon_2)\alpha\big(m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - 2\varepsilon_1\big) \tag{3.44}$$

for any $\varepsilon_2 \in (0,1)$, provided M is large enough. This completes the proof of Lemma 3.2, and hence of Proposition 3.1 for $Q \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$. $\square$

3.5 Step 5: Removing the assumption of ergodicity

Sections 3.1–3.4 contain the main ideas behind the proof of Proposition 3.1. In the present section we extend the bound from $\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$. This requires setting up a variant of the argument in Sections 3.1–3.4 in which the ergodic components of Q are "approximated with a common length scale on the letter level". This turns out to be technically involved and to fall apart into 6 substeps.

Let $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ have a non-trivial ergodic decomposition

$$Q = \int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})} Q'\, W_Q(dQ'), \tag{3.45}$$

where W_Q is a probability measure on $\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ (Georgii [13], Proposition 7.22). We may assume w.l.o.g. that $H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) < \infty$, otherwise we can simply employ the annealed bound. Thus, W_Q is in fact supported on $\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}}) \cap \{Q' \colon H(Q' \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) < \infty\}$.

Fix $\varepsilon > 0$. In the following steps, we will construct an open neighbourhood $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ of Q satisfying (3.1) (for technical reasons with ε replaced by some $\varepsilon' = \varepsilon'(\varepsilon)$ that becomes arbitrarily small as $\varepsilon \downarrow 0$).

3.5.1 Preliminaries

Observing that

$$m_Q = \int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})} m_{Q'}\, W_Q(dQ') < \infty, \qquad H(Q|q_{\rho,\nu}^{\otimes\mathbb{N}}) = \int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})} H(Q'|q_{\rho,\nu}^{\otimes\mathbb{N}})\, W_Q(dQ') < \infty, \tag{3.46}$$

we can find $K_0, K_1, m^* > 0$ and a compact set

$$\mathscr{C} \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \cap \mathrm{supp}(W_Q) \cap \{Q \colon H(\cdot|q_{\rho,\nu}^{\otimes\mathbb{N}}) \le K_0\} \tag{3.47}$$

such that

$$\sup\{H(\Psi_P \mid \nu^{\otimes\mathbb{N}}) \colon P \in \mathscr{C}\} \le K_1, \tag{3.48}$$

$$\sup\{m_P \colon P \in \mathscr{C}\} \le m^*, \tag{3.49}$$

$$\text{the family } \{\mathscr{L}_P(\tau_1) \colon P \in \mathscr{C}\} \text{ is uniformly integrable,} \tag{3.50}$$

$$W_Q(\mathscr{C}) \ge 1 - \varepsilon/2, \tag{3.51}$$

$$\int_{\mathscr{C}} H(Q'|q_{\rho,\nu}^{\otimes\mathbb{N}})\, W_Q(dQ') \ge H(Q|q_{\rho,\nu}^{\otimes\mathbb{N}}) - \varepsilon/2, \tag{3.52}$$

$$\int_{\mathscr{C}} m_{Q'} H(\Psi_{Q'}|\nu^{\otimes\mathbb{N}})\, W_Q(dQ') \ge m_Q H(\Psi_Q|\nu^{\otimes\mathbb{N}}) - \varepsilon/2. \tag{3.53}$$

In order to check (3.50), observe that $\mathbb{E}_Q[\tau_1] < \infty$ implies that there is a sequence (c_n) with $\lim_{n\to\infty} c_n = \infty$ such that

$$\mathbb{E}_Q\big[\tau_1 \mathbb{1}_{\{\tau_1 \ge c_n\}}\big] \le \frac{6}{\pi^2 n^3}\,\frac{\varepsilon}{6}, \quad n \in \mathbb{N}. \tag{3.54}$$

Put

$$\widehat{A}_n := \{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \colon \mathbb{E}_{Q'}\big[\tau_1 \mathbb{1}_{\{\tau_1 \ge c_n\}}\big] > 1/n\} \tag{3.55}$$

and $A := \cap_{n\in\mathbb{N}} (\widehat{A}_n)^c$. Each $\widehat{A}_n$ is open, hence A is closed, and by the Markov inequality we have

$$W_Q\Big(\{Q' \colon \mathbb{E}_{Q'}\big[\tau_1 \mathbb{1}_{\{\tau_1 \ge c_n\}}\big] > 1/n\}\Big) \le n\mathbb{E}_Q\big[\tau_1 \mathbb{1}_{\{\tau_1 \ge c_n\}}\big] \le \frac{6}{\pi^2 n^2}\,\frac{\varepsilon}{6}. \tag{3.56}$$

Thus,

$$W_Q(A^c) = W_Q\big(\cup_{n\in\mathbb{N}} (\widehat{A}_n)^c\big) \le \frac{\varepsilon}{6} \sum_{n\in\mathbb{N}} \frac{6}{\pi^2 n^2} = \frac{\varepsilon}{6}. \tag{3.57}$$

This implies that the mapping

$$Q' \mapsto m_{Q'} H(\Psi_{Q'} | \nu^{\otimes\mathbb{N}}) \quad \text{is lower semicontinuous on } \mathscr{C}. \tag{3.58}$$

Indeed, if $\mathrm{w}-\lim_{n\to\infty} Q'_n = Q''$ and $(Q'_n) \subset \mathscr{C}$, then $\lim_{n\to\infty} \mathbb{E}_{Q'_n}[\tau_1] = \lim_{n\to\infty} m_{Q'_n} = m_{Q''} = \mathbb{E}_{Q''}[\tau_1]$ and $\mathrm{w}-\lim_{n\to\infty} \Psi_{Q'_n} = \Psi_{Q''}$ by uniform integrability (see Birkner [3], Remark 7).

Furthermore, we can find $N_0, L_0 \in \mathbb{N}$ with $L_0 \le N_0$ and a finite set $\widetilde{W} \subset \widetilde{E}^{N_0}$ such that the following holds. Let

$$W := \Big\{ \pi_{L_0}(\theta^i \kappa(\zeta)) \colon\ \zeta = (\zeta^{(1)}, \dots, \zeta^{(N_0)}) \in \widetilde{W}, 0 \le i < |\zeta^{(1)}| \Big\} \tag{3.59}$$

be the set of words of length L_0 obtained by concatenating sentences from $\widetilde{W}$, possibly shifting the "origin" inside the first word and restricting to the first L_0 letters. Then for all $P \in \mathscr{D} \subset \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}) \cap \mathscr{C}$ that satisfy

$$\sum_{\zeta\in\widetilde{W}} P(\zeta) \;\ge\; 1 - \frac{\varepsilon}{3c_{\lceil 3/\varepsilon\rceil}}, \tag{3.60}$$

$$\frac{1}{N_0} \sum_{\zeta\in\widetilde{W}} P(\zeta) \log \frac{P(\zeta)}{q_{\rho,\nu}^{\otimes N_0}(\zeta)} \;\ge\; H(P \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - \varepsilon/2, \tag{3.61}$$

$$\frac{1}{L_0} \sum_{w\in W} \Psi_P(w) \log \frac{\Psi_P(w)}{\nu^{\otimes L_0}(w)} \;\ge\; H(\Psi_P \mid \nu^{\otimes\mathbb{N}}) - \varepsilon/2, \tag{3.62}$$

the following inequalities hold:

$$W_Q(\mathscr{D}) \;\ge\; 1 - 3\varepsilon/4, \tag{3.63}$$

$$\int_{\mathscr{D}} H(P \mid q_{\rho,\nu}^{\otimes\mathbb{N}})\, W_Q(dP) \;\ge\; H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - 3\varepsilon/4, \tag{3.64}$$

$$\int_{\mathscr{D}} m_P H(\Psi_P | \nu^{\otimes\mathbb{N}})\, W_Q(dP) \;\ge\; m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - 3\varepsilon/4. \tag{3.65}$$

We may choose the set $\widetilde{W}$ in such a way that

$$\delta_{\widetilde{W}} := \min\{q_{\rho,\nu}^{\otimes N_0}(\zeta) \colon \zeta \in \widetilde{W}\} \wedge \frac{\min\{\nu^{\otimes L_0}(\xi) \colon\ \xi \in W\}}{\max\{|\zeta^{(1)}| \colon\ \zeta \in \widetilde{W}\}} > 0. \tag{3.66}$$

3.5.2 Approximating with a given length scale on the letter level

For $\delta > 0$ and $L \in \mathbb{N}$, we say that $P \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ can be (δ, L)-*approximated* if there exists a finite subset $\mathscr{A}_P \subset \widetilde{E}^{\lceil L/m_P\rceil}$ of "P-typical" sentences, each consisting of $\approx L/m_P$ words, such that

$$P_{|\mathscr{F}_{\lceil L/m_P\rceil}}(\mathscr{A}_P) > 1 - \frac{\delta}{2}\delta_{\widetilde{W}} \Big(\min\big\{P(\zeta) \colon\ \zeta \in \widetilde{W}, P(\zeta) > 0\big\} \wedge \min\big\{\Psi_P(\xi) \colon\ \xi \in W, \Psi_P(\xi) > 0\big\}\Big) \tag{3.67}$$

and, for all $z = (y^{(1)}, \dots, y^{(\lceil L/m_P \rceil)}) \in \mathscr{A}_P$,

$$
\begin{aligned}
P(z) &\in \Big[\exp\big[-\lceil L/m_P\rceil(H(Q)+\delta)\big], \exp\big[-\lceil L/m_P\rceil(H(Q)-\delta)\big]\Big],\\
|\kappa(z)| &\in [L(1-\delta), L(1+\delta)],\\
P\big(K^{(\lceil L/m_P\rceil)} = z\big) &\in \Big[\exp\big[-L(H(\Psi_Q)+\delta)\big], \exp[-L(H(\Psi_Q)-\delta)]\Big],\\
\sum_{k=1}^{|\kappa(z)|} \log \nu(\kappa(z)_k) &\in [L(1-\delta), L(1+\delta)]\ \mathbb{E}_{\Psi_P}\big[\log \nu(X_1)\big],\\
\sum_{i=1}^{\lceil L/m_P\rceil} \log \rho(|y^{(i)}|) &\in [(L/m_P)(1-\delta), (L/m_P)(1+\delta)]\ \mathbb{E}_P\big[\log \rho(\tau_1)\big],\\
|\{z' \in \mathscr{A}_P\colon\ \kappa(z) = \kappa(z')\}| &\le \exp\big[(L/m_P)(H_{\tau|K}(P)+\delta)\big].
\end{aligned}
\tag{3.68}
$$

By the third and the fifth line of (3.68) we have, using (1.26),

$$
\mathbb{P}\big(X \text{ starts with some element of } \kappa(\mathscr{A}_P)\big) \le \exp\Big[-L(1-2\delta)H\big(\Psi_Q \mid \nu^{\otimes\mathbb{N}}\big)\Big]. \tag{3.69}
$$

For P that can be (δ, L)-approximated, define an open neighbourhood of P via

$$
\mathcal{U}_{(\delta,L)}(P) := \left\{P' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) : \frac{P'(z)}{P(z)} \in (1-\delta, 1+\delta)\ \ \forall\, z \in \mathscr{A}_P\right\}, \tag{3.70}
$$

where $\mathscr{A}_P = \mathscr{A}_P(\delta, L)$ is the set from (3.67–3.68). By the results of Section 3.1 and the above, for given $P \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}}) \cap \mathscr{C}$ and $\delta_0 > 0$ there exist $\delta' \in (0, \delta_0)$ and L_0 such that

$$
\forall\, L' \ge L_0\colon\ \ P \text{ can be } (\delta', L')\text{-approximated.} \tag{3.71}
$$

Assume that a given $P \in \mathscr{D}$ can be (δ, L)-approximated for some L such that $\lceil L/m_P\rceil \ge N_0$. We claim that then for any $P' \in \mathscr{D} \cap \mathcal{U}_{(\delta,L)}(P)$,

$$
\forall\, \zeta \in \widetilde{W}\colon\ P'(\zeta) \le \begin{cases} (1+2\delta)P(\zeta) & \text{if } P(\zeta) > 0,\\ \min\{q_{\rho,\nu}^{\otimes N_0}(\zeta')\colon\ \zeta' \in \widetilde{W}\} & \text{otherwise,}\end{cases} \tag{3.72}
$$

$$
\forall\, \xi \in W\colon\ m_{P'}\Psi_{P'}(\xi) \le \begin{cases} (1+2\delta)m_P\Psi_P(\xi) & \text{if } \Psi_P(\xi) > 0,\\ \min\{\nu^{\otimes L_0}(\xi')\colon\ \xi' \in W\} & \text{otherwise,}\end{cases} \tag{3.73}
$$

$$
m_{P'} \ge (1-3\delta)m_P. \tag{3.74}
$$

To verify (3.72), note that, for $\zeta \in \widetilde{W}$,

$$
\begin{aligned}
P'(\zeta) &\le \sum_{z\in\mathscr{A}_P\colon\ \pi_{N_0}(z)=\zeta} P'(z) + \sum_{z\in\widetilde{E}^{\lceil L/m_P\rceil}\setminus\mathscr{A}_P\colon\ \pi_{N_0}(z)=\zeta} P'(z)\\
&\le (1+\delta)\sum_{z\in\mathscr{A}_P:\pi_{N_0}(z)=\zeta} P(z) + P'\big(\widetilde{E}^{\lceil L/m_P\rceil}\setminus\mathscr{A}_P\big)\\
&\le (1+\delta)\Big[P(\zeta) + \frac{\delta}{2}\Big(\min\big\{P(\zeta)\colon\ \zeta \in \widetilde{W}, P(\zeta) > 0\big\} \wedge \min\{q_{\rho,\nu}^{\otimes N_0}(\zeta)\colon\ \zeta \in \widetilde{W}\}\Big)\Big].
\end{aligned}
$$

To verify (3.73), observe that, for $\xi \in W$ (recall the definition of $\Psi_{P'}$ from (1.10)),

$$\begin{aligned} m_{P'}\Psi_{P'}(\xi) &= \sum_{\zeta\in\widetilde{W}} P'(\zeta) \sum_{i=0}^{|\zeta^{(1)}|-1} \mathbb{1}_{\{\xi\}}\big(\pi_{L_0}(\theta^i\kappa(\zeta))\big) \\ &\le (1+\delta)m_P\Psi_P(\xi) + \sum_{\zeta\in\widetilde{W}:\ P(\zeta)=0} |\zeta^{(1)}|P'(\zeta) \end{aligned} \tag{3.75}$$

and that the sum in the second line above is bounded by

$$\max_{\eta\in\widetilde{W}} |\eta^{(1)}|P'\big(\widetilde{E}^{N_0}\setminus\mathscr{A}_P\big) \le (1+\delta)\frac{\delta}{2}\Big(\min\big\{\Psi_P(\xi)\colon\ \xi\in W, \Psi_P(\xi)>0\big\} \wedge \min\{\nu^{\otimes L_0}(\xi')\colon\ \xi'\in W\}\Big). \tag{3.76}$$

Lastly, to verify (3.74), note that

$$P'(\zeta) \ge (1-2\delta)P(\zeta) \quad \forall\,\zeta\in\widetilde{W} \tag{3.77}$$

(which can be proved in the same way as (3.72)), so that

$$m_{P'} = \sum_{y\in\widetilde{E}} |y|P'(y) \ge \sum_{\zeta\in\widetilde{W}} |\zeta^{(1)}|P'(\zeta) \ge (1-2\delta)\sum_{\zeta\in\widetilde{W}} |\zeta^{(1)}|P(\zeta). \tag{3.78}$$

Furthermore,

$$m_P \le \sum_{\zeta\in\widetilde{W}} |\zeta^{(1)}|P(\zeta) + c_{\lceil 3/\delta\rceil}P\big(\widetilde{E}^{N_0}\setminus\widetilde{E}\big) + \sum_{y\widetilde{E}:\ |y|>c_{\lceil 3/\delta\rceil}} |y|P(y). \tag{3.79}$$

Observing that the second and the third term on the right-hand side are each at most $\delta/3$, we find that (3.78–3.79) imply (3.74).

Finally, observe that (3.72–3.74) imply that, for any $P, P' \in \mathscr{D}$ such that P can be (δ, L)-approximated for some L with $\lceil L/m_P\rceil \ge N_0$ and $P' \in \mathcal{U}_{(\delta,L)}(P)$,

$$\begin{aligned} H(P' \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) &\le H(P \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + 2K_0\delta + \varepsilon/2, &(3.80)\\ m_{P'}H(\Psi_{P'} \mid \nu^{\otimes\mathbb{N}}) &\le m_P H(\Psi_P \mid \nu^{\otimes\mathbb{N}}) + 2K_1\delta + \varepsilon/2. &(3.81) \end{aligned}$$

Here, (3.80) follows from the observation

$$\begin{aligned} &H(P' \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - \frac{\varepsilon}{2} \\ &\le \frac{1}{N_0}\sum_{\zeta\in\widetilde{W}} P'(\zeta)\log\frac{P'(\zeta)}{q_{\rho,\nu}^{\otimes N_0}(\zeta)} \\ &\le \frac{1+2\delta}{N_0}\sum_{\zeta\in\widetilde{W}} P(\zeta)\log\frac{(1+2\delta)P(\zeta)}{q_{\rho,\nu}^{\otimes N_0}(\zeta)} + \frac{1}{N_0}\sum_{\zeta\in\widetilde{W}:P(\zeta)=0} P'(\zeta)\log\frac{\min\{q_{\rho,\nu}^{\otimes N_0}(\zeta')\colon \zeta'\in\widetilde{W}\}}{q_{\rho,\nu}^{\otimes N_0}(\zeta)} \\ &\le (1+2\delta)H(P \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + \frac{1+2\delta}{N_0}\log(1+2\delta). \end{aligned} \tag{3.82}$$

Similarly, observing that

$$
\begin{aligned}
& m_{P'} \sum_{\xi\in W} \Psi_{P'}(\xi) \log \frac{m_{P'}\Psi_{P'}(\xi)}{m_{P'}\nu^{\otimes L_0}(\xi)} \\
& \leq (1+2\delta) m_P \sum_{\xi\in W} \Psi_P(\xi) \log \frac{(1+2\delta)m_P\Psi_P(\xi)}{(1-3\delta)m_P\nu^{\otimes L_0}(\xi)} + m_{P'} \sum_{\xi\in W\colon\ \Psi_P(\xi)=0} \Psi_{P'}(\xi) \log \frac{\min\{\nu^{\otimes L_0}(\xi')\colon \xi'\in W\}}{\nu^{\otimes L_0}(\xi)} \\
& \leq (1+2\delta) L_0 m_P H(\Psi_P \mid \nu^{\otimes\mathbb{N}}) + (1+2\delta) m^* \log(1+6\delta),
\end{aligned}
\tag{3.83}
$$

we obtain (3.81) in view of (3.62).

3.5.3 Approximating the ergodic decomposition

In the previous subsection, we have approximated a given $P \in \mathcal{P}^{\mathrm{erg,fin}}$, i.e., we have constructed a certain neighbourhood of P w.r.t. the weak topology, which requires only conditions on the frequencies of sentences whose concatenations are $\approx L$ letters long. While the required L will in general vary with P, we now want to construct a compact $\mathscr{C}' \subset \mathscr{C}$ such that $W_Q(\mathscr{C}')$ is still close to 1 and all $P \in \mathscr{C}'$ can be approximated on the same scale L (on the letter level). To this end, let

$$\mathscr{D}_{\varepsilon',L'} := \big\{P \in \mathscr{D}\colon\ P \text{ can be } (\varepsilon', L')\text{-approximated}\big\}. \tag{3.84}$$

By (3.71), we have

$$\bigcup_{\substack{\varepsilon'\in(0,\varepsilon/2)\\ L'\in\mathbb{N}}} \mathscr{D}_{\varepsilon',L'} = \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}}) \cap \mathscr{C}, \tag{3.85}$$

so, in view of (3.51–3.53), we can choose

$$0 < \varepsilon_1 < \frac{\varepsilon}{8(1\vee K_0 \vee K_1)} \tag{3.86}$$

and $L \in \mathbb{N}$ such that

$$
\begin{aligned}
W_Q(\mathscr{D}_{\varepsilon_1,L}) &\geq 1-\varepsilon, && (3.87)\\
\int_{\mathscr{D}_{\varepsilon_1,L}} H(Q' \mid q_{\rho,\nu}^{\otimes\mathbb{N}})\, W_Q(dQ') &\geq H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - \varepsilon, && (3.88)\\
\int_{\mathscr{D}_{\varepsilon_1,L}} m_{Q'} H(\Psi_{Q'} \mid \nu^{\otimes\mathbb{N}})\, W_Q(dQ') &\geq m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - \varepsilon. && (3.89)
\end{aligned}
$$

For $P \in \mathscr{D}_{\varepsilon_1,L}$, let

$$\mathcal{U}'(P) := \left\{P' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon\ \frac{P'(z)}{P(z)} \in \Big(1-\frac{\varepsilon_1}{2}, 1+\frac{\varepsilon_1}{2}\Big)\ \ \forall\, z \in \mathscr{A}_P\right\}, \tag{3.90}$$

where $\mathscr{A}_P$ is the set from (3.67–3.68) that appears in the definition of $\mathcal{U}_{(\varepsilon_1,L)}(P)$. Note that $\mathcal{U}'(P) \subset \mathcal{U}_{(\varepsilon_1,L)}(P)$. Indeed, $\inf_{P\in\mathscr{D}_{\varepsilon_1,L}} \mathrm{dist}(\mathcal{U}'(P), \mathcal{U}_{(\varepsilon_1,L)}(P)^c) > 0$ if we metrize the weak topology. Consequently,

$$\mathscr{C}' := \mathscr{C} \cap \overline{\cup_{P\in\mathscr{D}_{\varepsilon_1,L}} \mathcal{U}'(P)}\ \big(\supset \mathscr{D}_{\varepsilon_1,L}\big) \tag{3.91}$$

is compact and satisfies $W_Q(\mathscr{C}') \geq 1-\varepsilon$, and

$$\mathscr{C}' \subset \bigcup_{P \in \mathscr{D}_{\varepsilon_1,L}} \mathcal{U}_{(\varepsilon_1,L)}(P) \tag{3.92}$$

is an open cover. By compactness there exist $R \in \mathbb{N}$ and (pairwise different) $Q_1,\dots,Q_R \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}}) \cap \mathscr{C}$ such that

$$\mathcal{U}_{(\varepsilon_1,L)}(Q_1) \cup \cdots \cup \mathcal{U}_{(\varepsilon_1,L)}(Q_R) \supset \mathscr{C}', \tag{3.93}$$

where $\mathcal{U}_{(\varepsilon_1,L)}(Q_r)$ is of the type (3.70) with a set $\mathscr{A}_r \subset \widetilde{E}^{M_r}$ satisfying (3.67–3.68) with P replaced by Q_r, and $M_r = \lceil L/m_{Q_r} \rceil$.

For $z \in \cup_{n\in\mathbb{N}} \widetilde{E}^n$ consider the probability measure on $[0,1]$ given by $\mu_{Q,z}(B) := W_Q(\{Q' \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})\colon\ Q'(z) \in B\})$, $B \subset [0,1]$ measurable. Observing that

$$\bigcup_{r=1}^{R} \bigcup_{z\in\mathscr{A}_r} \{u \in [0,1]\colon\ u \text{ is an atom of } \mu_{Q,z}\} \tag{3.94}$$

is at most countable, we can find $\varepsilon_2 \in [\varepsilon_1, \varepsilon_1+\varepsilon_1^2)$ (note that still $\varepsilon_2 < \varepsilon$) and $\widetilde{\delta} > 0$ such that

$$\begin{aligned} & W_Q\left(\left\{Q' \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})\colon \begin{array}{c} Q'(z)/Q_r(z) \in [1-\varepsilon_2-\widetilde{\delta}, 1-\varepsilon_2+\widetilde{\delta}] \cup [1+\varepsilon_2-\widetilde{\delta}, 1+\varepsilon_2+\widetilde{\delta}] \\ \text{for some } r \in \{1,\dots,R\} \text{ and } z \in \mathscr{A}_r \end{array} \right\}\right) \\ & \qquad \leq \frac{\varepsilon}{1 \vee K_0 \vee m^* K_1}. \end{aligned} \tag{3.95}$$

Define “disjointified” versions of the $\mathcal{U}_{(\varepsilon,L)}(Q_r)$ as follows. For $r = 1,\dots,R$, put iteratively

$$\widetilde{\mathcal{U}}_r := \left\{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon \begin{array}{c} Q'(z) \in Q_r(z)(1-\varepsilon_2, 1+\varepsilon) \text{ for all } z \in \mathscr{A}_r \text{ and for each } r' < r \text{ there} \\ \text{is } z' \in \mathscr{A}_{r'} \text{ such that } Q'(z') \notin Q_{r'}(z')[1-\varepsilon_2-\widetilde{\delta}, 1+\varepsilon_2+\widetilde{\delta}] \end{array} \right\}. \tag{3.96}$$

It may happen that some of the $\widetilde{\mathcal{U}}_r$ are empty or satisfy $W_Q(\widetilde{\mathcal{U}}_r) = 0$. We then (silently) remove these and re-number the remaining ones. Note that each $\widetilde{\mathcal{U}}_r$ is an open subset of $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ and

$$W_Q\big(\cup_{r=1}^R \widetilde{\mathcal{U}}_r\big) = \sum_{r=1}^R W_Q(\widetilde{\mathcal{U}}_r) \geq 1-2\varepsilon, \tag{3.97}$$

since $W_Q\big(\mathscr{C}' \setminus \cup_{r=1}^R \widetilde{\mathcal{U}}_r\big) \leq \varepsilon$.

For $r = 1,\dots,R$, we have, using (3.80–3.81) and the choice of ε_2 ($\leq 2\varepsilon_1$),

$$W_Q(\widetilde{\mathcal{U}}_r \cap \mathscr{D})\Big(H(Q_r \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + \varepsilon\Big) \geq \int_{\widetilde{\mathcal{U}}_r\cap\mathscr{D}} H(Q' \mid q_{\rho,\nu}^{\otimes\mathbb{N}})\, W_Q(dQ'), \tag{3.98}$$

$$W_Q(\widetilde{\mathcal{U}}_r \cap \mathscr{D})\Big(m_{Q_r} H(\Psi_{Q_r} \mid \nu^{\otimes\mathbb{N}}) + \varepsilon\Big) \geq \int_{\widetilde{\mathcal{U}}_r\cap\mathscr{D}} m_{Q'} H(\Psi_{Q'} \mid \nu^{\otimes\mathbb{N}})\, W_Q(dQ'), \tag{3.99}$$

so that altogether

$$\begin{aligned} & \sum_{r=1}^R W_Q(\widetilde{\mathcal{U}}_r) \Big\{ H(Q_r \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1) m_{Q_r} H(\Psi_{Q_r} \mid \nu^{\otimes\mathbb{N}}) \Big\} \\ & \geq H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1) m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - (3+3\alpha)\varepsilon. \end{aligned} \tag{3.100}$$

3.5.4 More layers: long sentences with the right pattern frequencies

For $z \in \cup_{n\in\mathbb{N}}\widetilde{E}^n$ and $\xi = (\xi^{(1)}, \dots, \xi^{(\widetilde{M})}) \in \widetilde{E}^M$ (with $M > |z|$), let

$$\mathrm{freq}_z(\xi) = \frac{1}{M}\big|\{1 \le i \le M - |z|\colon\ (\xi^{(i)}, \dots, \xi^{(i+|z|-1)}) = z\}\big| \tag{3.101}$$

be the empirical frequency of z in ξ. Note that, for any $P \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$, $z \in \cup_{n\in\mathbb{N}}\widetilde{E}^n$ and $\varepsilon > 0$, we have

$$\lim_{M\to\infty} P\Big(\big\{\xi \in \widetilde{E}^M\colon\ \mathrm{freq}_z(\xi) \in P(z)(1-\varepsilon, 1+\varepsilon)\big\}\Big) = 1 \tag{3.102}$$

and

$$\lim_{M\to\infty} P\Big(\big\{\xi \in \widetilde{E}^M\colon\ |\kappa(\xi)| \in M(m_P - \varepsilon, m_P + \varepsilon)\big\}\Big) = 1. \tag{3.103}$$

For $\widetilde{M} \in \mathbb{N}$ and $r \in \{1, \dots, R\}$, put

$$\begin{aligned} V_{r,\widetilde{M}} := \Big\{\xi \in \widetilde{E}^{\widetilde{M}}\colon \\ &|\kappa(\xi)| \in \widetilde{M}(m_{Q_r} - \varepsilon, m_{Q_r} + \varepsilon),\ \mathrm{freq}_z(\xi) \in Q_r(z)(1 - \varepsilon_2, 1 + \varepsilon_2) \ \text{ for all } z \in \mathscr{A}_r, \text{ and for each} \\ &\qquad r' < r \text{ there is a } z' \in \mathscr{A}_{r'} \text{ such that } \mathrm{freq}_{z'}(\xi) \notin Q_{r'}(z')[1 - \varepsilon_2 - \widetilde{\delta}, 1 + \varepsilon_2 + \widetilde{\delta}]\Big\}. \end{aligned} \tag{3.104}$$

Note that when $|E| < \infty$, also $|V_{r,\widetilde{M}}| < \infty$. Furthermore, $V_{r,\widetilde{M}} \cap V_{r',\widetilde{M}} = \emptyset$ for $r \neq r'$. For $\xi \in V_{r,\widetilde{M}}$, we have

$$\Big|\Big\{1 \le i \le \widetilde{M} - M_r + 1\colon\ \big(\xi^{(i)}, \xi^{(i+1)}, \dots, \xi^{(i+M_r-1)}\big) \in \mathscr{A}_r\Big\}\Big| \ge \widetilde{M}(1 - 2\varepsilon), \tag{3.105}$$

in particular, there are at least $K_r := \lfloor \widetilde{M}(1 - 3\varepsilon)/M_r \rfloor$ elements $z_1, \dots, z_{K_r} \in \mathscr{A}_r$ (not necessarily distinct) appearing in this order as disjoint subwords of ξ. The z_k's can for example be constructed in a "greedy" way, parsing ξ from left to right as in Section 3.2 (see, in particular, (3.21)). This implies, in particular, that

$$\begin{aligned} \prod_{i=1}^{\widetilde{M}} \rho(|\xi^{(i)}|) &\le \prod_{k=1}^{K_r} \prod_{w \text{ in } z_k} \rho(|w|) \le \Big(\exp\big[(1-\varepsilon)\widetilde{M}_r \mathbb{E}_{Q_r}[\log \rho(\tau_1)]\big]\Big)^{K_r} \\ &\le \exp\Big[(1 - 4\varepsilon)\widetilde{M}\mathbb{E}_{Q_r}[\log \rho(\tau_1)]\Big] \end{aligned} \tag{3.106}$$

if $\widetilde{M}$ is large enough. Furthermore, for each $r \in \{1, \dots, R\}$ and $\eta \in V_{r,\widetilde{M}}$, we have

$$\big|\big\{\zeta \in V_{r,\widetilde{M}}\colon\ \kappa(\zeta) = \kappa(\eta)\big\}\big| \le \exp\Big[\widetilde{M}(H_{\tau|K}(Q_r) + \delta_1)\Big], \tag{3.107}$$

where δ_1 can be made arbitrarily small by choosing ε small. (Note that the quantity on the left-hand side is the number of ways in which $\kappa(\eta)$ can be "re-cut" to obtain another element of $V_{r,\widetilde{M}}$.) In order to check (3.107), we note that any $\zeta \in V_{r,\widetilde{M}}$ must contain at least K_r disjoint subsentences from $\mathscr{A}_r$, and each $z \in \mathscr{A}_r \subset \widetilde{E}^{M_r}$ satisfies $|\kappa(z)| \ge L$. Hence there are at most

$$\binom{\widetilde{M}(m_{Q_r} + \varepsilon) - K_r(L-1)}{K_r} \le 2^{4\varepsilon\widetilde{M}m_{Q_r}} \le 2^{4\varepsilon m^* \widetilde{M}} \tag{3.108}$$

choices for the positions in the letter sequence $\kappa(\eta)$ where the concatenations of the disjoint subsentences from $\mathscr{A}_r$ can begin, and there are at most

$$\binom{\widetilde{M} - K_r(M_r - 1)}{K_r} \leq 2^{3\varepsilon \widetilde{M}} \tag{3.109}$$

choices for the positions in the word sequence ζ where the subsentences from $\mathscr{A}_r$ can begin. By construction (recall the last line of (3.68)), each $z \in \mathscr{A}_r$ can be "re-cut" in not more than $\exp[(L/m_{Q_r})(H_{\tau|K}(Q_r) + \varepsilon)]$ many ways. Combining these observations with the fact that

$$\Big(\exp\big[(L/m_{Q_r})(H_{\tau|K}(Q_r) + \varepsilon)\big]\Big)^{K_r} \leq \exp\Big[\frac{\widetilde{M}}{M_r} M_r (H_{\tau|K}(Q_r) + \varepsilon)\Big], \tag{3.110}$$

we get (3.107) with $\delta_1 := \varepsilon + 3\varepsilon \log 2 + 4\varepsilon m^* \log 2$.

We see from (3.102–3.103) and the definitions of $\widetilde{\mathscr{U}}_r$ and $V_{r,\widetilde{M}}$ that, for any $\varepsilon > 0$

$$\bigcup_{\widetilde{M} \in \mathbb{N}} \Big\{P \in \widetilde{\mathscr{U}}_r \colon\ P(V_{r,\widetilde{M}}) > 1 - \varepsilon\Big\} = \widetilde{\mathscr{U}}_r. \tag{3.111}$$

Hence we can choose $\widetilde{M}$ so large that

$$W_Q\Big(\Big\{P \in \widetilde{\mathscr{U}}_r \colon\ P(V_{r,\widetilde{M}}) > 1 - \frac{\varepsilon}{3}\Big\}\Big) > W_Q(\widetilde{\mathscr{U}}_r)\left(1 - \frac{\varepsilon}{2}\right), \quad r = 1, \dots, R. \tag{3.112}$$

For $M' > \widetilde{M}$ and $r = 1, \dots, R$, put

$$W_{r,M'} := \big\{\zeta \in \widetilde{E}^{M'} \colon\ \mathrm{freq}_{V_{r,\widetilde{M}}}(\zeta) > 1 - \varepsilon/2\big\}. \tag{3.113}$$

Note that for $r \neq r'$ (because $V_{r,\widetilde{M}} \cap V_{r',\widetilde{M}} = \emptyset$) there cannot be much overlap between $\zeta \in W_{r,M'}$ and $\eta \in W_{r',M'}$:

$$\max\{k \colon\ k\text{-suffix of } \zeta = k\text{-prefix of } \eta\} \leq \varepsilon M' \tag{3.114}$$

(here, the k-prefix of $\eta \in \widetilde{E}^n$, $k < n$, consists of the first k words, the k-suffix of the last k words). To see this, note that any subsequence of length k of ζ must contain at least $(k - \varepsilon M'/2)_+$ positions where a sentence from $V_{r,\widetilde{M}}$ starts, and any subsequence of length k of η must contain at least $(k - \varepsilon M'/2)_+$ positions where a sentence from $V_{r',\widetilde{M}}$ starts, so any k appearing in (3.114) must satisfy $2(k - \varepsilon M'/2)_+ \leq k$, which enforces $k \leq \varepsilon M'$. Now, (3.114) implies that we may choose M' so large that for $r = 1, \dots, R$,

$$\text{each } \zeta \in W_{r,M'} \text{ contains at least } (1 - \varepsilon)\frac{M'}{\widetilde{M}} \text{ disjoint subsentences from } V_{r,\widetilde{M}}. \tag{3.115}$$

For $P \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ with $P(V_{r,\widetilde{M}}) > 1 - \varepsilon/3$ we have

$$\lim_{M' \to \infty} P(W_{r,M'}) = 1, \tag{3.116}$$

and hence

$$\bigcup_{M' > \widetilde{M}} \Big\{P \in \widetilde{\mathscr{U}}_r \colon\ P(W_{r,M'}) > 1 - \varepsilon\Big\} \supset \Big\{P \in \widetilde{\mathscr{U}}_r \colon\ P(V_{r,\widetilde{M}}) > 1 - \varepsilon/3\Big\}, \tag{3.117}$$

and so we can choose M' so large that

$$W_Q\Big(\big\{P \in \widetilde{\mathcal{U}}_r\colon\ P(W_{r,M'}) > 1-\varepsilon\big\}\Big) > W_Q(\widetilde{\mathcal{U}}_r)(1-\varepsilon), \quad r = 1,\dots,R. \tag{3.118}$$

Now define

$$\mathcal{O}(Q) := \Big\{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon\ Q'(W_{r,M'}) > W_Q(\widetilde{\mathcal{U}}_r)(1-2\varepsilon),\ r = 1,\dots,R\Big\}. \tag{3.119}$$

Note that $\mathcal{O}(Q)$ is open in the weak topology on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, since it is defined in terms of requirements on certain finite marginals of Q', and that for $r = 1,\dots,R$,

$$Q(W_{r,M'}) = \int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})} Q'(W_{r,M'})\, W_Q(dQ') \geq \int_{\widetilde{\mathcal{U}}_r} Q'(W_{r,M'})\, W_Q(dQ') \geq \big(1-\varepsilon\big)^2 W_Q(\widetilde{\mathcal{U}}_r) \tag{3.120}$$

by (3.118), so that in fact $Q \in \mathcal{O}(Q)$.

3.5.5 Estimating the large deviation probability: good loops and filling loops

Consider a choice of "cut-points" $j_1 < \cdots < j_N$ as appearing in the sum in (3.23). Note that, by the definition of $\mathcal{O}(Q)$ (recall (3.16–3.17)),

$$R^N_{j_1,\dots,j_N}(X) \in \mathcal{O}(Q) \tag{3.121}$$

enforces

$$\big|\big\{1 \leq i \leq N-M'\colon\ (X|_{(j_{i-1},j_i]},\dots,X|_{(j_{i+M'-1},j_{i+M'}]}) \in W_{r,M'}\big\}\big| \geq N W_Q(\widetilde{\mathcal{U}}_r)(1-3\varepsilon), \quad r = 1,\dots,R, \tag{3.122}$$

when N is large enough. This fact, together with (3.114), enables us to pick at least

$$J := \sum_{r=1}^{R} \lceil (1-4\varepsilon)N/M' \rceil W_Q(\widetilde{\mathcal{U}}_r) \tag{3.123}$$

subsentences $\zeta_1,\dots,\zeta_J$ occurring as disjoint subsentences in this order on ξ_N such that

$$\big|\big\{1 \leq j \leq J\colon\ \zeta_j \in W_{r,M'}\big\}\big| > (1-4\varepsilon) W_Q(\widetilde{\mathcal{U}}_r)\,\frac{N}{M'}, \quad r = 1,\dots,R, \tag{3.124}$$

where we note that $J \geq (1-8\varepsilon)(N/M')$ by (3.97). Indeed, we can for example construct these ζ_j's iteratively in a "greedy" way, parsing through ξ_N from left to right and always picking the next possible subsentence from one of the R types whose count does not yet exceed $(1-4\varepsilon)W_Q(\widetilde{\mathcal{U}}_r)\,(N/M')$, as follows. Let $k_{s,r}$ be total number of subsentences of type r we have chosen after the s-th step ($k_{0,1} = \cdots = k_{0,R} = 0$). If in the s-th step we have picked $\zeta_s = (\xi_N^{(p)},\dots,\xi_N^{(p+M'-1)})$ at position p, then let

$$p' := \min\big\{i \geq p + M'\colon\ \text{at position } i \text{ in } \xi_N \text{ starts a sentence from } W_{u,M'} \text{ for some } u \in U_s\big\}, \tag{3.125}$$

where $U_s := \{r\colon k_{r,s} < (1-4\varepsilon)W_Q(\widetilde{\mathcal{U}}_r)\,(N/M')\}$, pick the next subsentence ζ_{s+1} starting at position p' (say, of type u) and increase the corresponding $k_{s+1,u}$. Repeat this until $k_{s,r} \geq (1-4\varepsilon)W_Q(\widetilde{\mathcal{U}}_r)\,(N/M')$ for $r = 1,\dots,R$.

In order to verify that this algorithm does not get stuck, let $\mathrm{rem}(s,r)$ be the "remaining" number of positions (to the right of the position where the word was picked in the s-th step) where a subsentence from $W_{r,M'}$ begins on ξ_N. By (3.122), we have

$$\mathrm{rem}(0,r) \geq NW_Q(\widetilde{\mathcal{U}}_r)(1-3\varepsilon). \tag{3.126}$$

If in the s-th step a subsentence of type r is picked, then we have $\mathrm{rem}(s+1,r) \geq \mathrm{rem}(s,r) - M'$, and for $r' \neq r$ we have $\mathrm{rem}(s+1,r') \geq \mathrm{rem}(s,r') - \varepsilon M'$ by (3.114). Thus,

$$\begin{aligned}\mathrm{rem}(s,r) &\geq \mathrm{rem}(0,r) - k_{s,r}M' - (s-k_{s,r})\varepsilon M'\\ &= \mathrm{rem}(0,r) - k_{s,r}(1-\varepsilon)M' - s\varepsilon M',\end{aligned} \tag{3.127}$$

which is > 0 as long as $k_{s,r} < (1-4\varepsilon)W_Q(\widetilde{\mathcal{U}}_r)\,(N/M')$ and $s < J$.

A. Combinatorial consequences. By (3.115) and (3.124), $R^N_{j_1,\dots,j_N}(X) \in \mathcal{O}(Q)$ implies that ξ_N contains at least

$$C := \sum_{r=1}^{R} \Big\lceil (1-4\varepsilon)W_Q(\widetilde{\mathcal{U}}_r)\frac{N}{M'}\Big\rceil \Big\lceil (1-\varepsilon)\frac{M'}{\widetilde{M}}\Big\rceil \quad \Big(\geq (1-10\varepsilon)\frac{N}{\widetilde{M}}\Big) \tag{3.128}$$

disjoint subsentences $\eta_1,\dots,\eta_C$ (appearing in this order in ξ_N) such that at least

$$\frac{N}{\widetilde{M}}(1-6\varepsilon)W_Q(\widetilde{\mathcal{U}}_r) \text{ of the } \eta_c\text{'s are from } V_{r,\widetilde{M}}, \quad r = 1,\dots,R. \tag{3.129}$$

Let $k_1,\dots,k_C$ ($k_{c+1} \geq k_c + \widetilde{M}$, $1 \leq c < C$) be the indices where the disjoint subsentences η_c start in ξ_N, i.e.,

$$\eta_c = \Big(\xi_N^{(j_{k_c})}, \xi_N^{(j_{k_c+1})}, \dots, \xi_N^{(j_{k_c+\widetilde{M}-1})}\Big) \in V_{r_c,\widetilde{M}}, \quad i = c,\dots,C, \tag{3.130}$$

and the r_c's must respect the frequencies dictated by the $W_Q(\widetilde{\mathcal{U}}_r)$'s as in (3.129). Thus, each choice $(j_1,\dots,j_N)$ yielding a non-zero summand in (3.23) leads to a triple

$$(\ell_1,\dots,\ell_C), (r_1,\dots,r_C), (\overline{\eta}_1,\dots,\overline{\eta}_C) \tag{3.131}$$

such that $\overline{\eta}_c \in \kappa(V_{r_c,\widetilde{M}})$, $\ell_{c+1} \geq \ell_c + |\overline{\eta}_c|$, the r_c's respect the frequencies as in (3.129), and

$$\text{the word } \overline{\eta}_c \text{ starts at position } \ell_c \text{ in } X \text{ for } c = 1,\dots,C. \tag{3.132}$$

As in Section 3.3, we call such triples *good*, the loops inside the subsentences η_i *good* loops, the others *filling* loops.

Fix a good triple for the moment. In order to count how many choices of $j_1 < \cdots < j_N$ can lead to this particular triple and to estimate their contribution, observe the following:

1. There are at most

$$\binom{N - C(\widetilde{M} - 1)}{C} \le \exp(\delta_1' N) \tag{3.133}$$

choices for the $k_1 < \cdots < k_C$, where δ_1' can be made arbitrarily small by choosing ε small and $\widetilde{M}$ large.

2. Once the k_c's are fixed, by (3.107) and (3.129) there are at most

$$\begin{aligned} &\prod_{r=1}^{R} \Big(\exp\big[\widetilde{M}(H_{\tau|K}(Q_r) + \delta_1)\big] \Big)^{\frac{N}{\widetilde{M}} W_Q(\widetilde{\mathcal{U}}_r)} \\ &\qquad = \exp\Big[N \sum_{r=1}^{R} W_Q(\widetilde{\mathcal{U}}_r)(H_{\tau|K}(Q_r) + \delta_1)\Big] \end{aligned} \tag{3.134}$$

choices for the good loops and, by (3.106), for each choice of the good loops the product of the $\rho(j_k - j_{k-1})$'s inside the good loops is at most

$$\begin{aligned} &\prod_{r=1}^{R} \Big(\exp\big[(1-4\varepsilon)\widetilde{M}\mathbb{E}_{Q_r}[\log \rho(\tau_1)]\big] \Big)^{\frac{N}{\widetilde{M}} W_Q(\widetilde{\mathcal{U}}_r)} \\ &\qquad = \exp\Big[N(1-4\varepsilon) \sum_{r=1}^{R} W_Q(\widetilde{\mathcal{U}}_r)\mathbb{E}_{Q_r}[\log \rho(\tau_1)]\Big]. \end{aligned} \tag{3.135}$$

3. For each choice of the k_c's, the contribution of the filling loops to the weight is

$$\begin{aligned} &\rho^{*(k_1-1)}(\ell_1 - 1) \prod_{c=1}^{C-1} \rho^{*(k_{c+1}-k_c-\widetilde{M})}(\ell_{c+1} - \ell_c - |\overline{\eta}_c|) \\ &\le\ (2^\alpha C_\rho \vee 1)^C k_1^{\alpha+1} \prod_{c=1}^{C-1} (k_{c+1} - k_c - \widetilde{M})^{\alpha+1} \prod_{c=1}^{C} \big((\ell_c - \ell_{c-1} - |\overline{\eta}_{c-1}|) \vee 1\big)^{-\alpha} \\ &\le\ (2^\alpha C_\rho \vee 1)^C \Big(\frac{N - C\widetilde{M}}{C}\Big)^{(\alpha+1)C} \prod_{c=1}^{C} \big((\ell_c - \ell_{c-1} - |\overline{\eta}_{c-1}|) \vee 1\big)^{-\alpha} \\ &\le\ e^{\delta_2' N} \prod_{c=1}^{C} \big((\ell_c - \ell_{c-1} - |\overline{\eta}_{c-1}|) \vee 1\big)^{-\alpha}, \end{aligned} \tag{3.136}$$

where δ_2' can be made arbitrarily small by choosing ε small and $\widetilde{M}$ large (and we interpret $\ell_0 = 0$, $|\overline{\eta}_0| = 0$). Here, we have used Lemma 2.3 in the first inequality, as well as the fact that the product $\prod_{c=1}^{C-1}(k_{c+1} - k_c - \widetilde{M})$ is maximal when all factors are equal in the second inequality.

Combining (3.133–3.136), we see that

$$
\begin{aligned}
&\mathbb{P}\big(R_N \in \mathcal{O}(Q)\big|X\big)\\
&\quad\le e^{(\delta_1'+\delta_2'+\delta_1)N}\exp\Big[N(1-4\varepsilon)\sum_{r=1}^{R} W_Q(\widetilde{\mathscr{U}}_r)\big(H_{\tau|K}(Q_r)+\mathbb{E}_{Q_r}[\log\rho(\tau_1)]\big)\Big]\\
&\qquad\times\sum_{\substack{(\ell_i),(r_i),(\overline{\eta}_i)\\ \text{good}}}\prod_{i=1}^{C}\big((\ell_i-\ell_{i-1}-|\overline{\eta}_{i-1}|)\vee 1\big)^{-\alpha}.
\end{aligned}
\tag{3.137}
$$

We claim that X-a.s.

$$
\begin{aligned}
&\limsup_{N\to\infty}\frac{1}{N}\log\sum_{\substack{(\ell_i),(r_i),(\overline{\eta}_i)\\ \text{good}}}\prod_{i=1}^{C}\big((\ell_i-\ell_{i-1}-|\overline{\eta}_{i-1}|)\vee 1\big)^{-\alpha}\\
&\qquad\le \delta_2-\alpha(1-4\varepsilon)\sum_{r=1}^{R}W_Q(\widetilde{\mathscr{U}}_r)m_{Q_r}H\big(\Psi_{Q_r}\mid\nu^{\otimes\mathbb{N}}\big),
\end{aligned}
\tag{3.138}
$$

where δ_2 can be made arbitrarily small by choosing ε small and L large. A proof of this is given below. Observe next that (3.137–3.138) (recall also (1.32)) yield that X-a.s. (with $\delta := \delta_1'+\delta_2'+\delta_1+\delta_2$)

$$
\begin{aligned}
&\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}\big(R_N\in\mathcal{O}(Q)\big|X\big)\\
&\le \delta-(1-4\varepsilon)\sum_{r=1}^{R}W_Q(\widetilde{\mathscr{U}}_r)\Big(H\big(Q_r\mid q\rho,\nu^{\otimes\mathbb{N}}\big)+\alpha m_{Q_r}H\big(\Psi_{Q_r}\mid\nu^{\otimes\mathbb{N}}\big)\Big)\\
&\le \delta+(1-4\varepsilon)\varepsilon(2+2\alpha)-(1-4\varepsilon)\int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})}H\big(Q'\mid q_{\rho,\nu}^{\otimes\mathbb{N}}\big)+(\alpha-1)m_{Q'}H\big(\Psi_{Q'}\mid\nu^{\otimes\mathbb{N}}\big)W_Q(dQ')\\
&= -(1-4\varepsilon)I^{\mathrm{fin}}(Q)+\delta+(1-4\varepsilon)\varepsilon(2+2\alpha)
\end{aligned}
\tag{3.139}
$$

(use (3.100) for the second inequality, and see (6.3) for the last equality), which completes the proof.

B. Coarse-graining X with R colours. It remains to verify (3.138), for which we employ a coarse-graining scheme similar to the one used in Section 3.4 (with block lengths $\lceil(1-\varepsilon_2)L\rceil$, etc.) To ease notation, we silently replace L by $(1-\varepsilon_2)L$ in the following. Split X into blocks of L consecutive letters, define a $\{0,1\}$-valued array $A_{i,r}$, $i\in\mathbb{N}$, $r\in\{1,\dots,R\}$ as in Section 3.4 inductively: For each r, put $A_{0,r}:=0$ and, given that $A_{0,r},A_{1,r},\dots,A_{l-1,r}$ have been assigned values, define A_l as follows:

(1) If $A_{l-1,r}=0$, then

$$
A_{l,r}:=\begin{cases}1, & \text{if in } X \text{ there is a word from } \kappa(\mathscr{A}_r) \text{ starting in } ((l-1)L,lL],\\ 0, & \text{otherwise.}\end{cases}
\tag{3.140}
$$

(2) If $A_{l-1,r} = 1$, then

$$A_l := \begin{cases} 1, & \text{if in } X \text{ there are two words from } \kappa(\mathscr{A}_r) \text{ starting in } ((l-2)L,(l-1)L], \\ & \text{respectively, } ((l-1)L, lL] \text{ and occurring disjointly,} \\ 0, & \text{otherwise.} \end{cases} \tag{3.141}$$

Put

$$p_r := L \exp\Big(-(1-\varepsilon) L H(\Psi_{Q_r} \mid \nu^{\otimes \mathbb{N}})\Big). \tag{3.142}$$

Arguing as in Section 3.4, we can couple the $(A_{i,r})_{i\in\mathbb{N},1\le r\le R}$ with an array $\overline{\omega} = (\omega_{i,r})_{i\in\mathbb{N},1\le r\le R}$ such that $A_{i,r} \le \omega_{i,r}$ and the sequence $\big((\omega_{i,1},\dots,\omega_{i,R})\big)_{i\in\mathbb{N}}$ is i.i.d. with $\mathbb{P}(\omega_{i,r}=1) = p_r$. In particular, for each r, $(\omega_{i,r})_{i\in\mathbb{N}}$ is a Bernoulli(p_r)-sequence. There may (and certainly will be if Ψ_{Q_r} and $\Psi_{Q_{r'}}$ are similar) an arbitrary dependence between the $\omega_{i,1},\dots,\omega_{i,R}$ for fixed i, but this will be harmless in the low-density limit we are interested in.

For $r \in \{1,\dots,R\}$, put $d_r := W_Q(\widetilde{\mathscr{U}}_r)(1-6\varepsilon)$, $D_r := \lceil (1-\varepsilon)\widetilde{M} m_{Q_r}/L\rceil$. If $\eta_c \in V_{r_c,\widetilde{M}}$, then

$$|\kappa(\eta_c)| \in \widetilde{M} m_{Q_{r_c}}(1-\varepsilon, 1+\varepsilon), \tag{3.143}$$

so $\kappa(\eta_c)$ covers at least D_{r_c} consecutive L-blocks of the coarse-graining. Furthermore, as η_c in turn contains at least $D_{r_c}(1-\varepsilon)$ disjoint subsentences from $\mathscr{A}_{r_c}$, we see that at least $D_{r_c}(1-\varepsilon)$ of these blocks must have $A_{k,r_c} = 1$. Thus, for fixed X, we read off from each good triple $(\ell_c), (r_c), (\overline{\eta}_c)$ numbers $m_1 < \cdots < m_C$ such that

$$\begin{aligned} & m_{c+1} \ge m_c + D_{r_c},\ c = 1,\dots,C-1, \\ & \big|\{m_c \le k < m_c + D_{r_c}\colon\ A_{k,r_c} = 1\}\big| \ge D_{r_c}(1-\varepsilon),\ c = 1,\dots,C, \\ & \big|\{1 \le c \le C\colon\ r_c = r\}\big| \ge d_r C,\ r = 1,\dots,R. \end{aligned} \tag{3.144}$$

where m_c is the number of the L-block that contains ℓ_c. Furthermore, note that for a given "coarse-graining" (m_c) and (r_c) satisfying (3.144), there are at most

$$L^C \Big(2\varepsilon \widetilde{M} \max_{r=1,\dots,R} m_{Q_r}\Big)^C \le \exp(\delta_3 N) \tag{3.145}$$

choices for ℓ_c and $\overline{\eta}_c$ that lead to a good triple $(\ell_c), (r_c), (\overline{\eta}_c)$ with this particular coarse-graining. Indeed, for each $c = 1,\dots,C$ there are at most L choices for ℓ_c and, since each $\eta \in V_{r_c,\widetilde{M}}$ satisfies

$$|\kappa(\eta)| \in \widetilde{M} m_{Q_{r_c}}(1-\varepsilon, 1+\varepsilon), \tag{3.146}$$

there are at most $2\varepsilon\widetilde{M} m_{Q_{r_c}}$ choices for $\overline{\eta}_c$ (note that once ℓ_c is fixed as a "starting point" for a word on X, choosing $\overline{\eta}_c$ in fact amounts to choosing an "endpoint"). Note that δ_3 can be made arbitrarily small by choosing ε small and $\widetilde{M}$ large. Finally, (3.145) and Lemma 3.3 yield (3.138). Indeed, since

$$\begin{aligned} \limsup_{N\to\infty} \frac{C}{N} &\le \frac{1}{\widetilde{M}}, \\ \sum_{r=1}^R d_r D_r \log p_r &\le -(1-8\varepsilon) \sum_{r=1}^R W_Q(\widetilde{U}_r) \frac{\widetilde{M} m_{Q_r}}{L} \Big(L H(\Psi_{Q_r} \mid \nu^{\otimes\mathbb{N}}) - \log L\Big), \end{aligned}$$

by choosing ε small, L and $\widetilde{M}$ large, and γ sufficiently close to $1/\alpha$, the right-hand side of (3.150) is smaller than the right-hand side of (3.138).

3.5.6 A multicolour version of the core lemma

The following is an extension of Lemma 2.1. Let $R \in \mathbb{N}$, $\overline{\omega}_i = (\omega_{i,1}, \dots, \omega_{i,R}) \in \{0,1\}^R$, and assume that $(\overline{\omega}_i)_{i\in\mathbb{N}}$ is i.i.d. with

$$\mathbb{P}(\omega_{i,r} = 1) = p_r, \quad i \in \mathbb{N},\ r = 1, \dots, R. \tag{3.147}$$

Note that there may be an arbitrary dependence between the $\omega_{i,r}$'s for fixed i. This will be harmless in the limit we are interested in below.

Lemma 3.3. *Let $\alpha \in (1,\infty)$, $\varepsilon > 0$, $(d_1, \dots, d_R) \in [0,1]^R$ with $\sum_{r=1}^R d_r \le 1$, $D_1, \dots, D_R \in \mathbb{N}$, $C \in \mathbb{N}$, put*

$$S_C(\overline{\omega}) := \sum\nolimits^*_{\substack{m_1,\dots,m_C\\ r_1,\dots,r_C}} \prod_{i=1}^{C} \big(m_i - m_{i-1} - D_{r_{i-1}}\big)^{-\alpha}, \tag{3.148}$$

where the sum $\sum^*$ extends over all pairs of C-tuples $m_0 := 0 < m_1 < \cdots < m_C$ from $\mathbb{N}^C$ and $(r_1, \dots, r_C) \in \{1, \dots, R\}^C$ satisfying the constraints

$$\begin{aligned} &m_{i+1} \ge m_i + D_{r_i},\\ &|\{1 \le i \le C\colon\ r_i = r\}| \ge d_r C,\ r = 1, \dots, R,\\ &|\{m_i \le k < m_i + D_{r_i}\colon\ \omega_{k,r_i} = 1\}| \ge D_{r_i}(1-\varepsilon),\ i = 1, \dots, C. \end{aligned} \tag{3.149}$$

Then $\overline{\omega}$-a.s.

$$\begin{aligned} &\limsup_{C\to\infty} \frac{1}{C} \log S_C(\overline{\omega})\\ &\le \inf_{\gamma\in(1/a,1)} \Big\{\frac{1}{\gamma}\Big(\log\zeta(a\gamma) + h(\underline{d}) + d_0 \log R + \big(\log 2\big)\sum_{r=1}^{R} d_r D_r + (1-\varepsilon)\sum_{r=1}^{R} d_r D_r \log p_r\Big)\Big\}, \end{aligned} \tag{3.150}$$

where $h(\underline{d}) := -\sum_{r=0}^R d_r \log d_r$ (with $d_0 := 1 - d_1 - \cdots - d_R$) is the entropy of $\underline{d}$, and $\phi(\varepsilon)$ is a function such that $\lim_{\varepsilon\downarrow 0} \phi(\varepsilon) = 0$.

Proof. The proof is a variation on the proof of Lemma 2.1. We again estimate fractional moments. For $\gamma \in (1/\alpha, 1)$, we have

$$\begin{aligned} &\mathbb{E}\big[(S_C)^\gamma\big]\\ &\le \sum\nolimits'_{r_1,\dots,r_C} \sum_{\substack{m_1,\dots,m_C\\ m_{i+1}\ge m_i+D_{r_i}\ \forall i}} \mathbb{P}\Big(\cap_{i=1}^C \big\{|\{k \in [m_i, m_i + D_{r_i} - 1]\colon\ \omega_{k,r_i} = 1\}| \ge (1-\varepsilon)D_{r_i}\big\}\Big)\\ &\qquad\qquad \times \prod_{i=1}^{C} (m_i - m_{i-1} - D_{r_{i-1}})^{-\alpha\gamma}, \end{aligned} \tag{3.151}$$

where the sum $\sum'$ extends over all $(r_1, \dots, r_C)$ satisfying the constraint in the second line of (3.149). Noting that

$$\begin{aligned} \mathbb{P}\Big(|\{k \in [m_i, m_i + D_{r_i} - 1]\colon\ \omega_{k,r_i} = 1\}| \ge (1-\varepsilon)D_{r_i}\Big) &= \sum_{m=(1-\varepsilon)D_{r_i}}^{D_{r_i}} \binom{D_{r_i}}{k} p_r^m (1-p_r)^{D_{r_i}-m}\\ &\le p_r^{(1-\varepsilon)D_{r_i}} 2^{D_{r_i}} \end{aligned}$$

and

$$\begin{aligned}&\left|\{(r_1,\dots,r_C)\in\{1,\dots,R\}^C\colon \text{ at least } d_rC \text{ of the } r_i=r,\ r=1,\dots,R\}\right|\\ &\quad\le\ R^{d_0C}\binom{C}{d_0C\ d_1C\ \dots\ d_RC}=\exp\Big[C\big(d_0\log R+h(\underline{d})+o(1)\big)\Big],\end{aligned}$$

we see from (3.151) that

$$\begin{aligned}\mathbb{E}\big[(S_C)^\gamma\big] &\le \exp\Big[C\big(d_0\log R+h(\underline{d})+o(1)\big)\Big]\times\prod_{r=1}^R\big(2p_r^{(1-\varepsilon)}\big)^{d_rCD_r}\\ &\qquad\times\sum_{\substack{m_1,\dots,m_C\\ m_{i+1}\ge m_i+D_{r_i}\ \forall i}}\prod_{i=1}^C(m_i-m_{i-1}-D_{r_{i-1}})^{-\alpha\gamma}\\ &= \exp C\Big[d_0\log R+h(\underline{d})+\log\zeta(a\gamma)+\textstyle\sum_{r=1}^R d_rD_r\log 2+(1-\varepsilon)\sum_{r=1}^R d_rD_rp_r\Big],\end{aligned}\tag{3.152}$$

which yields (3.150) as in the proof of Lemma 2.1. □

3.6 Step 6: Weakening the tail assumption

We finally show how to go from (3.3) to (1.1). Suppose that ρ satisfies (1.1) with a certain $\alpha\in(1,\infty)$. Then, for any $\alpha'\in(1,\alpha)$, there is a $C_\rho(\alpha')$ such that (3.3) holds for this α'. Hence, as shown in Sections 3.1–3.4, for any $\varepsilon>0$ we can find a neighbourhood $\mathcal{O}(Q)\subset\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ of Q such that

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}\big(R_N\in\mathcal{O}(Q)\mid X\big)\le -H(Q\mid q_{\rho,\nu}^{\otimes\mathbb{N}})-(\alpha'-1)\,m_Q\,H(\Psi_Q\mid\nu^{\otimes\mathbb{N}})+\frac{\varepsilon}{2}\qquad X-a.s.\tag{3.153}$$

The right-hand side is $\le -I^{\mathrm{fin}}(Q)+\varepsilon$ for α' sufficiently close to α, so that we again get (3.1). □

4 Lower bound

The following lower bound will be used in Section 5 to derive the lower bound in the definition of the LDP.

Proposition 4.1. *For any $Q\in\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ and any open neighbourhood $\mathcal{U}(Q)\subset\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ of Q,*

$$\liminf_{N\to\infty}\frac{1}{N}\log\mathbb{P}\big(R_N\in\mathcal{U}(Q)\mid X\big)\ge -I^{\mathrm{fin}}(Q)\quad X-a.s.\tag{4.1}$$

Proof. Suppose first that $Q\in\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$. Then, informally, our strategy runs as follows. In X, look for the first string of length $\approx Nm_Q$ that looks typical for Ψ_Q. Make the first jump long enough so as to land at the start of this string. Make the remaining $N-1$ jumps typical for Q. The probability of this strategy on the exponential scale is the conditional specific relative entropy of word lengths under Q w.r.t. $\rho^{\otimes\mathbb{N}}$ given the concatenation, i.e., $\approx\exp[-N(H_{\tau|K}(Q)+\mathbb{E}_Q[\log\rho(\tau_1)])]$, times the probability of the first long jump. In order to find a suitable string, we have to skip ahead in X a distance $\approx \exp[Nm_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}})]$. By (1.1), the probability of the first jump is therefore $\approx \exp[-N\alpha\, m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}})]$. In view of (1.16) and (1.32), this yields the claim. In the actual proof, it turns out to be technically simpler to employ a slightly different strategy, which has the same asymptotic cost, where we look not only for one contiguous piece of "Ψ_Q-typical" letters but for a sequence of $\lceil N/M \rceil$ pieces, each of length $\approx Mm_Q$. Then we let $N \to \infty$, followed by $M \to \infty$.

More formally, we choose for $\mathcal{O}(Q)$ an open neighborhood $\mathcal{O}' \subset \mathcal{O}$ of the type introduced in Section 3.2, and we estimate $\mathbb{P}(R_N \in \mathcal{O}' \mid X)$ from below by using (3.17–3.20).

Assume first that Q is ergodic. We can then assume that the neighbourhood $\mathcal{U}$ is given by

$$\mathcal{U} = \big\{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon\ (\pi_{L_u} Q')(\zeta_u) \in (a_u, b_u),\ u = 1, \dots, U\big\} \tag{4.2}$$

for some $U \in \mathbb{N}$, $L_1, \dots, L_U \in \mathbb{N}$, $0 \le a_u < b_u \le 1$ and $\zeta_u \in \widetilde{E}^{L_u}$, $u = 1, \dots, U$. As in Section 3.1, by ergodicity of Q we can find for each $\varepsilon > 0$ a sufficiently large $M \in \mathbb{N}$ and a set $\mathscr{A} = \{z_1, \dots, z_A\} \subset \widetilde{E}^M$ of "Q-typical sentences" satisfying (3.6–3.7) (with $\varepsilon_1 = \delta_1 = \varepsilon$, say), and additionally

$$\frac{1}{M}\big|\{0 \le j \le M - L_i\colon\ \pi_{L_u}(\widetilde{\theta}^j z_a) = \zeta_u\}\big| \in (a_i, b_i), \quad a = 1, \dots, A,\ u = 1, \dots, U. \tag{4.3}$$

Let $\mathscr{B} := \kappa(\mathscr{A})$. Then from (3.6–3.7) we have that, for each $b \in \mathscr{B}$,

$$|I_b| = |\{z \in \mathscr{A}\colon\ \kappa(z) = b\}| \ge \exp\big[M(H_{\tau|K}(Q) - 2\varepsilon)\big], \tag{4.4}$$

and

$$\mathbb{P}(X \text{ begins with some element of } \mathscr{B}) \ge \exp\big[-Mm_Q(H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) + 2\varepsilon)\big]. \tag{4.5}$$

Let

$$\begin{aligned} \sigma_1^{(M)} &:= \min\{i\colon\ \theta^i X \text{ begins with some element of } \mathscr{B}\}, \\ \sigma_l^{(M)} &:= \min\{i > \tau_{l-1} + M(m_Q + \varepsilon)\colon\ \sigma^i X \text{ begins with some element of } \mathscr{B}\}, \quad l = 2, 3, \dots. \end{aligned} \tag{4.6}$$

Restricting the sum in (3.23) over $0 < j_1 < \cdots < j_N < \infty$ such that $j_1 = \sigma_1^{(M)}$, $j_2 - j_1, \dots, j_M - j_{M-1}$ are the word lengths corresponding to the z_a's compatible with $\pi_{Mm_Q}(\theta^{\tau_1} X)$, $j_{M+1} = \sigma_2^{(M)}$, etc., we see that

$$\frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{U} \mid X) \ge H_{\tau|K}(Q) + \mathbb{E}_Q[\log \rho(\tau_1)] - 3\varepsilon - \alpha \frac{1}{N} \sum_{l=1}^{\lfloor N/M \rfloor} \log\big(\sigma_l^{(M)} - \sigma_{l-1}^{(M)}\big) \tag{4.7}$$

for N sufficiently large. Hence X-a.s.

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{U} \mid X) &\ge H_{\tau|K}(Q) + \mathbb{E}_Q[\log \rho(\tau_1)] - 3\varepsilon - \alpha \frac{1}{M} \mathbb{E}[\log \sigma_1^{(M)}] \\ &\ge H_{\tau|K}(Q) + \mathbb{E}_Q[\log \rho(\tau_1)] - \alpha m_Q (H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - 6\varepsilon \\ &= -I^{\mathrm{fin}}(Q) - 6\varepsilon, \end{aligned} \tag{4.8}$$

where we have used (4.5) in the second inequality. Now let $\varepsilon \downarrow 0$.

It remains to remove the restriction of ergodicity of Q, analogously to the proof of Birkner [3], Proposition 2. To that end, assume that $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ admits a non-trivial ergodic decomposition. Then, for each $\varepsilon > 0$, we can find $Q_1, \ldots, Q_R \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$, $\lambda_1, \ldots, \lambda_R \in (0,1)$, $\sum_{r=1}^{R} \lambda_r = 1$ such that $\lambda_1 Q_1 + \cdots + \lambda_R Q_R \in \mathcal{U}$ and

$$\sum_{i=1}^{R} \lambda_r I^{\mathrm{fin}}(Q_r) \leq I^{\mathrm{fin}}(Q) + \varepsilon \tag{4.9}$$

(for details see Birkner [3], p. 723; employ the fact that both terms in I^{fin} are affine). For each $r = 1, \ldots, R$, pick a small neighbourhood $\mathcal{U}_r$ of Q_r such that

$$Q_r' \in \mathcal{U}_r,\ r = 1, \ldots, R \quad \Longrightarrow \quad \sum_{i=1}^{R} \lambda_r Q_r' \in \mathcal{U}. \tag{4.10}$$

Using the above strategy for Q_1 for $\lambda_1 N$ loops, then the strategy for Q_2 for $\lambda_2 N$ loops, etc., we see that

$$\liminf_{N\to\infty} \frac{1}{N} \mathbb{P}(R_N \in \mathcal{U} \mid X) \geq -\sum_{i=1}^{R} \lambda_r I^{\mathrm{fin}}(Q_r) - 6\varepsilon \geq -I^{\mathrm{fin}}(Q) - 7\varepsilon. \tag{4.11}$$

□

5 Proof of Theorem 1.2

Proof. The proof comes in 3 steps. We first prove that, for each word length truncation level $\mathrm{tr} \in \mathbb{N}$, the family $\mathbb{P}([R_N]_{\mathrm{tr}} \in \cdot \mid X)$, $N \in \mathbb{N}$, X-a.s. satisfies an LDP on

$$\mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}}) = \left\{ Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \colon\ Q(|Y^{(1)}| \leq \mathrm{tr}) = 1 \right\} \tag{5.1}$$

(recall (1.11–1.13)) with a deterministic rate function $I^{\mathrm{fin}}([Q]_{\mathrm{tr}})$ (this is essentially the content of Propositions 4.1 and 3.1). Note that $[Q]_{\mathrm{tr}} = Q$ for $Q \in \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$, and that $\mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ is a closed subset of $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, in particular, a Polish space under the relative topology (which is again the weak topology). After we have given the proof for fixed tr, we let $\mathrm{tr} \to \infty$ and use a projective limit argument to complete the proof of Theorem 1.2.

1. Fix a truncation level $\mathrm{tr} \in \mathbb{N}$. Propositions 4.1 and 3.1 combine to yield the LDP on $\mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ in the following standard manner. Note that any $Q \in \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ satisfies $m_Q < \infty$.

1a. Let $\mathcal{O} \subset \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ be *open*. Then, for any $Q \in \mathcal{O}$, there is an open neighbourhood $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ of Q such that $\mathcal{O}(Q) \subset \mathcal{O}$. The latter inclusion, together with Proposition 4.1, yields

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big([R_N]_{\mathrm{tr}} \in \mathcal{O} \mid X\big) \geq -I^{\mathrm{fin}}(Q) \qquad X - a.s. \tag{5.2}$$

Optimising over $Q \in \mathcal{O}$, we get

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big([R_N]_{\mathrm{tr}} \in \mathcal{O} \mid X\big) \geq -\inf_{Q\in\mathcal{O}} I^{\mathrm{fin}}(Q) \qquad X - a.s. \tag{5.3}$$

Here, note that, since $\mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ is Polish, it suffices to optimise over a countable set generating the weak topology, allowing us to transfer the X-a.s. limit from points to sets (see, e.g., Comets [7], Section III).

1b. Let $\mathcal{K} \subset \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ be *compact*. Then there exist $M \in \mathbb{N}$, $Q_1, \dots, Q_M \in \mathcal{K}$ and open neighbourhoods $\mathcal{O}(Q_1), \dots, \mathcal{O}(Q_M) \subset \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ such that $\mathcal{K} \subset \cup_{m=1}^{M} \mathcal{O}(Q_m)$. The latter inclusion, together with Proposition 3.1, yields

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big([R_N]_{\mathrm{tr}} \in \mathcal{K} \mid X\big) \leq - \inf_{1\leq m\leq M} I^{\mathrm{fin}}(Q_m) + \varepsilon \qquad X-a.s. \qquad \forall\, \varepsilon > 0. \tag{5.4}$$

Extending the infimum to $Q \in \mathcal{K}$ and letting $\varepsilon \downarrow 0$ afterwards, we obtain

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big([R_N]_{\mathrm{tr}} \in \mathcal{K} \mid X\big) \leq - \inf_{Q\in\mathcal{K}} I^{\mathrm{fin}}(Q) \qquad X-a.s. \tag{5.5}$$

1c. Let $\mathcal{C} \subset \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ be *closed*. Because $Q \mapsto H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$ has compact level sets, for any $M < \infty$ the set $\mathcal{K}_M = \mathcal{C} \cap \{Q \in \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})\colon\ H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) \leq M\}$ is compact. Hence, doing *annealing* on X and using (5.5), we get

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big([R_N]_{\mathrm{tr}} \in \mathcal{C} \mid X\big) \leq \max\left\{-M, - \inf_{Q\in\mathcal{K}_M} I^{\mathrm{fin}}(Q)\right\} \qquad X-a.s. \tag{5.6}$$

Extending the infimum to $Q \in \mathcal{C}$ and letting $M \to \infty$ afterwards, we arrive at

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big([R_N]_{\mathrm{tr}} \in \mathcal{C} \mid X\big) \leq - \inf_{Q\in\mathcal{C}} I^{\mathrm{fin}}(Q) \qquad X-a.s. \tag{5.7}$$

Equations (5.3) and (5.7) complete the proof of the conditional LDP for $[R_N]_{\mathrm{tr}}$.

2. It remains to remove the truncation of word lengths. We know from Step 1 that, for every $\mathrm{tr} \in \mathbb{N}$, the family $\mathbb{P}([R_N]_{\mathrm{tr}} \in \cdot \mid X)$, $N \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}([\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}})$ with rate function I^{fin}. Consequently, by the Dawson-Gärtner projective limit theorem (see Dembo and Zeitouni [10], Theorem 4.6.1), the family $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with rate function

$$I^{\mathrm{que}}(Q) = \sup_{\mathrm{tr}\in\mathbb{N}} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}), \qquad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{5.8}$$

The sup may be replaced by a lim sup because the truncation may start at any level. For $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, we have $\lim_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) = I^{\mathrm{fin}}(Q)$ by Lemma A.1, and so we get the claim if we can show that lim sup can be replaced by a limit, which is done in Step 3. Note that I^{que} *inherits* from I^{fin} the properties qualifying it to be a rate function: this is part of the projective limit theorem. For I^{fin} these properties are proved in Section 6.

3. Since I^{que} is lower semi-continuous, it is equal to its lower semi-continuous regularisation

$$\widetilde{I}^{\mathrm{que}}(Q) := \sup_{\mathcal{O}(Q)} \inf_{Q'\in\mathcal{O}(Q)} I^{\mathrm{que}}(Q'), \tag{5.9}$$

where the supremum runs over the open neighborhoods of Q. For each $\mathrm{tr} \in \mathbb{N}$, $[Q]_{\mathrm{tr}} \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, while $\mathrm{w}-\lim_{\mathrm{tr}\to\infty}[Q]_{\mathrm{tr}} = Q$. So, in particular,

$$I^{\mathrm{que}}(Q) = \widetilde{I}^{\mathrm{que}}(Q) \leq \sup_{n} \inf_{\mathrm{tr}\geq n} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) = \liminf_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}), \tag{5.10}$$

implying that in fact

$$I^{\mathrm{que}}(Q) = \lim_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}), \quad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{5.11}$$

□

Lemma A.1 in Appendix A, together with (5.11), shows that $I^{\mathrm{que}}(Q) = I^{\mathrm{fin}}(Q)$ for $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, as claimed in the first line of (1.15).

6 Proof of Theorem 1.3

Proof. The proof comes in 5 steps.

1. Every $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ can be decomposed as

$$Q = \int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})} Q'\, W_Q(dQ') \tag{6.1}$$

for some unique probability measure W_Q on $\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ (Georgii [13], Proposition 7.22). If $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, then W_Q is concentrated on $\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$ and so, by (1.9–1.10),

$$m_Q = \int_{\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})} m_{Q'}\, W_Q(dQ'), \quad \Psi_Q = \int_{\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})} \frac{m_{Q'}}{m_Q}\, \Psi_{Q'}\, W_Q(dQ'). \tag{6.2}$$

Since $Q \mapsto H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$ and $\Psi \mapsto H(\Psi \mid \nu^{\otimes\mathbb{N}})$ are affine (see e.g. Deuschel and Stroock [12], Example 4.4.41), it follows from (1.16) and (6.1–6.2) that

$$I^{\mathrm{fin}}(Q) = \int_{\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})} I^{\mathrm{fin}}(Q')\, W_Q(dQ'). \tag{6.3}$$

Since $Q \mapsto W_Q$ is affine, (6.3) shows that I^{fin} is affine on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$.

2. Let $(Q_n)_{n\in\mathbb{N}} \subset \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ be such that $\mathrm{w}-\lim_{n\to\infty} Q_n = Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$. By Proposition 3.1, for any $\varepsilon > 0$ we can find an open neighbourhood $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ of Q such that

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in \mathcal{O}(Q) \mid X\big) \le -I^{\mathrm{fin}}(Q) + \varepsilon \quad X - a.s. \tag{6.4}$$

On the other hand, for n large enough so that $Q_n \in \mathcal{O}(Q)$, we have from Proposition 4.1 that

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in \mathcal{O}(Q) \mid X\big) \ge -I^{\mathrm{fin}}(Q_n) \quad X - a.s. \tag{6.5}$$

Combining (6.4–6.5), we get that, for any $\varepsilon > 0$,

$$\liminf_{n\to\infty} I^{\mathrm{fin}}(Q_n) \ge I^{\mathrm{fin}}(Q) - \varepsilon. \tag{6.6}$$

Now let $\varepsilon \downarrow 0$, to conclude that I^{fin} is lower semicontinuous on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ (recall also (5.11)).

3. From (1.16) we have

$$I^{\mathrm{fin}}(Q) \geq H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) \qquad \forall\, Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}) \tag{6.7}$$

Since $\{Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) \leq C\}$ is compact for all $C < \infty$ (see, e.g., Dembo and Zeitouni [10], Corollary 6.5.15), it follows that I^{fin} has compact level sets on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$.

4. As mentioned at the end of Section 5, I^{que} inherits from I^{fin} that it is lower semicontinuous and has compact level sets. In particular, I^{que} is the lower semicontinuous extension of I^{fin} from $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$. Moreover, since I^{fin} is affine on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ and I^{que} arises as the truncation limit of I^{fin} (recall (5.10)), it follows that I^{que} is affine on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$.

5. It is immediate from (1.15–1.16) that $q_{\rho,\nu}^{\otimes\mathbb{N}}$ is the unique zero of I^{que}.

□

7 Proof of Theorem 1.4

Proof. The extension is an easy generalisation of the proof given in Sections 3–4.

(a) Assume that ρ satisfies (1.1) with $\alpha = 1$. Since the LDP upper bound holds by the *annealed* LDP (compare (1.8) and (1.16)), it suffices to prove the LDP lower bound. To achieve this, we first show that for any $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ and $\varepsilon > 0$ there exists an open neighbourhood $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ of Q such that

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in \mathcal{O}(Q) \mid X\big) \geq -I^{\mathrm{ann}}(Q) - \varepsilon \qquad X\text{–a.s.} \tag{7.1}$$

After that, the extension from $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ follows the argument in Section 5.

In order to verify (7.1), observe that, by our assumption on $\rho(\cdot)$, for any $\alpha' > 1$ there exists a $C_{\alpha'} > 0$ such that

$$\frac{\rho(n)}{n^{\alpha'}} \geq C_{\alpha'} \qquad \forall\, n \in \mathrm{supp}(\rho). \tag{7.2}$$

Picking α' so close to 1 that $(\alpha'-1) m_Q H(\Psi_Q|\nu^{\otimes\mathbb{N}}) < \varepsilon/2$, we can trace through the proof of Proposition 4.1 in Section 4 to construct an open neighbourhood $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ of Q satisfying

$$\begin{aligned} &\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in \mathcal{O}(Q) \mid X\big) \\ &\qquad \geq -H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - (\alpha'-1) m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - \varepsilon/2 \geq -I^{\mathrm{ann}}(Q) - \varepsilon \quad X-a.s., \end{aligned} \tag{7.3}$$

which is (7.1).

(b) We only give a sketch of the argument. Assume $\alpha = \infty$ in (1.1). For $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, the lower bound (which is non-zero only when $Q \in \mathscr{R}_\nu$) follows from Birkner [3], Proposition 2, or can alternatively be obtained from the argument in Section 4. Now consider a $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with $m_Q = \infty$, $H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) < \infty$ and $\lim_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}}) = 0$, let $\mathcal{O}(Q) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ be an open neighbourhood of Q. For simplicity, we assume $\mathrm{supp}(\rho) = \mathbb{N}$. Fix $\varepsilon > 0$. We can find a sequence $\delta_N \downarrow 0$ such that

$$\max\Big\{ -\frac{1}{N}\log\rho(n)\colon\ n \leq \lceil N\delta_N\rceil\Big\} \leq \varepsilon. \tag{7.4}$$

Furthermore,

$$\frac{1}{N} h\left(Q_{|\mathscr{F}_N} \mid q_{\rho,\nu}^{\otimes N}\right) \geq H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - \varepsilon \tag{7.5}$$

for $N \geq N_0 = N_0(\varepsilon, Q)$, and we can find $\mathrm{tr}_0 \in \mathbb{N}$ such that

$$\frac{1}{N_0} h\left(([Q]_{\mathrm{tr}})_{|\mathscr{F}_{N_0}} \mid q_{\rho,\nu}^{\otimes N_0}\right) \geq \frac{1}{N_0} h\left(Q_{|\mathscr{F}_{N_0}} \mid q_{\rho,\nu}^{\otimes N_0}\right) - \varepsilon \tag{7.6}$$

for $\mathrm{tr} \geq \mathrm{tr}_0$. Hence

$$H([Q]_{\mathrm{tr}} \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) \geq H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - 2\varepsilon \quad \text{for } \mathrm{tr} \geq \mathrm{tr}_0. \tag{7.7}$$

We may also assume that $[Q]_{\mathrm{tr}} \in \mathcal{O}(Q)$ for $\mathrm{tr} \geq \mathrm{tr}_0$. For a given $N \geq N_0$, pick $\mathrm{tr}(N) \geq \mathrm{tr}_0$ so large that $m_{[Q]_{\mathrm{tr}(N)}} H(\Psi_{[Q]_{\mathrm{tr}(N)}} \mid \nu^{\otimes\mathbb{N}}) \leq \delta_N/2$. Using the strategy described at the beginning of Section 4, we can construct a neighbourhood $\mathcal{O}_N \subset \mathcal{O}(Q)$ of $[Q]_{\mathrm{tr}(N)}$ such that the conditional probability $\mathbb{P}(R_N \in \mathcal{O}_N | X)$ is bounded below by

$$\exp\big[-N(H([Q]_{\mathrm{tr}} \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) - \varepsilon)\big] \times \text{the cost of the first jump}, \tag{7.8}$$

where the first jump takes us to a region of size $\approx N m_{[Q]_{\mathrm{tr}(N)}}$ on which the medium looks "$\Psi_{[Q]_{\mathrm{tr}(N)}}$-typical". Since, in a typical medium, the size of the first jump will be

$$\approx \exp\big[N m_{[Q]_{\mathrm{tr}(N)}} H(\Psi_{[Q]_{\mathrm{tr}(N)}} \mid \nu^{\otimes\mathbb{N}})\big] \leq \exp[N\delta_N], \tag{7.9}$$

we obtain from (7.4) and (7.7–7.9) that

$$\mathbb{P}(R_N \in \mathcal{O}(Q)|X) \geq \exp\big[-N(H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + 4\varepsilon)\big] \tag{7.10}$$

for N large enough.

For the upper bound we can argue as follows: For $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ put

$$r(Q) := \limsup_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}(N)}} H(\Psi_{[Q]_{\mathrm{tr}(N)}} \mid \nu^{\otimes\mathbb{N}}). \tag{7.11}$$

Since ρ satisfies the bound (3.3) for any $\alpha > 1$, we obtain from the upper bound in Theorem 1.2 that the rate function at Q is at least

$$\limsup_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) = H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1) r(Q), \tag{7.12}$$

hence equals ∞ if $r(Q) > 0$. On the other hand, if $r(Q) = 0$, then this is simply the annealed bound.

□

8 Proof of Corollary 1.6

Proof. Without loss of generality we may assume that $E = \mathbb{N}$. For $\mathrm{c} \in \mathbb{N}$, let $\langle \cdot \rangle_c \colon E \to \langle E \rangle_c := \{1, \dots, c\}$ be the *letter truncation* map defined by

$$\langle x \rangle_\mathrm{c} := x \wedge \mathrm{c}, \qquad x \in E, \tag{8.1}$$

and extend this map to $\widetilde{E}$, $E^\mathbb{N}$, $\widetilde{E}^\mathbb{N}$, $\mathcal{P}^{\mathrm{inv}}(E^\mathbb{N})$ and $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^\mathbb{N})$ (similarly as in (1.11–1.13)). By Theorem 1.2, for each $c \in \mathbb{N}$ the family

$$\mathbb{P}(\langle R_N \rangle_\mathrm{c} \in \cdot \mid X), \quad N \in \mathbb{N}, \tag{8.2}$$

X-a.s. satisfies the LDP with deterministic rate function

$$I^{\mathrm{que}}_\mathrm{c}(Q) := H\big(Q \mid \langle q_{\rho,\nu}^{\otimes\mathbb{N}} \rangle_c\big) + (\alpha - 1) m_Q H\big(\Psi_Q \mid \langle \nu^{\otimes\mathbb{N}} \rangle_\mathrm{c}\big), \quad Q \in \mathcal{P}^{\mathrm{inv}}(\langle \widetilde{E} \rangle_c^\mathbb{N}), \tag{8.3}$$

where $\langle \widetilde{E} \rangle_c := \cup_{n\in\mathbb{N}} (\langle E \rangle_c)^n$. The letter truncations $\langle \cdot \rangle_\mathrm{c}$, $\mathrm{c} \in \mathbb{N}$, form a projective family. Hence, by the Dawson-Gärtner *projective limit theorem* (see Dembo and Zeitouni [10], Theorem 4.6.1), the family $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$, X-a.s. satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^\mathbb{N})$ with rate function

$$I^{\mathrm{que}}(Q) = \sup_{\mathrm{c} \in \mathbb{N}} I^{\mathrm{que}}_\mathrm{c}(\langle Q \rangle_c), \qquad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^\mathbb{N}). \tag{8.4}$$

However, the supremum equals the expression given in (1.15–1.16), because the specific relative entropies in the right-hand side of (8.3) are non-decreasing w.r.t. the letter truncation level, $m_Q = m_{\langle Q \rangle_c}$, $\langle \Psi_Q \rangle_\mathrm{c} = \Psi_{\langle Q \rangle_c}$, and the maps $\langle \cdot \rangle_\mathrm{c}$ and $[\cdot]_{\mathrm{tr}}$ commute. Thus, Theorem 1.2 indeed carries over. It is part of the projective limit theorem that I^{que} inherits from I^{que}_c, $c \in \mathbb{N}$, the properties qualifying it to be a rate function, so that also Theorem 1.3 carries over. $\square$

9 Proof of Theorems 1.8–1.9

9.1 Proof of Theorem 1.8

Proof. The idea is to put the problem into the framework of (1.1–1.5) and then apply Theorem 1.2. To that end, we pick

$$E := \mathbb{Z}^d, \qquad \widetilde{E} := \cup_{n\in\mathbb{N}} (\mathbb{Z}^d)^n, \tag{9.1}$$

and choose

$$\nu(u) := p(u), \quad u \in E, \qquad \rho(n) := \frac{p_n(0)}{G(0) - 1}, \quad n \in \mathbb{N}, \tag{9.2}$$

where

$$p(u) = p(0, u),\ u \in \mathbb{Z}^d, \quad p_n(u - v) = p^n(u, v),\ u, v \in \mathbb{Z}^d, \quad G(0) = \sum_{n=0}^{\infty} p_n(0), \tag{9.3}$$

the latter being the Green function at the origin.

Recalling (1.35), and writing

$$z^V = \big((z-1)+1\big)^V = 1 + \sum_{N=1}^{V} (z-1)^N \, \frac{V(V-1)\cdots(V-N+1)}{N!} \tag{9.4}$$

with

$$\frac{V(V-1)\cdots(V-N+1)}{N!} = \sum_{0<j_1<\cdots<j_N<\infty} \mathbb{1}_{\{S_{j_1}=S'_{j_1},\dots,S_{j_N}=S'_{j_N}\}}, \tag{9.5}$$

we have

$$\begin{aligned} \mathbb{E}\left[\, z^V \mid S \,\right] &= 1 + \sum_{N=1}^{\infty} (z-1)^N F_N^{(1)}(X), \\ \mathbb{E}\left[\, z^V \,\right] &= 1 + \sum_{N=1}^{\infty} (z-1)^N F_N^{(2)}, \end{aligned} \tag{9.6}$$

with

$$\begin{aligned} F_N^{(1)}(X) &:= \sum_{0<j_1<\cdots<j_N<\infty} \mathbb{P}(S_{j_1} = S'_{j_1}, \dots, S_{j_N} = S'_{j_N} \mid X), \\ F_N^{(2)} &:= \mathbb{E}\big[F_N^{(1)}(X)\big], \end{aligned} \tag{9.7}$$

where $X = (X_k)_{k\in\mathbb{N}}$ denotes the sequence of increments of S. (The upper indices 1 and 2 indicate the number of random walks being averaged over.)

The notation in (9.1–9.2) allows us to rewrite the first line of (9.7) as

$$\begin{aligned} F_N^{(1)}(X) &= \sum_{0<j_1<\cdots<j_N<\infty} \prod_{i=1}^{N} p_{j_i-j_{i-1}} \left(\sum_{k=j_{i-1}+1}^{j_i} X_k \right) \\ &= \sum_{0<j_1<\cdots<j_N<\infty} \prod_{i=1}^{N} \rho(j_i - j_{i-1}) \ \exp\left[\sum_{i=1}^{N} \log\left(\frac{p_{j_i-j_{i-1}}(\sum_{k=j_{i-1}+1}^{j_i} X_k)}{\rho(j_i - j_{i-1})} \right) \right] \end{aligned} \tag{9.8}$$

Let $Y^{(i)} = (X_{j_{i-1}+1}, \cdots, X_{j_i})$. Introduce $f\colon \widetilde{E} \to [0,\infty)$ by (recall (9.2))

$$f((x_1,\dots,x_n)) = \frac{p_n(x_1+\cdots+x_n)}{p_n(0)}\,[G(0)-1], \quad n\in\Lambda,\ x_1,\dots,x_n\in E, \tag{9.9}$$

with

$$\Lambda := \{n\in\mathbb{N}\colon \rho(n) = p_n(0) > 0\} \supset 2\mathbb{Z}, \tag{9.10}$$

let $R_N \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ be the empirical process of words defined in (1.5), and $\pi_1 R_N \in \mathcal{P}(\widetilde{E})$ the projection of R_N onto the first coordinate. Then we have

$$\begin{aligned} F_N^{(1)}(X) &= \mathbb{E}\left[\exp\left(\sum_{i=1}^{N} \log f(Y^{(i)}) \right) \middle|\, X \right] \\ &= \mathbb{E}\left[\exp\left(N \int_{\widetilde{E}} (\pi_1 R_N)(dy)\, \log f(y) \right) \middle|\, X \right]. \end{aligned} \tag{9.11}$$

The second line of (9.7) is obtained by averaging (9.11) over X:

$$F_N^{(2)} = \mathbb{E}\left[\exp\left(N \int_{\widetilde{E}} (\pi_1 R_N)(dy)\, \log f(y)\right)\right]. \tag{9.12}$$

Without conditioning on X, the sequence $(Y^{(i)})_{i\in\mathbb{N}}$ is i.i.d. with law (recall (1.4))

$$q_{\rho,\nu}^{\otimes\mathbb{N}} \quad \text{with} \quad q_{\rho,\nu}(x_1,\ldots,x_n) = \frac{p_n(0)}{G(0)-1} \prod_{k=1}^{n} p(x_k). \tag{9.13}$$

Next we note that f defined in (9.9) is bounded from above. Indeed, the Fourier representation of $p_n(x,y)$ reads

$$p_n(0,x) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} dk\, e^{-i(k\cdot x)}\, \frac{1}{1-\widehat{p}(k)} \tag{9.14}$$

with $\widehat{p}(k) = \sum_{x\in\mathbb{Z}^d} e^{i(k\cdot x)} p(0,x)$. Because $p(\cdot,\cdot)$ is symmetric, it follows that

$$\max_{x\in\mathbb{Z}^d} p^{2n}(0,x) = p^{2n}(0,0), \qquad \max_{x\in\mathbb{Z}^d} p^{2n+1}(0,x) \le p^{2n}(0,0), \qquad \forall\, n\in\mathbb{N}. \tag{9.15}$$

Consequently, $f((x_1,\ldots,x_n)) \le [p_{n-1}(0,0)/p_n(0,0)][G(0)-1]$, $n\in\Lambda$, which is bounded from above because of (1.34). The *annealed* LDP in Theorem 1.1, together with Varadhan's lemma applied to (9.12), therefore gives $z_2 = 1+\exp[-r_2]$ with

$$\begin{aligned} r_2 := \lim_{N\to\infty} \frac{1}{N} \log F_N^{(2)} &= \sup_{Q\in\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})} \left\{\int_{\widetilde{E}} \pi_1 Q(dy)\, \log f(y) - I^{\mathrm{ann}}(Q)\right\} \\ &= \sup_{q\in\mathcal{P}(\widetilde{E})} \left\{\int_{\widetilde{E}} q(dy)\, \log f(y) - h(q \mid q_{\rho,\nu})\right\} \end{aligned} \tag{9.16}$$

(recall (1.36) and (9.6)). The last equality stems from the fact that, on the set of Q's with a given marginal $\pi_1 Q = q$, the function $Q \mapsto I^{\mathrm{ann}}(Q) = H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$ has a unique minimiser $Q = q^{\otimes\mathbb{N}}$.

In order to carry out the second supremum in (9.16), we prove the following.

Lemma 9.1. *Let $Z := \sum_{y\in E} f(y) q_{\rho,\nu}(y)$. Then*

$$\int_{\widetilde{E}} q(dy)\, \log f(y) - h(q \mid q_{\rho,\nu}) = \log Z - h(q \mid q^*) \qquad \forall\, q\in\mathcal{P}(\widetilde{E}), \tag{9.17}$$

where $q^*(y) = f(y) q_{\rho,\nu}(y)/Z$, $y\in E$.

Proof. This follows from a straightforward computation. □

Inserting (9.17) into (9.16), we see that the suprema are uniquely attained at $q = q^*$ and $Q = (q^*)^{\otimes\mathbb{N}}$, and that $r_2 = \log Z$. From (9.9) and (9.13), we have

$$Z = \sum_{n\in\mathbb{N}} \sum_{x_1,\ldots,x_n\in\mathbb{Z}^d} p_n(x_1+\cdots+x_n) \prod_{k=1}^{n} p(x_k) = \sum_{n\in\mathbb{N}} p_{2n}(0) = G^{(2)}(0) - 1, \tag{9.18}$$

where we use that $\sum_{v\in\mathbb{Z}^d} p_m(u+v)p(v) = p_{m+1}(u)$, $u \in \mathbb{Z}^d$, $m \in \mathbb{N}$, and $G^{(2)}(0)$ is the Green function at the origin associated with $p^2(\cdot,\cdot)$. Hence the maximizer in (9.16) is

$$q^*(x_1,\dots,x_n) = \frac{p_n(x_1+\cdots+x_n)}{G^{(2)}(0)-1}\prod_{k=1}^{n} p(x_k). \tag{9.19}$$

Note that $z_2 = 1+\exp[-\log Z] = G^{(2)}(0)/[G^{(2)}(0)-1]$.

The *quenched* LDP in Theorem 1.2, together with Varadhan's lemma applied to (9.8), gives $z_1 = 1+\exp[-r_1]$ with

$$\begin{aligned} r_1 &:= \lim_{N\to\infty}\frac{1}{N}\log F_N^{(1)}(X) \\ &= \sup_{Q\in\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})}\left\{\int_{\widetilde{E}}\pi_1 Q(dy)\,\log f(y) - I^{\mathrm{que}}(Q)\right\} \qquad X-a.s., \end{aligned} \tag{9.20}$$

where $I^{\mathrm{que}}(Q)$ is given by (1.15–1.16).

To compare (9.20) with (9.16), we need the following lemma, the proof of which is deferred to Section 9.2.

Lemma 9.2. *Assume* (1.34)*. Let* $Q^* = (q^*)^{\otimes\mathbb{N}}$ *with* q^* *as in* (9.19)*. If* $m_{Q^*} < \infty$*, then* $I^{\mathrm{que}}(Q^*) > I^{\mathrm{ann}}(Q^*)$.

With the help of Lemma 9.2 we complete the proof of the existence of the gap as follows. Since $I^{\mathrm{que}}(Q) \ge I^{\mathrm{ann}}(Q)$, we have $r_1 \le r_2 < \infty$, and in order to prove the gap we are after, it suffices to show that $r_1 < r_2$. Since I^{que} has compact level sets, there exists a sequence $(Q_n)_{n\in\mathbb{N}} \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ such that

$$r_1 = \lim_{n\to\infty}\int_{\widetilde{E}}\pi_1 Q_n(dy)\,\log f(y) - I^{\mathrm{que}}(Q_n) \tag{9.21}$$

and $\mathrm{w}-\lim_{n\to\infty} Q_n = \widetilde{Q} \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$. Using that f is positive and bounded from above (and hence $\log f$ is negative after a shift), we have

$$\limsup_{n\to\infty}\int_{\widetilde{E}}\pi_1 Q_n(dy)\,\log f(y) \le \int_{\widetilde{E}}\pi_1\widetilde{Q}(dy)\,\log f(y) \tag{9.22}$$

by Fatou's Lemma. Furthermore, $\liminf_{n\to\infty} I^{\mathrm{que}}(Q_n) \ge I^{\mathrm{que}}(\widetilde{Q})$ by lower semicontinuity, so

$$r_1 = \int_{\widetilde{E}}\pi_1\widetilde{Q}(dy)\,\log f(y) - I^{\mathrm{que}}(\widetilde{Q}) \le \int_{\widetilde{E}}\pi_1\widetilde{Q}(dy)\,\log f(y) - I^{\mathrm{ann}}(\widetilde{Q}) \le r_2. \tag{9.23}$$

If $r_1 = r_2$, then $\widetilde{Q} = Q^*$, because the unconditional variational problem (9.16) has Q^* as its unique maximiser. But $I^{\mathrm{que}}(Q^*) > I^{\mathrm{ann}}(Q^*)$ by Lemma 9.2, so this is a contradiction, and we arrive at $r_1 < r_2$ as required. □

9.2 Proof of Lemma 9.2

Proof. Note that

$$q^*(E^n) = \sum_{x_1,\dots,x_n\in\mathbb{Z}^d} \frac{p_n(x_1+\cdots+x_n)}{G^{(2)}(0)-1} \prod_{k=1}^n p(x_k) = \frac{p_{2n}(0)}{G^{(2)}(0)-1}, \qquad n\in\mathbb{N}, \tag{9.24}$$

and hence, by assumption (1.35),

$$\lim_{n\to\infty} \frac{\log q^*(E^n)}{\log n} = -\alpha \tag{9.25}$$

and

$$m_{Q^*} = \sum_{n=1}^\infty n q^*(E^n) = \sum_{n=1}^\infty \frac{n p_{2n}(0)}{G^{(2)}(0)-1}. \tag{9.26}$$

We will show that

$$m_{Q^*} < \infty \quad \Longrightarrow \quad Q^* = (q^*)^{\otimes\mathbb{N}} \notin \mathcal{R}_\nu, \tag{9.27}$$

the set defined in (1.20). This implies $\Psi_{Q^*} \neq \nu^{\otimes\mathbb{N}}$ (recall (1.22)), and hence $H(\Psi_{Q^*}|\nu^{\otimes\mathbb{N}}) > 0$, implying the claim.

In order to verify (9.27), we compute the first two marginals of Ψ_{Q^*}. Using the symmetry of $p(\cdot,\cdot)$, we have

$$\Psi_{Q^*}(a) = \frac{1}{m_{Q^*}} \sum_{n=1}^\infty \sum_{j=1}^n \sum_{\substack{x_1,\dots,x_n\in\mathbb{Z}^d \\ x_j=a}} \frac{p_n(x_1+\cdots+x_n)}{G^{(2)}(0)-1} \prod_{k=1}^n p(x_k) = p(a)\, \frac{\sum_{n=1}^\infty n p_{2n-1}(a)}{\sum_{n=1}^\infty n p_{2n}(0)}. \tag{9.28}$$

Hence, $\Psi_{Q^*}(a) = p(a)$ for all $a\in\mathbb{Z}^d$ with $p(a)>0$ if and only if

$$a \mapsto \sum_{n=1}^\infty n\, p_{2n-1}(a) \text{ is constant on the support of } p(\cdot). \tag{9.29}$$

There are many $p(\cdot,\cdot)$'s for which (9.29) fails, and for these (9.27) holds. However, for simple random walk (9.29) does not fail, because $a\mapsto p_{2n-1}(a)$ is constant on the $2d$ neighbours of the origin, and so we have to look at the two-dimensional marginal.

Observe that $q^*(x_1,\dots,x_n) = q^*(x_{\sigma(1)},\dots x_{\sigma(n)})$ for any permutation σ of $\{1,\dots,n\}$. For $a,b\in\mathbb{Z}^d$, we have

$$\begin{aligned} m_{Q^*}\Psi_{Q^*}(a,b) &= \mathbb{E}_{Q^*}\left[\sum_{k=1}^{\tau_1} \mathbb{1}\left(\kappa(Y)_k = a, \kappa(Y)_{k+1} = b\right)\right] \\ &= \sum_{n=1}^\infty \sum_{n'=1}^\infty \sum_{x_1,\dots,x_{n+n'}} q^*(x_1,\dots,x_n)\; q^*(x_{n+1},\dots,x_{n+n'}) \sum_{k=1}^n \mathbb{1}_{(a,b)}(x_k,x_{k+1}) \\ &= q^*(x_1=a)\, q^*(x_1=b) + \sum_{n=2}^\infty (n-1) q^*\big(x_1=a, x_2=b\big). \end{aligned} \tag{9.30}$$

Since

$$\begin{aligned} q^*(x_1 = a) \\ &= \frac{p(a)^2}{G^{(2)}(0)-1} + \sum_{n=2}^{\infty} \sum_{x_2,\dots,x_n \in \mathbb{Z}^d} \frac{p_n(a + x_2 + \cdots + x_n)}{G^{(2)}(0)-1} p(a) \prod_{k=2}^{n} p(x_k) \\ &= \frac{p(a)}{G^{(2)}(0)-1} \sum_{n=1}^{\infty} p_{2n-1}(a) \end{aligned} \tag{9.31}$$

and

$$\begin{aligned} q^*\big(x_1 = a, x_2 = b\big) &= \mathbb{1}_{n=2} \frac{p(a)p(b)}{G^{(2)}(0)-1} p_2(a+b) \\ &\quad + \mathbb{1}_{n\geq 3} \frac{p(a)p(b)}{G^{(2)}(0)-1} \sum_{x_3,\dots,x_n \in \mathbb{Z}^d} p_n(a + b + x_3 + \cdots + x_n) \prod_{k=3}^{n} p(x_k) \\ &= \frac{p(a)p(b)}{G^{(2)}(0)-1} p_{2n-2}(a+b), \end{aligned} \tag{9.32}$$

we find

$$\Psi_{Q^*}(a,b) = \frac{p(a)p(b)}{\sum_{n=1}^{\infty} n p_{2n}(0)} \left(\left[\sum_{n=1}^{\infty} p_{2n-1}(a) \right] \left[\sum_{n=1}^{\infty} p_{2n-1}(b) \right] + \sum_{n=2}^{\infty} (n-1) p_{2n-2}(a+b) \right). \tag{9.33}$$

Pick $b = -a$ with $p(a) > 0$. Then, shifting n to $n-1$ in the last sum, we get

$$\Psi_{Q^*}(a,-a) - p(a)^2 = \frac{\left[\sum_{n=1}^{\infty} p_{2n-1}(a)\right]^2}{\sum_{n=1}^{\infty} n p_{2n}(0)} > 0. \tag{9.34}$$

This shows that consecutive letters are not uncorrelated under Ψ_{Q^*}, and implies that (9.27) holds as claimed. □

9.3 Upper bound on z_1

Unlike for z_2, no closed form expression is known for z_1. The arguments used to prove Theorem 1.8, which parallel those in Birkner [1], Chapter 5, imply that the value given in [1], Theorem 5, is in fact an upper bound.

Corollary 9.3. *Under the assumptions of Theorem* 1.8*,*

$$z_1 \leq 1 + \left(\sum_{n \in \mathbb{N}} e^{-h(p_n)} \right)^{-1}, \tag{9.35}$$

where $h(p_n) = -\sum_{x \in \mathbb{Z}^d} p_n(0,x) \log p_n(0,x)$ *is the entropy of* $p_n(0,\cdot)$.

Proof. Note that for $q \in \mathcal{P}(\widetilde{E})$ of the form

$$q(x_1,\dots,x_n) = \rho_q(n)\nu(x_1)\cdots\nu(x_n), \quad n \in \mathbb{N},\ x_1,\dots,x_n \in E, \tag{9.36}$$

for some $\rho_q \in \mathcal{P}(\mathbb{N})$, we have $I_1^{\mathrm{que}}(q) = h(\rho_q \mid \rho)$, as then the minimiser in the right-hand side of (1.23) is $Q = q^{\otimes\mathbb{N}}$. The claim therefore follows from (9.20) by choosing $Q = q^{\otimes\mathbb{N}}$, $\nu(x) = p(x)$, $x \in \mathbb{Z}^d$, and

$$\rho_q(n) = \frac{\exp[-h(p_n)]}{\sum_{m\in\mathbb{N}} \exp[-h(p_m)]}, \qquad n \in \mathbb{N}. \tag{9.37}$$

□

9.4 Proof of Theorem 1.9

The proof is a relatively minor extension of that of Theorem 1.8 in Sections 9.1–9.2.

Proof. The analogues of (9.4–9.7) are

$$z^{\widetilde{V}} = \sum_{N=0}^{\infty} (\log z)^N \, \frac{\widetilde{V}^N}{N!}, \tag{9.38}$$

with

$$\frac{\widetilde{V}^N}{N!} = \int_0^\infty dt_1 \cdots \int_{t_{N-1}}^\infty dt_N \; \mathbb{1}_{\{\widetilde{S}_{t_1}=\widetilde{S}'_{t_1},\dots,\widetilde{S}_{t_N}=\widetilde{S}'_{t_N}\}}, \tag{9.39}$$

and

$$\begin{aligned} \mathbb{E}\left[z^{\widetilde{V}} \mid S\right] &= \sum_{N=0}^{\infty} (\log z)^N \, F_N^{(1)}(\widetilde{S}), \\ \mathbb{E}\left[z^{\widetilde{V}}\right] &= \sum_{N=0}^{\infty} (\log z)^N \, F_N^{(2)}, \end{aligned} \tag{9.40}$$

with

$$\begin{aligned} F_N^{(1)}(\widetilde{S}) &:= \int_0^\infty dt_1 \cdots \int_{t_{N-1}}^\infty dt_N \; \mathbb{P}\left(\widetilde{S}_{t_1} = \widetilde{S}'_{t_1}, \dots, \widetilde{S}_{t_N} = \widetilde{S}'_{t_N} \mid \widetilde{S}\right), \\ F_N^{(2)} &:= \mathbb{E}\big[F_N^{(1)}(\widetilde{S})\big], \end{aligned} \tag{9.41}$$

where the conditioning in the first line is on the full continuous-time path $\widetilde{S} = (\widetilde{S}_t)_{t\geq 0}$. Our task is to compute

$$\begin{aligned} \widetilde{r}_1 &:= \lim_{N\to\infty} \frac{1}{N} \log F_N^{(1)}(\widetilde{S}), \qquad \widetilde{S} - a.s., \\ \widetilde{r}_2 &:= \lim_{N\to\infty} \frac{1}{N} \log F_N^{(2)}, \end{aligned} \tag{9.42}$$

and show that $\widetilde{r}_1 < \widetilde{r}_2$.

The idea is to average over the jump times of $\widetilde{S}$ while keeping its jumps fixed, thereby reducing the problem to the one for the discrete-time random walk treated in the proof of Theorem 1.9. For the first line in (9.41) this *partial annealing* gives an upper bound, while for the second line it is simply part of the averaging over $\widetilde{S}$. To that end, put $\sigma_0 := 0$, for $k \in \mathbb{N}$ put $\sigma_k := \inf\{t > \sigma_{k-1} \colon \widetilde{S}_t \neq \widetilde{S}_{\sigma_{k-1}}\}$, let

$$X^\natural = (X_k^\natural)_{k\in\mathbb{N}} \quad \text{with} \quad X_k^\natural := \widetilde{S}_{\sigma_k}, \tag{9.43}$$

and define

$$\begin{aligned} F_N^{(1)}(X^\natural) &:= \int_0^\infty dt_1 \cdots \int_{t_{N-1}}^\infty dt_N \; \mathbb{P}(\widetilde{S}_{t_1} = \widetilde{S}'_{t_1}, \dots, \widetilde{S}_{t_N} = \widetilde{S}'_{t_N} \mid X^\natural), \\ F_N^{(2)} &:= \mathbb{E}\big[F_N^{(1)}(X^\natural)\big], \end{aligned} \tag{9.44}$$

together with the critical values

$$\begin{aligned} \widetilde{r}_1^\natural &:= \lim_{N\to\infty} \frac{1}{N} \log F_N^{(1)}(X^\natural), \qquad X^\natural - a.s., \\ \widetilde{r}_2^\natural &:= \lim_{N\to\infty} \frac{1}{N} \log F_N^{(2)}. \end{aligned} \tag{9.45}$$

Clearly,

$$\widetilde{r}_1 \le \widetilde{r}_1^\natural \text{ and } \widetilde{r}_2 = \widetilde{r}_2^\natural, \tag{9.46}$$

which can be viewed as a result of "partial annealing", and so it suffices to show that $\widetilde{r}_1{}^\natural < \widetilde{r}_2^\natural$. To this end write out

$$\begin{aligned} &\mathbb{P}(\widetilde{S}_{t_1} = \widetilde{S}'_{t_1}, \dots, \widetilde{S}_{t_N} = \widetilde{S}'_{t_N} \mid X^\natural) \\ &\quad = \sum_{0 \le j_1 \le \cdots \le j_N < \infty} \left(\prod_{i=1}^N e^{-(t_i - t_{i-1})} \frac{(t_i - t_{i-1})^{j_i - j_{i-1}}}{(j_i - j_{i-1})!} \right) \\ &\qquad \sum_{0 \le j'_1 \le \cdots \le j'_N < \infty} \left(\prod_{i=1}^N e^{-(t_i - t_{i-1})} \frac{(t_i - t_{i-1})^{j'_i - j'_{i-1}}}{(j'_i - j'_{i-1})!} \right) \left[\prod_{i=1}^N p_{j'_i - j'_{i-1}} \left(\sum_{k=j_{i-1}+1}^{j_i} X_k^\natural \right) \right]. \end{aligned} \tag{9.47}$$

Integrating over $0 \le t_1 \le \cdots \le t_N < \infty$, we obtain

$$\begin{aligned} F_N^{(1)}(X^\natural) &= \sum_{0 \le j_1 \le \cdots \le j_N < \infty} \; \sum_{0 \le j'_1 \le \cdots \le j'_N < \infty} \\ &\prod_{i=1}^N \left[2^{-(j_i - j_{i-1}) - (j'_i - j'_{i-1}) - 1} \frac{[(j_i - j_{i-1}) + (j'_i - j'_{i-1})]!}{(j_i - j_{i-1})!(j'_i - j'_{i-1})!} \, p_{j'_i - j'_{i-1}} \left(\sum_{k=j_{i-1}+1}^{j_i} X_k^\natural \right) \right]. \end{aligned} \tag{9.48}$$

Abbreviating

$$\Theta_n(u) = \sum_{m=0}^\infty p_m(u) \, 2^{-n-m-1} \binom{n+m}{m}, \qquad n \in \mathbb{N} \cup \{0\}, \; u \in \mathbb{Z}^d, \tag{9.49}$$

we may rewrite (9.48) as

$$F_N^{(1)}(X^\natural) = \sum_{0 \le j_1 \le \cdots \le j_N < \infty} \prod_{i=1}^N \Theta_{j_i - j_{i-1}} \left(\sum_{k=j_{i-1}+1}^{j_i} X_k^\natural \right). \tag{9.50}$$

This expression is similar in form as the first line of (9.8), except that the order of the j_i's is not strict. However, defining

$$\widehat{F}_N^{(1)}(X^\natural) = \sum_{0 < j_1 < \cdots < j_N < \infty} \prod_{i=1}^N \Theta_{j_i - j_{i-1}} \left(\sum_{k=j_{i-1}+1}^{j_i} X_k^\natural \right), \tag{9.51}$$

we have

$$F_N^{(1)}(X^\natural) = \sum_{M=0}^{N} \binom{N}{M} \theta_0(0)^M \, \widehat{F}_{N-M}^{(1)}(X^\natural), \tag{9.52}$$

with the convention $\widehat{F}_0^{(1)}(X^\natural) \equiv 1$. Letting

$$\widehat{r}_1^\natural = \lim_{N\to\infty} \frac{1}{N} \log \widehat{F}_N^{(1)}(X^\natural), \qquad X^\natural - a.s., \tag{9.53}$$

and recalling (9.45), we therefore have the relation

$$\widetilde{r}_1^\natural = \log\left[\theta_0(0) + e^{\widehat{r}_1^\natural}\right], \tag{9.54}$$

and so it suffices to compute $\widehat{r}_1^\natural$.

Write

$$F_N^{(1)}(X^\natural) = \mathbb{E}\left[\exp\left(N \int_{\widetilde{E}} \log f^\natural(y)\,(\pi_1 R_N)(dy)\right) \middle|\, X^\natural\right], \tag{9.55}$$

where $f^\natural\colon\ \widetilde{E} \to [0,\infty)$ is defined by

$$f^\natural((x_1,\dots,x_n)) = \frac{\Theta_n(x_1+\cdots+x_n)}{p_n(0)}\,[G(0)-1], \quad n\in\mathbb{N},\, x_1,\dots,x_n \in E. \tag{9.56}$$

Equations (9.55–9.56) replace (9.8–9.9). We can now repeat the same argument as in (9.16–9.23), with the sole difference that f in (9.9) is replaced by $f^\natural$ in (9.56), and this, combined with Lemma 9.4 below, yields the gap $\widetilde{r}_1^\natural < \widetilde{r}_2^\natural$.

We first check that $f^\natural$ is bounded from above, which is necessary for the application of Varadhan's lemma. To that end, we insert the Fourier representation (9.14) into (9.49) to obtain

$$\theta_n(u) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} dk\, e^{-i(k\cdot u)}\,[2-\widehat{p}(k)]^{-n-1}, \qquad u\in\mathbb{Z}^d, \tag{9.57}$$

from which we see that $\theta_n(u) \le \theta_n(0)$, $u\in\mathbb{Z}^d$. Consequently,

$$f_n^\natural((x_1,\cdots,x_n)) \le \frac{\theta_n(0)}{p_n(0)}\,[G(0)-1], \qquad n\in\Lambda. \tag{9.58}$$

Next we note that

$$\lim_{n\to\infty} \frac{1}{n} \log\left[2^{-(a+b)n-1}\binom{(a+b)n}{an}\right] \begin{cases} =0, & \text{if } a=b,\\ <0, & \text{if } a\neq b.\end{cases} \tag{9.59}$$

From (1.34), (9.49) and (9.59) it follows that $\theta_n(0)/p_n(0) \le C < \infty$ for all $n\in\Lambda$, so that $f^\natural$ indeed is bounded from above.

Note that $X^\natural$ is the discrete-time random walk with transition kernel $p(\cdot,\cdot)$. The key ingredient behind $\widehat{r}_1^\natural < \widehat{r}_2^\natural$ is the analogue of Lemma 9.2, this time with $Q^* = (q^*)^{\otimes\mathbb{N}}$ and q^* given by

$$q^*(x_1,\dots,x_n) = \frac{\Theta_n(x_1+\cdots+x_n)}{G(0)-1} \prod_{k=1}^{n} p(x_k), \tag{9.60}$$

replacing (9.19).

Lemma 9.4. *Assume* (1.34)*. Let* $Q^* = (q^*)^{\otimes\mathbb{N}}$ *with* q^* *as in* (9.60)*. If* $m_{Q^*} < \infty$*, then* $I^{\mathrm{que}}(Q^*) > I^{\mathrm{ann}}(Q^*)$*.*

The analogue of (9.18) reads

$$\begin{aligned} Z^\natural &= \sum_{n\in\mathbb{N}} \sum_{x_1,\dots,x_n\in\mathbb{Z}^d} [\Theta_n(x_1+\cdots+x_n)] \prod_{k=1}^n p(x_k) \\ &= \sum_{n\in\mathbb{N}} \sum_{m=0}^\infty \Bigg\{ \sum_{x_1,\dots,x_n\in\mathbb{Z}^d} p_m(x_1+\cdots+x_n) \prod_{k=1}^n p(x_k) \Bigg\}\, 2^{-n-m-1}\binom{n+m}{m} \Bigg] \\ &= -\theta_0(0) + \sum_{n,m=0}^\infty p_{n+m}(0)\, 2^{-n-m-1}\binom{n+m}{m} \\ &= -\theta_0(0) + \frac{1}{2}\sum_{k=0}^\infty p_k(0) = -\theta_0 + \frac{G(0)}{2}. \end{aligned} \tag{9.61}$$

Consequently,

$$\log \widetilde{z}_2 = e^{-\widetilde{r}_2} = e^{-\widetilde{r}_2^\natural} = \frac{1}{\theta_0 + e^{\widetilde{r}_2^\natural}} = \frac{1}{\theta_0 + Z^\natural} = \frac{2}{G(0)}, \tag{9.62}$$

where we use (9.40), (9.42), (9.46), (9.54) and (9.61). □

9.5 Proof of Lemma 9.4

Proof. We must adapt the proof in Section 9.2 to the fact that q^* has a slightly different form, namely, $p_n(x_1+\cdots+x_n)$ is replaced by $\Theta_n(x_1+\cdots+x_n)$, which averages transition kernels. The computations are straightforward and are left to the reader. The analogues of (9.24) and (9.26) are

$$\begin{aligned} q^*(E^n) &= \frac{1}{G(0)-1}\sum_{m=0}^\infty p_{n+m}(0)\, 2^{-n-m-1}\binom{n+m}{m}, \\ m_{Q^*} &= \sum_{n\in\mathbb{N}} n q^*(E^n) = \tfrac{1}{4}\sum_{k=0}^\infty k p_k(0), \end{aligned} \tag{9.63}$$

while the analogues of (9.31–9.32) are

$$\begin{aligned} q^*(x_1 = a) &= \frac{p(a)}{G(0)-1}\,\tfrac{1}{2}\sum_{k=0}^\infty p_k(a)[1-2^{-k-1}], \\ q^*(x_1 = a, x_2 = b) &= \frac{p(a)p(b)}{G(0)-1}\left[\tfrac{1}{4}\sum_{k=0}^\infty k p_k(a+b) + \sum_{k=0}^\infty p_k(a+b)\, 2^{-k-3}\right]. \end{aligned} \tag{9.64}$$

Recalling (9.30), we find

$$\Psi_{Q^*}(a,-a) - p(a)^2 > 0, \tag{9.65}$$

implying that $\Psi_{Q^*} \neq \nu^{\otimes\mathbb{N}}$ (recall (9.2)), and hence $H(\Psi_{Q^*} \mid \nu^{\otimes\mathbb{N}}) > 0$, implying the claim. □

A Appendix: Continuity under truncation limits

The following lemma implies (1.17).

Lemma A.1. *For all $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$,*

$$\begin{aligned} \lim_{\mathrm{tr}\to\infty} H([Q]_{\mathrm{tr}} \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) &= H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}), \\ \lim_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\mathbb{N}}) &= m_Q H(\Psi_Q \mid \nu^{\mathbb{N}}). \end{aligned} \tag{A.1}$$

Proof. The proof is not quite standard, because Q and $[Q]_{\mathrm{tr}}$, respectively, Ψ_Q and $\Psi_{[Q]_{\mathrm{tr}}}$ are not "$\bar{d}$-close" when tr is large, so that we cannot use the fact that entropy is "$\bar{d}$-continuous" (see Shields [21]).

Lower semi-continuity yields $\liminf_{\mathrm{tr}\to\infty}$ l.h.s. $\geq$ r.h.s. for both limits, so we need only prove the reverse inequality. Note that, for all $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$,

$$H(Q) \leq h(Q_{|\mathscr{F}_1}) \leq h\big(\mathscr{L}_Q(\tau_1)\big) + m_Q \log|E| < \infty, \quad H(\Psi_Q) \leq \log|E| < \infty, \quad H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) < \infty. \tag{A.2}$$

For Z a random variable, we write $\mathscr{L}_Q(Z)$ to denote the law of Z under Q.

A.1 Proof of first half of (A.1)

Proof. Since $q_{\rho,\nu}^{\otimes\mathbb{N}}$ is a product measure, we have for, any $\mathrm{tr} \in \mathbb{N}$,

$$\begin{aligned} H([Q]_{\mathrm{tr}} \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) &= -H([Q]_{\mathrm{tr}}) - \mathbb{E}_{[Q]_{\mathrm{tr}}}[\log\rho(\tau_1)] - \mathbb{E}_{[Q]_{\mathrm{tr}}}\left[\sum_{i=1}^{\tau_1} \log\nu\left(Y_i^{(1)}\right)\right] \\ &= -H([Q]_{\mathrm{tr}}) - \mathbb{E}_Q[\log\rho(\tau_1\wedge\mathrm{tr})] - \mathbb{E}_Q\left[\sum_{i=1}^{\tau_1\wedge\mathrm{tr}} \log\nu\left(Y_i^{(1)}\right)\right]. \end{aligned} \tag{A.3}$$

By dominated convergence, using that $m_Q < \infty$ and $\log\rho(n) \leq C\log(n+1)$ for some $C < \infty$, we see that as $\mathrm{tr} \to \infty$ the last two terms in the second line converge to

$$-\mathbb{E}_Q\big[\log\rho(\tau_1)\big] - \mathbb{E}_Q\left[\sum_{i=1}^{\tau_1} \log\nu\left(Y_i^{(1)}\right)\right]. \tag{A.4}$$

Thus, it remains to check that

$$\lim_{\mathrm{tr}\to\infty} H([Q]_{\mathrm{tr}}) = H(Q). \tag{A.5}$$

Obviously, $H([Q]_{\mathrm{tr}}) \leq H(Q)$ for all $\mathrm{tr} \in \mathbb{N}$ (indeed, $h([Q]_{\mathrm{tr}|\mathscr{F}_N}) \leq h(Q_{|\mathscr{F}_N})$ for all $N, \mathrm{tr} \in \mathbb{N}$, because $[Q]_{\mathrm{tr}}$ is the image measure of Q under the truncation map). For the asymptotic converse, we argue as follows. A decomposition of entropy gives

$$h(Q_{|\mathscr{F}_N}) = h([Q]_{\mathrm{tr}|\mathscr{F}_N}) + \int_{[\widetilde{E}]_{\mathrm{tr}}^N} h\Big(\mathscr{L}_Q\big(\pi_N Y \mid \pi_N[Y]_{\mathrm{tr}} = z\big)\Big)\,(\pi_N[Q]_{\mathrm{tr}})(dz), \tag{A.6}$$

where π_N is the projection onto the first N words, and $\mathscr{L}_Q(\pi_N Y \mid \pi_N[Y]_{\rm tr} = z)$ is the conditional distribution of the first N words given their truncations. We have

$$h\Big(\mathscr{L}_Q\big(\pi_N Y \mid \pi_N[Y]_{\rm tr} = z\big)\Big) \le \sum_{i=1}^{N} h\Big(\mathscr{L}_Q\big(Y_i \mid \pi_N[Y]_{\rm tr} = z\big)\Big) \tag{A.7}$$

and

$$\begin{aligned}
&\int_{[\widetilde{E}]_{\rm tr}^N} h\Big(\mathscr{L}_Q\big(Y_i \mid \pi_N[Y]_{\rm tr} = z\big)\Big)\,(\pi_N[Q]_{\rm tr})(dz)\\
&\quad \le \int_{[\widetilde{E}]_{\rm tr}^N} h\Big(\mathscr{L}_Q\big(Y_i \mid [Y_i]_{\rm tr} = z_i\big)\Big)\,(\pi_N[Q]_{\rm tr})(dz)\\
&\quad = \int_{[\widetilde{E}]_{\rm tr}} h\Big(\mathscr{L}_Q\big(Y_1 \mid [Y_1]_{\rm tr} = y\big)\Big)\,(\pi_1[Q]_{\rm tr})(dy), \qquad 1 \le i \le N,
\end{aligned} \tag{A.8}$$

where the inequality in the second line comes from the fact that conditioning on less increases entropy, and the third line uses the shift-invariance. Combining (A.6–A.8) and letting $N \to \infty$, we obtain

$$H(Q) \le H([Q]_{\rm tr}) + \int_{[\widetilde{E}]_{\rm tr}} h\Big(\mathscr{L}_Q\big(Y_1 \mid [Y_1]_{\rm tr} = y\big)\Big)\,(\pi_1[Q]_{\rm tr})(dy), \tag{A.9}$$

and so it remains to check that the second term in the right-hand side vanishes as ${\rm tr} \to \infty$.

Note that this term equals (write ε for the empty word and $w \cdot w'$ for the concatenation of words w and w')

$$\begin{aligned}
&- \sum_{\substack{w \in \tilde{E}\\ \tau(w) = {\rm tr}}} [Q]_{\rm tr}(w) \sum_{w' \in \tilde{E} \cup \{\varepsilon\}} \frac{Q(w \cdot w')}{[Q]_{\rm tr}(w)} \log\left[\frac{Q(w \cdot w')}{[Q]_{\rm tr}(w)}\right]\\
&= - \sum_{\substack{w'' \in \tilde{E}\\ \tau(w'') \ge {\rm tr}}} Q(w'') \log Q(w'') + \sum_{\substack{w'' \in \tilde{E}\\ \tau(w'') \ge {\rm tr}}} Q(w'') \log [Q]_{\rm tr}([w'']_{\rm tr}).
\end{aligned} \tag{A.10}$$

But

$$0 \ge \sum_{\substack{w'' \in \tilde{E}\\ \tau(w'') \ge {\rm tr}}} Q(w'') \log [Q]_{\rm tr}([w'']_{\rm tr}) \ge \sum_{\substack{w'' \in \tilde{E}\\ \tau(w'') \ge {\rm tr}}} Q(w'') \log Q(w''), \tag{A.11}$$

and so the right-hand side of (A.10) vanishes as ${\rm tr} \to \infty$. □

A.2 Proof of second half of (A.1)

Note that $\lim_{{\rm tr}\to\infty} m_{[Q]_{\rm tr}} = m_Q$ and ${\rm w}-\lim_{{\rm tr}\to\infty} \Psi_{[Q]_{\rm tr}} = \Psi_Q$ by dominated convergence, implying that

$$\liminf_{{\rm tr}\to\infty} H(\Psi_{[Q]_{\rm tr}} \mid \nu^{\mathbb{N}}) \ge H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}). \tag{A.12}$$

So it remains to check the reverse inequality. Since $\nu^{\otimes\mathbb{N}}$ is product measure, we have

$$H(\Psi_{[Q]_{\rm tr}} \mid \nu^{\mathbb{N}}) = -H(\Psi_{[Q]_{\rm tr}}) - \frac{1}{m_{[Q]_{\rm tr}}} \mathbb{E}_Q\left[\sum_{i=1}^{\tau_1 \wedge {\rm tr}} \log \nu\left(Y_i^{(1)}\right)\right]. \tag{A.13}$$

By dominated convergence, as tr $\to \infty$ the second term converges to

$$\frac{1}{m_Q}\mathbb{E}_Q\left[\sum_{i=1}^{\tau_1}\log\nu\left(Y_i^{(1)}\right)\right] = \int_E \Psi_Q(dx)\,\log\nu(x). \tag{A.14}$$

Thus, it remains to check that

$$\lim_{\mathrm{tr}\to\infty} H(\Psi_{[Q]_{\mathrm{tr}}}) = H(\Psi_Q). \tag{A.15}$$

We will first prove (A.15) for ergodic Q, in which case $[Q]_{\mathrm{tr}}$, Ψ_Q, $\Psi_{[Q]_{\mathrm{tr}}}$ are ergodic (Birkner [3], Remark 5).

For $\Psi \in \mathcal{P}^{\mathrm{erg}}(E^{\mathbb{N}})$ and $\varepsilon \in (0,1)$, let

$$\mathcal{N}_n(\Psi,\varepsilon) = \min\left\{\#A\colon\ A\subset E^n, \Psi(A\times E^\infty)\geq\varepsilon\right\} \tag{A.16}$$

be the (n,ε) covering number of Ψ. For any $\varepsilon \in (0,1)$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\mathcal{N}_n(\Psi,\varepsilon) = H(\Psi) \tag{A.17}$$

(see Shields [21], Theorem I.7.4). The idea behind (A.15) is that there are $\approx \exp[nH(\Psi_Q)]$ "Ψ_Q-typical" sequences of length n, and that a "$\Psi_{[Q]_{\mathrm{tr}}}$-typical" sequence arises from a "Ψ_Q-typical" sequence by eliminating a fraction δ_{tr} of the letters, where $\delta_{\mathrm{tr}} \to 0$ as tr $\to \infty$. Hence $\mathcal{N}_n(\Psi_Q,\varepsilon)$ cannot be much larger than $\mathcal{N}_n(\Psi_{[Q]_{\mathrm{tr}}},\varepsilon)$ (on an exponential scale), implying that $H(\Psi_Q)-H(\Psi_{[Q]_{\mathrm{tr}}})$ must be small.

To make this argument precise, fix $\varepsilon > 0$ and pick N_0 so large that

$$Q\big(|\kappa(Y^{(1)},\dots,Y^{(N)})| \in Nm_Q[1-\varepsilon,1+\varepsilon]\big) > 1-\varepsilon \qquad \text{for } N\geq N_0. \tag{A.18}$$

Pick $\mathrm{tr}_0 \in \mathbb{N}$ so large that for $\mathrm{tr}\geq \mathrm{tr}_0$ and $N \geq N_0$,

$$Q\Big(\textstyle\sum_{i=1}^N(\tau_1-\mathrm{tr})_+ < N\varepsilon\Big) > 1-\varepsilon/2,\ Q\big(\tau_1\leq\mathrm{tr}\big) > 1-\varepsilon/2,\ m_{[Q]_{\mathrm{tr}}} > (1-\varepsilon)m_Q. \tag{A.19}$$

For $n \geq \lceil N_0/m_Q\rceil$, we will construct a set $B\subset E^n$ such that

$$\Psi_Q(B\times E^\infty)\geq \tfrac{1}{2}, \qquad |B| \leq \exp\left[n(H(\Psi_{[Q]_{\mathrm{tr}}})+\delta)\right], \tag{A.20}$$

where δ can be made arbitrarily small by choosing ε small in (A.18–A.19). Hence, by the asymptotic cover property (A.17), we have $H(\Psi_Q)\leq(1+\delta)H(\Psi_{[Q]_{\mathrm{tr}}})$ and

$$\liminf_{\mathrm{tr}\to\infty} H(\Psi_{[Q]_{\mathrm{tr}}}) \geq H(\Psi_Q), \tag{A.21}$$

completing the proof of (A.15).

We verify (A.20) as follows. Put $N := \lceil nm_Q(1+2\varepsilon)\rceil$. By (A.18–A.19) and the asymptotic cover property (A.17) for $\Psi_{[Q]_{\mathrm{tr}}}$, there is a set $A\subset \widetilde{E}^N$ such that

$$\mathbb{E}_Q\big[\tau_1 1\!\!1_A(Y^{(1)},\dots,Y^{(N)})\big] > (1-\varepsilon)m_Q \tag{A.22}$$

and

$$|\kappa(y^{(1)},\dots,y^{(N)})|\geq n(1+\varepsilon),\quad \tau(y^{(1)})\leq \mathrm{tr},\quad \sum_{i=1}^{N}(\tau(y^{(i)})-\mathrm{tr})_+<N\varepsilon,$$
$$\forall\,(y^{(1)},\dots,y^{(N)})\in A, \tag{A.23}$$

while the set

$$B':=\big\{\kappa([y^{(1)}]_{\mathrm{tr}},\dots,[y^{(N)}]_{\mathrm{tr}})|_{(0,\lceil(1-\varepsilon)n\rceil]}\colon\,(y^{(1)},\dots,y^{(N)})\in A\big\}\subset E^{\lceil(1-\varepsilon)n\rceil]} \tag{A.24}$$

satisfies

$$|B'|\leq \exp\big[n(H(\Psi_{[Q]_{\mathrm{tr}}})+\varepsilon)\big]. \tag{A.25}$$

Put

$$B:=\big\{\kappa(y^{(1)},\dots,y^{(N)})|_{(0,n]}\colon (y^{(1)},\dots,y^{(N)})\in A\big\}\subset E^n. \tag{A.26}$$

Observe that each $x'\in B'$ corresponds to at most

$$|E|^{\varepsilon n}\binom{n}{\varepsilon n}\leq \exp\big[-n(\varepsilon\log\varepsilon+(1-\varepsilon)\log(1-\varepsilon))+n\varepsilon\log|E|\big] \tag{A.27}$$

different $x\in B$, so that

$$|B|\leq |B'|\exp\big[-n(\varepsilon\log\varepsilon+(1-\varepsilon)\log(1-\varepsilon))+n\varepsilon\log|E|\big]. \tag{A.28}$$

We have

$$\begin{aligned} m_Q\Psi_Q(B\times E^\infty)&\geq \mathbb{E}_Q\left[\sum_{k=0}^{\tau_1-1}\mathbb{1}_{B\times E^\infty}\big(\theta^k\kappa(Y)\big)\mathbb{1}_A(Y^{(1)},\dots,Y^{(N)})\right]\\ &=\mathbb{E}_Q\left[\sum_{k=0}^{\tau_1\wedge\mathrm{tr}-1}\mathbb{1}_{B'\times E^\infty}\big(\theta^k\kappa([Y]_{\mathrm{tr}})\big)\mathbb{1}_A(Y^{(1)},\dots,Y^{(N)})\right]\\ &\geq \mathbb{E}_Q\left[\sum_{k=0}^{\tau_1\wedge\mathrm{tr}|-1}\mathbb{1}_{B'\times E^\infty}\big(\theta^k\kappa([Y]_{\mathrm{tr}})\big)\right]-\varepsilon m_Q\\ &=m_{[Q]_{\mathrm{tr}}}\Psi_{[Q]_{\mathrm{tr}}}(B'\times E^\infty)-\varepsilon m_Q, \end{aligned} \tag{A.29}$$

so that, finally,

$$\Psi_Q(B\times E^\infty)\geq \frac{m_{[Q]_{\mathrm{tr}}}}{m_Q}\,\Psi_{[Q]_{\mathrm{tr}}}(B'\times E^\infty)-\varepsilon\geq \tfrac12. \tag{A.30}$$

Combining (A.25), (A.28) and (A.30), we obtain (A.20) with

$$\delta=-\big(\varepsilon\log\varepsilon+(1-\varepsilon)\log(1-\varepsilon)\big)+\varepsilon\big(1+\log|E|\big). \tag{A.31}$$

Since $\limsup_{\mathrm{tr}\to\infty}H(\Psi_{[Q]_{\mathrm{tr}}})\leq H(\Psi_Q)$ by upper semi-continuity of H (see e.g. Georgii [13], Proposition. 15.14), this concludes the proof of (A.15) for ergodic Q.

For general $Q\in\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, we recall the ergodic decomposition formulas stated in (6.1–6.2). These yields

$$\Psi_{[Q]_{\mathrm{tr}}}=\int_{\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})}\frac{m_{[Q']_{\mathrm{tr}}}}{m_{[Q]_{\mathrm{tr}}}}\,\Psi_{[Q']_{\mathrm{tr}}}\,W_Q(dQ'), \tag{A.32}$$

and

$$H(\Psi_{[Q]_{\mathrm{tr}}}) = \int_{\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})} \frac{m_{[Q']_{\mathrm{tr}}}}{m_{[Q]_{\mathrm{tr}}}} H(\Psi_{[Q']_{\mathrm{tr}}}) \, W_Q(dQ'), \tag{A.33}$$

because specific relative entropy is affine. The integrand inside (A.33) is non-negative and, by the above, converges to $\frac{m_{Q'}}{m_Q} H(\Psi_{Q'})$ as $\mathrm{tr} \to \infty$. Hence, by Fatou's lemma,

$$\liminf_{\mathrm{tr}\to\infty} H(\Psi_{[Q]_{\mathrm{tr}}}) \geq \int_{\mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})} \frac{m_{Q'}}{m_Q} H(\Psi_{Q'}) \, W_Q(dQ') = H(\Psi_Q), \tag{A.34}$$

which concludes the proof. □

References

[1] M. Birkner, *Particle Systems with Locally Dependent Branching: Long-Time Behaviour, Genealogy and Critical Parameters*, Dissertation, Johann Wolfgang Goethe-Universität Frankfurt am Main, 2003.
`http://publikationen.ub.uni-frankfurt.de/volltexte/2003/314/`

[2] M. Birkner, A condition for weak disorder for directed polymers in random environment, Electron. Comm. Probab. 9 (2004) 22–25.

[3] M. Birkner, Conditional large deviations for a sequence of words, Stoch. Proc. Appl. 118 (2008) 703–729.

[4] M. Birkner and R. Sun, Annealed vs quenched critical points for a random walk pinning model, preprint 2008.

[5] E. Bolthausen, A note on the diffusion of directed polymers in a random environment, Commun. Math. Phys. 123 (1989) 529–534.

[6] A. Camanes and P. Carmona, Directed polymers, critical temperature and uniform integrability, preprint 2007.

[7] F. Comets, Large deviation estimates for a conditional probability distribution. Applications to random interaction Gibbs measures, Probab. Theory Relat. Fields 80 (1989) 407–432.

[8] F. Comets, T. Shiga and N. Yoshida, Directed polymers in random environment: path localization and strong disorder, Bernoulli 9 (2003) 705–723.

[9] F. Comets and N. Yoshida, Directed polymers in random environment are diffusive at weak disorder, Ann. Probab. 34 (2006) 1746–1770.

[10] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers, Boston, 1993.

[11] B. Derrida, G. Giacomin, H. Lacoin and F.L. Toninelli, Fractional moment bounds and disorder relevance for pinning models (2007), arXiv:0712.2515v1.

[12] J.-D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, Boston, 1989.

[13] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter, Berlin, 1988.

[14] A. Greven, Phase transition for the coupled branching process, Part I: The ergodic theory in the range of second moments, Probab. Theory Relat. Fields 87 (1991) 417–458.

[15] A. Greven, On phase transitions in spatial branching systems with interaction, in: *Stochastic Models* (L.G. Gorostiza and B.G. Ivanoff, eds.), CMS Conference Proceedings 26 (2000) 173–204.

[16] A. Greven and F. den Hollander, Phase transitions for the long-time behaviour of interacting diffusions, Ann. Probab. 35 (2007) 1250–1306.

[17] F. den Hollander, *Large Deviations*, Fields Institute Monographs 14, American Mathematical Society, Providence, RI, 2000.

[18] R. Holley and T. M. Liggett, Generalized potlatch and smoothing processes, Z. Wahrscheinlichkeitstheorie verw. Gebiete 55 (1981) 165-195.

[19] O. Kallenberg, Stability of critical cluster fields, Math. Nachr. 77 (1977) 7–45.

[20] C. Monthus and T. Garel, Freezing transition of the directed polymer in a 1+d random medium: Location of the critical temperature and unusual critical properties, Phys. Rev. E 74 (2006) 011101.

[21] P.C. Shields, *The Ergodic Theory of Discrete Sample Paths*, American Mathematical Society, Providence, RI, 1996.

[22] F. Spitzer, *Principles of Random Walk* (2nd. ed.), Springer, New York, 1976.

[23] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Statist. 36 (1965) 423–439.

[24] F. Toninelli, Disordered pinning models and copolymers: beyond annealed bounds, arXiv:0709.1629v1 (2007), to appear in Ann. Appl. Probab.

[25] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics 79, Springer, New York, 1982.