

# A GIBBSIAN MODEL FOR MESSAGE ROUTEING IN HIGHLY DENSE MULTIHOP NETWORKS

WOLFGANG KÖNIG<sup>1</sup> and ANDRÁS TÓBIÁS<sup>2</sup>

*WIAS Berlin and TU Berlin, and TU Berlin*

(13 August 2018)

**Abstract:** We investigate a probabilistic model for routeing of messages in relay-augmented multihop ad-hoc networks, where each transmitter sends one message to the origin. Given the (random) transmitter locations, we weight the family of random, uniformly distributed message trajectories by an exponential probability weight, favouring trajectories with low interference (measured in terms of signal-to-interference ratio) and trajectory families with little congestion (measured in terms of the number of pairs of hops using the same relay). Under the resulting Gibbs measure, the system targets the best compromise between entropy, interference and congestion for a common welfare, instead of an optimization of the individual trajectories.

In the limit of high spatial density of users, we describe the totality of all the message trajectories in terms of empirical measures. Employing large deviations arguments, we derive a characteristic variational formula for the limiting free energy and analyse the minimizer(s) of the formula, which describe the most likely shapes of the trajectory flow. The empirical measures of the message trajectories well describe the interference, but not the congestion; the latter requires introducing an additional empirical measure. Our results remain valid under replacing the two penalization terms by more general functionals of these two empirical measures.

*MSC 2010.* 60F10, 60G55, 60K30; 82B21; 90B15

*Keywords and phrases.* Gibbs distribution of trajectories, high-density limit, large deviations, empirical measure, number of incoming hops, weak topology, variational formula, minimizer, multihop ad-hoc network, message trajectories, signal-to-interference ratio, congestion

## 1. INTRODUCTION AND MAIN RESULTS

**1.1. Background and main goals.** In random networks, one of the prominent problems is the question how to conduct a message through the system in an optimal way. Optimality is often measured in terms of determining the shortest path from the transmitter to the recipient, or, if interference is considered, determining the path that yields the least interference. If many messages are considered at the same time, an additional aspect of optimality may be to achieve a minimal amount of congestion.

Many investigations concern the question just for one single transmitter/recipient pair, which is a question that every single participant faces. However, a strategy found in such a setting may lead to a selfish routeing, and it is quite likely that the totality of all these routeings for all the individuals is

---

<sup>1</sup>TU Berlin, Straße des 17. Juni 136, 10623 Berlin, and WIAS Berlin, Mohrenstraße 39, 10117 Berlin, [koenig@wias-berlin.de](mailto:koenig@wias-berlin.de)

<sup>2</sup>Berlin Mathematical School, TU Berlin, Straße des 17. Juni 136, 10623 Berlin, [tobias@math.tu-berlin.de](mailto:tobias@math.tu-berlin.de)

by far not optimal for the community of all the users [CCS16, Section 1]. Instead, the entire system may work even better if an optimal *compromise* is realized, by which we mean a joint strategy that leads to an optimum for the entire system, though possibly not for every participant.

In this paper, we present a probabilistic ansatz for describing a jointly optimal routing for an unbounded number of transmitter/recipient pairs, which takes into account the following three crucial properties of the family of message trajectories: entropy (i.e., counting complexity of the number of trajectory families satisfying certain properties), interference and congestion. That is, we consider a situation in which all the messages are directed through the system in a random way, such that each hop prefers a low interference, and such that the total amount of congestion is preferred to be low. Parameters control the strengths of influence of the three effects.

Let us describe our model in words. Let the users be located randomly as the sites of a Poisson point process, which we fix. Each user sends out precisely one message, which arrives at the (unique) base station, which is located at the origin. We consider the entire collection of possible trajectories of the messages through the system. We employ an ad-hoc relaying system with multiple hops, such that all the users act as relays for the handoffs of the messages. The maximal number of hops is  $k_{\max} \in \mathbb{N}$  for each message. Each  $k$ -hop message trajectory (with  $k \in \{1, \dots, k_{\max}\}$  itself random) is random and *a priori* uniformly distributed. The family of all trajectories is *a priori* independent.

Now, the probability distribution of this family is given in terms of a Gibbs ansatz by introducing two exponential weight terms. (That is, we define a quenched measure on trajectories given the locations of the users.) The first one weights the total amount of interference, measured in terms of the signal-to-interference ratio for each hop. The second one weights the total congestion, i.e., the number of times that any two trajectories use the same relay. Under the arising measure, there is a competition between all the three decisive effects of the trajectory family: entropy, interference and congestion. Furthermore, the users form a random environment for the family, which not only determines the starting sites of all the trajectories, but also has a decisive effect on interference and congestion. While the latter has a smoothing effect on the fine details of the spatial distribution of all the trajectories, the effect of the former is not so clear to predict, as the superposition of signals have a very non-local influence.

Our main interest is in understanding the spatial distribution of the totality of all the message trajectories under the Gibbs distribution. The measure under consideration is a highly complex object, as it depends on all the user locations and on many detailed properties and quantities. However, we make a substantial step towards a thorough understanding by deriving an asymptotic formula for the logarithmic behaviour of the normalization constant in the limit of a high spatial density of the users. The limiting situation is then described in terms of a large deviations rate function and a variational formula, whose minimizers describe the optimal joint choices of the trajectories. This formula is deterministic and depends only on general spatial considerations, not on the individual users. These are our main results in this paper.

The main objects in terms of which we achieve this description are the empirical measures of the trajectories of the messages sent out by the users, disintegrated with respect to the lengths and rescaled to finite asymptotic size. These measures turn out to converge in the weak topology in the high-density limit that we consider in this paper. The counting complexity of the statistics of the message trajectories can be written in terms of multinomial expressions and afterwards, in the limit of finer and finer decompositions of the space, approximated in terms of relative entropies, using Stirling's formula. The interference term can also be handled in a standard way [HJKP18], since it is a continuous function of the collection of empirical measures of message trajectories.

However, a key finding of our work is that the congestion term is a highly discontinuous function of these empirical measures. Indeed, one cannot express its limiting behaviour in terms of these measures. Instead, one needs to substantially enlarge the probability space of trajectories and to

introduce another collection of empirical measures, the ones of the locations of users (relays) who receive given numbers of incoming messages (counted with multiplicity if a message trajectory hits the same relay multiple times). The congestion expression then turns out to be a lower semicontinuous function of these empirical measures, and hence the limiting congestion term is still expressible in terms of the weak limits of these measures. Again, using explicit combinatorics and Stirling's formula, we arrive at explicit entropic terms describing the statistics of these measures. These two families of empirical measures together enable us to describe all the properties of the message trajectories that we are interested in. We establish a full large deviation principle for the tuple of all these measures with an explicit rate function and obtain in particular their convergence towards the minimizer(s) of a characteristic variational formula. We also derive their positivity properties and characterize them in terms of Euler-Lagrange equations. Unfortunately, due to the complexity of the congestion term, we are not able to decide about the uniqueness of the minimizer.

Nevertheless, in the special case when congestion is not penalized, the minimizer turns out to be unique, and we obtain an explicit expression that is amenable to further investigation. In certain limiting regimes, we can derive a good understanding of decisive quantities of the system, like the typical number of hops, the typical length of a hop and the typical shape of a trajectory as a function of the distance between the transmission site and the origin. We expect that such properties of the system are similar if congestion is also penalized, as the effect of congestion is *a priori* not spatial, but combinatorial. We decided to analyse such questions in a separate work [KT18], as they have a strongly analytic, rather than probabilistic, nature. The present paper includes a short summary of the results of [KT18].

The main purpose of the present paper is to provide the mathematical framework of large deviations for the quenched trajectory distribution, given the user locations, in the high-density limit. We also provide a discussion of the relevance of the model for telecommunication theory. A connection of our work to traffic theory is outlined in [KT18], which paper also includes numerical examples for the case when congestion is not penalized. Hence, in the present paper, we will formulate the model in a more general, slightly abstract, way in order to bring its mathematical essence to the surface. That is, we consider a random complete geometric graph in a compact subset of  $\mathbb{R}^d$  (where the vertices are the users and the edges are the straight line segments between any two users), and a distribution of trajectories that has an interaction (the interference) with all the locations of the nodes and suppresses local clumping (the congestion).

1.1.1. *The high-density limit for multihop networks.* The quality of service in large multihop ad-hoc networks has received particular interest in the last years. In order to be able to derive a clear picture, one has to make a certain approximation in limiting settings. Two mathematical settings are frequently used: the high-density limit (sometimes called also a hydrodynamic limit or a mean-field limit), where the number of users in a compact fixed area diverges, and the thermodynamic limit, where the area diverges as well, such that the number of users per space unit remains fixed. The former models a situation like at concerts, demonstrations or sports events, while the latter one models large-area networks with moderate user density.

A number of papers on this subject use large deviations methods. This has several advantages. Indeed, the corresponding large deviation principles often come with a law of large numbers for certain empirical measures, and with exponential bounds on the probabilities of deviations from the limit. This suggests that the qualitative behaviour of the network is close to the limit already for relatively moderate values of the diverging parameter. These methods lead in the high-density setting to much more handy formulas (see e.g. [HJKP18, HJP16, HJ17]) than in the thermodynamic limit (see e.g. [HJKP16]). This is why we decided to analyse our Gibbsian model in the high-density setting.

1.1.2. *Related literature.* Apart from the potential value for understanding a new type of message routing models in telecommunication, the present paper provides also some interesting mathematical research on topological fine properties of random paths in random a environment in a high-density setting, a subject that received a lot of interest for various types of such processes over the last decades. We remind the reader on a number of investigations of the intersection properties of random walks and Brownian motions (both self-intersections and mutual intersections) in highly dense settings, see the monograph [Ch09] and some particular investigations in [KM02, KM13]; in all these works, one is interested in large deviations properties of suitable empirical measures, and the lack of continuity of the path properties is the main difficulty. Let us mention that the main aspect of the approach in [KM02] is the same as in the present paper: an approximation of combinatorics in finer and finer decompositions of the space by entropic terms. Another line of research in which similar questions arise is a mean-field variant of a spatial version of Bose-Einstein statistics, like in [AK08], where the statistics of the empirical measures of a diverging number of Brownian bridges with symmetrized initial-terminal condition is analysed in terms of a large deviation principle in the weak topology. While [AK08] works with the same method as we in the present paper (spatial discretization with limiting fineness), [T08] showed that a method based entirely on the notion of entropy is able to derive such results in a more general framework.

1.1.3. *Organization of the remainder of this paper.* We introduce the model and necessary notation in Section 1.2, present our main results in Sections 1.3 (the limiting free energy of the model), 1.4 (the description of the minimizer(s)), 1.5 (the large deviation principle and the convergence of the empirical measures), and 1.6 (results in case congestion is not penalized), we interpret the minimizer(s) and summarize the results of [KT18] about their qualitative properties in Section 1.7, and we discuss and comment our findings in Section 1.8. The remaining sections are devoted to the proofs: in Section 2 we prepare for the proofs by introducing our methods and deriving formulas for the probability terms, in Section 3 we put all this together to a proof of the limiting free energy, the large deviation principle and the convergence of the empirical measures, in Section 4 we analyse the minimizer(s) of the characteristic variational formula, and in Section 5 we extend the proofs to the case when congestion is not penalized.

1.2. **The Gibbsian model.** We introduce now the mathematical setting. For any topological space  $V$ , let  $\mathcal{M}(V)$  denote the set of all finite nonnegative Borel measures on  $V$ , which we equip with the weak topology. We are working in  $\mathbb{R}^d$  with some fixed  $d \in \mathbb{N}$ . Our model is defined as follows. Let  $W \subset \mathbb{R}^d$  be compact, the territory of our telecommunication system, containing the origin  $o$  of  $\mathbb{R}^d$ .

1.2.1. *Users.* Let  $\mu \in \mathcal{M}(W)$  be an absolutely continuous measure on  $W$  with  $\mu(W) > 0$ . Note that we do not require that  $\text{supp}(\mu) = W$ . For  $\lambda > 0$ , we denote by  $X^\lambda$  a Poisson point process in  $W$  with intensity measure  $\lambda\mu$ . The points  $X_i \in X^\lambda$  are interpreted as the *locations of the users* in the system, while the origin  $o$  of  $\mathbb{R}^d$  is the single *base station*. We assume that  $X^\lambda = \{X_i : i \in I^\lambda\}$  with  $I^\lambda = \{1, \dots, N(\lambda)\}$  and  $(N(\lambda))_{\lambda>0}$  a homogeneous Poisson process on  $\mathbb{N}_0$  with intensity  $\mathbb{E}[N(1)] = \mu(W)$ , and  $(X_i)_{i \in \mathbb{N}}$  is an i.i.d. sequence of  $W$ -distributed random variables with distribution  $\mu(\cdot)/\mu(W)$  defined on one probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Since  $\mu$  has a density, all points  $X_i$  are mutually different with probability one. Further,  $X^\lambda$  is increasing in  $\lambda$ , and its *empirical measure*, normalized by  $1/\lambda$ ,

$$L_\lambda = \frac{1}{\lambda} \sum_{i \in I^\lambda} \delta_{X_i}, \quad (1.1)$$

converges towards  $\mu$  almost surely as  $\lambda \rightarrow \infty$ .

These assumptions on the users can be relaxed, see Section 1.8.4.

1.2.2. *Message trajectories.* We now introduce the collection of trajectories sent out from the users to  $o$ , i.e., for uplink communication. For any  $i \in I^\lambda$ , we call a vector of the form

$$S^i = (S_{-1}^i = K_i, S_0^i = X_i, S_1^i \in X^\lambda, \dots, S_{K_i-1}^i \in X^\lambda, S_{K_i}^i = o) \in \bigcup_{k \in \mathbb{N}} \left( \{k\} \times \{X_i\} \times W^{k-1} \times \{o\} \right), \quad (1.2)$$

a *message trajectory* from  $X_i$  to  $o$  with  $K_i$  hops. That is,  $S^i$  starts from  $X_i$  and ends in  $o$  after  $K_i$  hops from user to user in  $X^\lambda$ . Hence, each user sends exactly one message to  $o$ , and each user has the function of a relay. We fix a number  $k_{\max} \in \mathbb{N}$  and write  $\mathcal{S}_{k_{\max}}^i(X^\lambda)$  for the set of all possible realizations of the random variable  $S^i$  with  $K_i \leq k_{\max}$ , i.e., with no more than  $k_{\max}$  hops. Hence, elements  $s^i = (s_{-1}^i, s_0^i, s_1^i, \dots, s_{s_{-1}^i-1}^i, s_{s_{-1}^i}^i)$  of  $\mathcal{S}_{k_{\max}}^i(X^\lambda)$  satisfy  $s_{-1}^i \in \{1, \dots, k_{\max}\}$ ,  $s_0^i = X_i$  and  $s_{s_{-1}^i}^i = o$ . We write  $\mathcal{S}_{k_{\max}}(X^\lambda) = \times_{i \in I^\lambda} \mathcal{S}_{k_{\max}}^i(X^\lambda)$  for the set of all possible realizations of the families  $S = (S^i)_{i \in I^\lambda}$ . We use the notation  $[k] = \{1, \dots, k\}$  for  $k \in \mathbb{N}$ . The assumption that we choose a finite upper bound  $k_{\max}$  on the number of hops will be discussed in Section 1.8.7.

Given  $i \in I^\lambda$ , we consider each trajectory  $S^i$  in (1.2) as an  $\mathcal{S}_{k_{\max}}^i(X^\lambda)$ -valued random variable. Its *a priori* measure is defined by the formula

$$s^i \mapsto \frac{1}{N(\lambda)^{s_{-1}^i-1}}, \quad s^i \in \mathcal{S}_{k_{\max}}^i(X^\lambda). \quad (1.3)$$

That is, its restriction to  $\{s^i \in \mathcal{S}_{k_{\max}}^i(X^\lambda) : s_{-1}^i = k\}$  is the uniform distribution on the set of all  $k$ -hop trajectories from  $X_i$  to  $o$  for any  $k \in [k_{\max}]$ , and its total mass is equal to  $k_{\max}$ . Recall that it fixes the starting point  $X_i$  and the terminal point  $o$ .

Under our joint *a priori* measure, all the trajectories are independent; indeed, it gives the value

$$s = (s^i)_{i \in I^\lambda} \mapsto \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{s_{-1}^i-1}} \quad (1.4)$$

to the configuration  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$ . Thus, it gives a total mass of  $k_{\max}^{N(\lambda)}$  to  $\mathcal{S}_{k_{\max}}(X^\lambda)$ .

1.2.3. *Gibbsian trajectory distribution.* In this section, we define the central object of this study: a Gibbs distribution on the set  $\mathcal{S}_{k_{\max}}(X^\lambda)$  of collections of trajectories. After providing the abstract definitions, in Section 1.2.4 we sketch the key example that is relevant for application in telecommunication. The general conditions on the ingredients of the Gibbs distribution in this section arise naturally from the properties of this example.

We introduce the following notation. For  $k \in \mathbb{N}$ , elements of the product space  $W^k = W^{\{0,1,\dots,k-1\}}$  are denoted as  $(x_0, \dots, x_{k-1})$ . For  $l = 0, \dots, k-1$ , the  $l$ -th marginal of a measure  $\nu_k \in \mathcal{M}(W^k)$  is denoted by  $\pi_l \nu_k \in \mathcal{M}(W)$ , i.e.,  $\pi_l \nu_k(A) = \nu_k(W^{\{0,\dots,l-1\}} \times A \times W^{\{l+1,\dots,k-1\}})$  for any Borel set  $A \subseteq W$ .

For fixed  $k \in [k_{\max}]$  and for a collection of trajectories  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$ , we define

$$R_{\lambda,k}(s) = \frac{1}{\lambda} \sum_{i \in I^\lambda : s_{-1}^i = k} \delta_{(s_0^i, \dots, s_{k-1}^i)}, \quad (1.5)$$

the empirical measures of all the  $k$ -hop trajectories, which is an element of  $\mathcal{M}(W^k)$ . By the assumption that each user sends out exactly one message, we have

$$\sum_{k=1}^{k_{\max}} \pi_0 R_{\lambda,k}(s) = L_\lambda. \quad (1.6)$$

For  $k \in [k_{\max}]$ , we choose a continuous function  $f_k: \mathcal{M}(W) \times W^k \rightarrow \mathbb{R}$  that is bounded from below. Using (1.6), we put

$$\mathfrak{S}(s) = \lambda \sum_{k=1}^{k_{\max}} \left\langle R_{\lambda,k}(s)(\cdot), f_k(L_\lambda, \cdot) \right\rangle = \sum_{k=1}^{k_{\max}} \sum_{i \in I^\lambda: s_{-1}^i = k} f_k(L_\lambda, s_0^i, \dots, s_{k-1}^i), \quad (1.7)$$

where we write  $\langle \nu, f \rangle$  for the integral of the function  $f$  against the measure  $\nu$ . Moreover, we define

$$m_i(s) = \sum_{j \in I^\lambda} \sum_{l=1}^{s_{-1}^j - 1} \mathbb{1}\{s_l^j = s_0^i\}, \quad i \in I^\lambda, \quad (1.8)$$

as the number of incoming hops into the user (relay)  $s_0^i = X_i$  of any of the trajectories.

We pick a function  $\eta: \mathbb{N}_0 \rightarrow \mathbb{R}$ , bounded from below such that  $\lim_{m \rightarrow \infty} \eta(m)/m = \infty$ . Then we put

$$\mathfrak{M}(s) = \sum_{i \in I^\lambda} \eta(m_i(s)). \quad (1.9)$$

Now, we define

$$P_{\lambda, X^\lambda}^{\gamma, \beta}(s) := \frac{1}{Z_{\lambda, \beta}^{\gamma, \beta}(X^\lambda)} \left( \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{s_{-1}^i - 1}} \right) \exp \left\{ -\gamma \mathfrak{S}(s) - \beta \mathfrak{M}(s) \right\}, \quad (1.10)$$

where  $\gamma > 0$  and  $\beta > 0$  are parameters. This is the Gibbs distribution with independent reference measure given in (1.4), subject to two exponential weights with the terms (1.7) and (1.9). Here

$$Z_{\lambda, \beta}^{\gamma, \beta}(X^\lambda) = \sum_{r \in \mathcal{S}_{k_{\max}}(X^\lambda)} \left( \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{r_{-1}^i - 1}} \right) \exp \left\{ -\gamma \mathfrak{S}(r) - \beta \mathfrak{M}(r) \right\} \quad (1.11)$$

is the normalizing constant, which we will refer to as *partition function*. Note that  $P_{\lambda, X^\lambda}^{\gamma, \beta}(\cdot)$  is random conditional on  $X^\lambda$ , and it is a probability measure on  $\mathcal{S}_{k_{\max}}(X^\lambda)$ . In the jargon of statistical mechanics, it is a *quenched* measure, which we will consider almost surely with respect to the process  $(X^\lambda)_{\lambda > 0}$ . In the *annealed* setting, one would average out over  $(X^\lambda)_{\lambda > 0}$ , see Section 1.8.5.

**1.2.4. The key example: penalization of interference and congestion.** In this section, we sketch the most important example for the exponents  $\mathfrak{S}$  and  $\mathfrak{M}$  in (1.10), where  $\mathfrak{S}$  registers interference and  $\mathfrak{M}$  expresses congestion in a telecommunication network. Analysing the qualitative properties of the network with these choices of  $\mathfrak{M}$  and  $\mathfrak{S}$  is the main topic of our accompanying paper [KT18], see Section 1.7.2.

Now we introduce interference. We choose a *path-loss function*, which describes the propagation of signal strength over distance. This is a monotone decreasing, continuous function  $\ell: [0, \infty) \rightarrow (0, \infty)$ . An example used in practice is  $\ell(r) = \min\{1, r^{-\alpha}\}$ , for some  $\alpha > 0$ , see e.g. [GT08, Section II.], for further examples see [BB09, Section 2.3.1]. The *signal-to-interference ratio (SIR)* of a transmission from  $X_i \in X^\lambda$  to  $x \in W$  in the presence of the users in  $X^\lambda$  is given as

$$\text{SIR}(X_i, x, L_\lambda) = \frac{\ell(|X_i - x|)}{\frac{1}{\lambda} \sum_{j \in I^\lambda} \ell(|X_j - x|)}. \quad (1.12)$$

We call the denominator of the r.h.s of (1.12) the *interference* at  $x$ . The definition (1.12) comes from [HJKP18]. It is adapted to the high-density setting, and it differs from the usual definition [GK00] of SIR in the following way. The sum in the interference in (1.12) is multiplied by  $1/\lambda$ , and it contains also the term  $\ell(|X_i - x|)$ . For a justification of these differences, see [KT18, Section 6.1.1].

Now, given a trajectory configuration  $s = (s^i)_{i \in I^\lambda} \in \mathcal{S}_{k_{\max}}(X^\lambda)$ , we put

$$\mathfrak{G}(s) = \sum_{i \in I^\lambda} \sum_{l=1}^{s_{l-1}^i} \text{SIR}(s_{l-1}^i, s_l^i, L_\lambda)^{-1} = \lambda \sum_{k=1}^{k_{\max}} \int_{W^k} R_{\lambda,k}(s)(dx_0, \dots, dx_{k-1}) \sum_{l=1}^k \frac{\int_W \ell(|y - x_l|) L_\lambda(dy)}{\ell(|x_{l-1} - x_l|)}, \quad (1.13)$$

where for  $k \in [k_{\max}]$ , we write  $x_k = o$ . Then (1.13) is a special case of (1.7) with

$$f_k(\nu, x_0, \dots, x_{k-1}) = \sum_{l=1}^k \frac{\int_W \ell(|y - x_l|) \nu(dy)}{\ell(|x_{l-1} - x_l|)}, \quad x_k = o, k \in [k_{\max}]. \quad (1.14)$$

Next, we introduce congestion. We define  $\eta(m) = m(m-1)$ , and, as in (1.9), we put

$$\mathfrak{M}(s) = \sum_{i \in I^\lambda} \eta(m_i(s)) = \sum_{i \in I^\lambda} m_i(s)(m_i(s) - 1), \quad s \in \mathcal{S}_{k_{\max}}(X^\lambda). \quad (1.15)$$

Note that  $\eta(m_i(s)) = m_i(s)(m_i(s) - 1)$  is the number of ordered pairs of hops arriving at the relay  $X_i = s_0^i$ . We will explain and motivate these choices in Section 1.8.1.

In the downlink scenario, instead of users sending messages to the base station, the base station sends exactly one message to each of the users, using the same relaying rules. One can define a Gibbsian model analogously, now for trajectories from  $o$  to  $X_i$  instead of the other way around. For this, the interference term and the congestion term have to be redefined in an obvious way. We are certain that analogues of all our results are true and can be proved in the same way, hence we abstained from spelling them out.

For possible extensions of this model involving time dependence or users transmitting multiple messages, see Section 1.8.6.

**1.3. The limiting free energy.** The main goal of this paper is the description of this model in the limit  $\lambda \rightarrow \infty$  in the quenched setting. Our first result describes the limiting *free energy*, i.e., the exponential behaviour of the partition function  $Z_\lambda^{\gamma, \beta}(X^\lambda)$ . One expects that this is entirely governed by the large deviations behaviour of the empirical measures  $((R_{\lambda,k})(S)_{k \in [k_{\max}]})_{\lambda > 0}$ . This expectation is supported by the fact that, for  $i \in I^\lambda$  and  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$ , we can express  $m_i(s)$  defined in (1.8) in terms of  $(R_{\lambda,k}(s))_{k \in [k_{\max}]}$  as follows

$$m_i(s) = \lambda \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k_{\max}} \pi_l R_{\lambda,k}(s)(\{s_0^i\}). \quad (1.16)$$

Surprisingly, it turns out that the limiting free energy cannot be described entirely in terms of these measures. The reason is that the function in (1.16) that maps them onto  $m_i(s)$  is highly discontinuous in the limit  $\lambda \rightarrow \infty$ ; even a proper formulation of such continuity would be awkward since both  $i$  and  $s$  depend on  $\lambda$ .

One therefore needs to substantially extend the probability space and to choose an additional family of empirical measures such that the congestion term  $\mathfrak{M}(s)$  can be written as a (lower semi-)continuous function of these measures in the limit  $\lambda \rightarrow \infty$ . A natural choice of such a family is the one of the measures

$$P_{\lambda,m}(s) = \frac{1}{\lambda} \sum_{i \in I^\lambda: m_i(s)=m} \delta_{s_0^i}, \quad m \in \mathbb{N}_0. \quad (1.17)$$

Then for  $m \in \mathbb{N}_0$ ,  $P_{\lambda,m}(s) \in \mathcal{M}(W)$  is the empirical measure of the users  $s_0^i$  whose number of incoming hops equals  $m$ . For any  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$  the following hold

$$(i) \quad \sum_{k=1}^{k_{\max}} \pi_0 R_{\lambda,k}(s) = L_\lambda, \quad (ii) \quad \sum_{m=0}^{\infty} P_{\lambda,m}(s) = L_\lambda, \quad (iii) \quad \sum_{m=0}^{\infty} m P_{\lambda,m}(s) = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l R_{\lambda,k}(s). \quad (1.18)$$

Condition (i) expresses our assumption that each user transmits precisely one message, (ii) says that each user serves as a relay for precisely  $m$  message trajectories for precisely one  $m \in \mathbb{N}_0$ , and (iii) says that the relays can be calculated in two ways: according to the number of incoming hops and according to the index of the hop of a trajectory that uses it. Moreover, we can write (1.9) in terms of  $(P_{\lambda,m}(s))_{m \in \mathbb{N}_0}$  as follows

$$\mathfrak{M}(s) = \sum_{i \in I^\lambda} \eta(m_i(s)) = \lambda \sum_{m=0}^{\infty} \eta(m) P_{\lambda,m}(s)(W).$$

We note that the function  $\mathcal{M}(W)^{\mathbb{N}_0} \rightarrow \mathbb{R} \cup \{\infty\}$ ,  $(\xi_m)_{m \in \mathbb{N}_0} \mapsto \sum_{m=0}^{\infty} \eta(m) \xi_m(W)$  is lower semicontinuous, and even continuous on  $\{(\xi_m)_{m \in \mathbb{N}_0} : \sum_{m=0}^{\infty} \eta(m) \xi_m(W) \leq \alpha\}$  for any  $\alpha \in \mathbb{R}$ .

The limiting free energy will be described in terms of the following kind of families of measures, and it will turn out that all subsequential limits of the families  $((R_{\lambda,k}(S))_{k=1}^{k_{\max}}, (P_{\lambda,m}(S))_{m=0}^{\infty})$  in the quenched limit  $\lambda \rightarrow \infty$  are of this kind.

**Definition 1.1.** *An admissible trajectory setting is a collection of measures  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$  with  $\nu_k \in \mathcal{M}(W^k)$  for all  $k$  and  $\mu_m \in \mathcal{M}(W)$  for all  $m$ , satisfying the following properties.*

$$(i) \quad \sum_{k=1}^{k_{\max}} \pi_0 \nu_k = \mu, \quad (ii) \quad \sum_{m=0}^{\infty} \mu_m = \mu, \quad (iii) \quad M := \sum_{m=0}^{\infty} m \mu_m = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k. \quad (1.19)$$

The measure  $\nu_k$  is the measure of the  $k$ -hop trajectories and  $\mu_m$  the measure of the users that receive precisely  $m$  incoming hops; note that there is no reason that they be normalized (like for  $\mu$ ). Observe that in (1.18),  $L_\lambda$ ,  $R_{\lambda,k}(s)$  and  $P_{\lambda,m}(s)$  play the role of  $\mu$ ,  $\nu_k$  and  $\mu_m$ , respectively. In particular, after replacing  $\mu$  by  $L_\lambda$ ,  $((R_{\lambda,k}(s))_{k \in [k_{\max}]}, (P_{\lambda,m}(s))_{m \in \mathbb{N}_0})$  satisfies the definition of an admissible trajectory setting. See Section 1.8 for more explanations and interpretations, moreover for a modified version of our model where the assumption (i) is relaxed. By

$$\mathcal{H}_V(\nu | \tilde{\nu}) = \begin{cases} \int_V d\nu \log \frac{d\nu}{d\tilde{\nu}} - \nu(V) + \tilde{\nu}(V), & \text{if the density } \frac{d\nu}{d\tilde{\nu}} \text{ exists,} \\ +\infty & \text{otherwise,} \end{cases} \quad (1.20)$$

we denote the relative entropy [GZ93, Section 2.3] of a Borel measure  $\nu$  with respect to another Borel measure  $\tilde{\nu}$  on a measurable subset  $V$  of  $\mathbb{R}^n$ ,  $n \in \mathbb{N}$ .

For an admissible trajectory setting  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$  we define

$$S(\Psi) = \sum_{k=1}^{k_{\max}} \int_{W^k} d\nu_k \tilde{f}_k, \quad \text{where } \tilde{f}_k(x_0, \dots, x_{k-1}) = f_k(\mu, x_0, \dots, x_{k-1}), \quad (1.21)$$

and

$$M(\Psi) = \sum_{m=0}^{\infty} \eta(m) \mu_m(W) \quad (1.22)$$

and

$$I(\Psi) = \sum_{k=1}^{k_{\max}} \mathcal{H}_{W^k}(\nu_k | \mu \otimes M^{\otimes(k-1)}) + \sum_{m=0}^{\infty} \mathcal{H}_W(\mu_m | \mu c_m) + \mu(W) \left( 1 - \sum_{k=1}^{k_{\max}} M(W)^{k-1} \right) - \frac{1}{e}, \quad (1.23)$$



where we recall  $M = \sum_{m \in \mathbb{N}_0} m \mu_m$  from Definition 1.1(iii),  $\eta$  defined before (1.9), and  $c_m = \exp(-1/(e\mu(W)))(e\mu(W))^{-m}/m!$  are the weights of the Poisson distribution with parameter  $1/(e\mu(W))$ . In Section 1.8.2, we argue that  $I(\Psi)$  is well-defined as an element of  $(-\infty, \infty]$  and  $\Psi \mapsto I(\Psi)$  is a lower semicontinuous function that is bounded from below, and we provide an interpretation for  $I(\cdot)$ . A tedious but elementary calculation shows that  $I$  is convex. In Section 1.5,  $I$  will turn out to govern the large deviations of the trajectory configuration. The terms  $S(\cdot)$  and  $M(\cdot)$  are analogues of the penalty terms  $\mathfrak{S}(\cdot)$  and  $\mathfrak{M}(\cdot)$  in the high-density setting, respectively.

We fix all the parameters  $W, \mu, k_{\max}, f_k, \eta, \gamma$  and  $\beta$  of the model. Our first main result is the following.

**Theorem 1.2** (Quenched exponential rate of the partition function). *For  $\mathbb{P}$ -almost all  $\omega \in \Omega$ ,*

$$\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma, \beta}(X^{\lambda}(\omega)) = - \inf_{\Psi \text{ admissible trajectory setting}} (I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)). \quad (1.24)$$

See Section 1.8 for a discussion and Section 3.4 for the proof.

**1.4. Description of the minimizers.** From the variational formula in (1.24), descriptive information about the typical behaviour of the network can be deduced, especially in the case of the specific choice of  $\mathfrak{M}$  and  $\mathfrak{S}$  from Section 1.2.4, see Sections 1.5, 1.7 and 1.8. Hence, it is important to derive the Euler-Lagrange equations and to characterize the minimizers in most explicit terms. Our main results in this respect are the following. Note that the case  $k_{\max} = 1$  is trivial.

**Proposition 1.3** (Characterization of the minimizer(s)). *Let  $k_{\max} > 1$ . The infimum in the variational formula in (1.24) is attained, and every minimizer  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$  has the following form.*

$$\nu_k(dx_0, \dots, dx_{k-1}) = \mu(dx_0) A(x_0) \prod_{l=1}^{k-1} (C(x_l) M(dx_l)) e^{-\gamma \tilde{f}_k(x_0, \dots, x_{k-1})}, \quad k \in [k_{\max}], \quad (1.25)$$

$$\mu_m(dx) = \mu(dx) B(x) \frac{(C(x)\mu(W))^{-m}}{m!} e^{-\beta \eta(m)}, \quad m \in \mathbb{N}_0, \quad (1.26)$$

where  $A, B, C: W \rightarrow [0, \infty)$  are functions such that the conditions in (1.19) are satisfied.

The proof of Proposition 1.3 is in Section 4.

While explicit formulas for the functions  $A$  and  $B$  can, given the function  $C$ , easily be derived from (i) and (ii) in (1.19) (see (4.11)), the condition for  $C$  coming from (iii) is deeply involved and cannot be easily solved intrinsically; see (4.13) – (4.15). We have no argument for its existence to offer other than via proving the existence of a minimizer  $\Psi$  and deriving the Euler-Lagrange equations. By convexity of  $I, S$  and  $M$ , every solution  $\Psi$  to these equations is a minimizer. Even the uniqueness of  $C$  is unknown to us. We will interpret the equations (4.9)–(4.10) in Section 1.7.1. The equations (1.25)–(1.26) become more explicit in case  $\beta = 0$ , and in this case, uniqueness of the minimizer holds, see Section 1.6.

In case  $k_{\max} = 1$ , the only admissible trajectory setting is  $\Psi = (\nu_1, (\mu_m)_{m \in \mathbb{N}_0})$  with  $\mu_0 = \nu_1 = \mu$  and  $\mu_m = 0$  otherwise, therefore this  $\Psi$  minimizes (1.24).

**1.5. Large deviations for the empirical trajectory measure.** Actually, the minimizers of the variational formula in (1.24) receive a rigorous interpretation in terms of important objects that describe the network. As we have already mentioned, the family of empirical measures

$$\Psi_{\lambda}(s) = ((R_{\lambda, k}(s))_{k \in [k_{\max}]}, (P_{\lambda, m}(s))_{m \in \mathbb{N}_0}) \quad (1.27)$$

satisfies the definition of an admissible trajectory setting, apart from the fact that in Definition 1.1,  $\mu$  has to be replaced by  $L_\lambda$  everywhere, where we recall that  $L_\lambda$  converges to  $\mu$  almost surely as  $\lambda \rightarrow \infty$ . According to our remarks after Definition 1.1,  $R_{\lambda,k}(s)$  and  $P_{\lambda,m}(s)$  play the roles of  $\nu_k$  and  $\mu_m$ , respectively, in an admissible trajectory setting, which explains this term. Furthermore, for  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$ , we can express the term  $\mathfrak{M}$  as

$$\mathfrak{M}(s) = \lambda M(\Psi_\lambda(s)).$$

Moreover, for the continuous penalization term we have

$$\mathfrak{S}(s) \approx \lambda S(\Psi_\lambda(s)), \quad (1.28)$$

where we typically do not have an identity, because  $\mathfrak{S}(s) = \lambda \sum_{k=1}^{k_{\max}} \int_{W^k} dR_{\lambda,k}(s) f_k(L_\lambda, \cdot)$ , which is usually not equal to  $\lambda S(\Psi_\lambda(s)) = \lambda \sum_{k=1}^{k_{\max}} \int_{W^k} dR_{\lambda,k}(s) f_k(\mu, \cdot)$ . However, since  $L_\lambda \Rightarrow \mu$  almost surely, this difference vanishes in the limit, see Proposition 3.2.

We consider now the distribution of  $\Psi_\lambda(S)$  with  $S$  distributed under the product reference measure introduced in (1.4), normalized to a probability measure,  $P_{\lambda, X^\lambda}^{0,0}$ ; note that the normalization  $Z_\lambda^{0,0}(X^\lambda)$  is equal to  $k_{\max}^{N(\lambda)}$ . Our next main result, Theorem 1.4, is a large deviation principle (LDP; see (1.30)–(1.31)) and the convergence towards the minimizers of the variational formula.

**Theorem 1.4** (LDP and convergence for the empirical measures). *The following statements hold almost surely with respect to  $(X^\lambda)_{\lambda>0}$ .*

(i) *The distribution of  $\Psi_\lambda(S)$  under  $P_{\lambda, X^\lambda}^{0,0}$  satisfies an LDP as  $\lambda \rightarrow \infty$  with scale  $\lambda$  on the set*

$$\mathcal{A} = \left( \prod_{k=1}^{k_{\max}} \mathcal{M}(W^k) \right) \times \mathcal{M}(W)^{\mathbb{N}_0} \quad (1.29)$$

*with rate function given by  $\mathcal{A} \ni \Psi \mapsto I(\Psi) + \mu(W) \log k_{\max}$ , which we define as  $\infty$  if  $\Psi$  is not an admissible trajectory setting.*

(ii) *For any  $\gamma, \beta \in (0, \infty)$ , the distribution of  $\Psi_\lambda(S)$  under  $P_{\lambda, X^\lambda}^{\gamma, \beta}$  converges towards the set of minimizers of the variational formula in (1.24).*

For the reader's convenience, we recall that the LDP states that the rate function  $I + \mu(W) \log k_{\max}$  is lower semicontinuous and

$$\limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) \in F) \leq - \inf_F (I + \mu(W) \log k_{\max}), \quad (1.30)$$

$$\liminf_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) \in G) \geq - \inf_G (I + \mu(W) \log k_{\max}), \quad (1.31)$$

for any closed set  $F$  and any open set  $G$  in  $\mathcal{A}$ . See [DZ98] for more on large deviations theory. On  $\mathcal{A}$ , we consider the product topology that is induced by weak convergence in each factor; this is equal to coordinatewise weak convergence, see Section 3.3 for more details. Convergence of a distribution towards a set is defined by requiring that for any neighbourhood of the set, the probability of not being in the neighbourhood vanishes.

The proof of Theorem 1.4(i) is carried out in Section 3.5, using Lemma 4.1. Assertion (ii) is a simple consequence of (i), since the functionals  $S$  and  $M$  are bounded and continuous on the set  $B_C = \{\Psi \in \mathcal{A} : M(\Psi) \leq C\}$  for any  $C$ , and  $B_C$  is compact in  $\mathcal{A}$  (see Lemma 4.1). Denoting the level sets of the rate function  $I + \mu \log k_{\max}$  by  $\Phi_\alpha = \{\Psi \in \mathcal{A} : I(\Psi) + \mu(W) \log k_{\max} \leq \alpha\}$  for  $\alpha \in \mathbb{R}$ , we see that  $\Phi_\alpha \cap B_C$  is compact for any  $\alpha$  and  $C$ . Thus, Varadhan's lemma can be applied to prove Assertion (ii).

**1.6. Dropping the congestion term.** Proposition 1.3 yields a rather implicit description of the minimizers of (1.24) in the case  $\beta, \gamma > 0$ . The cardinality of the set of minimizers is also unclear. However, in the special case  $\beta = 0$ , where the congestion functional  $\mathfrak{M}$  (1.9) is absent, the situation is much better. Indeed, it turns out that the minimizer is unique and is explicitly given in terms of the parameters of the model. Especially for the specific choice of Section 1.2.4 where  $\mathfrak{S}$  penalizes interference (1.13), on base of this knowledge, we are able in [KT18] to derive a number of relevant qualitative properties of the trajectories, see Section 1.7.2 for a summary.

In what follows, we call  $\Sigma = (\nu_k)_{k \in [k_{\max}]}$  with  $\nu_k \in \mathcal{M}(W^k)$  for all  $k \in [k_{\max}]$  an *asymptotic routeing strategy* if we have  $\sum_{k=1}^{k_{\max}} \pi_0 \nu_k = \mu$ . In (1.19) we see that the first coordinate,  $\Sigma$ , of an admissible trajectory setting  $\Psi$  is an asymptotic routeing strategy, and in (1.21) we see that  $S(\Psi)$  depends only on  $\Sigma$ . We will therefore write  $S(\Sigma)$  for  $S(\Psi)$ . Further, we write  $M = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k$ , in accordance with (1.19) but with no regard to the measures  $(\mu_m)_{m \in \mathbb{N}_0}$ . We define an entropic term  $J$  for asymptotic routeing strategies as follows.

$$J(\Sigma) = \sum_{k=1}^{k_{\max}} \mathcal{H}_{W^k}(\nu_k \mid \mu^{\otimes k}) - \sum_{k=2}^{k_{\max}} \mu(W)^k + M(W) \log \mu(W). \quad (1.32)$$

Similarly to  $I$  in (1.23),  $J$  describes counting complexity in the high-density limit, but without reference to the measures  $(\mu_m)_m$ .

The following proposition summarizes the analogues of Theorem 1.2, Proposition 1.3 and Theorem 1.4 in case  $\beta = 0$ , after dropping all the measures  $\mu_m$ .

**Proposition 1.5.** *The following statements hold almost surely with respect to  $(X^\lambda)_{\lambda > 0}$ .*

(1) *The distribution of*

$$\Sigma_\lambda(S) = (R_{\lambda,k}(S))_{k \in [k_{\max}]} \quad (1.33)$$

*under  $P_{\lambda, X^\lambda}^{0,0}$  satisfies an LDP as  $\lambda \rightarrow \infty$  with scale  $\lambda$  on the set  $\mathcal{A}_0 = \prod_{k=1}^{k_{\max}} \mathcal{M}(W^k)$  with rate function given by  $\mathcal{A}_0 \ni \Psi \mapsto J(\Sigma) + \mu(W) \log k_{\max}$ , which we define as  $\infty$  if  $\Sigma$  is not an asymptotic routeing strategy. Further, the rate function has compact level sets.*

(2) *For any  $\gamma \in (0, \infty)$ ,*

$$\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma,0}(X^\lambda) = - \inf_{\Sigma \text{ asymptotic routeing strategy}} (J(\Sigma) + \gamma S(\Sigma)) \quad (1.34)$$

(3) *Let  $\gamma > 0$  and  $k_{\max} > 1$ . The variational formula on the r.h.s. of (1.34) exhibits a unique minimizer  $\Sigma = (\nu_k)_{k \in [k_{\max}]}$  given as*

$$\nu_k(dx_0, \dots, dx_{k-1}) = \mu(dx_0) A(x_0) \prod_{l=1}^{k-1} \frac{\mu(dx_l)}{\mu(W)} e^{-\gamma \tilde{f}_k(x_0, \dots, x_{k-1})}, \quad k \in [k_{\max}], \quad (1.35)$$

where

$$\frac{1}{A(x_0)} = \sum_{k=1}^{k_{\max}} \int_{W^{k-1}} \prod_{l=1}^{k-1} \frac{\mu(dx_l)}{\mu(W)} e^{-\gamma \tilde{f}_k(x_0, \dots, x_{k-1})}, \quad x_0 \in W. \quad (1.36)$$

(4) *For any  $\gamma \in (0, \infty)$ , the distribution of  $\Sigma_\lambda(S)$  under  $P_{\lambda, X^\lambda}^{\gamma,0}$  converges to the minimizer of the variational formula in (1.34).*

Proposition 1.5 is proved in Section 5. An interpretation of the equations (1.35)–(1.36) can be found in Section 1.7.1.

Let us explain in what way Proposition 1.5 is the special case of the aforementioned results for  $\beta = 0$  and in what way it differs. It is true that the LDP in Assertion (1) directly follows from the

LDP in Theorem 1.4(i) via the contraction principle [DZ98, Theorem 4.2.1] for the projection map  $(\Sigma, (\mu_m)_m) \mapsto \Sigma$ , however, with rate function given by

$$J(\Sigma) = \inf_{(\mu_m)_{m \in \mathbb{N}_0} : \Psi = (\Sigma, (\mu_m)_{m \in \mathbb{N}_0}) \text{ admissible trajectory setting}} I(\Psi). \quad (1.37)$$

It is an elementary but tedious task to identify this as in (1.32) by identifying

$$\mu_m(dx) = \mu(dx) \frac{\left(\frac{M(dx)}{\mu(dx)}\right)^m}{m!} e^{-M(dx)/\mu(dx)}, \quad m \in \mathbb{N}_0, \quad (1.38)$$

as the unique minimizer on the right-hand side of (1.37), given  $M = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k$ . However, we chose an alternative route for proving the LDP with explicit identification of  $J$ , which is a variant of the proof of Theorem 1.4(i). From (1.37) it is clear that the variational formula in (1.34) is indeed the special case of (1.24) for  $\beta = 0$ , i.e.,

$$\inf_{\Sigma \text{ asymptotic routing strategy}} (J(\Sigma) + \gamma S(\Sigma)) = \inf_{\Psi \text{ admissible trajectory setting}} (I(\Psi) + \gamma S(\Psi)). \quad (1.39)$$

Note also that the minimizer  $\Psi$  is unique. This raises the additional question whether or not the measures  $(P_{\lambda, m}(S))_{m \in \mathbb{N}_0}$  converge to the minimizer  $(\mu_m)_{m \in \mathbb{N}_0}$  in (1.38) under  $P_{\lambda, X^\lambda}^{\gamma, 0}$  for  $M$  corresponding to the minimal  $\nu_k$ 's of (1.35). Since the congestion term, which gave rise to a strong compactness argument, is now absent, this question cannot immediately be decided using large deviations arguments, but we nevertheless believe it is true. Moreover, this compactness property was also used in the proof of Theorem 1.2, which is another reason that we had to redo the proofs of Proposition 1.5(2) and (3), given our proof of Assertion (1).

## 1.7. Interpretation and qualitative properties of the minimizer(s).

**1.7.1. Interpretation of the minimizer(s).** In case  $\beta, \gamma > 0$ , Proposition 1.3 tells us quite some information about the limiting trajectory distribution and the limiting spatial distribution of users with a given number of incoming hops under the measure  $P_{\lambda, X^\lambda}^{\gamma, \beta}$ . Indeed, both have densities that are  $\mu^{\otimes k}$ -almost everywhere positive. It is remarkable that the  $k$ -hop trajectories follow a distribution that comes from choosing independently all the  $k$  sites with measures that do not depend on  $k$  (the starting point according to  $A(x) \mu(dx)$  and all the other  $k - 1$  sites each according to  $C(x) M(dx)$ ), exponentially weighted with the term  $\gamma \tilde{f}_k$ . Furthermore, all the measures of the users receiving  $m$  incoming hops superpose each other on the full set  $\text{supp}(\mu)$ , and at each space point  $x$ , this number  $m$  is distributed according to some Poisson distribution, exponentially weighted with the term  $\beta \eta(m)$ .

As for the case  $\beta = 0, \gamma > 0$ , we have a unique minimizer, which exhibits all the properties enumerated for  $\beta, \gamma > 0$ . In the  $k$ -hop trajectories, the starting point is chosen according to  $A(x) \mu(dx)$  and all the other  $k - 1$  sites according to the measure  $\mu(dx)/\mu(W)$ , weighted with  $\gamma \tilde{f}_k$ . Moreover, the number of incoming hops at a given relay at the site  $x \in W$  is Poisson distributed with parameter equal to  $M(dx)/\mu(dx)$ .

**1.7.2. Qualitative properties of the minimizer.** In our accompanying paper [KT18], we analyse the joint routing behaviour of our Gibbsian system in the high-density limit. We investigate qualitative properties of the minimizer of the variational formula in (1.24), such as the typical number of hops, the typical length of a hop and the typical shape of a trajectory, in case  $\gamma > \beta = 0$  and the interference term is chosen as in Section 1.2.4. In this case, the minimizer is unique and has the form (cf. (1.35))

$$\nu_k(dx_0, \dots, dx_{k-1}) = \mu(dx_0) A(x_0) \prod_{l=1}^{k-1} \frac{\mu(dx_l)}{\mu(W)} e^{-\gamma \sum_{l=1}^k \frac{\int_W \ell(|y-x_l|) \mu(dy)}{\ell(|x_{l-1}-x_l|)}}, \quad k \in [k_{\max}], \quad (1.40)$$

where

$$\frac{1}{A(x_0)} = \sum_{k=1}^{k_{\max}} \int_{W^{k-1}} \prod_{l=1}^{k-1} \frac{\mu(dx_l)}{\mu(W)} e^{-\gamma \sum_{l=1}^k \frac{\int_W \ell(|y-x_l|) \mu(dy)}{\ell(|x_{l-1}-x_l|)}}, \quad x_0 \in W. \quad (1.41)$$

Further, the empirical measures of trajectories  $\Sigma_\lambda(S) = (R_{\lambda,k}(S))_{k \in [k_{\max}]}$  converge to this minimizer under the Gibbs distribution  $P_{\lambda, X^\lambda}^{\gamma, 0}$ , almost surely as  $\lambda \rightarrow \infty$ . Thus, qualitative information about  $\Sigma = (\nu_k)_{k \in [k_{\max}]}$  yields information about these empirical measures for large  $\lambda$ .

In order to obtain clear pictures and neat results, one needs to analyse the minimizer in extreme regimes. That is, one has to carry out another limit after the high-density limit has been taken. We consider the following regimes: (1) large communication areas, coupled with large transmitter–receiver distances and large numbers of hops, (2) strong penalization of interference, (3) high local density of the intensity measure on a subset of  $W$ . In these regimes, the probability of deviating from the typical behaviour decays exponentially fast. This indicates that the behaviour of the minimizers is close to the limiting one already for moderate values of the diverging parameter(s). In regime (2), this indication is supported by numerical examples [KT18, Section 8]. We now survey the main results of [KT18] about regimes (1), (2), (3) in words; for further details, see [KT18, Sections 3, 4, 5], respectively.

In regimes (1) and (2), the typical trajectory quickly approaches the straight line between transmitter and receiver, and the probability of macroscopic deviations from this straight line decays exponentially fast. Interestingly, in regime (1), the typical number of hops tends to infinity in a sub-linear way, thus the typical length of a hop tends to infinity. In regime (3), we analyse the global and local repulsive effect of a particularly highly populated subset of  $W$ . Here, we replace the intensity measure  $\mu$  by  $\mu^a = \mu + a \text{Leb}|_\Delta$  for some  $\Delta \subseteq W$  with  $\text{Leb}(\Delta) > 0$ , and we let  $a \rightarrow \infty$ . The global effect is that if  $\ell$  is close to constant on  $W$ , then  $M^a(W)$  tends to zero exponentially fast as  $a \rightarrow \infty$ . Here,  $M^a$  is the measure  $M$  of all relaying hops (cf. (1.19)) corresponding to the minimizer (1.40), in case  $\mu$  is replaced by  $\mu^a$ . As for local effects, we discuss under what conditions it becomes unlikely for small  $\Delta$  for a user to choose a relay inside a small neighbourhood of  $\Delta$  than one outside a larger neighbourhood of  $\Delta$ .

## 1.8. Discussion.

1.8.1. *The interference term and the congestion term.* The interference term  $\mathfrak{S}(s)$  in (1.13) quantifies the joint quality of the transmissions of the messages in terms of a sum of an individual interference term over all the  $N(\lambda)$  trajectories and over all of their hops. The reason why we choose the *reciprocals* of the SIRs is that the *bandwidth* used for a transmission is defined [SPW07] as

$$\frac{\varrho}{\log_2(1 + \text{SIR}(\cdot))}, \quad (1.42)$$

where  $\varrho$  is the data transmission rate, and SIR is defined as in (1.12) but without the factor of  $1/\lambda$  in the denominator. This quantity is of order  $1/\lambda$  for  $\lambda$  large, under the assumption that  $L_\lambda \implies \mu$ . Thus, in the high-density setting  $\lambda \rightarrow \infty$ , (1.42) can be approached well by (a constant times) the reciprocals of the SIR, since  $\log(1+x) \sim x$  as  $x \rightarrow 0$ .

The reason that we took the sum over all the hops of a trajectory is that [SPW07, Section 3] suggests that in case of multihop communication, the used bandwidth equals the sum of the used bandwidth values corresponding to the individual hops. See [BC12] for another (single hop) setting where the sum of inverse values of SIRs appears as a cost function to be minimized.

The congestion term  $\mathfrak{M}(s)$  in (1.15) counts the number of ordered pairs of incoming hops arriving at the relays in the system. This is certainly an important characteristics of the quality of service, as too high an accumulation of many messages at relays results in a delay. Hence, it is natural to suppress the occurrence of such events, in order to increase the value of the model for realistic modeling.

An important property of this term is that it introduces dependence between the trajectories of different messages, unlike the interference term. Indeed, while  $\mathfrak{S}(s)$  can be decomposed into a sum of terms depending on the respective trajectories, each summand in  $\mathfrak{M}(s)$  involves many different trajectories. This is not only true in the special case of penalizing interference and congestion, but in general in our setting introduced in Section 1.2.3.

1.8.2. *The entropy term.* Let us now explain some important properties of the entropy term

$$I(\Psi) = \sum_{k=1}^{k_{\max}} \mathcal{H}_{W^k}(\nu_k \mid \mu \otimes M^{\otimes(k-1)}) + \sum_{m=0}^{\infty} \mathcal{H}_W(\mu_m \mid \mu c_m) + \mu(W) \left( 1 - \sum_{k=1}^{k_{\max}} M(W)^{k-1} \right) - \frac{1}{e} \quad (1.43)$$

defined in (1.23).

According to (i) and (iii) in (1.19), we have that  $M(W) \leq (k_{\max} - 1)\mu(W)$ , further, the first term on the right-hand side of (1.23) is bounded from below. Moreover, since by (ii) in (1.19), we have

$$\sum_{m \in \mathbb{N}_0} c_m = \sum_{m \in \mathbb{N}_0} \frac{\mu_m(W)}{\mu(W)} = 1, \quad (1.44)$$

it follows that  $\sum_{m=0}^{\infty} \mathcal{H}_W(\mu_m \mid \mu c_m)$  is nonnegative. These together with elementary properties of the relative entropy [DZ98, Section 6.2] imply that  $I(\Psi)$  is well-defined as an element of  $(-\infty, \infty]$  and  $\Psi \mapsto I(\Psi)$  is a lower semicontinuous function that is bounded from below. More precisely, the LDP in Theorem 1.4 implies that the infimum of  $I(\Psi)$  over admissible trajectory settings equals  $-\mu(W) \log k_{\max}$ , which equals the almost sure limit of  $1/\lambda$  times the logarithm of the total mass  $k_{\max}^{N(\lambda)}$  of the joint *a priori* measure (1.4).

Let us now provide an interpretation of  $I(\cdot)$ . Let  $\lambda > 0$  and  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$ . Recall the empirical measure family  $\Psi_\lambda(s) = ((R_{\lambda,k}(s))_{k \in [k_{\max}]}, (P_{\lambda,m}(s))_{m \in \mathbb{N}_0})$  from (1.27) and the constraints (1.18), which are similar to the ones (1.19) but with  $\mu$  replaced by the rescaled empirical measure  $L_\lambda$  everywhere.

Informally speaking, for  $\lambda > 0$  large,  $I(\Psi)$  asymptotically describes the following crucial counting term:

$$I(\Psi) \approx -\frac{1}{\lambda} \log \frac{\#\left\{s \in \mathcal{S}_{k_{\max}}(X^\lambda) : R_{\lambda,k}(s) \approx \nu_k, \forall k \in [k_{\max}] \text{ and } P_{\lambda,m}(s) \approx \mu_m, \forall m \in \mathbb{N}_0\right\}}{N(\lambda)^{\sum_{i \in I^\lambda} (\tilde{s}_i - 1)}}, \quad (1.45)$$

where  $\tilde{s}$  in the denominator is an arbitrarily chosen element of the set in the numerator. In this way, it fully describes the distribution of  $\Psi_\lambda(s)$  on an exponential level.

A major part of the proof of Theorem 1.2 consists in making (1.45) rigorous and verifying the corresponding formal statement. In the beginning of Section 2, we will argue why taking “=” signs instead of “ $\approx$ ” in the numerator of the right-hand side of (1.45) is not applicable. Instead, first, in Section 2.1, we will introduce a spatial discretization procedure and formulate a rigorous discrete version of these “ $\approx$ ” relations for fixed  $\lambda$  and fixed fineness parameter of the discretization. In Section 2.2, we will derive explicit combinatorial formulas for the cardinality of the trajectories in this setting. Next, in Section 3.1, we will conclude that  $1/\lambda$  times the logarithm of the quotient of the obtained counting complexity and the term  $N(\lambda)^{\sum_{i \in I^\lambda} (\tilde{s}_i - 1)}$  tends to  $I(\Psi)$  in the limit  $\lambda \rightarrow \infty$  followed by the fineness parameter of the discretization tending to zero.

The characterization of  $I(\Psi)$  that arises directly from this argument is not exactly (1.43) but another, however identical expression (3.3). The reason why we chose (1.43) as the definition of  $I(\Psi)$  is that it is given in terms of objects that are commonly used in large deviations theory: relative entropies, multiples of total masses of the corresponding measures plus an additive constant. In particular, we find it natural in (1.43) that each  $\mu_m$  is compared to the intensity measure  $\mu$  multiplied by the weight of a Poisson distribution at  $m$ . Indeed, in case  $\beta = 0$ , for the minimizer  $\Psi$  of the variational formula

(1.35),  $(\frac{d\mu_m}{d\mu}(x))_{m \in \mathbb{N}_0}$  is the Poisson distribution with parameter  $\frac{dM}{d\mu}(x)$  for each  $x \in W$ . Roughly speaking, the relative entropies plus the linear term  $\mu(W)(1 - \sum_{k=1}^{k_{\max}} M(W)^{k-1})$  arise from  $1/\lambda$  times the logarithmic rates of certain multinomial expressions in the above mentioned discretized counting procedure, after carrying out the limit  $\lambda \rightarrow \infty$  followed by the fineness parameter tending to zero.

1.8.3. *Rotation symmetry.* Let us assume that  $\beta = 0$  (cf. Section 1.6), and let us consider the special setting of Section 1.2.4 where  $\mathfrak{S}$  penalizes interference. If  $W = \overline{B_r(o)} \subset \mathbb{R}^d$  is a closed ball and  $\mu$  is invariant under rotations, then the measures  $(\nu_k)_k$  in (1.35) are also invariant under rotations of the entire trajectory, i.e., for any orthogonal  $d \times d$ -matrix  $O$ , we have that  $\nu_k(dx_0, \dots, dx_{k-1}) = \nu_k^O(dx_0, \dots, dx_{k-1}) \equiv \nu_k(d(Ox_0), \dots, d(Ox_{k-1}))$  for any  $k \in [k_{\max}]$ . This is easily seen by an inspection of the formulas for the entropy term  $J$  in (1.32) and for the interference term  $S$  in (1.21), as the function  $(x, y) \mapsto \int_W \ell(|z - y|) \mu(dz) / \ell(|x - y|)$  is invariant under multiplication of both arguments with the same orthogonal matrix.

1.8.4. *Non-Poissonian users.* In fact, the main results of this paper hold for any collection of (random or non-random) point processes  $((X_i)_{i=1, \dots, N(\lambda)})_{\lambda > 0}$  on  $W$  for which  $L_\lambda = \frac{1}{\lambda} \sum_{i=1}^{N(\lambda)} \delta_{X_i}$  converges weakly (almost surely, if random) to  $\mu$  as  $\lambda \rightarrow \infty$ . Neither the independence or monotonicity in  $\lambda$ , nor the Poissonity of  $(N(\lambda))_{\lambda > 0}$  is used for the proofs. For example, our results remain also true for the deterministic set  $X^\lambda = W \cap (\frac{1}{\lambda}\mathbb{Z}^d)$  and  $\mu$  the Lebesgue measure on  $W$ .

1.8.5. *The annealed setting.* Of mathematical interest might also be the annealed setting, where we average also over the locations of the users. In order to get an interesting result, we have to assume that  $L_\lambda$  satisfies a large deviation principle on the set  $\mathcal{M}(W)$  with some good rate function  $H$ . (In the case of a Poisson point process with intensity measure  $\lambda\mu$ ,  $H$  would be [HJP16, Proposition 3.6] the relative entropy with respect to  $\mu$ , see (1.20).) Then the large- $\lambda$  exponential rate of the annealed free energy should be equal to the negative infimum over  $\mu_0 \in \mathcal{M}(W)$  of  $H(\mu_0)$  plus the quenched rate function terms from the right-hand side of (1.24) with  $\mu$  replaced by  $\mu_0$  everywhere. Also our other results on the LDP and the form of the minimizer(s) should have some analogue, which we do not spell out.

1.8.6. *Extensions of the model of Section 1.2.4.* The main goal of this paper is to set up a model where message routeing happens probabilistically, with a uniform *a priori* distribution weighted by two penalization terms of different nature. Our results demonstrate how the interplay between entropy and energy leads to an orderly behaviour of the system in the high-density limit on its own.

Section 1.2.4 provides an example of the two penalization terms that can be interpreted in terms of well-known objects in telecommunication. In Section 1.7.2, we summarized qualitative properties of the minimizer of the variational formula in this special setting for  $\beta = 0$ . This special case can be viewed as a snap-shot type model, where there is no time-dependence but all hops of all transmissions happen simultaneously at the same time. Further, we assume that each user submits exactly one message. However, one can relax these assumptions and make the model more realistic in various ways.

First, one easily sees from the proofs in Sections 2–(5) that Theorems 1.2 and 1.4 as well as Proposition 1.5 can be extended to cases where users send out no message or multiple messages. This models the standard situation in which large messages are cut into many smaller ones, who independently find their ways through the system. For this, the trajectory probability space has to be enlarged: to each user  $X_i \in X^\lambda$ , we attach the number  $P_i \in \mathbb{N}_0$  of transmitted messages, and for each  $j \in \{1, \dots, P_i\}$ , there is an independent trajectory  $X_i \rightarrow o$ . The empirical trajectory measure  $R_{\lambda,k}(\cdot)$  must be augmented by these trajectories. The main additional assumption then is that  $\sum_{k=1}^{k_{\max}} \pi_0 R_{\lambda,k}(S)$  converges to some measure  $\mu_0 \in \mathcal{M}(W)$  with  $0 \neq \mu_0 \ll \mu$ . (Also the case that  $\mu_0$  is not absolutely continuous

with respect to  $\mu$  is interesting, but will need additional work.) The interference term  $\mathfrak{S}(\cdot)$  introduced in (1.13) also has to be changed. According to [BB09, Sections 2.3.1, 5.1], the SIR of the transmission of one of the  $P_i$  messages from  $X_i$  to  $x \in W$  should be defined as

$$\frac{\ell(|X_i - x|)}{\frac{1}{\lambda} \sum_{j \in I^\lambda} \ell(|X_j - x|) P_j}.$$

One could also incorporate (possibly random) sizes of the messages, which would require an additional enlargement of the trajectory space.

Second, the model of Section 1.2.4 can be made time-dependent. If one, e.g., introduces  $k_{\max}$  discrete time slots indexed by  $[k_{\max}]$ , and assumes the  $l$ th hop of any message trajectory to happen at time  $l$  for any  $l \in [k_{\max}]$ , then the interference of a transmission at time  $l$  is obtained from the starting points of all hops that happen at the same time, see [GK00, Section I.A]. The SIR is defined analogously to (1.12) but with this notion of interference, which depends on the entire message trajectories rather than only on the users. Further, the congestion term can also be adapted to the time-dependent situation via counting numbers of incoming hops at each time step separately. This variant of the model is indeed mathematically not far from the model treated in the present paper; for example the entropy term does not change.

More realistic and mathematically much more demanding time-dependent versions of our model can be set up in various ways; for example, one could allow for a much longer time horizon (for example, of order  $\lambda$ , and then dropping the factor of  $\lambda$  in the interference term), which must come with the possibility of messages standing still for many separate time units. Furthermore, one could allow users to transmit multiple messages over time. One could also introduce mobility of users similarly to [HJKP18]. The new notion of SIR comes with significant changes in the behaviour of the system in the high-density limit, and we decided to defer such investigations to a later work.

**1.8.7. Allowing an unbounded number of hops.** If the upper bound  $k_{\max}$  for the length of the trajectories is dropped, then the *a priori* measure defined in (1.4) has infinite total mass, and therefore the entropy function  $I$  is not bounded from below. However, since the function  $f_k$  in (1.7) is bounded from below, for any  $\gamma > 0, \beta \geq 0$ , the total probability mass of all the  $k$ -hop trajectories under the Gibbs distribution is upper bounded by some geometrically decaying term in  $k$ . Hence, the definition of the model is no problem for  $k_{\max} = \infty$  and  $\gamma > 0, \beta \geq 0$ . We believe that all our results of Section 1 about its limiting behaviour as  $\lambda \rightarrow \infty$  remain essentially true (apart from the LDP under the *a priori* measure). However, proofs will require an additional cutting argument, which might become rather nasty. Our belief that the results remain unchanged is supported by the fact that also the minimizing objects  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^\infty)$  defined in Proposition 1.3 enjoy a geometric upper bound for  $\nu_k(W^k)$  in  $k$ . Thus, these measures are also well-defined and form the set of minimizers of the variational formula (1.24) in case  $k_{\max} = \infty$ . The assertions of [KT18] corresponding to the large-distance limit, which we explained in Section 1.7.2, are proved also for  $k_{\max} = \infty$ , see [KT18, Section 3.3.3].

**1.8.8. Relation to an optimization problem via Monte Carlo Markov chains.** In the light of the Section 1.8.1, it is certainly interesting to minimize the cost function  $s \mapsto \gamma \mathfrak{S}(s) + \beta \mathfrak{M}(s)$  for fixed  $\gamma, \beta \in (0, \infty)$ . (For game-theoretic properties of this optimization problem, we refer the reader to [KT18, Section 7].) Computationally, this is in general a hard problem for high densities  $\lambda$  because the cardinality of  $\mathcal{S}_{k_{\max}}(X^\lambda)$  increases super-exponentially in  $N(\lambda) \asymp \lambda$ . Thus, computing all values of  $s \mapsto \gamma \mathfrak{S}(s) + \beta \mathfrak{M}(s)$  and then extracting the maximum is only feasible for small  $\lambda$ .

Now, our Gibbs distribution opens the possibility to optimize this cost function via the well-known approach of *simulated annealing*. Furthermore, for  $\lambda$  large, it is substantially less complex to realize the Gibbs distribution using Monte Carlo Markov chains than to directly minimize the cost function. Thus,



our Gibbsian ansatz provides an easier implementable joint strategy for the routeing of all messages, preferring trajectory collections having low cost function values, already for moderate values of  $\gamma, \beta$ .

Indeed, the recent master's thesis of Morgenstern [M18] investigates the computational complexity of realizing the Gibbs distribution  $P_{\lambda, X^\lambda}^{\gamma, \beta}$  numerically using Monte Carlo Markov chains, for  $\lambda, \beta, \gamma > 0$ . The author finds irreducible and aperiodic Markov chains on the state space  $\mathcal{S}_{k_{\max}}(X^\lambda)$ , both of Gibbs sampler and Metropolis types, having the Gibbs distribution as their stationary distribution. These chains converge towards  $P_{\lambda, X^\lambda}^{\gamma, \beta}$  as the number of Markovian steps tends to infinity. Using these chains, the number of operations needed in order to simulate the Gibbs distribution up to a given error  $\varepsilon > 0$  in total variation distance is at most exponential in  $\lambda$ . This is much more efficient than evaluating all the trajectory collections. In a variant of the Gibbsian model where each user can receive at most a given number  $m_{\max} \in \mathbb{N}_0$  incoming hops, the number of necessary operations is even polynomial in  $\lambda$ .

These Monte Carlo Markov chains can also be used in order to find the optimum of the cost function  $s \mapsto \gamma \mathfrak{G}(s) + \beta \mathfrak{M}(s)$  for a fixed  $\lambda$  and a fixed realization of  $X^\lambda$ , using simulated annealing. Here, one lets the transition probability of the  $t$ -th step of the chain depend on  $t$  via replacing  $(\gamma, \beta)$  by  $(\gamma_t, \beta_t)$  such that  $\gamma_t, \beta_t \rightarrow \infty$  sufficiently slowly as  $t \rightarrow \infty$ . [M18, Theorem 7.1] shows that if one chooses  $\beta_t = \frac{\beta}{\gamma} \gamma_t \leq c_0 \log t$  for a suitably chosen  $c_0 = c_0(\lambda) \asymp \lambda/N(\lambda)^2$ , then the Markov chain converges to the uniform distribution on the set of minimizers of the cost function.

## 2. THE DISTRIBUTION OF THE EMPIRICAL MEASURES

Having seen in Section 1.5 that the Gibbsian model can be entirely described in terms of the trajectory setting  $\Psi_\lambda(s)$ , i.e., of the crucial empirical measures  $R_{\lambda, k}(s)$  and  $P_{\lambda, m}(s)$  defined in (1.5)–(1.17), we now consider the question how to describe their distributions. We have to quantify the number of message trajectory families  $s$  that give the same family of empirical measures. The plain and short (but wrong) answer is

$$\sum_{s \in \mathcal{S}_{k_{\max}}(X^\lambda): R_{\lambda, k}(s) = \nu_k \forall k, P_{\lambda, m}(s) = \mu_m \forall m} \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{s_i - 1}} \approx e^{-\lambda I(\Psi)}, \quad (2.1)$$

where we recall  $I(\Psi)$  from (1.23) and recall that  $\Psi = ((\nu_k)_{k \in [k_{\max}]}, (\mu_m)_{m \in \mathbb{N}_0})$ . From such an assertion, it is indeed not far to conclude Theorem 1.2, but the problem is that this statement is not true like this. Actually, there are very many  $\Psi$ 's such that the left-hand side is equal to zero, for example if any of the  $\nu_k$ 's or  $\mu_m$ 's has values outside  $\frac{1}{\lambda} \mathbb{N}_0$ . However, if we do not consider single  $\Psi$ 's, but open sets of  $\Psi$ 's, then the idea behind (2.1) is sustainable. Therefore, we proceed in a standard way by decomposing the area  $W$  into finitely many subsets and count the message trajectories only according to the discretization sets that they visit. In Section 2.1 we introduce necessary notation for carrying out this strategy, and we comment on the relevance of the discretization procedure. Next, in Section 2.2, we derive explicit formulas for the distribution of the empirical measures in this discretization.

For the purpose of the present paper, where we consider the high-density limit  $\lambda \rightarrow \infty$ , we later need to take this limit and afterwards the limit as the fineness parameter  $\delta$  of the decomposition of  $W$  goes to zero. The outcome of these parts of the procedure is formulated in Proposition 3.1. In Proposition 3.2 the consequences for the interference term and for the congestion term are formulated.

**2.1. Our discretization procedure.** Let us now head towards the formulation of the discretization procedure. We proceed by triadic spatial discretization of the Poisson point process  $(X^\lambda)_{\lambda > 0}$ , similarly to the approach of [HJKP18]. To be more precise, we perform the following discretization argument. Note that we may assume that our communication area  $W$  is taken as  $W = [-r, r]^d$ , by accordingly extending  $\mu$  trivially. We write  $\mathbb{B} = \{3^{-n} | n \in \mathbb{N}_0\}$ . For  $\delta \in \mathbb{B}$ , we define the set

$$W_\delta = \{[x - r\delta, x + r\delta]^d : x \in (2r\delta\mathbb{Z})^d \cap W\}$$

of congruent sub-cubes of  $W$  of side length  $2r\delta$  and centres in  $(2r\delta\mathbb{Z})^d$ . Note that  $W_\delta$  is a finite set,  $o$  is a centre of an element of  $W_\delta$  and any intersection of two distinct elements of  $W_\delta$  has zero Lebesgue measure. Elements of  $W_\delta$  will be called  $\delta$ -subcubes. We will assume that for all  $\delta \in \mathbb{B}$ , the  $\delta$ -subcubes are canonically numbered as  $W_1^\delta, \dots, W_{\delta^{-d}}^\delta$ , which can be done e.g. according to the increasing lexicographic order of the midpoints of the subcubes. Now, for Lebesgue-almost every  $x \in W$ , for all  $\delta \in \mathbb{B}$  there exists a unique  $W_j^\delta$  that contains  $x$ ; let us denote this  $W_j^\delta$  by  $W_\delta^x$ , and the set of all  $x \in W$  for which  $W_\delta^x$  is well-defined by  $W_\mathbb{B}$ .

Now, if  $\nu \in \mathcal{M}(W)$ , then for any  $\delta \in \mathbb{B}$ , we define  $\nu^\delta(\cdot) = \nu(\cdot \mid \mathcal{F}_\delta) \in \mathcal{M}(W)$  as the conditional version of  $\nu$  given  $\mathcal{F}_\delta = \sigma(W_\delta)$ , that is, the measure on  $W$  that has in each box  $W_i^\delta$  a constant Lebesgue density and mass equal to  $\nu(W_i^\delta)$ . Since  $\mathcal{F}_\delta \subset \mathcal{F}_{\delta'}$  for  $\delta, \delta' \in \mathbb{B}$  with  $\delta' < \delta$ , we see that  $(\nu^\delta)^{\delta'} = (\nu^{\delta'})^\delta = \nu^\delta$  by the tower property. We also write  $L_\lambda^\delta := (L_\lambda)^\delta$  for  $\lambda > 0$  and  $\delta \in \mathbb{B}$ , where the empirical measure  $L_\lambda$  was defined in (1.1). We proceed analogously for  $W^k$ ,  $k \in [k_{\max}]$  instead of  $W$ . Note that  $\nu^\delta \implies \nu$  as  $\delta \downarrow 0$ , which can be shown by a martingale convergence argument, since the union of all the  $\mathcal{F}_\delta$  generate the Borel- $\sigma$ -field on  $W$ .

Now we are able to define what a *standard setting* is, the interpretation of which will be given right after the definition. Roughly speaking, the measures  $\nu_k$  and  $\mu_m$  appearing in its definition will later play the role of the measures appearing in (2.1), their  $\delta$ -approximations are defined as above, and their  $(\delta, \lambda)$ -versions approach them in the limit  $\lambda \rightarrow \infty$ , followed by  $\delta \downarrow 0$ . The latter ones satisfy the constraints of (1.18) restricted to  $\mathcal{F}_\delta^{\otimes k}$  respectively  $\mathcal{F}_\delta$ , hence, the  $\nu_k$  and  $\mu_m$  will later turn out to be eligible for the variational problem in (1.24) (under some mild additional assumption, see Lemma 2.5).

**Definition 2.1.** *A standard setting is a collection of measures*

$$\begin{aligned} \underline{\Psi} = & \left( (\nu_k)_{k=1}^{k_{\max}}, ((\nu_k^\delta)_{k=1}^{k_{\max}})_{\delta \in \mathbb{B}}, ((\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}})_{\delta \in \mathbb{B}, \lambda > 0}, \right. \\ & \left. (\mu_m)_{m=0}^\infty, ((\mu_m^\delta)_{m=0}^\infty)_{\delta \in \mathbb{B}}, ((\mu_m^{\delta, \lambda})_{m=0}^\infty)_{\delta \in \mathbb{B}, \lambda > 0}, (\mu^{\delta, \lambda})_{\delta \in \mathbb{B}, \lambda > 0} \right) \end{aligned} \quad (2.2)$$

with the following properties: for any  $\delta, \delta' \in \mathbb{B}$ ,  $\lambda > 0$ ,  $k \in [k_{\max}]$ ,  $m \in \mathbb{N}_0$  and  $i, i_0, \dots, i_{k-1} = 1, \dots, \delta^{-d}$ , respectively,  $\nu_k \in \mathcal{M}(W^k)$  and  $\mu_m \in \mathcal{M}(W)$ , and

- (1)  $\mu^{\delta, \lambda} = L_\lambda^\delta$ .
- (2)  $\nu_k^{\delta, \lambda} \in \mathcal{M}(W^k)$ . Further,  $\sum_{k=1}^{k_{\max}} \pi_0 \nu_k^{\delta, \lambda} = \mu^{\delta, \lambda}$ , moreover  $\lambda \nu_k^{\delta, \lambda}(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta) \in \mathbb{N}_0$ .
- (3) If  $\delta' \leq \delta$ , then  $\nu_k^{\delta', \lambda}(\cdot \mid \mathcal{F}_\delta^{\otimes k}) = \nu_k^{\delta, \lambda}(\cdot)$ .
- (4)  $\nu_k^{\delta, \lambda} \xrightarrow{\lambda \rightarrow \infty} \nu_k^\delta$ .
- (5)  $\mu_m^{\delta, \lambda} \in \mathcal{M}(W)$  with the property that  $\sum_{m=0}^\infty \mu_m^{\delta, \lambda} = \mu^{\delta, \lambda}$ , moreover  $\lambda \mu_m^{\delta, \lambda}(W_i^\delta) \in \mathbb{N}_0$ .
- (6)  $\sum_{m=0}^\infty m \mu_m^{\delta, \lambda} = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}$ .
- (7) If  $\delta' \leq \delta$ , then  $\mu_m^{\delta', \lambda}(\cdot \mid \mathcal{F}_\delta) = \mu_m^{\delta, \lambda}(\cdot)$ .
- (8)  $\mu_m^{\delta, \lambda} \xrightarrow{\lambda \rightarrow \infty} \mu_m^\delta$ .

**Remark 2.2.** *Immediate properties of a standard setting  $\Psi$  are the following.*

- (A) If  $\delta' \leq \delta$ , then  $\mu^{\delta', \lambda}(\cdot \mid \mathcal{F}_\delta) = \mu^{\delta, \lambda}(\cdot)$ .
- (B)  $\mu^{\delta, \lambda} \xrightarrow{\lambda \rightarrow \infty} \mu^\delta$  since  $L_\lambda \implies \mu$  as  $\lambda \rightarrow \infty$ .
- (C)  $\mu^\delta(\cdot) = \mu(\cdot \mid \mathcal{F}_\delta)$ . In particular,  $\mu^\delta \xrightarrow{\delta \downarrow 0} \mu$ .
- (D)  $\nu_k^\delta(\cdot) = \nu_k(\cdot \mid \mathcal{F}_\delta^{\otimes k})$ . In particular,  $\nu_k^\delta \xrightarrow{\delta \downarrow 0} \nu_k$ .
- (E)  $\mu_m^\delta(\cdot) = \mu_m(\cdot \mid \mathcal{F}_\delta)$ . In particular,  $\mu_m^\delta \xrightarrow{\delta \downarrow 0} \mu_m$ .

**Remark 2.3** (Standard settings and message trajectories). *The properties of Remark 2.2 explain the meaning of the  $\delta$ -indexed coordinates of a standard setting. Let us now interpret how one can*

obtain the  $(\delta, \lambda)$ -dependent coordinates of a standard setting starting from a fixed trajectory collection  $s \in \mathcal{S}_{k_{\max}}(X^\lambda)$ . Let

$$P_\lambda(s) = \frac{1}{\lambda} \sum_{i \in I^\lambda} \delta_{s_i} \quad (2.3)$$

denote the empirical measure of the starting sites of the trajectories, then  $P_\lambda(s) = L_\lambda$ , by our assumption that each user is picked precisely once in such a configuration. Hence, its  $\delta$ -discretized version  $P_\lambda^\delta(s)$  equals  $\mu^{\delta, \lambda}$ . Let us now choose  $\nu_k^{\delta, \lambda} = R_{\lambda, k}^\delta(s)$  (recall (1.5)). Then the requirement  $\sum_{k=1}^{k_{\max}} \pi_0 \nu_k^{\delta, \lambda} = \mu^{\delta, \lambda} = L_\lambda^\delta$  in (2) holds by (1.18). Further, let us choose  $\mu_m^{\delta, \lambda} = P_{\lambda, m}(s)$  (recall (1.17)). Then the constraints  $\sum_{m=0}^\infty \mu_m^{\delta, \lambda} = \mu^{\delta, \lambda}$  in (5) and  $\sum_{m=0}^\infty m \mu_m^{\delta, \lambda} = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}$  in (6) also hold by (1.18). Note that all the other requirements of Definition 2.1 are satisfied because of the tower property respectively of the convergence of  $L_\lambda$  towards  $\mu$ .

In the proof of Theorem 1.2, it will be essential to verify that, for certain standard settings,  $\sum_{m=0}^\infty m \mu_m^\delta(W)$  converges to  $\sum_{m=0}^\infty m \mu_m(W)$  as  $\delta \downarrow 0$ , which is not implied by Definition 2.1. However, similarly to the de la Vallée Poussin theorem about uniform integrability, the super-linear increase of  $m \mapsto \eta(m)$  yields the following handy criterion.

**Definition 2.4.** A controlled standard setting is a standard setting  $\underline{\Psi}$  as in (2.2) with the following extra property:

$$\lim_{\lambda \rightarrow \infty} \sum_{m=0}^\infty \eta(m) \mu_m^{\delta, \lambda}(W) = \sum_{m=0}^\infty \eta(m) \mu_m^\delta(W) < \infty, \quad \text{for all } \delta \in \mathbb{B}. \quad (2.4)$$

Note that by part (D) of Remark 2.2, we have  $\sum_{k=1}^{k_{\max}} k \nu_k^\delta(W^k) = \sum_{k=1}^{k_{\max}} k \nu_k(W^k)$  for any standard setting. Using this, we verify the following lemma.

**Lemma 2.5.** Let  $\underline{\Psi}$  be a controlled standard setting as in (2.2). Then  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^\infty)$  is an admissible trajectory setting.

*Proof.* Part (2) of Definition 2.1 claims that for all  $\delta \in \mathbb{B}$  and  $\lambda > 0$  we have  $\sum_{k=1}^{k_{\max}} \pi_0 \nu_k^{\delta, \lambda} = \mu^{\delta, \lambda}$ . By parts (B) and (C) of Remark 2.2, we have  $\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \nu_k^{\delta, \lambda} = \nu_k$  in the weak topology of  $\mathcal{M}(W^k)$ , for any fixed  $k \in [k_{\max}]$ . Similarly, by part (4) of Definition 2.1 and part (D) of Remark 2.2, we have  $\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \mu_m^{\delta, \lambda} = \mu_m$  in the weak topology of  $\mathcal{M}(W)$ . Moreover, since taking marginals is a continuous operation, also  $\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \pi_0 \nu_k^{\delta, \lambda} = \pi_0 \nu_k$  for all  $k$  in the weak topology of  $\mathcal{M}(W)$ . Thus, we have (i) in (1.19) for  $(\nu_k)_{k=1}^{k_{\max}}$ . In order to see that (ii) holds for  $(\mu_m)_{m=0}^\infty$ , one can additionally use part (5) of Definition 2.1, together with (2.4) and dominated convergence. Finally, by part (6) of Definition 2.1, (2.4) in Definition 2.4, the fact that  $\lim_{m \rightarrow \infty} \eta(m)/m = \infty$  and dominated convergence, we see that for any controlled setting  $\underline{\Psi}$ , we also have

$$\sum_{m=0}^\infty m \mu_m = \lim_{\delta \downarrow 0} \sum_{m=0}^\infty m \mu_m^\delta = \lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \sum_{m=0}^\infty m \mu_m^{\delta, \lambda} = \lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda} = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k \quad (2.5)$$

in the weak topology of  $\mathcal{M}(W)$ . This implies (iii) in (1.19) for  $\Psi$ . Hence,  $\Psi$  is an admissible trajectory setting.  $\square$

A converse of Lemma 2.5 also holds, in the following sense: for any admissible trajectory setting  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^\infty)$ , there exists a standard setting containing it, which can be chosen controlled if  $\sum_{m \in \mathbb{N}_0} \eta(m) \mu_m(W) < \infty$ . This will be the content of Proposition 3.3, a preliminary result for Theorems 1.2 and 1.4(i).

**2.2. The distribution of the empirical measures.** In this section, we describe the combinatorics of the system. For a standard setting  $\underline{\Psi}$  as in Definition 2.1, let us introduce the configuration set

$$J^{\delta,\lambda}(\underline{\Psi}) = \left\{ s \in \mathcal{S}_{k_{\max}}(X^\lambda) \mid R_{\lambda,k}^\delta(s) = \nu_k^{\delta,\lambda} \forall k, \quad P_{\lambda,m}^\delta(s) = \mu_m^{\delta,\lambda} \forall m \right\} \quad (2.6)$$

for fixed  $\delta \in \mathbb{B}$  and  $\lambda > 0$ . In words,  $J^{\delta,\lambda}(\underline{\Psi})$  is the set of families of trajectories such that the  $\delta$ -coarsenings of the empirical measures of the trajectories and the hop numbers are given by the respective measures in the setting  $\underline{\Psi}$ . Note that  $J^{\delta,\lambda}(\underline{\Psi})$  depends only on the  $\delta$ - $\lambda$  depending measures in the collection  $\underline{\Psi}$ .

In case  $\mu^{\delta,\lambda}(W) > 0$ , we will refer to the entity  $s_0^i$ ,  $i = 1, \dots, \lambda\mu^{\delta,\lambda}(W)$ , as the  $i$ th user or  $i$ th transmitter, the entity  $s^i$ ,  $i = 1, \dots, \lambda\mu^{\delta,\lambda}(W)$ , as the trajectory of the  $i$ th user,  $s_{-1}^i$  as the length (number of hops) of  $s^i$ ,  $s_l^i$  as the  $l$ -th relay of  $s^i$  (for  $l = 1, \dots, s_{-1}^i - 1$ ), finally  $m_i(s)$  as the number of incoming hops at the relay  $s_0^i$ .

The combinatorics of computing  $\#J^{\delta,\lambda}(\underline{\Psi})$  is given as follows.

**Lemma 2.6** (Cardinality of  $J^{\delta,\lambda}(\underline{\Psi})$ ). *For any  $\delta, \lambda > 0$ , and for any standard setting  $\underline{\Psi}$ ,*

$$\#J^{\delta,\lambda}(\underline{\Psi}) = N_{\delta,\lambda}^1(\underline{\Psi}) \times N_{\delta,\lambda}^2(\underline{\Psi}) \times N_{\delta,\lambda}^3(\underline{\Psi}), \quad (2.7)$$

where

$$N_{\delta,\lambda}^1(\underline{\Psi}) = \prod_{i=1}^{\delta-d} \left( \binom{\lambda\mu^{\delta,\lambda}(W_i^\delta)}{((\lambda\nu_k^{\delta,\lambda}(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta))_{i_1, \dots, i_{k-1}=1})_{k=1}^{\delta-d}} \right), \quad (2.8)$$

$$N_{\delta,\lambda}^2(\underline{\Psi}) = \prod_{i=1}^{\delta-d} \left( \binom{\lambda\mu^{\delta,\lambda}(W_i^\delta)}{(\lambda\mu_m^{\delta,\lambda}(W_i^\delta))_{m \in \mathbb{N}_0}} \right), \quad (2.9)$$

$$N_{\delta,\lambda}^3(\underline{\Psi}) = \prod_{i=1}^{\delta-d} \frac{(\lambda \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta,\lambda}(W_i^\delta))!}{\prod_{m=0}^{\infty} m! \lambda \mu_m^{\delta,\lambda}(W_i^\delta)} = \prod_{i=1}^{\delta-d} \frac{(\lambda \sum_{m=0}^{\infty} m \mu_m^{\delta,\lambda}(W_i^\delta))!}{\prod_{m=0}^{\infty} m! \lambda \mu_m^{\delta,\lambda}(W_i^\delta)}. \quad (2.10)$$

*Proof.* We proceed in three steps by counting first the trajectories, registering only the partition sets  $W_i^\delta$  that they travel through, second, for each  $m \in \mathbb{N}_0$ , the sets of relays in each partition set that receive precisely  $m$  ingoing hops and finally the choices of the relays for each hop in each partition set. Since every choice in the three steps can be freely combined with the other ones, the product of the three cardinalities is equal to the number of all trajectory configurations with the requested coarsened empirical measures.

(A) *Number of the transmitters of trajectories passing through given sequences of  $\delta$ -subcubes.* For each configuration  $s \in J^{\delta,\lambda}(\underline{\Psi})$  defined in (2.6), in each  $\delta$ -subcube  $W_i^\delta$ ,  $i = 1, \dots, \delta-d$ , there are  $\lambda\mu^{\delta,\lambda}(W_i^\delta)$  users. Out of them exactly  $\lambda\nu_k^{\delta,\lambda}(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta)$  take exactly  $k$  hops, having their first relay in  $W_{i_1}^\delta$ , their second in  $W_{i_2}^\delta$  etc. and their  $(k-1)$ st relay in  $W_{i_{k-1}}^\delta$ , for any  $k \in [k_{\max}]$  and  $i_1, \dots, i_{k-1} = 1, \dots, \delta-d$ . Such choices in different sub-cubes  $W_i^\delta$  corresponding to the transmitters are independent. Thus, the total number of such choices equals the number  $N_{\delta,\lambda}^1(\underline{\Psi})$  defined in (2.8). Note that for  $i = 1, \dots, \delta-d$ ,

$$\sum_{k=1}^{k_{\max}} \sum_{i_1, \dots, i_{k-1}=1}^{\delta-d} \nu_k^{\delta,\lambda}(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta) = \sum_{k=1}^{k_{\max}} \pi_0 \nu_k^{\delta,\lambda}(W_i^\delta) = \mu^{\delta,\lambda}(W_i^\delta),$$

where we used part (2) of Definition 2.1; hence the multinomial expressions in (2.8) are well-defined.

- (B) *Number of incoming hops.* In this step, for any  $\delta$ -subcube  $W_i^\delta$ , we count all the possible ways to distribute the incoming hops among the relays (= users)  $X_j \in W_i^\delta$ , under the two constraints that in  $W_i^\delta$  there are  $\lambda\mu^{\delta,\lambda}(W_i^\delta)$  potential relays, and for any  $m \in \mathbb{N}_0$ , exactly  $\lambda\mu_m^{\delta,\lambda}(W_i^\delta)$  receive precisely  $m$  incoming hops. Such choices are clearly independent of each other for different  $\delta$ -subcubes. Hence, the total number of such choices equals the number  $N_{\delta,\lambda}^2(\underline{\Psi})$  defined in (2.9). Again, the constraint (5) from Definition 2.1 implies that the multinomial expression (2.9) is well-defined. Clearly, all choices in this part are independent of the choices in part (A).
- (C) *Number of assignments of the hops to the relays.* Assume that we have chosen one possible choice in part (A) and one possible choice in part (B). We now derive the number of possible ways of distributing, for any  $i$ , all the incoming hops in  $W_i^\delta$  among the relays in  $W_i^\delta$ . Let us call this number  $M_i$ , then we know from part (A) that  $M_i = \lambda \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta,\lambda}(W_i^\delta)$ , since each such hop is the  $l$ -th of some of the trajectories for some  $l$ . The cardinality of the set of relays in  $W_i^\delta$  is equal to  $\lambda\mu^{\delta,\lambda}(W_i^\delta) = \lambda \sum_{m=0}^{\infty} \mu_m^{\delta,\lambda}(W_i^\delta)$ , and in part (B) we decomposed it into sets, indexed by  $m$ , in which each relay receives precisely  $m$  ingoing hops. Let us call such a relay an  $m$ -relay. Think of each such relay as being replaced by precisely  $m$  copies (in particular those with  $m = 0$  are discarded), then we have  $\lambda \sum_{m=0}^{\infty} m\mu_m^{\delta,\lambda}(W_i^\delta)$  virtual relays in  $W_i^\delta$ . (Note that this is equal to  $M_i$  by (6).) Now, if all these  $m$  copies of the  $m$ -relays were distinguishable, then the number of ways to distribute the  $M_i$  ingoing hops to the relays would be simply equal to  $M_i!$ . However, since these  $m$  copies are identical, we overcount by a factor of  $m!$  for any  $m$ -relay. This means that the number of hops into  $W_i^\delta$  is equal to  $M_i! / \prod_{m=0}^{\infty} (m!)^{\lambda\mu_m^{\delta,\lambda}(W_i^\delta)}$ . Since all these cardinalities can freely be combined with each other, we have deduced that the number of possible choices is equal to the number  $N_{\delta,\lambda}^3(\underline{\Psi})$  defined in (2.10).

We also see that all the choices in the three parts are independent of each other, i.e., can be freely combined with each other and yield different combinations. Hence, we arrived at the assertion.  $\square$

### 3. THE LIMITING FREE ENERGY AND THE LDP: PROOF OF THEOREMS 1.2 AND 1.4

In this section, we prove Theorems 1.2 and 1.4(i), that is, we derive the variational formula in (1.24) for the high-density (i.e.,  $\lambda \rightarrow \infty$ ) exponential rate of the partition function, and we verify the LDP for the empirical measures. Our first step is to derive the large- $\lambda$  exponential rate of the combinatorial formulas for the empirical measures of Lemma 2.6 in Section 3.1. Furthermore, in Section 3.2 we formulate and prove how the interference term and the congestion term behave in the limits  $\lambda \rightarrow \infty$ , followed by  $\delta \downarrow 0$ . In Section 3.3, given an admissible trajectory setting, we construct a standard setting containing it. Using all these, in Section 3.4 we prove Theorem 1.2, and in Section 3.5, we complete the proof of Theorem 1.4(i).

For the rest of this section, we fix the set  $\Omega_1 \subset \Omega$  of full  $\mathbb{P}$ -measure on which we do our quenched investigations:

$$\Omega_1 = \left\{ \omega \in \Omega : X_i(\omega) \in W_{\mathbb{B}} \ \forall i \in \mathbb{N}, \right. \\ \left. \lim_{\lambda \rightarrow \infty} \frac{\#\{i \in I^\lambda(\omega) : X_i(\omega) \in W_j^\delta\}}{\lambda} = \mu(W_j^\delta), \ \forall j = 1, \dots, \delta^{-d}, \ \forall \delta \in \mathbb{B} \right\}. \quad (3.1)$$

That  $\mathbb{P}(\Omega_1) = 1$  holds follows immediately from the Restriction Theorem [K93, Section 2.2] combined with the Poisson Law of Large Numbers [K93, Section 4.2] and the fact that  $\mu$  is absolutely continuous.

**3.1. The asymptotics of the combinatorics.** Let us fix a controlled standard setting  $\underline{\Psi}$  as in (2.2). Fix any  $\omega \in \Omega_1$ , and let the quantities  $I^\lambda$  and  $X^\lambda$  refer to this  $\omega$ . Denote

$$N_{\delta,\lambda}^0(\underline{\Psi}) = \prod_{i=1}^{\delta-d} \prod_{k=1}^{k_{\max}} \prod_{l=1}^{k-1} N(\lambda)^{\lambda \pi_l \nu_k^{\delta,\lambda}(W_i^\delta)}. \quad (3.2)$$

For a measurable subset  $V$  of  $\mathbb{R}^d$  and  $\nu, \tilde{\nu} \in \mathcal{M}(V)$ , let us write  $H_V(\nu|\tilde{\nu}) = \int_V d\nu \log \frac{d\nu}{d\tilde{\nu}}$  if the density  $\frac{d\nu}{d\tilde{\nu}}$  exists and  $H_V(\nu|\tilde{\nu}) = \infty$  otherwise. (The difference between  $H_V(\nu|\tilde{\nu})$  and the relative entropy  $\mathcal{H}_V(\nu|\tilde{\nu})$  defined in (1.20) is the additive term  $\tilde{\nu}(V) - \nu(V)$ .) Let us recall  $M = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k = \sum_{m=0}^{\infty} m \mu_m$  from (1.19) and  $c_m = \exp(-1/(e\mu(W)))(e\mu(W))^{-m}/m!$  from (1.23). Note that the rate function  $I$  defined in (1.23) has also the representation

$$I(\Psi) = \sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu^{\otimes k}) - H_W(M|\mu) + \sum_{m=0}^{\infty} H_W(\mu_m | \mu c_m) - \frac{1}{e}, \quad (3.3)$$

which we are going to use here. The equivalence between (3.3) and (1.23) will be verified in Section A of the Appendix. Recall (1.44), which implies that the third term in (3.3) is invariant under replacing  $H$  by  $\mathcal{H}$ . We now identify the large- $\lambda$  exponential rate of the cardinality of  $J^{\delta,\lambda}(\underline{\Psi})$  both on the scale  $\lambda \log \lambda$  and  $\lambda$ :

**Proposition 3.1** (Exponential rates of counting terms). *Let  $\underline{\Psi}$  be a controlled standard setting. Let us write  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$ . We have*

$$\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \frac{\#J^{\delta,\lambda}(\underline{\Psi})}{N_{\delta,\lambda}^0(\underline{\Psi})} = -I(\Psi),$$

as an identity in  $[0, \infty]$ . Moreover if  $I(\Psi) < \infty$ , then

$$\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda \log \lambda} \log \#J^{\delta,\lambda}(\underline{\Psi}) = M(W) < \infty,$$

almost surely.

*Proof.* Recall that  $\Psi$  is an admissible trajectory setting, according to Lemma 2.5. In particular,  $I(\Psi) \in (-\infty, \infty]$  is well-defined.

We use Stirling's formula  $\lambda! = (\lambda/e)^\lambda e^{o(\lambda)}$  in the limit  $\mathbb{N} \ni \lambda \rightarrow \infty$ , which leads to

$$\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \binom{a^{(\lambda)}}{a_1^{(\lambda)}, \dots, a_n^{(\lambda)}} = - \sum_{i=1}^n a_i \log \frac{a_i}{a}, \quad (3.4)$$

for any integers  $a_1^{(\lambda)}, \dots, a_n^{(\lambda)}$  that sum up to  $a^{(\lambda)}$  and satisfy  $\frac{1}{\lambda} a_i^{(\lambda)} \xrightarrow{\lambda \rightarrow \infty} a_i$  for  $i = 1, \dots, n$  with positive numbers  $a_1, \dots, a_n$  satisfying  $\sum_{i=1}^n a_i = a$ .

From (2.8) we obtain that

$$\begin{aligned} I_\delta^1(\underline{\Psi}) &= - \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log N_{\delta,\lambda}^1(\underline{\Psi}) \\ &= \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{i_1, \dots, i_{k-1}=1}^{\delta-d} \nu_k^\delta(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta) \log \frac{\nu_k^\delta(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta)}{\mu^\delta(W_i^\delta)}, \end{aligned}$$

where we also used that all the measures  $\nu_k^{\delta,\lambda}$  and  $\mu^{\delta,\lambda}$  converge as  $\lambda \rightarrow \infty$  to  $\nu_k^\delta$  and  $\mu^\delta$ , respectively.

Now we add the term  $\prod_{l=1}^{k-1} \mu^\delta(W_{i_l}^\delta)$  both in the numerator and the denominator under the logarithm and separate these two terms. In the former, we write its logarithm as  $\sum_{l=1}^{k-1} \log \mu^\delta(W_{i_l}^\delta)$ ,

interchange this sum on  $l$  with all the other sums on the  $i_0, \dots, i_{k-1}$  and write the sums over  $i_0, \dots, i_{l-1}, i_{l+1}, \dots, i_{k-1}$  in terms of the  $l$ -th marginal measure of  $\nu_k^\delta$ . This gives

$$\begin{aligned} I_\delta^1(\underline{\Psi}) &= \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{i_1, \dots, i_{k-1}=1}^{\delta-d} \nu_k^\delta(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta) \log \frac{\nu_k^\delta(W_i^\delta \times W_{i_1}^\delta \times \dots \times W_{i_{k-1}}^\delta)}{\mu^\delta(W_i^\delta) \prod_{l=1}^{k-1} \mu^\delta(W_{i_l}^\delta)} \\ &\quad + \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^\delta(W_i^\delta) \log \mu^\delta(W_i^\delta). \end{aligned} \quad (3.5)$$

In the same way as for  $I_1^\delta$ , we obtain

$$I_\delta^2(\underline{\Psi}) = - \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log N_{\delta, \lambda}^2(\underline{\Psi}) = \sum_{i=1}^{\delta-d} \sum_{m=0}^{\infty} \mu_m^\delta(W_i^\delta) \log \frac{\mu_m^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)}. \quad (3.6)$$

Using (3.1), on  $\Omega_1$  we have that the asymptotic behaviour of (3.2) is the following

$$N_{\delta, \lambda}^0(\underline{\Psi}) = N(\lambda)^\lambda \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}(W_i^\delta) = (\lambda \mu(W))^{\lambda(1+o(1))} \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}(W_i^\delta).$$

On the other hand, also by Stirling's formula, we can identify the large- $\lambda$  rate of the quotient of the counting terms in (2.10) and (3.2) as follows:

$$\begin{aligned} I_\delta^{3,0}(\underline{\Psi}) &= - \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \frac{N_{\delta, \lambda}^3(\underline{\Psi})}{N_{\delta, \lambda}^0(\underline{\Psi})} \\ &= - \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \prod_{i=1}^{\delta-d} \frac{\left( \frac{1}{e \mu(W)} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}(W_i^\delta) \right)^\lambda \sum_{k'=1}^{k_{\max}} \sum_{l'=1}^{k'-1} \pi_{l'} \nu_{k'}^{\delta, \lambda}(W_i^\delta)}{\prod_{m=0}^{\infty} m! \mu_m(W_i^\delta)} \\ &= - \sum_{i=1}^{\delta-d} \sum_{k'=1}^{k_{\max}} \sum_{l'=1}^{k'-1} \pi_{l'} \nu_{k'}^\delta(W_i^\delta) \left( \log \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^\delta(W_i^\delta) - (1 + \log \mu(W)) \right) \\ &\quad + \sum_{i=1}^{\delta-d} \sum_{m=0}^{\infty} \mu_m^\delta(W_i^\delta) \log(m!), \end{aligned} \quad (3.7)$$

where for the last term we used the fact that  $\underline{\Psi}$  is controlled (see also Lemma 2.5), together with dominated convergence. We can summarize the sum of the terms in (3.5), (3.6) and (3.7) as

$$\begin{aligned} - \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \frac{\#J^{\delta, \lambda}(\underline{\Psi})}{N_{\delta, \lambda}^0(\underline{\Psi})} &= I_\delta^1(\underline{\Psi}) + I_\delta^2(\underline{\Psi}) + I_\delta^{3,0}(\underline{\Psi}) \\ &= \sum_{k=1}^{k_{\max}} \sum_{i_0, \dots, i_{k-1}=1}^{\delta-d} \nu_k^\delta(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta) \log \frac{\nu_k^\delta(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta)}{\prod_{l=0}^{k-1} \mu^\delta(W_{i_l}^\delta)} \\ &\quad + \sum_{i=1}^{\delta-d} \sum_{m=0}^{\infty} \mu_m^\delta(W_i^\delta) \left( \log \frac{\mu_m^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} + m(1 + \log \mu(W)) + \log(m!) \right) \\ &\quad - \sum_{i=1}^{\delta-d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)}, \end{aligned} \quad (3.8)$$

where in the first line on the right-hand side we changed the summing index  $i$  into  $i_0$ . Since we have

$$\sum_{m=0}^{\infty} \mu_m^\delta(W) = \sum_{m=0}^{\infty} \mu_m(W) = \mu(W),$$

and thus

$$\sum_{i=1}^{\delta^{-d}} \sum_{m=0}^{\infty} \mu_m^\delta(W_i^\delta) \left( \log \frac{\mu_m^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} + m(1 + \log \mu(W)) + \log(m!) \right) = \sum_{m=0}^{\infty} \mu_m^\delta(W_i^\delta) \log \frac{\mu_m^\delta(W_i^\delta)}{c_m \mu^\delta(W_i^\delta)} - \frac{1}{e},$$

we obviously arrived at the discrete version of the entropy terms in (3.3), more precisely, the entropy of the measures in (3.3) with respect to the  $\sigma$ -field  $\mathcal{F}_\delta$ , respectively  $\mathcal{F}_\delta^{\otimes k}$ . Now, according to [G11, Proposition (15.6)], the limit of these entropies as  $\delta \downarrow 0$  is equal to their corresponding continuous version, i.e., the right-hand side of (3.8) converges to  $I(\Psi)$ . The first part of Proposition 3.1 follows.

Moreover, if  $I(\Psi) < \infty$ , then we have by continuity

$$\begin{aligned} \lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda \log \lambda} \log \#J^{\delta, \lambda}(\underline{\Psi}) &= \lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda \log \lambda} \log N_{\delta, \lambda}^0(\underline{\Psi}) \\ &= \lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \sum_{i=1}^{\delta^{-d}} \pi_l \nu_k^{\delta, \lambda}(W_i^\delta) = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k(W) = \sum_{k=1}^{k_{\max}} (k-1) \nu_k(W^k) \in [0, \infty), \end{aligned}$$

where in the last identity we used that by Fubini's theorem,  $\pi_0 \nu_k(W) = \nu_k(W^k)$  holds for all  $k$ . Hence, using that  $\Psi$  is an admissible trajectory setting, we conclude the second part of Proposition 3.1.  $\square$

**3.2. Approximations for the penalization terms.** The limiting relations between the penalization terms depending on the numbers of incoming hops in (1.9) and (1.22), and between the continuous penalization terms in (1.7) and (1.21) are given as follows.

**Proposition 3.2.** *Let  $\underline{\Psi}$  be a controlled standard setting. Let us write  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^\infty)$  for the admissible trajectory setting contained in  $\underline{\Psi}$ . Then, almost surely,*

$$\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \sup_{s \in J^{\delta, \lambda}(\underline{\Psi})} \left| \frac{1}{\lambda} \mathfrak{M}(s) - M(\Psi) \right| = 0, \quad (3.9)$$

and

$$\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \sup_{s \in J^{\delta, \lambda}(\underline{\Psi})} \left| \frac{1}{\lambda} \mathfrak{S}(s) - S(\Psi) \right| = 0. \quad (3.10)$$

*Proof.* Throughout the proof, we perform our analysis on  $\Omega_1$ . First, we show (3.9). Consider some  $s \in J^{\delta, \lambda}(\underline{\Psi})$  for  $\lambda > 0$  and  $\delta \in \mathbb{B}$ . Additionally assume that  $s_i^l \in W_{\mathbb{B}}$  for all  $i \in I^\lambda$  and  $l = 0, \dots, k$  (which is always the case for  $s = S = (S^i)_{i \in I^\lambda}$  on  $\Omega_1$ ).

Then  $P_\lambda^\delta(s) = \mu^{\delta, \lambda}$  and  $P_{\lambda, m}^\delta(s) = \mu_m^{\delta, \lambda}$  for all  $m \in \mathbb{N}_0$ , see the definition (2.6) of  $J^{\delta, \lambda}(\underline{\Psi})$ , (2.2) and (2.3). Recall that  $m_i(s)$  is the number of ingoing messages at relay  $X_i$  for the trajectory configuration  $s$ . Hence we have

$$\begin{aligned} \mathfrak{M}(s) &= \sum_{i \in I^\lambda} \eta(m_i(s)) = \sum_{m=0}^{\infty} \eta(m) \#\{i \in I^\lambda : m_i(s) = m\} = \sum_{m=0}^{\infty} \eta(m) P_{\lambda, m}^\delta(s)(W) \\ &= \sum_{m=0}^{\infty} \eta(m) P_{\lambda, m}^\delta(s)(W) = \lambda \sum_{m=0}^{\infty} \eta(m) \mu_m^{\delta, \lambda}(W), \end{aligned}$$

for all such  $s$ . Note that by part (8) of Definition 2.1 and part (E) of Remark 2.2, we obtain that  $\mu_m^{\delta, \lambda}$  tends to  $\mu_m$  as  $\lambda \rightarrow \infty$  followed by  $\delta \downarrow 0$ . Now, (2.4) in Definition 2.4, together with the fact that the total mass of  $\mu_m^{\delta, \lambda}$  equals the one of  $\mu_m$  for any  $m$ , implies the assertion in (3.9).

We continue with verifying (3.10). Let us fix an arbitrary controlled standard setting  $\underline{\Psi}$ . Our goal is to prove that (3.10) holds for this  $\underline{\Psi}$ . Using that, for an admissible trajectory setting  $\Psi =$



$((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$ ,  $S(\Psi)$  depends only on  $(\nu_k)_{k=1}^{k_{\max}}$ , we have for any  $\lambda > 0$ ,  $\delta \in \mathbb{B}$ ,  $s \in \mathcal{J}^{\delta, \lambda}(\underline{\Psi})$  and  $k \in [k_{\max}]$

$$\frac{1}{\lambda} \mathfrak{S}(s) - S(\Psi) = \langle R_{\lambda, k}(s), f_k(L_{\lambda}, \cdot) \rangle - \langle \nu_k, f_k(\mu, \cdot) \rangle.$$

In the rest of Section 3, we will often have to verify convergence of certain (sequences of) measures in the (coordinatewise) weak topology. In order to keep our arguments clear and short, for  $k \in \mathbb{N}$ , we fix a metric  $d_k(\cdot, \cdot)$  on  $\mathcal{M}(W^k)$  that generates the weak topology on this space. It turns out to be convenient to choose  $d_k$  to be the *Lipschitz bounded metric* [DZ98, Section D.2] on  $\mathcal{M}(W^k)$ , that is,

$$d_k(\nu_k^1, \nu_k^2) = \sup\{|\langle \nu_k^1, f \rangle - \langle \nu_k^2, f \rangle| : f \in \text{Lip}_1(W^k)\} \quad (3.11)$$

for all  $k$ , where  $\text{Lip}_1(W^k)$  is the set of Lipschitz continuous functions taking  $W^k$  to  $\mathbb{R}$  with Lipschitz parameter less than or equal to 1 and with uniform bound 1. We have for  $k \in [k_{\max}]$

$$\begin{aligned} \left| \frac{1}{\lambda} \mathfrak{S}(s) - S(\Psi) \right| &= \left| \langle R_{\lambda, k}(s), f_k(L_{\lambda}, \cdot) \rangle - \langle \nu_k, f_k(\mu, \cdot) \rangle \right| \\ &\leq \left| \langle R_{\lambda, k}(s), f_k(L_{\lambda}, \cdot) \rangle - \langle \nu_k^{\delta, \lambda}, f_k(L_{\lambda}, \cdot) \rangle \right| + \left| \langle \nu_k^{\delta, \lambda}, f_k(L_{\lambda}, \cdot) \rangle - \langle \nu_k^{\delta, \lambda}, f_k(L_{\lambda}^{\delta}, \cdot) \rangle \right| \\ &\quad + \left| \langle \nu_k^{\delta, \lambda}, f_k(L_{\lambda}^{\delta}, \cdot) \rangle - \langle \nu_k^{\delta, \lambda}, f_k(\mu, \cdot) \rangle \right| + \left| \langle \nu_k^{\delta, \lambda}, f_k(\mu, \cdot) \rangle - \langle \nu_k, f_k(\mu, \cdot) \rangle \right|. \end{aligned} \quad (3.12)$$

Now, we claim that all the four terms on the right-hand side tend to 0 in the limit  $\lambda \rightarrow \infty$  followed by  $\delta \downarrow 0$ . Indeed, for the first term, let  $g \in \text{Lip}^1(W^k)$ . Then we have

$$\begin{aligned} \left| \langle R_{\lambda, k}(s), g \rangle - \langle \nu_k^{\delta, \lambda}, g \rangle \right| &= \left| \int_{W^k} g(y) R_{\lambda, k}(s)(dy) - \int_{W^k} g(y) R_{\lambda, k}^{\delta}(s)(dy) \right| \\ &\leq \sum_{i_0, \dots, i_{k-1}=1}^{\delta-d} \sup_{y, z \in W_{i_0}^{\delta} \times \dots \times W_{i_{k-1}}^{\delta}} |g(y) - g(z)| R_{\lambda, k}(s)(W_{i_0}^{\delta} \times \dots \times W_{i_{k-1}}^{\delta}) \\ &\leq \delta \sqrt{dk} R_{\lambda, k}(s)(W) \leq \delta \sqrt{dk} L_{\lambda}(W), \end{aligned}$$

which tends to 0 as  $\lambda \rightarrow \infty$  followed by  $\delta \downarrow 0$ . It follows that  $R_{\lambda, k}(s) - \nu_k^{\delta, \lambda}$  tends weakly to 0 as  $\lambda \rightarrow \infty$  followed by  $\delta \downarrow 0$ . We note that for any  $\alpha > 0$ , the restriction of  $f_k$  to  $\mathcal{M}_{\leq \alpha}(W) \times W^k$  is bounded, where we wrote  $\mathcal{M}_{\leq \alpha}(V)$  for the set of measures on the space  $V$  with total mass  $\leq \alpha$ . Indeed, since  $W$  is compact,  $\mathcal{M}_{\leq \alpha}(W)$  with the weak topology is also a compact, metrizable space by Prohorov's theorem. Thus, the continuous function  $f_k : \mathcal{M}_{\leq \alpha}(W) \times W^k \rightarrow \mathbb{R}$  is uniformly continuous, and therefore it is bounded. Now, since eventually  $L_{\lambda} \in \mathcal{M}_{\leq 2\mu(W)}(W)$ , the first term on the right-hand side of (3.12) tends to 0.

As for the second term, note that for any  $\delta \in \mathbb{B}$ ,  $L_{\lambda} - L_{\lambda}^{\delta}$  tends to  $\mu - \mu^{\delta}$  as  $\lambda \rightarrow \infty$ , which tends to 0 as  $\delta \downarrow 0$ . Thus, by the fact that  $f_k$  is continuous and bounded on  $\mathcal{M}_{\leq 2\mu(W)}(W) \times W^k$  and eventually  $L_{\lambda}, L_{\lambda}^{\delta}, \nu_k^{\delta, \lambda} \in \mathcal{M}_{\leq 2\mu(W)}(W)$  for all  $\delta \in \mathbb{B}$ , the second term also tends to 0 as first  $\lambda \rightarrow \infty$  and afterwards  $\delta \downarrow 0$ . An analogous argument applies for the third term, using that  $L_{\lambda}^{\delta}$  converges to  $\mu$  as first  $\lambda \rightarrow \infty$  and then  $\delta \downarrow 0$  by part (B) of Remark 2.2 and the definition of  $\mu^{\delta}$ ,  $\delta \in \mathbb{B}$ . The fourth term tends to zero since it easily follows from part (4) of Definition 2.1 and part (D) of Remark 2.2 that  $\nu_k^{\delta, \lambda}$  converges weakly to  $\nu_k$ , and  $f_k$  is continuous as bounded on  $\mathcal{M}_{\leq 2\mu(W)}(W) \times W^k$ . We conclude (3.10) and Proposition 3.2.  $\square$

**3.3. Existence of standard settings.** Recall that we equip  $\mathcal{A}$  defined in (1.29) with the product topology of the weak topologies of the factors  $\mathcal{M}(W^k)$ ,  $\mathcal{M}(W)$ , and that this is the topology of coordinatewise weak convergence. For  $k \in \mathbb{N}$ , let  $d_k(\cdot, \cdot)$  be the Lipschitz bounded metric (3.11) on

$\mathcal{M}(W^k)$ , which generates the weak topology on this space. Then,

$$d_0(\Psi^1, \Psi^2) = \sum_{k=1}^{k_{\max}} d_k(\nu_k^1, \nu_k^2) + \sum_{m=0}^{\infty} 2^{-m} d_1(\mu_m^1, \mu_m^2), \quad \Psi^1, \Psi^2 \in \mathcal{A}, \quad (3.13)$$

is a metric on  $\mathcal{A}$  that generates the product topology. For  $\varrho > 0$  and  $\Psi \in \mathcal{A}$ , let us write  $B_\varrho(\Psi) = \{\Psi' \in \mathcal{A} : d_0(\Psi', \Psi) < \varrho\}$  for the open  $\varrho$ -ball around  $\Psi$ . We have the following.

**Proposition 3.3.** *On  $\Omega_1$ , for any admissible trajectory setting (see Definition 1.1),  $\Psi = ((\nu_k)_k, (\mu_m)_m)$ , there exists a standard setting  $\underline{\Psi}$  containing it. If  $\sum_m \eta(m) \mu_m(W) < \infty$ , then  $\underline{\Psi}$  can be chosen to be a controlled standard setting.*

*Proof.* We fix an admissible trajectory setting  $\Psi$  and construct  $\underline{\Psi}$  as follows. As is required in Definition 2.1, the measures  $\nu_k^\delta$  for  $k \in [k_{\max}]$  and  $\mu_m^\delta$  for  $m \in \mathbb{N}_0$  are the  $\delta$ -coarsenings of the measures  $\nu_k$  and  $\mu_m$ , respectively, and  $\mu^{\delta, \lambda} = L_\lambda^\delta$ . Now for  $\delta \in \mathbb{B}$  and  $\lambda > 0$ , pick some measures  $\nu_k^{\delta, \lambda}$  and  $\mu_m^{\delta, \lambda}$  with values in  $\frac{1}{\lambda} \mathbb{N}_0$  such that the requirements (2)  $\sum_{k=1}^{k_{\max}} \pi_0 \nu_k^{\delta, \lambda} = \mu^{\delta, \lambda}$ , (5)  $\sum_{m=0}^{\infty} \mu_m^{\delta, \lambda} = \mu^{\delta, \lambda}$  and (6)  $\sum_{m=0}^{\infty} m \mu_m^{\delta, \lambda} = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}$  of Definition 2.1 are met, such that  $\nu_k^{\delta, \lambda} \implies \nu_k^\delta$  and  $\mu_m^{\delta, \lambda} \implies \mu_m^\delta$  as  $\lambda \rightarrow \infty$  and such that the collection  $\underline{\Psi}$  of all these measures is a standard setting containing  $\Psi$ , which is controlled if  $\sum_m \eta(m) \mu_m(W) < \infty$ .

We claim that this can be done by taking suitable up- and downroundings of the numbers

$$\nu_k^{\delta, \lambda}(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta) = \nu_k^\delta(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta) \frac{L_\lambda^\delta(W_{i_0}^\delta)}{\mu^\delta(W_{i_0}^\delta)} \mathbb{1}\{\mu^\delta(W_{i_0}^\delta) > 0\}, \quad k \in [k_{\max}], \quad (3.14)$$

for all  $i_0, \dots, i_{k-1} = 1, \dots, \delta^{-d}$ , and dividing by  $\lambda$ , analogously for the  $\mu_m$ 's. Now, using the  $d$ -metric defined in (3.13), we prove that the convergences required in Definition 2.1 hold for such  $\underline{\Psi}$ .

First, we prove the convergence of the  $\delta$ -coarsenings  $\Psi^\delta = ((\nu_k^\delta)_k, (\mu_m^\delta)_m)$  to  $\Psi$  in the  $d_0$ -metric. We claim that for any  $\varrho > 0$ , there exists  $\delta_0 \in \mathbb{B}$  such that  $\Psi^\delta \in B_\varrho(\Psi)$  for all  $\mathbb{B} \ni \delta \leq \delta_0$ . Indeed, for  $k \in [k_{\max}]$ ,  $\nu_k \in \mathcal{M}(W^k)$  and  $\delta \in \mathbb{B}$  we see that the distance between  $\nu_k$  and its  $\delta$ -coarsening is of order  $\delta$  with respect to the Lipschitz bounded metric:

$$\begin{aligned} d_k(\nu_k, \nu_k^\delta) &= \sup_{f \in \text{Lip}_1(W^k)} \left| \sum_{i_0, \dots, i_{k-1}=1}^{\delta^{-d}} \left( \int_{W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta} f(x) \nu_k(dx) - \int_{W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta} f(x) \nu_k^\delta(dx) \right) \right| \\ &\leq \sup_{f \in \text{Lip}_1(W^k)} \sum_{i_0, \dots, i_{k-1}=1}^{\delta^{-d}} \sup_{x, y \in W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta} |f(x) - f(y)| \nu_k(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta) \leq \delta \nu_k(W^k) \sqrt{k d}, \end{aligned}$$

where we wrote  $x = (x_0, \dots, x_{k-1})$ ; and analogously for  $\mu_m$ . Thus, we have

$$d_0(\Psi, \Psi^\delta) \leq \delta \sqrt{d} \left[ \sum_{k=1}^{k_{\max}} \nu_k(W^k) \sqrt{k} + \sum_{m=0}^{\infty} \mu_m(W) 2^{-m} \right].$$

Since  $\sum_{m=0}^{\infty} \mu_m(W) < \infty$  by (ii) in (1.19), there exists a constant  $C$ , only depending on  $\Psi$ , such that  $\Psi^\delta \in B_\varrho(\Psi)$  for any  $\delta \leq C\varrho$ .

Second, we ignore the up- or downroundings in the construction of  $\Psi$  and prove the following. For  $\delta \in \mathbb{B}$  and  $\lambda > 0$ , let  $\Psi^{\delta, \lambda}$  be the collection of the measures introduced in (3.14). We claim that on  $\Omega_1$ , we have

$$\limsup_{\lambda \rightarrow \infty} d_0(\Psi^\delta, \Psi^{\delta, \lambda}) = 0.$$

Indeed, for any  $k \in [k_{\max}]$  and  $i_0, \dots, i_{k-1} = 1, \dots, \delta^{-d}$ ,  $d_k(\nu_k^\delta, \nu_k^{\delta, \lambda})$  is bounded from above by

$$\sup_{f \in \text{LiP}_1(W^k)} \sum_{i_0, \dots, i_{k-1}=1}^{\delta^{-d}} \nu_k^\delta(W_{i_0}^\delta \times \dots \times W_{i_{k-1}}^\delta) \left| \frac{L_\lambda^\delta(W_{i_0}^\delta)}{\mu^\delta(W_{i_0}^\delta)} - 1 \right| \|f\|_\infty \leq \nu_k^\delta(W^k) \max_{i_0=1}^{\delta^{-d}} \left| \frac{L_\lambda^\delta(W_{i_0}^\delta)}{\mu^\delta(W_{i_0}^\delta)} - 1 \right|. \quad (3.15)$$

Similarly, for any  $\delta \in \mathbb{B}$  and  $m \in \mathbb{N}_0$ ,  $d_1(\mu_m^\delta, \mu_m^{\delta, \lambda})$  vanishes in the limit  $\lambda \rightarrow \infty$ . Thus,

$$d_0(\Psi^\delta, \Psi^{\delta, \lambda}) \leq \left( \sum_{k=1}^{k_{\max}} \nu_k^\delta(W^k) + \sum_{m=0}^{\infty} 2^{-m} \mu_m^\delta(W) \right) \max_{i_0=1}^{\delta^{-d}} \left| \frac{L_\lambda^\delta(W_{i_0}^\delta)}{\mu^\delta(W_{i_0}^\delta)} - 1 \right|,$$

which tends to 0 on  $\Omega_1$  as  $\lambda \rightarrow \infty$ , according to (3.1).

Now, if we add the suitable up- and downroundings, we only change distances in the  $d$ -metric by an error term of order  $1/\lambda$ , which vanishes as  $\lambda \rightarrow \infty$ . This implies that  $\underline{\Psi}$  is a standard setting. It also follows easily that if  $\sum_m \eta(m) \mu_m(W) < \infty$ , then  $\underline{\Psi}$  is controlled.  $\square$

### 3.4. Proof of Theorem 1.2.

Abbreviate

$$\mathfrak{Y}(r) = \left( \prod_{i \in I^\lambda} N(\lambda)^{-(r^i - 1)} \right) \exp \left\{ -\gamma \mathfrak{S}(r) - \beta \mathfrak{M}(r) \right\}, \quad \lambda > 0, r \in \mathcal{S}_{k_{\max}}(X^\lambda),$$

and note that the partition function is given as

$$Z_\lambda^{\gamma, \beta}(X^\lambda) = \sum_{r \in \mathcal{S}_{k_{\max}}(X^\lambda)} \mathfrak{Y}(r). \quad (3.16)$$

Then Theorem 1.2 says that its large- $\lambda$  negative exponential rate is given as the infimum of  $I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)$ , taken over all admissible trajectory settings  $\Psi$ . Throughout the proof, we assume that the configuration  $X^\lambda = X^\lambda(\omega)$  comes from some  $\omega \in \Omega_1$  defined in (3.1).

Having proved Propositions 3.1, 3.2 and 3.3, our strategy to prove Theorem 1.2 is the following. First, Proposition 3.3 gives a standard way of constructing from an admissible trajectory setting  $\Psi$  satisfying  $I(\Psi) + \gamma S(\Psi) + \beta M(\Psi) < \infty$  a controlled standard setting  $\underline{\Psi}$  that contains  $\Psi$ . Then the lower bound for the partition function is easily given in terms of the objects that are contained in any such  $\underline{\Psi}$  and using the logarithmic asymptotics for their combinatorics from Propositions 3.1 and 3.2 and finally taking the infimum over all such  $\Psi$ , respectively  $\underline{\Psi}$ . The upper bound needs more care, since the entire sum over  $r$  has to be handled. First of all, we show that the sum can be restricted for all  $\lambda > 0$ , modulo some error term that is negligible on the exponential scale, to the sum of those configurations whose congestion exponent is at most  $C\lambda$  for some appropriate large constant  $C > 0$ . This sum can be decomposed, for any  $\delta \in \mathbb{B}$ , to sums on configurations coming from a particular choice of empirical measures on the  $\delta$ -partitions of  $W$ . The number of these empirical measures and the sum on the partitions is negligible in the limit  $\lambda \rightarrow \infty$ , and the asymptotics of the sums on  $r$  in these partitions can be evaluated with the help of our spatial discretization procedure, using arguments of the proofs of Propositions 3.1 and 3.2 in the limit  $\lambda \rightarrow \infty$ , followed by  $\delta \downarrow 0$ . Using these, we arrive at the formula (1.24).

Let us give the details. We start with the proof of the lower bound. For any admissible trajectory setting  $\Psi$  such that  $I(\Psi) + \gamma S(\Psi) + \beta M(\Psi) < \infty$ , we pick a controlled standard setting  $\underline{\Psi}$  as in Proposition 3.3 and recall the configuration class  $J^{\delta, \lambda}(\underline{\Psi})$  from (2.6). Then, for any  $\lambda > 0$  and  $\delta \in \mathbb{B}$ ,

$$Z_\lambda^{\gamma, \beta}(X^\lambda) \geq \sum_{r \in J^{\delta, \lambda}(\underline{\Psi})} \mathfrak{Y}(r) \geq \frac{\#J^{\delta, \lambda}(\underline{\Psi})}{\sup_{r \in J^{\delta, \lambda}(\underline{\Psi})} \prod_{i \in I^\lambda} N(\lambda)^{-(r^i - 1)}} \exp \left\{ - \sup_{r \in J^{\delta, \lambda}(\underline{\Psi})} (\gamma \mathfrak{S}(r) + \beta \mathfrak{M}(r)) \right\}. \quad (3.17)$$

Hence,

$$\begin{aligned}
\liminf_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma, \beta}(X^{\lambda}) &\geq \liminf_{\delta \downarrow 0} \liminf_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \frac{\#J^{\delta, \lambda}(\underline{\Psi})}{\sup_{r \in J^{\delta, \lambda}(\underline{\Psi})} \prod_{i \in I^{\lambda}} N(\lambda)^{-(r_i - 1)}} \\
&\quad - \gamma \limsup_{\delta \downarrow 0} \limsup_{\lambda \rightarrow \infty} \sup_{r \in J^{\delta, \lambda}(\underline{\Psi})} \frac{1}{\lambda} \mathfrak{S}(r) - \beta \limsup_{\delta \downarrow 0} \limsup_{\lambda \rightarrow \infty} \sup_{r \in J^{\delta, \lambda}(\underline{\Psi})} \frac{1}{\lambda} \mathfrak{M}(r) \\
&= -\mathbf{I}(\underline{\Psi}) - \gamma \mathbf{S}(\underline{\Psi}) - \beta \mathbf{M}(\underline{\Psi}).
\end{aligned} \tag{3.18}$$

In the last step we also used Propositions 3.1 and 3.2 together with the fact that  $\underline{\Psi}$  is controlled. Now take the supremum over all such  $\underline{\Psi}$  on the r.h.s. of (3.18) to conclude that the lower bound in (1.24) holds.

The upper bound of Theorem 1.2 requires some additional work. We start from (3.16). For  $C > 0$  we have

$$Z_{\lambda}^{\gamma, \beta}(X^{\lambda}) = \sum_{r \in \mathcal{S}_{k_{\max}}(X^{\lambda}): \mathfrak{M}(r) \leq \lambda C} \mathfrak{Y}(r) + \sum_{r \in \mathcal{S}_{k_{\max}}(X^{\lambda}): \mathfrak{M}(r) > \lambda C} \mathfrak{Y}(r). \tag{3.19}$$

Since the total mass of our *a priori* measure has a bounded large- $\lambda$  exponential rate (see Section 1.2.2) and  $\mathfrak{S}$ ,  $\mathfrak{M}$  are bounded from below, we see that

$$\limsup_{C \rightarrow \infty} \limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \sum_{r \in \mathcal{S}_{k_{\max}}(X^{\lambda}): \mathfrak{M}(r) > \lambda C} \mathfrak{Y}(r) = -\infty.$$

Thus, for  $C$  sufficiently large, the exponential rate of  $Z_{\lambda}^{\gamma, \beta}(X^{\lambda})$  is equal to the one of the first term on the right-hand side of (3.19). We additionally require  $C$  so large that

$$\inf_{\Psi \text{ adm. traj. setting, } \mathbf{M}(\Psi) \leq C} (\mathbf{I}(\Psi) + \gamma \mathbf{S}(\Psi) + \beta \mathbf{M}(\Psi)) = \inf_{\Psi \text{ adm. traj. setting}} (\mathbf{I}(\Psi) + \gamma \mathbf{S}(\Psi) + \beta \mathbf{M}(\Psi)). \tag{3.20}$$

Let us write  $\mathcal{S}_{k_{\max}, C}(X^{\lambda}) = \{r \in \mathcal{S}_{k_{\max}}(X^{\lambda}): \mathfrak{M}(r) \leq \lambda C\}$  and  $Z_{\lambda}^{\gamma, \beta, C}(X^{\lambda}) = \sum_{r \in \mathcal{S}_{k_{\max}, C}(X^{\lambda})} \mathfrak{Y}(r)$ . The upper bound of Theorem 1.2 follows as soon as we show that

$$\limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma, \beta, C}(X^{\lambda}) \leq - \inf_{\Psi \text{ admissible trajectory setting, } \mathbf{M}(\Psi) \leq C} (\mathbf{I}(\Psi) + \gamma \mathbf{S}(\Psi) + \beta \mathbf{M}(\Psi)). \tag{3.21}$$

For fixed  $\lambda > 0$  and  $\delta \in \mathbb{B}$ , let us say that a collection of measures  $\Psi^{\delta, \lambda} = ((\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}}, (\mu_m^{\delta, \lambda})_{m=0}^{\infty})$  lies in  $G(\delta, \lambda) = G(\delta, \lambda)(X^{\lambda})$  if all these measures take values in  $\frac{1}{\lambda} \mathbb{N}_0$  only and satisfy the constraints  $\sum_{k=1}^{k_{\max}} \pi_0 \nu_k^{\delta, \lambda} = L_{\lambda}^{\delta}$ ,  $\sum_{m=0}^{\infty} \mu_m^{\delta, \lambda} = L_{\lambda}^{\delta}$  and  $\sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda} = \sum_{m=0}^{\infty} m \mu_m^{\delta, \lambda}$ . We will write  $J^{\delta, \lambda}(\Psi^{\delta, \lambda})$  for the set  $J^{\delta, \lambda}(\underline{\Psi})$  defined in (2.6). Then the union of  $J^{\delta, \lambda}(\Psi^{\delta, \lambda})$  over all  $\Psi^{\delta, \lambda}$  with  $\sum_{m=0}^{\infty} \eta(m) \mu_m^{\delta, \lambda}(W) \leq C$  is equal to

$$\{(R_{\lambda, k}^{\delta}(r))_{k \in [k_{\max}]}, (P_{\lambda, m}^{\delta}(r))_{m \in \mathbb{N}_0}\}: r \in \mathcal{S}_{k_{\max}, C}(X^{\lambda}),$$

since these three equations characterize the tuple of the measures  $(R_{\lambda, k}^{\delta}(S))_{k=1}^{k_{\max}}$  and  $(P_{\lambda, m}^{\delta}(S))_{m=0}^{\infty}$  if  $(S^i)_{i \in I^{\lambda}} \in \mathcal{S}_{k_{\max}, C}(X^{\lambda})$ .

Using this, we can estimate, for any  $\delta \in \mathbb{B}$ ,

$$Z_{\lambda}^{\gamma, \beta, C}(X^{\lambda}) = \sum_{\Psi^{\delta, \lambda} \in G(\delta, \lambda): \mathbf{M}(\Psi^{\delta, \lambda}) \leq C} \sum_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \mathfrak{Y}(r) \leq \#G(\delta, \lambda) \sup_{\Psi^{\delta, \lambda} \in G(\delta, \lambda): \mathbf{M}(\Psi^{\delta, \lambda}) \leq C} \sum_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \mathfrak{Y}(r). \tag{3.22}$$

Hence,

$$\begin{aligned}
 & \limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma, \beta, C}(X^{\lambda}) \\
 & \leq \limsup_{\delta \downarrow 0} \limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \#G(\delta, \lambda) \\
 & \quad + \limsup_{\delta \downarrow 0} \limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \sup_{\Psi^{\delta, \lambda} \in G(\delta, \lambda): \mathbb{M}(\Psi^{\delta, \lambda}) \leq C} \left[ \frac{\#J^{\delta, \lambda}(\Psi^{\delta, \lambda})}{\inf_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \prod_{i \in I^{\lambda}} N(\lambda)^{-(r^i - 1)}} \right] \\
 & \quad - \gamma \liminf_{\delta \downarrow 0} \liminf_{\lambda \rightarrow \infty} \inf_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \frac{1}{\lambda} \mathfrak{S}(r) - \beta \liminf_{\delta \downarrow 0} \liminf_{\lambda \rightarrow \infty} \inf_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \frac{1}{\lambda} \mathfrak{M}(r).
 \end{aligned} \tag{3.23}$$

According to Lemma 3.4 below, the first term on the right-hand side is equal to zero. Now pick a sequence  $(\delta_n)_n$  and for each  $n$  a sequence  $(\lambda_{n,j})_j$  along which the superior limits as  $n \rightarrow \infty$ , respectively  $j \rightarrow \infty$ , are realized. Now pick, for any  $n$  and  $j$ , a maximizer  $\tilde{\Psi}^{\delta_n, \lambda_{n,j}}$ . Pick  $\lambda_0$  so large that  $N(\lambda) \leq 2\mu(W)\lambda$  for all  $\lambda \geq \lambda_0$ . Hence,

$$\bigcup_{\lambda > \lambda_0, \delta \in \mathbb{B}} G(\delta, \lambda) \subseteq \left( \prod_{k=1}^{k_{\max}} \mathcal{M}_{\leq 2\mu(W)}(W^k) \right) \times \mathcal{M}_{\leq 2\mu(W)}(W)^{\mathbb{N}_0}, \tag{3.24}$$

where we recall that  $\mathcal{M}_{\leq \alpha}(V)$  is the set of measures on a space  $V$  with total mass  $\leq \alpha$ . Note that  $\mathcal{M}_{\leq 2\mu(W)}(W^k)$  is compact in the weak topology of  $\mathcal{M}(W^k)$  for any  $k$ , according to Prohorov's theorem.

Without loss of generality (using two diagonal sequence arguments), we can assume that for all  $n \in \mathbb{N}$ ,  $\tilde{\Psi}^{\delta_n, \lambda_{n,j}}$  converges coordinatewise weakly to a collection of measures  $\tilde{\Psi}^{\delta_n} = ((\tilde{\nu}_k^{\delta_n})_{k=1}^{k_{\max}}, (\tilde{\mu}_m^{\delta_n})_{m=0}^{\infty})$  as  $j \rightarrow \infty$ , and  $\tilde{\Psi}^{\delta_n}$  converges coordinatewise weakly to a collection of measures  $\tilde{\Psi}$  as  $n \rightarrow \infty$ . Then, it is clear that  $\tilde{\Psi}$  satisfies (i) from (1.19), and also that

$$\lim_{n \rightarrow \infty} \lim_{j \rightarrow \infty} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \tilde{\nu}_k^{\delta_n, \lambda_{n,j}} = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \tilde{\nu}_k.$$

In order to see that (iii) holds for  $\tilde{\Psi}$ , it remains to show that  $\lim_{n \rightarrow \infty} \lim_{j \rightarrow \infty} \sum_{m=0}^{\infty} m \tilde{\mu}_m^{\delta_n, \lambda_{n,j}} = \sum_{m=0}^{\infty} m \tilde{\mu}_m$ . For  $N \in \mathbb{N}$  and for any continuous function  $f: W \rightarrow \mathbb{R}$ , we estimate

$$\left| \left\langle \sum_{m=0}^{\infty} m (\tilde{\mu}_m^{\delta_n, \lambda_{n,j}} - \tilde{\mu}_m), f \right\rangle \right| \leq \sum_{m=0}^N m \left| \langle \tilde{\mu}_m^{\delta_n, \lambda_{n,j}} - \tilde{\mu}_m, f \rangle \right| + \sum_{m=N+1}^{\infty} \|f\|_{\infty} m \left| \tilde{\mu}_m^{\delta_n, \lambda_{n,j}}(W) - \tilde{\mu}_m(W) \right|.$$

The first term on the r.h.s. clearly tends to 0 as  $j \rightarrow \infty$ , followed by  $n \rightarrow \infty$ , for any fixed  $N$ . The second term can further be estimated from above as follows

$$\|f\|_{\infty} \sum_{m > N} \eta(m) \left( \sup_{\tilde{m} > N} \frac{\tilde{m}}{\tilde{\eta}(\tilde{m})} \right) (\tilde{\mu}_m^{\delta_n, \lambda_{n,j}}(W) + \tilde{\mu}_m(W)) \leq 2\|f\|_{\infty} \left( \sup_{\tilde{m} > N} \frac{\tilde{m}}{\tilde{\eta}(\tilde{m})} \right) C.$$

By the assumption that  $(\eta(N)/N) \rightarrow \infty$  as  $N \rightarrow \infty$ , the right-hand side tends to 0. One can analogously show that  $\sum_{m=0}^{\infty} \tilde{\mu}_m^{\delta_n, \lambda_{n,j}}$  tends to  $\sum_{m=0}^{\infty} \tilde{\mu}_m$  as  $j \rightarrow \infty$  followed by  $n \rightarrow \infty$ , and hence condition (ii) from (1.19) holds. Also we have  $\sum_{m=0}^{\infty} \eta(m) \tilde{\mu}_m(W) \leq C$ . Altogether,  $\tilde{\Psi}$  is an admissible trajectory setting.

Now, using the arguments of the proofs of Propositions 3.1 and 3.2 (which also involve the coarsened limits  $\tilde{\Psi}^{\delta_n}$  for fixed  $n \in \mathbb{N}$ ) for the subsequential limits  $j \rightarrow \infty$  followed by  $n \rightarrow \infty$ , we conclude that

$$\lim_{n \rightarrow \infty} \lim_{j \rightarrow \infty} \frac{\#J^{\delta_n, \lambda_{n,j}}(\tilde{\Psi}^{\delta_n, \lambda_{n,j}})}{\inf_{r \in J^{\delta_n, \lambda_{n,j}}(\tilde{\Psi}^{\delta_n, \lambda_{n,j}})} \prod_{i \in I^{\lambda_{n,j}}} N(\lambda_{n,j})^{-(r^i - 1)}} = \mathbb{I}(\tilde{\Psi})$$

and, using the boundedness and continuity of each  $f_k$  on  $\mathcal{M}_{\leq 2\mu(W)}(W) \times W^k$ ,

$$\lim_{n \rightarrow \infty} \lim_{j \rightarrow \infty} \inf_{r \in J^{\delta n, \lambda n, j}(\tilde{\Psi}^{\delta n, \lambda n, j})} \frac{1}{\lambda_{n,j}} \mathfrak{S}(r) = \mathfrak{S}(\tilde{\Psi}).$$

Furthermore, the lower semicontinuity of  $\mathcal{M}(W)^{\mathbb{N}_0} \rightarrow (-\infty, \infty]$ ,  $(\nu_m)_{m \in \mathbb{N}_0} \mapsto \sum_{m \in \mathbb{N}_0} \eta(m) \nu_m(W)$ , together with Fatou's lemma implies that

$$-\beta \liminf_{n \rightarrow \infty} \liminf_{j \rightarrow \infty} \inf_{r \in J^{\delta n, \lambda n, j}(\tilde{\Psi}^{\delta n, \lambda n, j})} \frac{1}{\lambda_{n,j}} \mathfrak{M}(r) \leq -\beta \mathfrak{M}(\tilde{\Psi}). \quad (3.25)$$

Thus, we conclude that (3.21) (and therefore the upper bound in Theorem 1.2) holds, as soon as Lemma 3.4 is formulated and verified. This we do now. Recall that we are working with fixed  $\omega \in \Omega_1$ , and that the notion of  $G(\delta, \lambda)$  depends on  $\omega$  via  $G(\delta, \lambda) = G(\delta, \lambda)(X^\lambda(\omega))$ .

**Lemma 3.4.** *For any  $\delta \in \mathbb{B}$  and  $\omega \in \Omega_1$ , we have*

$$\limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \#G(\delta, \lambda) = 0.$$

*Proof.* For  $\lambda > 0$ , let  $G_1(\delta, \lambda)$  denote the set of  $(\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}}$  satisfying part (2) from Definition 2.1. It is easily seen that its cardinality increases only polynomially in  $\lambda$ . Now, given  $(\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}} \in G_1(\delta, \lambda)$ , we will give an upper bound for the number of  $(\mu_m^{\delta, \lambda})_{m=0}^{\infty}$  such that the pair of these tuples is in  $G(\delta, \lambda)$ . This is much more demanding, since there is *a priori* no upper bound for  $m$ . We will provide a  $\lambda$ -dependent one.

For any  $\lambda > 0$ ,  $\Psi^{\delta, \lambda} \in G(\delta, \lambda)$  and  $j = 1, \dots, \delta^{-d}$  we have that

$$\lambda \sum_{m=0}^{\infty} m \mu_m^{\delta, \lambda}(W_j^\delta) = \lambda \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}(W_j^\delta) \leq (k_{\max} - 1)N(\lambda),$$

and the numbers  $\mu_0^{\delta, \lambda}(W_j^\delta), \dots, \mu_{(k_{\max}-1)N(\lambda)}^{\delta, \lambda}(W_j^\delta)$ , are  $\frac{1}{\lambda}$  times nonnegative integers. In particular,  $\mu_m^{\delta, \lambda}(W_j^\delta) = 0$  for  $m > (k_{\max} - 1)N(\lambda)$ .

Let  $\varepsilon > 0$  be fixed. We claim that for all sufficiently large  $\lambda > 0$ , there are not more than  $\varepsilon N(\lambda) \sim \varepsilon \lambda \mu(W)$  nonzero ones out of these quantities. Indeed, if there were at least  $\lceil \varepsilon N(\lambda) \rceil$  nonzero ones, denoted  $\mu_{m_0}^{\delta, \lambda}(W_j^\delta), \dots, \mu_{m_{\lceil \varepsilon N(\lambda) \rceil - 1}}^{\delta, \lambda}(W_j^\delta)$  with  $0 \leq m_0 < m_1 < \dots < m_{\lceil \varepsilon N(\lambda) \rceil - 1} \leq (k_{\max} - 1)N(\lambda)$ , then we could estimate

$$\begin{aligned} (k_{\max} - 1)N(\lambda) &\geq \sum_{m=0}^{(k_{\max}-1)N(\lambda)} \lambda m \mu_m^{\delta, \lambda}(W_j^\delta) \geq \sum_{i=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m_i \mu_{m_i}^{\delta, \lambda}(W_j^\delta) \mathbb{1}\{\mu_{m_i}^{\delta, \lambda}(W_j^\delta) > 0\} \\ &= \sum_{i=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m_i \mu_{m_i}^{\delta, \lambda}(W_j^\delta) \mathbb{1}\{\mu_{m_i}^{\delta, \lambda}(W_j^\delta) \geq \frac{1}{\lambda}\} \geq \sum_{i=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} m_i \geq \sum_{m=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} m \sim \frac{1}{2}(\varepsilon N(\lambda))(\varepsilon N(\lambda) - 1), \end{aligned}$$

which is a contradiction for all  $\lambda > 0$  sufficiently large.

Now,  $\#G(\delta, \lambda)$  can be estimated as follows. Let us first fix  $(\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}} \in G_1(\delta, \lambda)$ , i.e., satisfying part (2) from Definition 2.1, and let us count the number of  $(\mu_m^{\delta, \lambda})_{m=0}^{(k_{\max}-1)N(\lambda)}$  such that  $((\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}}, (\mu_m^{\delta, \lambda})_{m=0}^{(k_{\max}-1)N(\lambda)})$  lies in  $G(\delta, \lambda)$ . Out of the  $k_{\max} \delta^{-d} N(\lambda)$  quantities  $\mu_0^{\delta, \lambda}(W_j^\delta), \dots, \mu_{(k_{\max}-1)N(\lambda)}^{\delta, \lambda}(W_j^\delta)$ ,  $j = 1, \dots, \delta^{-d}$ , at most  $\lceil \varepsilon N(\lambda) \rceil \delta^{-d}$  are nonzero. The number of ways to choose them equals  $\binom{k_{\max} N(\lambda) \delta^{-d}}{\lceil \varepsilon N(\lambda) \rceil \delta^{-d}}$ . Having chosen  $\lceil \varepsilon N(\lambda) \rceil \delta^{-d}$  potentially nonzero ones so that the remaining  $k_{\max} \delta^{-d} N(\lambda) - \lceil \varepsilon N(\lambda) \rceil \delta^{-d}$  ones are equal to zero, according to part (5) of Definition 2.1 we note that the potentially nonzero ones sum up to  $N(\lambda)$ , and each one has a value in

$\frac{1}{\lambda}\mathbb{N}_0$ . For this, there are at most  $\binom{N(\lambda)+\lceil\varepsilon N(\lambda)\rceil\delta^{-d}-1}{\lceil\varepsilon N(\lambda)\rceil\delta^{-d}-1}$  combinations, for any choice of the set of the potentially nonzero ones. Since  $\omega \in \Omega_1$ , we have  $N(\lambda)(\omega) = \lambda(\mu(W) + o(1))$  as  $\lambda \rightarrow \infty$  (where the  $o(1)$  term depends on  $\omega$ ). Therefore, using Stirling's formula as in (3.4), we have the following estimate in the limit  $\lambda \rightarrow \infty$

$$\begin{aligned} \#G(\delta, \lambda) &\leq \#G_1(\delta, \lambda) \binom{k_{\max}N(\lambda)\delta^{-d}}{\lceil\varepsilon N(\lambda)\rceil\delta^{-d}} \binom{N(\lambda) + \lceil\varepsilon N(\lambda)\rceil\delta^{-d} - 1}{\lceil\varepsilon N(\lambda)\rceil\delta^{-d} - 1} \\ &= e^{o(\lambda)} \exp\left(-\lambda\mu(W)\left((k_{\max} - \varepsilon)\delta^{-d} \log \frac{(k_{\max} - \varepsilon)\delta^{-d}}{k_{\max}\delta^{-d}} + \varepsilon\delta^{-d} \log \frac{\varepsilon\delta^{-d}}{k_{\max}\delta^{-d}}\right)\right) \\ &\quad \times \exp\left(-\lambda\mu(W)\left(\varepsilon\delta^{-d} \log \frac{\varepsilon\delta^{-d}}{1 + \varepsilon\delta^{-d}} + \log \frac{1}{1 + \varepsilon\delta^{-d}}\right)\right) \end{aligned}$$

Letting  $\varepsilon \downarrow 0$ , we conclude that  $\limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \#G(\delta, \lambda) = 0$ .  $\square$

**3.5. The large deviation principle: proof of Theorem 1.4(i).** In this section, we prove Theorem 1.4(i). The combinatorial essence of this theorem has already been proved in Proposition 3.1, including the relations with  $\delta$ -coarsenings. What remains to be done is to relate this to the coordinatewise weak convergence on  $\mathcal{A}$ . We will be able to use some of the arguments of Section 3.4.

The lower semicontinuity of  $I + \mu(W) \log k_{\max}$  was already discussed in Section 1.3, the nonnegativity in Section 1.5. These together mean that  $I + \mu(W) \log k_{\max}$  is a rate function.

We proceed with the proof of the lower bound. Let  $G \subseteq \mathcal{A}$  be open. If  $\inf_G I = \infty$ , then there is nothing to show, therefore let us assume that there exists  $\Psi \in G$  with  $I(\Psi) < \infty$ . According to Proposition 3.3, there is a standard setting  $\underline{\Psi}$  containing  $\Psi$ . Since  $G$  is open, there exists  $\varrho > 0$  such that  $B_\varrho(\Psi) \subseteq G$ . Let us choose  $\delta_0 \in \mathbb{B}$  and, for any  $\mathbb{B} \ni \delta \leq \delta_0$ , some  $\lambda_0 = \lambda_0(\delta) > 0$  such that  $\Psi^\delta, \Psi^{\delta, \lambda} \in B_\varrho(\Psi)$  for any  $\lambda > \lambda_0$ . Now we can estimate, for these  $\delta$  and  $\lambda$ ,

$$\begin{aligned} P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) \in G) &\geq P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) \in B_\varrho(\Psi)) \geq P_{\lambda, X^\lambda}^{0,0}((\Psi_\lambda(S))^\delta = \Psi^{\delta, \lambda}) \\ &= \frac{1}{Z_\lambda^{0,0}(X^\lambda)} \sum_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \frac{1}{\prod_{i \in I^\lambda} N(\lambda)^{r^i - 1}} \geq \frac{\#J^{\delta, \lambda}(\Psi^{\delta, \lambda})}{k_{\max}^{N(\lambda)} \sup_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \prod_{i \in I^\lambda} N(\lambda)^{r^i - 1}}. \end{aligned}$$

Now, using Proposition 3.1 and the fact that  $N(\lambda)/\lambda \rightarrow \mu(W)$ , we obtain

$$\liminf_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) \in G) \geq -\mu(W) \log k_{\max} - I(\Psi).$$

Note that  $\underline{\Psi}$  is not necessarily controlled because  $M(\Psi) < \infty$  is not guaranteed. However, since for all  $\delta \in \mathbb{B}$ ,  $s = 1, \dots, \delta^{-d}$ ,  $\lambda > 0$ ,  $\mu_m^{\delta, \lambda}(W_s^\delta)/\mu_m^\delta(W_s^\delta)$  does not depend on  $m$ , we easily see that Proposition 3.1 holds for this  $\underline{\Psi}$  as well. Now, take the supremum over  $\Psi \in G \cap \{I < \infty\}$  to conclude that the lower bound holds.

We continue with the upper bound. Let  $F \subseteq \mathcal{A}$  be closed. Let us choose an increasing sequence  $(\lambda_n)_{n \in \mathbb{N}}$  of positive numbers along which the limit superior in (1.30) is realized. For  $\lambda > 0$ , let us put

$$O(\lambda) = \{\Psi \in \mathcal{A} : P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) = \Psi) > 0\}.$$

If for all but finitely many  $n \in \mathbb{N}$  we have  $F \cap O(\lambda_n) = \emptyset$ , then

$$\limsup_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log P_{\lambda, X^\lambda}^{0,0}(\Psi_\lambda(S) \in F) = -\infty. \quad (3.26)$$

Therefore, without loss of generality, we can assume that  $O(\lambda_n) \cap F$  is non-empty for all  $n \in \mathbb{N}$ . For  $\delta \in \mathbb{B}$  and  $A \subset \mathcal{A}$ , let us write  $A^\delta = \{\Psi^\delta : \Psi \in A\}$ , where  $\Psi^\delta$  is the coordinatewise  $\delta$ -coarsened version

of  $\Psi$ . Then we have

$$\begin{aligned} \mathbb{P}_{\lambda_n, X^{\lambda_n}}^{0,0}(\Psi_{\lambda_n}(S) \in F) &= \mathbb{P}_{\lambda_n, X^{\lambda_n}}^{0,0}(\Psi_{\lambda_n}(S) \in F \cap O(\lambda_n)) = \mathbb{P}_{\lambda_n, X^{\lambda_n}}^{0,0}((\Psi_{\lambda_n}(S))^\delta \in (F \cap O(\lambda_n))^\delta) \\ &\leq \#(F \cap O(\lambda_n))^\delta \sup_{\Psi \in F \cap O(\lambda_n)} \frac{\#J^{\delta, \lambda_n}(\Psi^\delta)}{k_{\max}^{N(\lambda_n)} \inf_{r \in J^{\delta, \lambda_n}(\Psi^\delta)} \prod_{i \in I^{\lambda_n}} N(\lambda_n)^{r^i - 1}}. \end{aligned} \quad (3.27)$$

It is clear that  $(F \cap O(\lambda_n))^\delta \subseteq G(\delta, \lambda_n) = (O(\lambda_n))^\delta$  for all  $n \in \mathbb{N}$  and  $\delta \in \mathbb{B}$ , where  $G(\delta, \lambda_n)$  was defined in Section 3.4. Hence, by Lemma 3.4,

$$\limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\lambda_n} \log \#(F \cap O(\lambda_n))^\delta = 0.$$

It remains to show that

$$\limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\lambda_n} \log \left[ \sup_{\Psi \in F \cap O(\lambda_n)} \frac{\#J^{\delta, \lambda_n}(\Psi^\delta)}{\inf_{r \in J^{\delta, \lambda_n}(\Psi^\delta)} \prod_{i \in I^{\lambda_n}} N(\lambda_n)^{r^i - 1}} \right] \leq - \inf_{\Psi \in F} \mathbb{I}(\Psi). \quad (3.28)$$

One can do this analogously to the proof of the upper bound of Theorem 1.2 starting from (3.23). Indeed, using Prohorov's theorem together with a diagonal sequence argument, we find  $\Psi^* \in \mathcal{A}$  that the maximizer in (3.27) converges to along a subsequence of  $\delta$ 's and  $\lambda_n$ 's. The limit lies in  $F$  because  $F$  is closed. Using the lower semicontinuity of  $\mathbb{I}$  together with Fatou's lemma, we conclude that the left-hand side of (3.28) is not larger than  $-\mathbb{I}(\Psi^*)$ , which itself is not larger than  $-\inf_F \mathbb{I}$ . This finishes the proof of the upper bound in Theorem 1.4(i).

#### 4. ANALYSIS OF THE MINIMIZERS

This section is devoted to the proof of Proposition 1.3. In particular, in Section 4.1, we show that the infimum in (1.24) is attained and, for any minimizer  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^\infty)$ , for any  $k \in [k_{\max}]$ ,  $\mu^{\otimes k}$  is absolutely continuous with respect to  $\nu_k$  and  $\mu$  is absolutely continuous with respect to each  $\mu_m$ . Further, for all  $k \in [k_{\max}]$  and  $m \in \mathbb{N}_0$ , there exist constants  $c_k > 0$ ,  $k \in [k_{\max}]$ , and  $c'_m > 0$ ,  $m \in \mathbb{N}_0$  such that  $\nu_k(A) \geq c_k \mu^{\otimes k}(A)$  for all  $A \subseteq W^k$  measurable and  $\mu_m(A') \geq c'_m(A')$  for all  $A' \subseteq W$  measurable. This is a prerequisite for perturbing the minimizer in many admissible directions. In Section 4.2 we finish the proof of Proposition 1.3 by deriving the Euler–Lagrange equations. For the rest of the section, we fix all parameters  $W, \mu, \gamma, \beta$  and  $k_{\max}$ . Moreover, we use the following representation of  $\mathbb{I}$  from (1.23).

$$\mathbb{I}(\Psi) = \sum_{k=1}^{k_{\max}} H_{W^k}(\mu \otimes M^{\otimes(k-1)}) + \sum_{m=0}^{\infty} H_W(\mu_m \mid \mu) - \mu_m(W) \log \frac{(e\mu(W))^{-m}}{m!}. \quad (4.1)$$

**4.1. Existence and positivity of the minimizers.** We start with the following lemma, which follows almost immediately from the arguments of the proof of the upper bound of Theorem 1.2 in Section 3.4.

**Lemma 4.1.** *The set of minimizers of the variational formula in (1.24) is non-empty, compact and convex.*

*Proof.* Recall that the three functionals  $\mathbb{I}$ ,  $\mathbb{S}$ ,  $\mathbb{M}$  are lower semicontinuous and convex. Furthermore, it is clear that we can restrict the infimum in (1.24) to those  $\Psi$  that satisfy also  $\mathbb{M}(\Psi) \leq C$  for any sufficiently large  $C$ . But, as we have seen in Section 3.4, this set of  $\Psi$ 's is compact. From this, all our assertions easily follow.  $\square$



Now we prove that, for each minimizer  $\Psi$ ,  $\mu^{\otimes k}$  is absolutely continuous with respect to  $\nu_k$  and  $\mu$  is absolutely continuous with respect to each  $\mu_m$ , and the corresponding Radon–Nikodym derivatives are even bounded away from 0. (Note that the opposite absolute continuities are true by finiteness of the relative entropies in (1.23).) We need to show this only for  $k_{\max} > 1$ , as we explained after Proposition 1.3. Let us start with verifying the absolute continuities.

**Lemma 4.2.** *If  $k_{\max} > 1$  and  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$  is a minimizer of (1.24), then  $\mu^{\otimes k} \ll \nu_k$  for any  $k \in [k_{\max}]$ , and  $\mu \ll \mu_m$  for any  $m \in \mathbb{N}_0$ .*

*Proof.* The essence of the proof is the following. The functionals  $M(\cdot)$  and  $S(\cdot)$  are linear in each  $\mu_m$  respectively  $\nu_k$ , as well as the third term in  $I(\cdot)$  in (1.23) in each  $\mu_m$ . On the other hand, the function  $x \mapsto x \log x$  has the slope  $-\infty$  at  $x \downarrow 0$ . We show the following assertions about the minimizer  $\Psi$  step by step as follows. Recall that  $M = \sum_{m \in \mathbb{N}_0} m \mu_m = \sum_{k \in [k_{\max}]} \sum_{l=1}^{k-1} \pi_l \nu_k$ . We write  $\geq$  and  $>$ , respectively, between measures in  $\mathcal{M}(W^k)$  if their difference lies in  $\mathcal{M}(W^k)$ , respectively in  $\mathcal{M}(W^k) \setminus \{0\}$ .

Fix a measurable set  $A \subset W$  such that  $\mu(A) > 0$ . Then we have:

- (1)  $M(A) > 0$ .
- (2) for any  $m_1 < m_0 < m_2$  such that  $\mu_{m_1}(A) > 0$  and  $\mu_{m_2}(A) > 0$ , also  $\mu_{m_0}(A) > 0$ .
- (3)  $\mu_0(A) > 0$ .
- (4)  $\mu_m(A) > 0$  for any  $m \geq k_{\max}$ .
- (5)  $\nu_k(A^k) > 0$  for any  $k \in [k_{\max}]$ .

Indeed, these steps are verified respectively as follows. In each of the steps, for  $\varepsilon \in (0, 1)$ , we construct an admissible trajectory setting  $\Psi^\varepsilon = ((\nu_k^\varepsilon)_{k=1}^{k_{\max}}, (\mu_m^\varepsilon)_{m=0}^{\infty})$  such that  $I(\Psi^\varepsilon) + \gamma S(\Psi^\varepsilon) + \beta M(\Psi^\varepsilon) < I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)$  for sufficiently small  $\varepsilon > 0$ , and therefore  $\Psi$  is not a minimizer of (1.24).

- (1) If  $M(A) = 0$ , then in particular  $\mu_0(A) = \nu_1(A) = \mu(A)$  and  $\mu_m(A) = 0$  for all  $m > 0$ . Also,  $\pi_1 \nu_2(A) = \nu_2(W \times A) = 0$ , according to the definition of  $M$ .

Let us define  $\Psi^\varepsilon$  as follows:  $\nu_2^\varepsilon = (1 - \varepsilon)\nu_2 + \varepsilon(\mu^{\otimes 2})/\mu(W)$ ,  $\nu_k^\varepsilon = (1 - \varepsilon)\nu_k$  for  $k \neq 2$ ,  $\mu_1^\varepsilon = (1 - \varepsilon)\mu_1 + \varepsilon\mu$  and  $\mu_m^\varepsilon = (1 - \varepsilon)\mu_m$  for  $m \neq 1$ . Then we compute and estimate the three terms of the entropy  $I(\Psi)$  as follows.

$$\begin{aligned}
 & \sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k^\varepsilon \mid \mu \otimes (M^\varepsilon)^{\otimes(k-1)}) \\
 & \leq \sum_{k=1}^{k_{\max}} H_{W \times (W \setminus A)^{k-1}}((1 - \varepsilon)\nu_k \mid \mu \otimes (M^\varepsilon)^{\otimes(k-1)}) + H_{W \times A}\left(\frac{\varepsilon \mu^{\otimes 2}}{\mu(W)} \mid \varepsilon \mu^{\otimes 2}\right) + \mathcal{O}(\varepsilon) \\
 & \leq \sum_{k=1}^{k_{\max}} H_{W^k}(\mu \otimes M^{\otimes(k-1)}) + \mathcal{O}(\varepsilon),
 \end{aligned}$$

furthermore

$$\begin{aligned}
 & \sum_{m=0}^{\infty} H_W(\mu_m^\varepsilon \mid \mu) - \mu_m^\varepsilon(W) \log \frac{(\varepsilon \mu(W))^{-m}}{m!} \\
 & \leq H_W((1 - \varepsilon)\mu_m + \varepsilon\mu \mid \mu) - \mu_m(W) \log \frac{(\varepsilon \mu(W))^{-m}}{m!} + \mu(A)\varepsilon \log \varepsilon + \mathcal{O}(\varepsilon).
 \end{aligned} \tag{4.2}$$

For the second term we used the convexity of the relative entropy in the form

$$H_W((1 - \varepsilon)\mu_1 + \varepsilon\mu \mid \mu) \leq (1 - \varepsilon)H_W(\mu_1 \mid \mu) \leq H_W(\mu_1 \mid \mu) + \mathcal{O}(\varepsilon). \tag{4.3}$$

This in turn follows from [HJKP18, Lemmas 3.10, 3.11], which implies that, for any  $k \in \mathbb{N}$ ,  $\xi, \eta \in \mathcal{M}(W^k)$  with  $\eta \neq 0$  and  $\xi \ll \eta$ ,

$$\left| H_{W^k}(\xi | \eta) - H_{W^k}((1-\varepsilon)\xi | \eta) \right| \underset{\varepsilon \downarrow 0}{\asymp} \varepsilon.$$

It follows that, as  $\varepsilon \downarrow 0$ ,

$$\mathbf{I}(\Psi^\varepsilon) + \gamma \mathbf{S}(\Psi^\varepsilon) + \beta \mathbf{M}(\Psi^\varepsilon) - [\mathbf{I}(\Psi) + \gamma \mathbf{S}(\Psi) + \beta \mathbf{M}(\Psi)] \leq \mathcal{O}(\varepsilon) + \mu(A)\varepsilon \log \varepsilon, \quad (4.4)$$

which is negative for all sufficiently small  $\varepsilon > 0$ . Thus,  $\Psi$  is not a minimizer.

- (2) If  $M(A) > 0$  but  $\mu_{m_1}(A) > 0$ ,  $\mu_{m_2}(A) > 0$  and  $\mu_{m_0}(A) = 0$  for some  $m_1 < m_0 < m_2$ , then let  $\nu_k^\varepsilon = \nu_k$  for all  $k \in [k_{\max}]$  and let  $\mu_{m_0}^\varepsilon = (1-\varepsilon)\mu_{m_0} + \varepsilon(\alpha_1\mu_{m_1} + \alpha_2\mu_{m_2})$ ,  $\mu_{m_1}^\varepsilon = (1-\alpha_1\varepsilon)\mu_{m_1}$ ,  $\mu_{m_2}^\varepsilon = (1-\varepsilon\alpha_2)\mu_{m_2}$ , where  $\alpha_1, \alpha_2 \in (0, 1)$  are such that  $\alpha_1 + \alpha_2 = 1$  and  $m_1\alpha_1 + m_2\alpha_2 = m_0$ . Then,  $\Psi^\varepsilon$  is an admissible trajectory setting with  $M^\varepsilon = M$ . It follows similarly to the previous computation that  $\mathbf{I}(\Psi^\varepsilon) + \gamma \mathbf{S}(\Psi^\varepsilon) + \beta \mathbf{M}(\Psi^\varepsilon) < \mathbf{I}(\Psi) + \gamma \mathbf{S}(\Psi) + \beta \mathbf{M}(\Psi)$  for all sufficiently small  $\varepsilon > 0$ . However, instead of (4.2), we have

$$\begin{aligned} & \sum_{m=0}^{\infty} H_W(\mu_m^\varepsilon | \mu) - \mu_m^\varepsilon(W) \log \frac{(e\mu(W))^{-m}}{m!} \\ & \leq \sum_{m=0}^{\infty} H_W(\mu_m | \mu) - \mu_m(W) \log \frac{(e\mu(W))^{-m}}{m!} + (\alpha_1\mu_{m_1}(A) + \alpha_2\mu_{m_2}(A))\varepsilon \log \varepsilon + \mathcal{O}(\varepsilon), \end{aligned}$$

as  $\varepsilon \downarrow 0$ .

- (3) If  $M(A) > 0$  but  $\mu_0(A) = 0$ , let  $\nu_k^\varepsilon = (1-\varepsilon)\nu_k$  for all  $1 < k \leq k_{\max}$ ,  $\mu_m^\varepsilon = (1-\varepsilon)\mu_m$  for all  $m > 0$ ,  $\mu_0^\varepsilon = \varepsilon\mu + (1-\varepsilon)\mu_0$  and  $\nu_1^\varepsilon = (1-\varepsilon)\nu_1 + \varepsilon\mu$ . It is again sufficient to consider the entropy terms in  $\mathbf{I}$ . The summands on  $k > 1$  can be estimated as follows.

$$\begin{aligned} \sum_{k=2}^{k_{\max}} H_{W^k}(\nu_k^\varepsilon | \mu \otimes (M^\varepsilon)^{(k-1)}) &= \sum_{k=2}^{k_{\max}} H_{W^k}((1-\varepsilon)\nu_k | (1-\varepsilon)^{k-1}\mu \otimes M^{k-1}) \\ &\leq \sum_{k=2}^{k_{\max}} H_{W^k}(\nu_k | \mu \otimes M^{(k-1)}) + \mathcal{O}(\varepsilon). \end{aligned}$$

The summand for  $k = 1$  can be estimated with the help of (4.3). For the summand for  $m = 0$ , we have

$$\begin{aligned} H_W(\mu_0^\varepsilon | \mu) &= H_{W \setminus A}((1-\varepsilon)\mu_0 + \varepsilon\mu | \mu) + \mu(A)\varepsilon \log \varepsilon \\ &\leq H_{W \setminus A}((1-\varepsilon)\mu_0 | \mu) + \mu(A)\varepsilon \log \varepsilon + \mathcal{O}(\varepsilon) = H_W(\mu_0 | \mu) + \mu(A)\varepsilon \log \varepsilon + \mathcal{O}(\varepsilon). \end{aligned}$$

while the remaining sum is handled as follows.

$$\begin{aligned} & \sum_{m=1}^{\infty} H_W(\mu_m^\varepsilon | \mu) - \mu_m^\varepsilon(W) \log \frac{(e\mu(W))^{-m}}{m!} \\ &= \sum_{m=1}^{\infty} H_W((1-\varepsilon)\mu_m | \mu) - \mu_m(W) \log \frac{(e\mu(W))^{-m}}{m!} + \mathcal{O}(\varepsilon). \end{aligned}$$

Thus, (4.4) holds also here, which implies the claim.

- (4) If  $M(A) > 0$  but  $\mu_{m_0}(A) = 0$  for some  $m_0 \geq k_{\max}$ , let  $\mu_{m_0}^\varepsilon = (1-\varepsilon)\mu_{m_0} + \varepsilon M/m_0$ ,  $\mu_m^\varepsilon = (1-\varepsilon)\mu_m$  for  $m \notin \{0, m_0\}$ , and  $\nu_k^\varepsilon = \nu_k$  for all  $k \in [k_{\max}]$ . Then,

$$\sum_{m=1}^{\infty} m\mu_m^\varepsilon = (1-\varepsilon) \sum_{m=1}^{\infty} m\mu_m + \frac{\varepsilon m_0}{m_0} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k.$$

On the other hand, we have

$$\mu - \sum_{m=1}^{\infty} \mu_m^\varepsilon \geq \mu - (1 - \varepsilon) \sum_{m=1}^{\infty} \mu_m - \frac{\varepsilon(k_{\max} - 1)}{m_0} \mu \geq (1 - \varepsilon)\mu - (1 - \varepsilon) \sum_{m=1}^{\infty} \mu_m = (1 - \varepsilon)\mu_0.$$

Therefore, if we put  $\mu_0^\varepsilon = \mu - \sum_{m=1}^{\infty} \mu_m^\varepsilon$ , then  $\mu_0^\varepsilon \geq (1 - \varepsilon)\mu_0 \geq 0$ , in particular  $\Psi^\varepsilon$  is an admissible trajectory setting. Now we can proceed analogously to (3) to conclude that  $\mathbf{I}(\Psi^\varepsilon) + \gamma\mathbf{S}(\Psi^\varepsilon) + \beta\mathbf{M}(\Psi^\varepsilon) < \mathbf{I}(\Psi) + \gamma\mathbf{S}(\Psi) + \beta\mathbf{M}(\Psi)$  for sufficiently small  $\varepsilon > 0$ .

The proof of (5) is very similar to the ones of (2), (3) and (4), thus we leave it to the reader.  $\square$

The proof of the following lemma is similar to the one of Lemma 4.2, therefore we omit its proof.

**Lemma 4.3.** *If  $k_{\max} > 1$  and  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$  is a minimizer of (1.24), then for each  $k \in [k_{\max}]$ , there exists  $c_k > 0$  such that for all  $A \subseteq W^k$  measurable,  $\nu_k(A) \geq c_k \mu^{\otimes k}(A)$  holds. Similarly, for each  $m \in \mathbb{N}_0$ , there exists  $c'_m > 0$  such that for all  $A' \subseteq W$  measurable,  $\mu_m(A') \geq c'_m \mu(A')$  holds.*

**4.2. Deriving the Euler–Lagrange equations.** In this section, we finish the proof of Proposition 1.3. According to the results of Section 4.1, now we see that (1.24) exhibits at least one minimizer, and all minimizers have almost everywhere positive Lebesgue density on the corresponding powers of  $\text{supp } \mu$ . Knowing this, we now carry out the perturbation analysis for the minimizer(s) of the optimization problem in (1.24) and derive the shape of the minimizers in most explicit terms.

We use the method of Lagrange multipliers in the framework of a perturbation argument. Let  $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$  minimize (1.24). Fix any collection of signed measures  $\Phi = ((\tau_k)_{k=1}^{k_{\max}}, (\sigma_m)_{m=0}^{\infty})$  such that only finitely many  $\sigma_m$ 's are different from zero, the Radon–Nikodym derivative  $\frac{d\tau_k}{d\mu^{\otimes k}}$  is a simple function for each  $k$ , also  $\frac{d\sigma_m}{d\mu}$  is a simple function for each  $m$ , further they satisfy the following constraints:

$$(i) \quad \sum_{k=1}^{k_{\max}} \pi_0 \tau_k = 0, \quad (ii) \quad \sum_{m=0}^{\infty} \sigma_m = 0, \quad (iii) \quad M_\Phi := \sum_{m=0}^{\infty} m \sigma_m = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \tau_k. \quad (4.5)$$

Then it follows from Lemma 4.3 that, for any  $\varepsilon \in \mathbb{R}$  with sufficiently small  $|\varepsilon|$ ,  $\Psi + \varepsilon\Phi = ((\nu_k + \varepsilon\tau_k)_{k=1}^{k_{\max}}, (\mu_m + \varepsilon\sigma_m)_{m=0}^{\infty})$  is a collection of (non-negative!) measures that satisfies (1.19) and is therefore admissible in the variational formula in (1.24). That (1.19) is satisfied follows easily from (4.5). Furthermore, using the notation of Section 4.1, the non-negativity follows from the fact that each  $\tau_k$  respectively each  $\sigma_m$  is a finite linear combination of measures of the form  $\mathbb{1}_A d\mu^{\otimes k}$  with  $A \subset W^k$  respectively of the form  $\mathbb{1}_B d\mu$  with  $B \subset W$ , and we have  $\mathbb{1}_A d\mu^{\otimes k} \leq c_k^{-1} \mathbb{1}_A \nu_k$  respectively  $\mathbb{1}_B d\mu \leq c'_m \mathbb{1}_B d\mu_m$ . Since only finitely many such summands are involved, there is a constant  $C > 0$  such that  $|\tau_k| \leq C\nu_k$  and  $|\sigma_m| \leq C\mu_m$  for any  $k \in [k_{\max}]$  and  $m \in \mathbb{N}_0$ , and hence it suffices to take  $|\varepsilon| < 1/C$ .

From minimality, we deduce that

$$0 = \left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} \left( \mathbf{I}(\Psi + \varepsilon\Phi) + \gamma\mathbf{S}(\Psi + \varepsilon\Phi) + \beta\mathbf{M}(\Psi + \varepsilon\Phi) \right). \quad (4.6)$$

We calculate the latter two terms as

$$\left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} \left( \gamma\mathbf{S}(\Psi + \varepsilon\Phi) + \beta\mathbf{M}(\Psi + \varepsilon\Phi) \right) = \gamma \sum_{k \in [k_{\max}]} \langle \tau_k, \tilde{f}_k \rangle + \beta \sum_{m \in \mathbb{N}_0} \eta(m) \sigma_m(W),$$

where, as before, we used the notation  $\langle \nu, f \rangle$  for the integral of a function  $f$  with respect to a measure  $\nu$ . Abbreviating  $M = \sum_{k \in [k_{\max}]} \sum_{l=1}^{k-1} \pi_l \nu_k$  and using the representation (3.3) of  $\mathbf{I}(\cdot)$ , we see that

$$\left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} \mathbf{I}(\Psi + \varepsilon\Phi) = \sum_{k \in [k_{\max}]} \left\langle \tau_k, 1 + \log \frac{d\nu_k}{d\mu^{\otimes k}} \right\rangle - \left\langle M_\Phi, 1 + \log \frac{dM}{d\mu} \right\rangle + \sum_{m \in \mathbb{N}_0} \left\langle \sigma_m, 1 + \log \frac{d\mu_m}{d(c_m \mu)} \right\rangle, \quad (4.7)$$

where we recall that  $c_m = e^{-1/(e\mu(W))} (e\mu(W))^{-m} / m!$ . Summarizing, we obtain from (4.6) that

$$0 = \left\langle \Phi, ((h_k)_{k \in [k_{\max}]}, (g_m)_{m \in \mathbb{N}_0}) \right\rangle, \quad (4.8)$$

where

$$h_k = \gamma \tilde{f}_k + 2 - k + \log \frac{d\nu_k}{d(\mu \otimes M^{\otimes(k-1)})} \quad \text{and} \quad g_m = \beta \eta(m) + 1 + \log \frac{d\mu_m}{d\mu} - \log \frac{(e\mu(W))^{-m}}{m!}.$$

We conceive  $\Phi$  as an element of the vector space

$$\mathcal{A}_{\pm} = \prod_{k \in [k_{\max}]} \mathcal{M}_{\pm}(W^k) \times \mathcal{M}_{\pm}(W)^{\mathbb{N}_0}$$

where  $\mathcal{M}_{\pm}$  is the set of signed measures equipped with the weak topology, and  $((h_k)_{k \in [k_{\max}]}, (g_m)_{m \in \mathbb{N}_0})$  as a function on  $\prod_{k \in [k_{\max}]} W^k \times W^{\mathbb{N}_0}$ . The condition in (4.5) means that  $\Phi$  is perpendicular to any function in

$$\begin{aligned} \mathcal{F} = & \left\{ ((\varphi_k)_{k \in [k_{\max}]}, (\psi_m)_{m \in \mathbb{N}_0}) : \varphi_k : W^k \rightarrow \mathbb{R}, \psi_m : W \rightarrow \mathbb{R} \text{ bounded and measurable for any } k, m, \right. \\ & \exists \tilde{A}, \tilde{B}, \tilde{C} : W \rightarrow \mathbb{R} : \varphi_k(x_0, \dots, x_{k-1}) = \tilde{A}(x_0) + \sum_{l=1}^{k-1} \tilde{C}(x_l), \\ & \left. \text{and } \psi_m(x) = \tilde{B}(x) - m\tilde{C}(x) \text{ for } x, x_0, \dots, x_{k-1} \in W \right\}. \end{aligned}$$

We have shown that, if  $\Phi$  is perpendicular to any simple function in  $\mathcal{F}$ , then it is also perpendicular to  $((h_k)_{k \in [k_{\max}]}, (g_m)_{m \in \mathbb{N}_0})$ . Since  $\mathcal{F}$  is a closed linear subspace of  $\mathcal{A}_{\pm}$ , it follows that it contains this element. That is, there are three functions  $\tilde{A}, \tilde{B}, \tilde{C}$  on  $W$  such that, for any  $k$  respectively  $m$ ,

$$h_k(x_0, \dots, x_{k-1}) = \tilde{A}(x_0) + \sum_{l=1}^{k-1} \tilde{C}(x_l) \quad \text{and} \quad g_m(x) = \tilde{B}(x) - m\tilde{C}(x), \quad x, x_0, \dots, x_{k-1} \in W.$$

Using an obvious substitution, this is equivalent to the existence of three positive functions  $A, B, C$  such that

$$\nu_k(dx_0, \dots, dx_{k-1}) = \mu(dx_0) A(x_0) \prod_{l=1}^{k-1} (C(x_l) M(dx_l)) e^{-\gamma \tilde{f}_k(x_0, \dots, x_{k-1})}, \quad k \in [k_{\max}], \quad (4.9)$$

$$\mu_m(dx) = \mu(dx) B(x) \frac{(C(x)\mu(W))^{-m}}{m!} e^{-\beta \eta(m)}, \quad m \in \mathbb{N}_0. \quad (4.10)$$

From (i) and (ii) in (1.19), we can identify  $A$  and  $B$  as

$$\frac{1}{A(x_0)} = \sum_{k \in [k_{\max}]} \int_{W^{k-1}} \prod_{l=1}^{k-1} (C(x_l) M(dx_l)) e^{-\gamma \tilde{f}_k(x_0, \dots, x_{k-1})}, \quad (4.11)$$

$$\frac{1}{B(x)} = \sum_{m \in \mathbb{N}_0} \frac{(C(x)\mu(W))^{-m}}{m!} e^{-\beta \eta(m)}. \quad (4.12)$$

Furthermore, condition (iii) says that

$$\frac{1}{C(x)} = \frac{1}{C(x)} \frac{\mu(dx)}{M(dx)} \varphi\left(\frac{1}{C(x)\mu(W)}\right) = \Gamma(C \, dM, x), \quad x \in W, \quad (4.13)$$

where  $\varphi(\alpha) = \sum_{m \in \mathbb{N}_0} m \frac{\alpha^m}{m!} e^{-\beta \eta(m)} / \sum_{m \in \mathbb{N}_0} \frac{\alpha^m}{m!} e^{-\beta \eta(m)}$  for  $\alpha \in [0, \infty)$  and

$$\Gamma(d\tilde{M}, x) = \int_W \mu(dx_0) \frac{\sum_{k \in [k_{\max}]} \int_{W^{k-2}} \prod_{l=1}^{k-2} \tilde{M}(dx_l) F_k(x_0, x_1, \dots, x_{k-2}, x)}{\sum_{k \in [k_{\max}]} \int_{W^{k-1}} \prod_{l=1}^{k-1} \tilde{M}(dx_l) e^{-\gamma \tilde{f}_k(x_0, \dots, x_{k-1})}}, \quad (4.14)$$

where

$$F_k(x_0, x_1, \dots, x_{k-2}, x) = \sum_{l=1}^{k-1} e^{-\gamma \tilde{f}_k(x_0, y^l)}, \quad (4.15)$$

$y^l$  is the vector of length  $k - 1$ , consisting of  $x_1, \dots, x_{k-2}$ ; augmented by  $x$  at the  $l$ -th place, and  $\tilde{M}(dx) = C(x)M(dx)$ . This ends our derivation of the Euler–Lagrange equations for any minimizer  $\Psi$  of (1.24).

This description of  $C$  and  $M$  is rather implicit and involved, therefore we cannot offer any simple criterion for the uniqueness of the minimizers of (1.24). Also, the question of continuity of the tilting functions  $A$ ,  $B$  and  $C$  is open.

Since  $I + \gamma S + \beta M$  is convex, it follows that any admissible trajectory setting  $\Psi$  satisfying (4.9)–(4.15) is a minimizer of (1.24).

## 5. PROOF OF PROPOSITION 1.5

We proceed analogously to Sections 2 and 3, and thus we start with part (2), i.e., with verifying (1.34). We use the discretization argument from Section 2.1 again. We now provide the definition of a *transmission setting*, the analogue of Definition 2.1 of a standard setting with no reference to users receiving given numbers of incoming hops.

**Definition 5.1.** *A transmission setting is a collection of measures*

$$\underline{\Sigma} = \left( \Sigma = (\nu_k)_{k=1}^{k_{\max}}, ((\nu_k^\delta)_{k=1}^{k_{\max}})_{\delta \in \mathbb{B}}, ((\nu_k^{\delta, \lambda})_{k=1}^{k_{\max}})_{\delta \in \mathbb{B}, \lambda > 0}, (\mu^{\delta, \lambda})_{\delta \in \mathbb{B}, \lambda > 0} \right) \quad (5.1)$$

such that for any  $\delta, \delta' \in \mathbb{B}$ ,  $\lambda > 0$ ,  $k \in [k_{\max}]$  and  $s, s_0, \dots, s_{k-1} = 1, \dots, \delta^{-d}$ , respectively,  $\nu_k \in \mathcal{M}(W^k)$ , and parts (1), (2), (3) and (4) of Definition 2.1 hold.

Recall that Definition 5.1 implies parts (A), (B), (C) and (D) of Remark 2.2. Further, it is easy to see that for any transmission setting  $\underline{\Sigma}$ ,  $\Sigma$  is an asymptotic routing strategy.

The following lemma describes the combinatorics of the choices of message trajectories in the system. We recall the empirical measures  $(R_{\lambda, k}(s))_{k \in [k_{\max}]}$  from (1.5).

**Lemma 5.2.** *Let  $\underline{\Sigma}$  be a transmission setting. For  $\delta \in \mathbb{B}$  and  $\lambda > 0$  let*

$$K^{\delta, \lambda}(\underline{\Sigma}) = \{s \in \mathcal{S}_{k_{\max}}(X^\lambda) : R_{\lambda, k}^\delta(s) = \nu_k^{\delta, \lambda} \ \forall k = 1, \dots, k_{\max}\}.$$

Then we have  $\#K^{\delta, \lambda}(\underline{\Sigma}) = N_{\delta, \lambda}^1(\underline{\Sigma}) \times N_{\delta, \lambda}^4(\underline{\Sigma})$ , where  $N_{\delta, \lambda}^1(\underline{\Sigma})$  equals  $N_{\delta, \lambda}^1(\underline{\Psi})$  from (2.8) for any standard setting  $\underline{\Psi}$  containing  $\underline{\Sigma}$ , and

$$N_{\delta, \lambda}^4(\underline{\Sigma}) = \prod_{j=1}^{\delta^{-d}} (\lambda \mu^{\delta, \lambda}(W_j^\delta))^{\lambda \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k^{\delta, \lambda}(W_j^\delta)}.$$

*Proof.* We proceed in two steps by counting first the trajectories, registering only the partition sets  $W_i^\delta$  that they travel through, second, the choices of the relays for each hop in each partition set. Since every choice in the two steps can be freely combined with the other one, the product of the two cardinalities is equal to the number of all trajectory configurations with the prescribed coarsened empirical measures.

(A) *Number of the transmitters of trajectories passing through given sequences of  $\delta$ -subcubes.* This is equal to the corresponding quantity in the proof of Lemma 2.6, hence it equals  $N_{\delta, \lambda}^1(\underline{\Sigma})$ .

(B) *Number of assignments of the hops to the relays.* For each  $i = 1, \dots, \delta^{-d}$ , there are  $\lambda \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k(W_i^\delta)$  incoming hops arriving to the relays in  $W_i^\delta$  in total. Each incoming hop arriving at  $W_i^\delta$  can choose any of the  $\lambda \mu^{\delta, \lambda}(W_i^\delta)$  users as the corresponding relay. Such choices between different hops in  $W_i^\delta$  are independent, moreover all the choices in  $W_i^\delta$  are independent from all the choices in  $W_j^\delta$  for  $j \neq i$ . It follows that the number of assignments equals  $N_{\delta, \lambda}^4(\underline{\Sigma})$ .

We also see that all the choices in the two parts are independent of each other, i.e., they can be freely combined with each other and yield different combinations. Hence, we arrived at the assertion.  $\square$

Using the arguments of the proof of Proposition 3.1, the next lemma immediately follows.

**Lemma 5.3.** *Let  $\underline{\Sigma}$  be a transmission setting. Then*

$$\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \frac{\#K^{\delta, \lambda}(\underline{\Sigma})}{N_{\delta, \lambda}^0(\underline{\Sigma})} = -J(\underline{\Sigma}) \in (-\infty, \infty], \quad (5.2)$$

where  $N_{\delta, \lambda}^0(\underline{\Sigma})$  equals  $N_{\delta, \lambda}^0(\underline{\Psi})$  from (3.2) for any standard setting  $\underline{\Psi}$  containing  $\underline{\Sigma}$ . Moreover, if the r.h.s. of (5.2) is finite, then

$$\lim_{\delta \downarrow 0} \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda \log \lambda} \log \#K^{\delta, \lambda}(\underline{\Sigma}) = M(W) = \sum_{k=1}^{k_{\max}} (k-1) \nu_k(W).$$

Now, the identity in (1.34) follows from the proof of Theorem 1.2, using transmission settings instead of standard settings and replacing Proposition 3.1 by our Lemma 5.3. There is one more major change in the proof. Indeed, instead of the compactness of  $\{\Psi: \Psi \text{ adm. trajectory setting, } M(\Psi) \leq y\}$  for all  $y \geq 0$  in Section 3.4 and the fact that any level set of  $I + \gamma S + \beta M$  is contained in a larger level set of  $M$ , one shall use the following argument. Using that  $S$  is continuous on the set of asymptotic routing strategies, and that  $J$  is lower semicontinuous, bounded from below and it has compact level sets [DZ98, Section 6.2], it follows that each level set of  $J + \gamma S$  is included in a larger level set of  $S$ . Now, for all  $y \in \mathbb{R}$ , the set  $\{\Sigma: \Sigma \text{ asymptotic routing strategy, } S(\Sigma) \leq y\}$  is compact, because it is closed and contained in the set  $\{\Sigma: \Sigma \text{ asymptotic routing strategy, } \sum_{k=1}^{k_{\max}} k \nu_k(W^k) \leq y'\}$  for all sufficiently large  $y' \in \mathbb{R}$ , and such sets are compact by Prohorov's theorem. These together allow us to conclude (1.34).

From this, parts (1) and (4) of Proposition 1.5 can be derived analogously to how Theorem 1.4 was derived from Theorem 1.2 in Section 3.5. The additional fact that the rate function  $J + \mu(W) \log k_{\max}$  has compact level sets holds because relative entropies with respect to fixed reference measures have compact level sets [DZ98, Section 6.2].

Lastly, we verify (3), i.e., we prove that (1.35) is the unique minimizer of (1.34). The fact that the set of minimizers of the variational formula on the right-hand side of (1.34) is non-empty, compact and convex follows similarly to Lemma 4.1, again by Prohorov's theorem and the compactness of the sets  $\{\Sigma: \Sigma \text{ asymptotic routing strategy, } S(\Psi) \leq y\}$ ,  $y > 0$ . Further, an argument analogous to Lemmas 4.2 and 4.3 shows that for all minimizers  $\Sigma = (\nu_k)_{k \in [k_{\max}]}$ , we have that  $\nu_k \ll \mu^{\otimes k} \ll \nu_k$  and  $\frac{d\nu_k}{d\mu^{\otimes k}}$  is bounded away from zero, for all  $k \in [k_{\max}]$ . Deriving the Euler–Lagrange equations similarly to Section 4.2, it follows that (1.35)–(1.36) hold for any minimizer  $\Sigma = (\nu_k)_{k \in [k_{\max}]}$  of (1.34). This also implies that the minimizer  $\Sigma$  is unique. Thus, we conclude Proposition 1.5.  $\square$

## APPENDIX A. REPRESENTATIONS OF THE ENTROPY TERM

We defined the entropy term  $\Psi \mapsto I(\Psi)$  via the formula (1.23), which we interpreted in Section 1.8.2. It is easy to see that (1.23) is equivalent to the representation in (4.1), which we used for analytical

investigations. Now we show that (1.23) is equivalent to (3.3), which arises from the combinatorics in Section 3.1.

Recall that for  $k \in \mathbb{N}$  and  $\xi, \eta \in \mathcal{M}(W^k)$ , we have  $\mathcal{H}_{W^k}(\xi|\eta) = H_{W^k}(\xi|\eta) - \xi(W^k) + \eta(W^k)$ . Further, for an admissible trajectory setting  $\Psi = ((\nu_k)_{k \in [k_{\max}]}, (\mu_m)_{m \in \mathbb{N}_0})$ , recall the measure  $M = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu_k = \sum_{m=0}^{\infty} m \mu_m$ . Starting from the definition of  $I(\cdot)$ , in (1.23), we compute

$$\begin{aligned} I(\Psi) &= \sum_{k=1}^{k_{\max}} \mathcal{H}_{W^k}(\nu_k | \mu \otimes M^{\otimes(k-1)}) + \sum_{m=0}^{\infty} \mathcal{H}_W(\mu_m | \mu c_m) + \mu(W) \left(1 - \sum_{k=1}^{k_{\max}} M(W)^{k-1}\right) - \frac{1}{e} \\ &= \sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu \otimes M^{\otimes(k-1)}) - \sum_{k=1}^{k_{\max}} \nu_k(W^k) + \mu(W) \sum_{k=1}^{k_{\max}} M(W)^{k-1} + \sum_{m=0}^{\infty} \mathcal{H}_W(\mu_m | \mu c_m) \\ &\quad + \mu(W) \left(1 - \sum_{k=1}^{k_{\max}} M(W)^{k-1}\right) - \frac{1}{e} \\ &= \sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu \otimes M^{\otimes(k-1)}) + \sum_{m=0}^{\infty} H_W(\mu_m | \mu c_m) - \frac{1}{e}, \end{aligned}$$

where we used (1.44), and the fact that  $\sum_{k=1}^{k_{\max}} \nu_k(W^k) = \mu(W)$  by (i) in (1.19). By the definition of the measure  $M$ , it suffices to show that

$$\sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu \otimes M^{\otimes(k-1)}) = \sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu^{\otimes k}) - H_W(M | \mu). \quad (\text{A.1})$$

Clearly, if any of the sides of (A.1) is infinite, then so is the other side. Else, we verify (A.1) as follows

$$\begin{aligned} &\sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu \otimes M^{\otimes(k-1)}) \\ &= \sum_{k=1}^{k_{\max}} \int_{W^k} d\nu_k(x_0, \dots, x_{k-1}) \left[ \log \frac{d\nu_k}{d\mu^{\otimes k}}(x_0, \dots, x_{k-1}) - \log \frac{d(\mu \otimes M^{\otimes(k-1)})}{d\mu^{\otimes k}}(x_0, \dots, x_{k-1}) \right] \\ &= \sum_{k=1}^{k_{\max}} \int_{W^k} d\nu_k(x_0, \dots, x_{k-1}) \left[ \log \frac{d\nu_k}{d\mu^{\otimes k}}(x_0, \dots, x_{k-1}) - \log \left( \prod_{l=1}^{k-1} \frac{dM}{d\mu}(x_l) \right) \right] \\ &= \sum_{k=1}^{k_{\max}} \int_{W^k} d\nu_k(x_0, \dots, x_{k-1}) \log \frac{d\nu_k}{d\mu^{\otimes k}}(x_0, \dots, x_{k-1}) - \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \int_W \pi_l \nu_k(dx_l) \log \frac{dM}{d\mu}(x_l) \\ &= \sum_{k=1}^{k_{\max}} H_{W^k}(\nu_k | \mu^{\otimes k}) - H_W(M | \mu). \end{aligned}$$

**Acknowledgements.** The authors thank B. Jahnel, C. Hirsch, M. Klimm, S. Morgenstern and M. Renger for interesting discussions and comments. The second author acknowledges support by the Berlin Mathematical School (BMS) and the DFG RTG 1845 ‘Stochastic Analysis with Applications in Biology, Finance and Physics’.

## REFERENCES

- [AK08] S. ADAMS and W. KÖNIG, Large deviations for many Brownian bridges with symmetrised initial-terminal condition. *Probab. Theory Relat. Fields* **142:1**, 79–124 (2008).
- [BB09] F. BACCELLI and B. BLASZCZYSHYN, *Stochastic Geometry and Wireless Networks: Volume I: Theory*. Now Publishers Inc. (2009).

- [BC12] F. BACCELLI and C. S. CHEN, Self-Optimization in Mobile Cellular Networks: Power Control and User Association, *IEEE International Conference on Communications*, Cape Town (2010).
- [CCS16] R. COLINI-BALDESCHI, ROBERTO COMINETTI and MARCO SCARSINI, On the Price of Anarchy of Highly Congested Nonatomic Network Games, *arXiv:1605.03081* (2016).
- [Ch09] X. CHEN, *Random Walk Intersections: Large Deviations and Related Topics*. Mathematical Surveys and Monographs, AMS, Vol. 157, Providence, RI. (2009).
- [DZ98] A. DEMBO and O. ZEITOUNI, *Large Deviations Techniques and Applications*, 2nd edition, Springer, Berlin (1998).
- [GK00] P. GUPTA and P. R. KUMAR, The Capacity of Wireless Networks, *IEEE Trans. Inform. Theory* **46:2**, (2000).
- [GT08] A. J. GANESH and G. L. TORRISI, Large Deviations of the Interference in a Wireless Communication Model, *IEEE Trans. Inform. Theory* **54:8**, (2008).
- [G11] H-O. GEORGII, *Gibbs Measures and Phase Transitions*, De Gruyter, 2nd edition (2011).
- [GZ93] H-O. GEORGII and H. ZESSIN, Large deviations and the maximum entropy principle for marked point random fields, *Probab. Theory Relat. Fields* **96:2**, 177–204 (1993).
- [HJKP16] C. HIRSCH, B. JAHNEL, P. KEELER and R. PATTERSON, Large-deviation principles for connectable receivers in wireless networks, *Advances in Applied Probability*, **48**, 1061–1094 (2016). *arXiv:1506.00576*.
- [HJKP18] C. HIRSCH, B. JAHNEL, P. KEELER and R. PATTERSON, Large deviations in relay-augmented wireless networks, *Queueing Systems* **88**, 349–387 (2018). *arXiv:1510.04146*.
- [HJP16] C. HIRSCH, B. JAHNEL and R. PATTERSON, Space-time large deviations in capacity-constrained relay networks, *Latin American Journal of Probability and Statistics* **15**. 587–615 (2018), *arXiv:1609.06856*.
- [HJ17] C. HIRSCH and B. JAHNEL, Large deviations for the capacity in dynamic spatial relay networks, *arXiv:1712.03763* (2017).
- [K93] J. F. C. KINGMAN, *Poisson Processes*, Oxford University Press, New York (1993).
- [KM02] W. KÖNIG and P. MÖRTERS, Brownian intersection local times: upper tail asymptotics and thick points. *Ann. Probab.* **30**, 1605–1656 (2002).
- [KM13] W. KÖNIG and C. MUKHERJEE, Large deviations for Brownian intersection measures, *Commun. Pure Appl. Math.* **66:2**, 263–306 (2013).
- [KT18] W. KÖNIG and A. TÓBIÁS, Routing properties in a Gibbsian model for highly dense multihop networks, *arXiv:1801.04985v2* (2018).
- [M18] S. MORGENSTERN, *Markov Chain Monte Carlo for Message Routing*, master’s thesis, TU Berlin (2018).
- [SPW07] H. SONG, M. PENG and W. WANG, Node Selection in Relay-based Cellular Networks, *IEEE 2007 International Symposium on Microwave, Antenna, Propagation, and EMC Technologies For Wireless Communications*.
- [T08] J. TRASHORRAS, Large Deviations for Symmetrised Empirical Measures. *J. Theoret. Probab.* **21**, 397–412 (2008).