

## SMOOTHED ANALYSIS OF BALANCING NETWORKS\*

TOBIAS FRIEDRICH<sup>1</sup>, THOMAS SAUERWALD<sup>2</sup>, AND DAN VILENCHIK<sup>3</sup>

ABSTRACT. In a balancing network each processor has an initial collection of unit-size jobs (tokens) and in each round, pairs of processors connected by balancers split their load as evenly as possible. An excess token (if any) is placed according to some predefined rule. As it turns out, this rule crucially affects the performance of the network. In this work we propose a model that studies this effect. We suggest a model bridging the uniformly-random assignment rule, and the arbitrary one (in the spirit of smoothed-analysis). We start with an arbitrary assignment of balancer directions and then flip each assignment with probability  $\alpha$  independently. For a large class of balancing networks our result implies that after  $\mathcal{O}(\log n)$  rounds the discrepancy is  $\mathcal{O}((1/2 - \alpha) \log n + \log \log n)$  with high probability. This matches and generalizes known upper bounds for  $\alpha = 0$  and  $\alpha = 1/2$ . We also show that a natural network matches the upper bound for any  $\alpha$ .

## 1. INTRODUCTION

In this work we are concerned with two topics whose name contains the word “smooth”, but in totally different meaning. The first is *balancing (smoothing) networks*, the second is *smoothed analysis*. Let us start by introducing these two topics, and then introduce our contribution – interrelating the two.

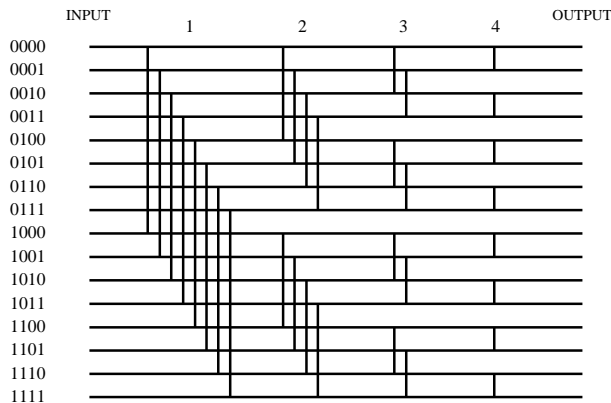
**1.1. Balancing (smoothing) networks.** In the standard abstraction of *smoothing* (balancing) networks [2], processors are modeled as the vertices of a graph and connection between them as edges. Each process has an initial collection of unit-size jobs (which we call tokens). Tokens are routed through the network by transmitting tokens along the edges according to some local rule. We measure the quality of such a balancing procedure by the maximum difference between the number of tokens at any two vertices at the end.

<sup>1</sup> MAX-PLANCK-INSTITUT FÜR INFORMATIK, SAARBRÜCKEN, GERMANY

<sup>2</sup> SIMON FRASER UNIVERSITY, BURNABY, CANADA

<sup>3</sup> DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA LOS ANGELES, CA, USA

\* A conference version [10] appeared in the 36th International Colloquium on Automata, Languages and Programming (ICALP 2009). This work was done while the first two authors were postdoctoral fellows at International Computer Science Institute Berkeley supported by the German Academic Exchange Service (DAAD) and the third author was a postdoctoral fellow at the Computer Science Division of the University of California Berkeley.



**Figure 1.** The network  $\text{CCC}_{16}$ .

The local scheme of communication we study is a *balancer* gate: the number of tokens is split as evenly as possible between the communicating vertices with the excess token (if such remains) routed to the vertex towards which the balancer points. More formally, the balancing network consists of  $n$  vertices  $v_1, v_2, \dots, v_n$ , and  $m$  matchings (either perfect or not)  $M_1, M_2, \dots, M_m$ . We associate with every matching edge a balancer gate (that is, we think of the edges as directed edges). At the beginning of the first iteration,  $x_j$  tokens are placed in vertex  $v_j$ , and at every iteration  $r = 1, \dots, m$ , the vertices of the network perform a balancing operation according to the matching  $M_r$  (that is, vertices  $v_i$  and  $v_j$  interact if  $(v_i, v_j) \in M_r$ ).

One motivation for considering smoothing networks comes from the server-client world. Each token represents a client request for some service; the service is provided by the servers residing at the vertices. Routing tokens through the network must ensure that all servers receive approximately the same number of tokens, no matter how unbalanced the initial number of tokens is (cf. [2]). More generally, smoothing networks are attractive for multiprocessor coordination and load balancing applications where low-contention is a requirement; these include *producers-consumers* [11] and distributed numerical computations [3]. Together with *counting networks*, smoothing networks have been studied quite extensively since introduced in the seminal paper of Aspnes et al. [2].

Herlihy and Tirthapura [12, 13] initiated the study of the CCC *network* (cube-connected-cycles, see Figure 1) as a smoothing network<sup>1</sup>. For the special case of the CCC, sticking to previous conventions, we adopt a “topographical” view of the network, thus calling the vertices *wires*, and looking at the left-most side of the network as the “input” and the right-most as

<sup>1</sup>Actually, they considered the so-called *block network*. However, it was observed in [16] that the block network is isomorphic to the CCC-network and therefore we will stick to the latter in the following.

the “output”. In the CCC, two wires at layer  $\ell$  are connected by a balancer if the respective bit strings of the wires differ exactly in bit  $\ell$ . The CCC is a canonical network in the sense that it has the smallest possible depth (number of rounds) of  $\log n$  as a smaller depth cannot ensure any discrepancy independent of the initial one. Moreover, it has been used in more advanced constructions such as the *periodic (counting) network* [2, 6].

As it turns out, the initial setting of the balancers’ directions is crucial. Two popular options are an arbitrary orientation or a uniformly random one. A maximal discrepancy of  $\log n$  was established for the  $\text{CCC}_n$  for an arbitrary initial orientation [13]. For a random initial orientation of the  $\text{CCC}_n$ , [12] show a discrepancy of at most  $2.36\sqrt{\log n}$  for the  $\text{CCC}_n$  (this holds *whp*<sup>2</sup> over the random initialization), which was improved by Mavronicolas and Sauerwald [16] to  $\log \log n + \mathcal{O}(1)$  (and a matching lower bound).

Results for more general networks have been derived in Rabani, Sinclair, and Wanka [17] for arbitrary orientations. For expander graphs, they show an  $\mathcal{O}(\log n)$ -discrepancy after  $\mathcal{O}(\log(Kn))$  rounds when  $K$  is the discrepancy of the initial load vector. This was recently strengthened assuming the orientations are set randomly and in addition the matchings themselves are chosen randomly [9]. Specifically, for expander graphs one can achieve within the same number of rounds a discrepancy of only  $\mathcal{O}(\log \log n)$ .

**1.2. Smoothed analysis.** Let us now turn to the second meaning of “smoothed”. Smoothed analysis comes to bridge between the random instance, which typically has a very specific “unrealistic” structure, and the completely arbitrary instance, which in many cases reflects just the worst case scenario, and is thus over-pessimistic in general. In the smoothed analysis paradigm, first an adversary generates an input instance, then this instance is randomly perturbed.

The smoothed analysis paradigm was introduced by Spielman and Teng in 2001 [19] to help explain why the simplex algorithm for linear programming works well in practice but not in (worst-case) theory. They considered instances formed by taking an arbitrary constraint matrix and perturbing it by adding independent Gaussian noise with variance  $\varepsilon$  to each entry. They showed that, in this case, the shadow-vertex pivot rule succeeds in expected polynomial time. Independently, Bohman, Frieze, and Martin [4] studied the issue of Hamiltonicity in a dense graph when random edges are added. In the context of graph optimization problems we can also mention [8, 14], in the context of  $k$ -SAT [5, 7], and in various other problems [1, 15, 18, 20].

Our work joins this long series of papers studying perturbed instances in a variety of problems. Specifically in our setting we study the following question: what if the balancers were not set completely adversarially but also not in a completely random fashion. Besides the mathematical and analytical challenge that such a problem poses, in real network applications one may not always assume that the random source is unbiased, or in some cases

---

<sup>2</sup>Writing *whp* we mean with probability tending to 1 as  $n$  goes to infinity.

one will not be able to quantitatively measure the amount of randomness involved in the network generation. Still it is desirable to have an estimate of the typical behavior of the network. Although we do not claim that our smoothed-analysis model captures all possible behaviors, it does give a rigorous and tight characterization of the tradeoff between the quality of load balancing and the randomness involved in setting the balancers' directions, under rather natural probabilistic assumptions.

As far as we know, no smoothed analysis framework was suggested to a networking related problem. Formally, we suggest the following framework.

**1.3. The Model.** We define both the smoothed-analysis aspect of the model, and the load-balancing one. For the load balancing part, our model is similar to (and, as we will shortly explain, a generalization of) the periodic balancing circuits studied in [17]. We think of the balancing network in terms of an  $n$ -vertex graph. The processors in the network are the vertices of the graph, and balancers connecting processors are the (directed) edges of the graph.

Before we proceed, since what follows is somewhat heavy on notation and indices, it will be helpful for the reader to bear in mind the following legend: we use superscripts (in round brackets) to denote a time stamp, and subscripts to denote an index. In subscripts, we use the vertices of the graph as indices (thus assuming some ordering of the vertex set). For example,  $\mathbf{A}_{u,v}^{(i)}$  stands for the  $(u, v)$ -entry in matrix  $\mathbf{A}^{(i)}$ , which corresponds to time/round  $i$ .

Let  $M^{(1)}, \dots, M^{(T)}$  be an arbitrary sequence of  $T$  (not necessarily perfect) matchings. With each matching  $M^{(i)}$  we associate a matrix  $\mathbf{P}^{(i)}$  with  $\mathbf{P}_{uv}^{(i)} = 1/2$  if  $u$  and  $v$  are matched in  $M^{(i)}$ ,  $\mathbf{P}_{uu}^{(i)} = 1/2$  if  $u$  is matched in  $M^{(i)}$ ,  $\mathbf{P}_{uu}^{(i)} = 1$  if  $u$  is not matched in  $M^{(i)}$ , and  $\mathbf{P}_{uv}^{(i)} = 0$  otherwise.

In round  $i$ , every two vertices matched in  $M^{(i)}$  perform a balancing operation. That is, the sum of the number of tokens in both vertices is split evenly between the two, with the remaining token (if exists) directed to the vertex pointed by the matching edge.

**Remark 1.1.** In *periodic balancing networks* (see [17] for example) an ordered set of  $d$  (usually perfect) matchings is fixed. Every round of balancing is a successive application of the  $d$  matchings. Our model is a (slight) generalization of the latter.

Let us now turn to the smoothed-analysis part. Given a balancing network consisting of a set  $T$  of directed matchings, an  $\alpha$ -**perturbation** of the network is a flip of direction for every edge with probability  $\alpha$  independently of all other edges. Setting  $\alpha = 0$  gives the completely "adversarial model", and  $\alpha = 1/2$  is the uniform random case.

**Remark 1.2.** For our results, it suffices to consider  $\alpha \in [0, 1/2]$ . The case  $\alpha > 1/2$  can be reduced to the case  $\alpha \leq 1/2$  by flipping the initial orientation

of all balancers and taking  $1 - \alpha$  instead of  $\alpha$ . It is easy to see that both distributions are identical.

**1.4. Our Contribution.** For a load vector  $\mathbf{x}$ , its discrepancy is defined to be  $\max_{u,v} |\mathbf{x}_u - \mathbf{x}_v|$ . We use  $e_u$  to denote the unit vector whose all entries are 0 except the  $u^{\text{th}}$ . For a matrix  $A$ ,  $\lambda(A)$  stands for the second largest eigenvalue of  $A$  (in absolute value). Unless stated otherwise,  $\|z\|$  stands for the  $\ell_2$ -norm of the vector  $z$ . In the following, we will assume an ordering of the vertices from 1 to  $n$ . When we write  $(u, v) \in E$ , we refer to the case where  $u$  and  $v$  are connected by an undirected edge and  $u < v$ .

**Theorem 1.1.** *Let  $G$  be a balancing network with matchings  $M^{(1)}, \dots, M^{(T)}$ . For any two time stamps  $t_1, t_2$  satisfying  $t_1 < t_2 \leq T$ , and any input vector with initial discrepancy  $K$ , the discrepancy at time step  $t_2$  in  $\alpha$ -perturbed  $G$  is whp at most*

$$(t_2 - t_1) + 3 \left( \frac{1}{2} - \alpha \right) t_1 + \Lambda_1 + \Lambda_2,$$

where

$$\begin{aligned} \Lambda_1 &= \max_{w \in V} \sqrt{\log n \sum_{i=1}^{t_1} \sum_{[u:v] \in M^{(i)}} \left( (e_u - e_v)^T \left( \prod_{j=i+1}^{t_2} \mathbf{P}^{(j)} \right) e_w \right)^2}, \\ \Lambda_2 &= \lambda \left( \prod_{i=1}^{t_2} \mathbf{P}^{(i)} \right) \sqrt{n} K. \end{aligned}$$

Before we proceed let us motivate the result stated in Theorem 1.1. There are two factors that affect the discrepancy: the fact that tokens are indivisible (and therefore the balancing operation may not be “perfect”), and how many balancing rounds are there. On the one hand, the more rounds there are the more balancing operations are carried, and the smoother the output is. On the other hand, the longer the process runs, its susceptibility to rounding errors and arbitrary placement of excess tokens increases. This is however only a seemingly tension, as indeed the more rounds there are, the smoother the output is. Nevertheless, in the analysis (at least as we carry it), this tension plays part. Specifically, optimizing over these two contesting tendencies is reflected in the choice of  $t_1$  and  $t_2$ .  $\Lambda_2$  is the contribution resulting from the number of balancing rounds being bounded, and  $\Lambda_1$ , along with the first two terms, account for the indivisibility of the tokens. In the cases that will interest us,  $t_1, t_2$  will be chosen so that  $\Lambda_1, \Lambda_2$  will be low-order terms compared to the first two terms.

Our Theorem 1.1 also implies the following results:

- For the aforementioned periodic setting Theorem 1.1 implies the following: after  $\mathcal{O}(\log(Kn)/\nu)$  rounds ( $\nu = (1 - \lambda(\mathbf{Q}))^{-1}$ ),  $\mathbf{Q}$  is the so-called round matrix which corresponds to one period,  $K$  the initial discrepancy) the discrepancy is whp at most

$$\mathcal{O} \left( \frac{d \log(Kn)}{\nu} \cdot \left( \frac{1}{2} - \alpha \right) + \frac{d \log \log n}{\nu} \right).$$

Setting  $\alpha = 0$  (and assuming  $K$  is polynomial in  $n$ ) we get the result of [17]<sup>3</sup>, and for  $\alpha = 1/2$  we get the result of [9]. (The restriction on  $K$  being polynomial can be lifted but at the price of more cumbersome expressions in Theorem 1.1. Arguably, the interesting cases are anyway when the total number of tokens, and in particular  $K$ , is polynomial). Complete details in Section 4.

- For the  $\text{CCC}_n$ , after  $\log n$  rounds the discrepancy is *whp* at most

$$3 \left( \frac{1}{2} - \alpha \right) \log n + \log \log n + \mathcal{O}(1).$$

Full details in Section 5.

Let us now turn to the lower bound. Here we consider the all-up-orientation of the balancers of a  $\text{CCC}_n$  meaning that before the  $\alpha$ -perturbation, all balancers are directed to the wire with a smaller number.

**Theorem 1.2.** *Consider a  $\text{CCC}_n$  with the all-up orientation of the balancers and assume that the number of tokens at each wire is uniformly distributed over  $\{0, 1, \dots, n-1\}$  (independently at each wire). The discrepancy of the  $\alpha$ -perturbed network is *whp* at least*

$$\max \left\{ \left( \frac{1}{2} - \alpha \right) \log n - 2 \log \log n, (1 - o(1)) (\log \log n)/2 \right\}.$$

Two more points to note regarding the lower bound:

- For  $\alpha = 0$  our lower bound matches the experimental findings of [12]. The authors examined CCC's of size up to  $2^{24}$  where all balancers are set in the same direction and the number of tokens at each input is a random number between 1 and 100,000. Their observation was that the average discrepancy is close to  $(\log n)/2$  (which matches our lower bound with  $\alpha = 0$ ).
- The input distribution that we use for the lower bound is arguably more natural than the tailored and somewhat artificial ones used in previous lower bound proofs [13, 16].

Finally, we state a somewhat more technical result that we obtain, which lies in the heart of the proof of the lower bound and sheds light on the mechanics of the CCC in the average case input. In what follows, for a balancer  $\mathbf{b}$ , we let  $\text{Odd}(\mathbf{b})$  be an indicator function which is 1 if  $\mathbf{b}$  had an excess token. By  $\mathcal{B}_i$  we denote the set of balancers that affect output wire  $i$  (that is, there is a path through consecutive layers from the balancers to the output wire  $i$ ).  $\mathcal{B}_i(\ell)$  is the restriction of  $\mathcal{B}_i$  to balancers in layer  $\ell$ .

**Lemma 1.3.** *Consider a  $\text{CCC}_n$  network with any fixed orientation of the balancers. Assume that the number of tokens at each wire is uniformly distributed over  $\{0, 1, \dots, n-1\}$  (independently at each wire). Then every balancer  $\mathbf{b}$  in layer  $\ell$ ,  $1 \leq \ell \leq \log n$ , satisfies the following properties:*

<sup>3</sup>We point out that in the original statement in [17, Corollary 5], only the number of periods is counted. Hence, in their statement the number of rounds is by a factor of  $d$  smaller.

- $\Pr[\text{Odd}(\mathbf{b}) = 1] = 1/2$ ,
- moreover, for every wire  $i$ ,  $\{\text{Odd}(\mathbf{b}) \mid \mathbf{b} \in \mathcal{B}_i\}$  is a set of independent random variables.

The proof of the lemma is given in Section 2.3. Let us remark at this point that the lemma holding under such weak conditions is rather surprising. First, it is valid regardless of the given orientation. Secondly, the  $\text{Odd}$ 's of the balancers that affect the *same* output wire are independent. While this is obvious for balancers that are in the same layer, it seems somewhat surprising for balancers in subsequent layers that are connected.

**1.5. Paper's Organization.** The remainder of the paper is organized as follows. We set out with the proof of Theorem 1.2 in Section 2, preceding the proof of Theorem 1.1 in Section 3. The reason is that the lower bound is concerned with the  $\text{CCC}_n$  (a special case of our general model). The techniques used in the proof of the lower bound serve as a good introduction to the more complicated proof of the upper bound. In Sections 4 and 5 we show how to derive the special cases of the periodic balancing network and the  $\text{CCC}_n$  from Theorem 1.1. Finally we present experimental results that we obtain for the  $\text{CCC}_n$  in Section 6.

## 2. PROOF OF THE LOWER BOUND

As we mentioned before, for the special case of the  $\text{CCC}_n$  we adopt a “topographical” view of the network: calling the vertices *wires*, the time steps *layers*, the left-most side of the network the “input” and the right-most the “output”.

The proof outline is the following. Given an input vector  $\mathbf{x}$  (uniformly distributed over the range  $\{0, \dots, n-1\}$ ), we shall calculate the expected divergence from the average load  $\mu = \|\mathbf{x}\|_1/n$ . The expectation is taken over both the smoothing operation and the input. After establishing the “right” order of divergence (in expectation) we shall prove a concentration result. One of the main keys to estimating the expectation is Lemma 1.3 saying that if the input is uniformly distributed as above, then for every balancer  $\mathbf{b}$ ,  $\Pr[\text{Odd}(\mathbf{b}) = 1] = 1/2$  (the probability is taken only over the input).

Before proceeding with the proof, let us introduce some further notation. Let  $y_1$  be the number of tokens exiting on the top output wire of the network. For any balancer  $\mathbf{b}$ ,  $\Psi(\mathbf{b})$  is an indicator random variable which takes the value  $-1/2$  if the balancer  $\mathbf{b}$  was perturbed, and  $1/2$  otherwise (Recall that all balancers are pointing up before the perturbation takes place).

Using the “standard” backward (recursive) unfolding (see also [12, 16] for a concrete derivation for the  $\text{CCC}_n$ ) we obtain that,

$$y_1 = \mu + \sum_{\ell=1}^{\log n} 2^{-\log n + \ell} \sum_{\mathbf{b} \in \mathcal{B}_1(\ell)} \text{Odd}(\mathbf{b}) \cdot \Psi(\mathbf{b}).$$

The latter already implies that the discrepancy of the entire network is at least

$$y_1 - \mu = \sum_{\ell=1}^{\log n} 2^{-\log n + \ell} \sum_{\mathbf{b} \in \mathcal{B}_1(\ell)} \text{Odd}(\mathbf{b}) \cdot \Psi(\mathbf{b}),$$

because there is at least one wire whose output has at most  $\mu$  tokens (a further improvement of a factor of 2 will be obtained by considering additionally the bottom output wire and proving that on this wire only a small number of tokens exit).

Write  $y_1 - \mu = \sum_{\ell=1}^{\log n} S_\ell$ , defining for each layer  $1 \leq \ell \leq \log n$ ,

$$(1) \quad S_\ell := 2^{-\log n + \ell} \sum_{\mathbf{b} \in \mathcal{B}_1(\ell)} \text{Odd}(\mathbf{b}) \cdot \Psi(\mathbf{b}).$$

**2.1. Proof of  $(\frac{1}{2} - \alpha) \log n - 2 \log \log n$ .** We now turn to bounding the expected value of  $S_\ell$ . Using the following facts: (a) the  $\text{Odd}(\mathbf{b})$  and  $\Psi(\mathbf{b})$  are independent (b) Lemma 1.3 which gives  $\mathbf{E}[\text{Odd}(\mathbf{b})] = 1/2$ , (c) the simple fact that  $\mathbf{E}[\Psi(\mathbf{b})] = \frac{1}{2} - \alpha$  and (d) the fact that in layer  $\ell$  there are  $2^{\log n - \ell}$  balancers which affect output wire 1 (this is simply by the structure of the  $\text{CCC}_n$ ), we get

$$\begin{aligned} \mathbf{E}[S_\ell] &= 2^{-\log n + \ell} \sum_{\mathbf{b} \in \mathcal{B}_1(\ell)} \frac{1}{2} \cdot \left(\frac{1}{2} - \alpha\right) \\ &= 2^{-\log n + \ell} \cdot 2^{\log n - \ell} \cdot \frac{1}{2} \cdot \left(\frac{1}{2} - \alpha\right) \\ &= \frac{1}{2} \left(\frac{1}{2} - \alpha\right). \end{aligned}$$

This in turn gives that

$$\mathbf{E}[y_1 - \mu] = \mathbf{E}\left[\sum_{\ell=1}^{\log n} S_\ell\right] = \frac{1}{2} \left(\frac{1}{2} - \alpha\right) \log n.$$

Our next goal is to claim that typically the discrepancy behaves like the expectation; in other words, a concentration result. Specifically, we apply Hoeffdings bound to each layer  $S_\ell$  separately. It is applicable as the random variables  $2^{-\log n + \ell} \cdot \text{Odd}(\mathbf{b}) \cdot \Psi(\mathbf{b})$  are independent for balancers within the same layer (such balancers concern disjoint sets of input wires, and the input to the network was chosen independently for each wire). For the bound to be useful we need the range of values for the random variables to be small. Thus, in the probabilistic argument, we shall be concerned only with the first  $\log n - \log \log n$  layers (the last  $\log \log n$  layers we shall bound deterministically). We use the following Hoeffding bound:

**Lemma 2.1** (Hoeffdings Bound). *Let  $Z_1, Z_2, \dots, Z_n$  be a sequence of independent random variables with  $Z_i \in [a_i, b_i]$  for each  $i$ . Then for any number  $\varepsilon \geq 0$ ,*

$$\Pr \left[ \left| \sum_{i=1}^n Z_i - \mathbf{E} \left[ \sum_{i=1}^n Z_i \right] \right| \geq \varepsilon \right] \leq 2 \cdot \exp \left( - \frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2} \right).$$



For any random variable  $X$ , let  $\mathbf{Range}[X]$  be the difference between the maximum and minimum value that  $X$  can attain. For a balancer  $\mathbf{b} \in \mathcal{B}_1(\ell)$  we plug in,

$$\begin{aligned} Z_{\mathbf{b}} &= 2^{-\log n + \ell} \cdot \mathbf{Odd}(\mathbf{b}) \cdot \Psi(\mathbf{b}), \\ \varepsilon &= 2^{(\ell - \log n + \log \log n)/2}, \\ \mathbf{Range}[Z_{\mathbf{b}}]^2 &= \left(2^{\ell - \log n}\right)^2, \end{aligned}$$

and the sum is over  $2^{\log n - \ell}$  balancers in layer  $\ell$ . Therefore,

$$\begin{aligned} \Pr \left[ |S_{\ell} - \mathbf{E}[S_{\ell}]| \geq 2^{(\ell - \log n + \log \log n)/2} \right] \\ \leq 2 \exp \left( - \frac{2 \cdot 2^{\ell - \log n + \log \log n}}{2^{\ell - \log n}} \right) \\ \leq n^{-1}. \end{aligned}$$

In turn, with probability at least  $1 - \log n/n$  (take the union bound over at most  $\log n$   $S_{\ell}$  terms):

$$\sum_{\ell=1}^{\log n - \log \log n} S_{\ell} \geq \frac{1}{2} \left( \frac{1}{2} - \alpha \right) (\log n - \log \log n) - \sum_{\ell=1}^{\log n - \log \log n} 2^{(\ell - \log n + \log \log n)/2}.$$

The second term is just a geometric series with quotient  $\sqrt{2}$ , and therefore can be bounded by  $\frac{1}{1-1/\sqrt{2}} < 4$ .

For the last  $\log \log n$  layers, we have that for every  $\ell$ ,  $|S_{\ell}|$  cannot exceed  $\frac{1}{2}$ , and therefore their contribution, in absolute value is at most  $\frac{1}{2} \log \log n$ . Wrapping it up, *whp*

$$y_1 - \mu = \sum_{\ell=1}^{\log n} S_{\ell} \geq \frac{1}{2} \left( \frac{1}{2} - \alpha \right) (\log n - \log \log n) - 4 - \frac{1}{2} \log \log n.$$

The same calculation implies that the number of tokens at the *bottom*-most output wire deviates from  $\mu$  in the same way (just in the opposite direction), so

$$y_n - \mu \leq -\frac{1}{2} \left( \frac{1}{2} - \alpha \right) (\log n - \log \log n) + 4 + \frac{1}{2} \log \log n.$$

Hence, the discrepancy is *whp* lower bounded by (using the union bound over the top and bottom wire, and not claiming independence)

$$\begin{aligned} y_1 - y_n &\geq \left( \frac{1}{2} - \alpha \right) \log n - 8 - \left( \frac{3}{2} - \alpha \right) \log \log n \\ &\geq \left( \frac{1}{2} - \alpha \right) \log n - 2 \log \log n. \end{aligned}$$

**2.2. Proof of  $(1 - o(1)) \log \log n/2$ .** The proof here goes along similar lines to Section 2.1, only that now we choose the set of balancers we apply it to more carefully. By the structure of the  $\text{CCC}_n$ , the last  $x$  layers form the parallel cascade of  $n/2^x$  independent CCC subnetworks each of which has  $2^x$  wires (by independent we mean that the sets of balancers are disjoint).

We call a subnetwork *good* if after an  $\alpha$ -perturbation of the all-up initial orientation, all the balancers affecting the top (or bottom) output wire were not flipped (that is, still point up).

The first observation that we make is that *whp* (for a suitable choice of  $x$ , to be determined shortly) at least one subnetwork is good. Let us prove this fact.

The number of balancers affecting the top (or bottom) wire in one of the subnetworks is  $\sum_{\ell=1}^x 2^{\ell-1} \leq 2^x$ . In total, there are no more than  $2^{x+1}$  affecting both wires. The probability that none of these balancers was flipped is (using our assumption  $\alpha \leq 1/2$ )  $(1 - \alpha)^{2^{x+1}} \geq 2^{-2^{x+1}}$ . Choosing  $x = \log \log n - 2$ , this probability is at least  $n^{-1/2}$ ; there are at least  $n/\log n$  such subnetworks, thus the probability that none is good is at most

$$\left(1 - n^{-1/2}\right)^{n/\log n} = o(1).$$

Fix one good subnetwork and let  $\mu'$  be the average load at the input to that subnetwork. Repeating the arguments from Section 2.1 (with  $\alpha = 0$ ,  $\log n$  re-scaled to  $x = \log \log n - 2$ , and now using the second item in Lemma 1.3 which guarantees that the probability of  $\text{Odd}(\cdot) = 1$  is still  $1/2$ , for any orientation of the balancers) gives that in the top output wire of the subnetwork there are *whp* at least  $\mu' + (\log \log n)/4 - \mathcal{O}(\log \log \log n)$  tokens, while on the bottom output wire there are *whp* at most  $\mu' - (\log \log n)/4 + \mathcal{O}(\log \log \log n)$  tokens. Using the union bound, the discrepancy is *whp* at least their difference, that is, at least  $(\log \log n)/2 - \mathcal{O}(\log \log \log n)$ .

**2.3. Proof of Lemma 1.3.** The following observation is the key idea in proving Lemma 1.3. Recall that  $\mathcal{B}_j$  stands for the set of balancers that affect wire  $j$ . For a balancer  $\mathbf{b}$  in layer  $\ell$  let  $A(\mathbf{b})$  describe an assignment of  $\text{Odd}(\mathbf{b}')$  values for all balancers  $\mathbf{b}'$  in preceding layers that affect  $\mathbf{b}$  (that is, there is a path from  $\mathbf{b}'$  to  $\mathbf{b}$  through consecutive layers).

**Lemma 2.2.** *Consider a  $\text{CCC}_n$  network with any orientation of the balancers. Assume that the number of tokens at each wire is uniformly distributed over  $\{0, 1, \dots, n-1\}$ . Consider a balancer  $\mathbf{b}$  in layer  $\ell$  with  $1 \leq \ell \leq \log n$  and let  $x_1, x_2$  denote the two input wires that go into  $\mathbf{b}$ . Then for any assignment  $A(\mathbf{b})$ ,  $x_i \bmod (n/2^{\ell-1}) \mid A(\mathbf{b})$  is uniformly distributed over  $\{0, 1, \dots, (n/2^{\ell-1}) - 1\}$ , for  $i = 1, 2$ .*

The lemma easily implies Lemma 1.3: Consider a balancer  $\mathbf{b}$  in layer  $\ell$ , with  $1 \leq \ell \leq \log n$  with two inputs  $x_1, x_2$ . Lemma 2.2 implies that  $x_i \bmod (n/2^{\ell-1})$  is uniformly distributed over  $\{0, 1, \dots, n/2^{\ell-1} - 1\}$ , and in particular is uniform over that range mod 2. Hence both  $x_1, x_2$  are odd/even with probability  $1/2$ . Further observe that by the structure of the CCC network, the input wires  $x_1, x_2$  depend on disjoint sets of wires and balancers. By the two latter facts it follows that the sum  $x_1 + x_2$  is odd (or even) with probability  $1/2$ , or in other words  $\Pr[\text{Odd}(\mathbf{b}) = 1] = 1/2$ . The independence part follows from the fact that this is true for every conditioning

$A(\mathbf{b})$  on balancers from previous layers that affect  $\mathbf{b}$  (by Lemma 2.2) and the fact that balancers in the same layer that affect the same output wire are independent by construction, as those balancers depend on disjoint sets of balancers and disjoint sets of input wires.

*Proof of Lemma 2.2.* We prove the lemma by induction on  $\ell$ , the layer of the balancer. The base case is immediate: the input to a balancer in layer 1 is just the original input, which is by definition distributed uniformly over  $\{0, 1, \dots, n/2^{1-1} - 1\}$  which is simply  $\{0, 1, \dots, n - 1\}$ . Assume the lemma is true for all balancers in layer  $\ell$  and consider a balancer in layer  $\ell + 1$ . Let  $x_1, x_2$  be its two input wires. By the structure of the CCC, the value on each wire is determined by a disjoint set of balancers and preceding wires, therefore we can treat, w.l.o.g., only  $x_1$ . Let  $A_1(\mathbf{b})$  be the part of  $A(\mathbf{b})$  that affects  $x_1$ . ( $A_1(\mathbf{b})$  would be the set of balancers so that there is a path from them to  $x_1$ . Since the initial load on the input wires is chosen independently at every wire, and the sets of balancers affecting  $x_1$  and  $x_2$  are disjoint, indeed only  $A_1(b)$  affects  $x_1$ , and the same applies for  $A_2(b)$  and  $x_2$ ). Thus our goal is to calculate

$$\Pr \left[ x_1 \equiv k \bmod (n/2^\ell) \mid A_1(\mathbf{b}) \right].$$

Let  $\mathbf{b}'$  be the balancer in layer  $\ell$  whose one outlet is  $x_1$ , and let  $x'_1, x'_2$  be its two inputs. Recall that  $x_1 = \lfloor (x'_1 + x'_2)/2 \rfloor$ , and a possible  $+1$  addition in case this sum is odd and the balancer points in the direction of  $x_1$ . Furthermore it is easy to verify that for every  $a$  (assume  $a$  is even, if odd write  $a + 1$  instead)

$$a/2 \equiv k \bmod (n/2^\ell) \Leftrightarrow a \equiv 2k \bmod (n/2^{\ell-1}).$$

Therefore for the event  $x_1 \equiv k \bmod (n/2^\ell)$  to occur, either

- $x'_1 + x'_2 \equiv 2k \bmod (n/2^{\ell-1})$ , or
- $x'_1 + x'_2 \equiv 2k - 1 \bmod (n/2^{\ell-1})$  (assuming w.l.o.g. that the balancer points towards  $x_1$ , otherwise the sum equals  $2k + 1$ ).

Let us consider the first case.

$$\begin{aligned} & \Pr \left[ x'_1 + x'_2 \equiv 2k \bmod (n/2^{\ell-1}) \mid A_1(b) \right] \\ &= \Pr \left[ \bigvee_{i=0}^{n/2^{\ell-1}-1} x'_1 \equiv i \bmod (n/2^{\ell-1}) \wedge x'_2 \equiv 2k - i \bmod (n/2^{\ell-1}) \mid A_1(b) \right]. \end{aligned}$$

Now observe that the values of  $x'_1$  and  $x'_2$  are determined independently, as again, by the structure of the CCC, they involve disjoint sets of balancers and input wires. Similarly to  $A_1(b)$  we can define  $A'_1(b)$  and  $A'_2(b)$  which correspond to the parts of  $A$  affecting  $x'_1$  and  $x'_2$ . By the structure of the CCC,  $A'_1(b)$  and  $A'_2(b)$  depend on a disjoint set of input wires (and balancers). As the input is chosen independently for every wire,  $A'_2(b)$  does not affect

$x'_1$  and similarly  $A'_1(b)$  does not affect  $x'_2$ . Thus the latter reduces to

$$\sum_{i=0}^{n/2^{\ell-1}-1} \Pr \left[ x'_1 \equiv i \bmod \left( n/2^{\ell-1} \right) \mid A'_1(b) \right] \\ \cdot \Pr \left[ x'_2 \equiv 2k - i \bmod \left( n/2^{\ell-1} \right) \mid A'_2(b) \right].$$

By the induction hypothesis (applied to the  $x'_i$  at layer  $\ell - 1$ ), for every  $i$ , each of  $\Pr \left[ x'_1 \equiv i \bmod \left( n/2^{\ell-1} \right) \mid A'_1(b) \right]$  and  $\Pr \left[ x'_2 \equiv 2k - i \bmod \left( n/2^{\ell-1} \right) \mid A'_2(b) \right]$  is uniformly distributed over the range  $\{0, \dots, n/2^{\ell-1} - 1\}$ , and therefore in particular the entire expression does not depend on  $k$ , or, put differently is the same for every choice of  $k$ . The same argument holds for the case  $x'_1 + x'_2 = 2k - 1 \bmod \left( n/2^{\ell-1} \right)$ . This completes the proof.  $\square$

### 3. PROOF OF THE UPPER BOUND

We shall derive our bound by measuring the difference between the number of tokens at any vertex and the average load (as we did in the proof of the lower bound for the  $\text{CCC}_n$ ). Specifically we shall bound  $\max_i |y_i^{(t)} - \mu|$ ,  $y_i^{(t)}$  being the number of tokens at vertex  $i$  at time  $t$  (we use  $\mathbf{y}^{(t)} = (y_i^{(t)})_{i \in V}$  for the vector of loads at time  $t$ ). There are two contributions to the divergence from  $\mu$  (which we analyze separately):

- The divergence of the idealized process from  $\mu$  due to its finiteness.
- The divergence of the actual process from the idealized process due to indivisibility.

The idea to compare the actual process to an idealized one was suggested in [17] and was combined with convergence results of Markov chains. Though we were inspired by the basic setup from [17] and the probabilistic analysis from [9], our setting differs in a crucial point: when dealing with the case  $0 < \alpha < 1/2$ , we get a delicate mixture of the deterministic and the random model. For example, the random variables in our analysis are not symmetric anymore which leads to additional technicalities (cf. Lemma 3.2).

Formally, let  $\xi^{(t)}$  be the load vector of the idealized process at time  $t$ , then by the triangle inequality ( $\mathbf{1}$  is the all-one vector)

$$\|\mathbf{y}^{(t)} - \mu \mathbf{1}\|_{\infty} \leq \|\mathbf{y}^{(t)} - \xi^{(t)}\|_{\infty} + \|\xi^{(t)} - \mu \mathbf{1}\|_{\infty}.$$

**Proposition 3.1.** *Let  $G$  be a balancing network with matchings  $M^{(1)}, \dots, M^{(T)}$ . Then,*

- $\|\xi^{(t_2)} - \mu \mathbf{1}\|_{\infty} \leq \Lambda_2$ ,
- *whp over the  $\alpha$ -perturbation operation,*  
 $\|\mathbf{y}^{(t_2)} - \xi^{(t_2)}\|_{\infty} \leq (t_2 - t_1) + 3 \left( \frac{1}{2} - \alpha \right) t_1 + \Lambda_1.$

Theorem 1.1 then follows. The proof of the first part of the proposition consists of standard spectral arguments and is given in Section 3.1 for completeness. The proof the second part is more involved and is given in Section 3.2.

**3.1. Proof of Proposition 3.1: Bounding  $\|\xi^{(t_2)} - \mu \mathbf{1}\|_\infty$ .** Letting  $\xi^{(0)}$  be the initial load vector, it is easily seen that

$$\xi^{(t_2)} = \xi^{(0)} \mathbf{P}^{(1)} \mathbf{P}^{(2)} \dots \mathbf{P}^{(t_2)},$$

where  $\mathbf{P}^{(i)}$  is the matrix corresponding to matching  $M^{(i)}$  (as defined in Section 1.3). For simplicity let us abbreviate

$$\mathbf{P}^{[i,j]} := \mathbf{P}^{(i)} \mathbf{P}^{(i+1)} \mathbf{P}^{(i+2)} \dots \mathbf{P}^{(j)}.$$

Since  $\mathbf{P}^{[1,t_2]}$  is real valued and symmetric (as each of the  $\mathbf{P}^{(i)}$ 's is), it has  $n$  real-valued eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  whose corresponding eigenvectors  $v_1, \dots, v_n$  form an orthogonal basis of  $\mathbb{R}^n$ . Next we observe that

$$\xi^{(t)} - \mu \mathbf{1} = \xi^{(0)} \mathbf{P}^{[1,t_2]} - \mu \mathbf{1} \mathbf{P}^{[1,t_2]} = (\xi^{(0)} - \mu \mathbf{1}) \mathbf{P}^{[1,t_2]},$$

since  $\mathbf{1}$  is an eigenvector of  $\mathbf{P}^{[1,t_2]}$  corresponding to  $\lambda_1 = 1$ . Furthermore  $\xi^{(0)} \cdot \mathbf{1}$  is just the total (initial) number of tokens, and therefore by definition we get  $(\xi^{(0)} - \mu \mathbf{1}) \cdot \mathbf{1} = 0$ . Finally, let us project  $\xi^{(0)} - \mu \mathbf{1}$  onto  $v_1, \dots, v_n$ , that is, write  $\xi^{(0)} - \mu \mathbf{1} = \sum_{i=2}^n c_i v_i$  ( $c_1 = 0$  as we said). For our goal to bound  $\|\xi^{(t_2)} - \mu \mathbf{1}\|_\infty$ , it suffices to bound  $\|\xi^{(t_2)} - \mu \mathbf{1}\|$  (recall that  $\|\cdot\|$  refers to the  $\ell_2$ -norm) as for every vector  $z$ ,  $\|z\|_\infty \leq \|z\|$ . By the above,

$$\begin{aligned} \|\xi^{(t_2)} - \mu \mathbf{1}\| &= \|(\xi^{(0)} - \mu \mathbf{1}) \mathbf{P}^{[1,t_2]}\| \\ &= \left\| \sum_{i=2}^n c_i v_i \cdot \mathbf{P}^{[1,t_2]} \right\| \\ &= \left\| \sum_{i=2}^n c_i \lambda_i v_i \right\|. \end{aligned}$$

Recall that  $\lambda(\mathbf{P}^{[1,t_2]})$  denotes the second largest eigenvalue of  $\mathbf{P}^{[1,t_2]}$  in absolute value. By the definition of the  $\ell_2$ -norm, and using the fact that the  $v_i$ 's form an orthogonal basis, the latter equals

$$\begin{aligned} \left( \sum_{i=2}^n c_i^2 \lambda_i^2 \|v_i\|^2 \right)^{1/2} &\leq \lambda(\mathbf{P}^{[1,t_2]}) \cdot \left( \sum_{i=2}^n c_i^2 \|v_i\|^2 \right)^{1/2} \\ &= \lambda(\mathbf{P}^{[1,t_2]}) \cdot \|\xi^{(0)} - \mu \mathbf{1}\| \\ &\leq \lambda(\mathbf{P}^{[1,t_2]}) K \sqrt{n}. \end{aligned}$$

**3.2. Proof of Proposition 3.1: Bounding  $\|\mathbf{y}^{(t_2)} - \xi^{(t_2)}\|_\infty$ .** The proof of this part resembles in nature the proof of Theorem 1.2. Assuming an ordering of  $G$ 's vertices, for a balancer  $\mathbf{b}$  in round  $t$ ,  $\mathbf{b} = (u, v)$ ,  $u < v$ , we set  $\Phi_{u,v}^{(t)} = 1$  if the initial direction (before the perturbation) is  $u \rightarrow v$  and  $-1$  otherwise (in the lower bound we considered the all-up orientation thus we had no use of these variables). As in Section 2, for a balancer  $\mathbf{b} = (u, v)$  in round  $t$ , the random variable  $\Psi_{u,v}^{(t)}$  is  $-1/2$  if the balancer is perturbed and  $1/2$  otherwise. Using these notations we define a *rounding vector*  $\rho^{(t)}$ , which accounts for the rounding errors in step  $t$ . Formally,

$$\rho_u^{(t)} = \begin{cases} \text{Odd}(y_u^{(t-1)} + y_v^{(t-1)}) \cdot \Psi_{u,v}^{(t)} \cdot \Phi_{u,v}^{(t)} & \text{if } u \text{ and } v \text{ are matched in } M^{(t)} \text{ and } u < v, \\ -\text{Odd}(y_u^{(t-1)} + y_v^{(t-1)}) \cdot \Psi_{v,u}^{(t)} \cdot \Phi_{v,u}^{(t)} & \text{if } u \text{ and } v \text{ are matched in } M^{(t)} \text{ and } u > v, \\ 0 & \text{if } u \text{ is unmatched.} \end{cases}$$

Now we can write the actual process as follows:

$$(2) \quad \mathbf{y}^{(t)} = \mathbf{y}^{(t-1)} \mathbf{P}^{(t)} + \rho^{(t)}.$$

Let  $M_{\text{Even}}^{(t)}$  be the set of balancers at time  $t$  with no excess token, and  $M_{\text{Odd}}^{(t)}$  the ones with. We can rewrite  $\rho^{(t)}$  as follows:

$$\rho^{(t)} = \sum_{(u,v) \in M_{\text{Odd}}^{(t)}} \Psi_{u,v}^{(t)} \cdot \Phi_{u,v}^{(t)} \cdot (e_u - e_v).$$

Unfolding equation (2), yields then

$$\mathbf{y}^{(t)} = \mathbf{y}^{(0)} \mathbf{P}^{[1,t]} + \sum_{i=1}^t \rho^{(i)} \mathbf{P}^{[i+1,t]},$$

where  $\mathbf{P}^{[2,1]} = \mathbf{I}$ . Observe that  $\mathbf{y}^{(0)} \mathbf{P}^{[1,t]}$  is just  $\xi^{(t)}$  (as  $\xi^{(0)} = \mathbf{y}^{(0)}$ ), and therefore

$$\begin{aligned} \mathbf{y}^{(t)} - \xi^{(t)} &= \sum_{i=1}^t \rho^{(i)} \mathbf{P}^{[i+1,t]} \\ &= \sum_{i=1}^t \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \cdot \Phi_{u,v}^{(i)} \cdot (e_u - e_v) \cdot \mathbf{P}^{[i+1,t]}. \end{aligned}$$

In turn,

$$(3) \quad (\mathbf{y}^{(t)} - \xi^{(t)})_w = \sum_{i=1}^t \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \cdot \Phi_{u,v}^{(i)} \cdot \left( \mathbf{P}_{u,w}^{[i+1,t]} - \mathbf{P}_{v,w}^{[i+1,t]} \right).$$

Our next task is to bound equation (3) to receive the desired term from Proposition 3.1. We do that similar in spirit to the way we went around in Section 2.1. We break this sum into its first  $t_1$  summands (whose expected sum we calculate and to which we apply a large-deviation-bound). The remaining  $(t - t_1)$  terms are bounded deterministically.

One major difficulty in the general setting is that Lemma 1.3 (which was crucial in the proof of Theorem 1.2) does not hold in general as its proof makes substantial use of the special structure of the CCC.

Equation 3 with  $t = t_2$  yields

$$(x^{(t_2)} - \xi^{(t_2)})_w = \sum_{i=1}^{t_2} \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} \left( \mathbf{P}_{u,w}^{[i+1,t_2]} - \mathbf{P}_{v,w}^{[i+1,t_2]} \right).$$

With  $e_{u,v}$  denoting the row-vector with  $+1$  at  $u$ -th column and  $-1$  at  $v$ -th column and zeros elsewhere, we can rewrite and split this equation as follows:

$$\begin{aligned} (x^{(t_2)} - \xi^{(t_2)})_w &= \sum_{i=1}^{t_1} \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} (e_{u,v} \mathbf{P}^{[i+1,t_2]} e_w) \\ &\quad + \sum_{i=t_1+1}^{t_2} \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} (e_{u,v} \mathbf{P}^{[i+1,t_2]} e_w). \end{aligned}$$

Clearly,

$$\begin{aligned} &\sum_{i=t_1+1}^{t_2} \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} (e_{u,v} \mathbf{P}^{[i+1,t_2]} e_w) \\ &= \sum_{i=t_1+1}^{t_2} \left( \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} e_{u,v} \right) \cdot (\mathbf{P}^{[i+1,t_2]} e_w). \end{aligned}$$

Observe that  $\sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} e_{u,v}$  is a vector all of whose entries are bounded by 1 in absolute value. Since  $\mathbf{P}^{[i+1,t_2]}$  is a stochastic matrix, the sum of all entries of the  $w$ -th column of  $\mathbf{P}^{[i+1,t_2]}$ , which is  $|\mathbf{P}^{[i+1,t_2]} e_w|_1$ , is exactly one, hence

$$\left| \sum_{i=t_1+1}^{t_2} \left( \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} e_{u,v} \right) \cdot (\mathbf{P}^{[i+1,t_2]} e_w) \right| \leq (t_2 - t_1).$$

It remains to bound

$$W_{\text{Odd}} := \sum_{i=1}^{t_1} \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} (e_{u,v} \mathbf{P}^{[i+1,t_2]} e_w).$$

Because  $W_{\text{Odd}}$  is not necessarily a sum of independent random variables, it will be more convenient to work with the following quantity (which is a sum of independent random variables, as it assumes that every balancer gets an excess token),

$$W := \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} (e_{u,v} \mathbf{P}^{[i+1,t_2]} e_w).$$

So our strategy is first to bound the deviation of  $W$  from its mean by Hoeffdings Bound and then apply the following lemma (whose proof is in Section 3.3), which justifies using  $W$  instead of  $W_{\text{Odd}}$ .

**Lemma 3.2.** Fix  $0 \leq \alpha \leq \frac{1}{2}$ . For all  $t_1, t_2$  with  $t_1 < t_2 \leq T$  and arbitrary weights  $w_{u,v}^{(i)} \in \mathbb{R}$ , let

$$\begin{aligned} W_{\text{Odd}} &:= \sum_{i=t_1}^{t_2} \sum_{(u,v) \in M_{\text{Odd}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} w_{u,v}^{(i)}, \\ W_{\text{Even}} &:= \sum_{i=t_1}^{t_2} \sum_{(u,v) \in M_{\text{Even}}^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} w_{u,v}^{(i)}, \\ W &:= \sum_{i=t_1}^{t_2} \sum_{(u,v) \in M^{(i)}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} w_{u,v}^{(i)}. \end{aligned}$$

Then for any  $\delta > 0$ ,

$$\Pr [|W_{\text{Odd}}| \geq \delta + 2 \max\{|\mathbf{E}[W^+]|, |\mathbf{E}[W^-]|\}] \leq 16 \Pr [|W| \geq \delta],$$

where

$$\begin{aligned} W^+ &:= \sum_{i=t_1}^{t_2} \sum_{\substack{(u,v) \in M^{(i)} \\ \Phi_{u,v}^{(i)} w_{u,v}^{(i)} > 0}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} w_{u,v}^{(i)}, \\ W^- &:= \sum_{i=t_1}^{t_2} \sum_{\substack{(u,v) \in M^{(i)} \\ \Phi_{u,v}^{(i)} w_{u,v}^{(i)} < 0}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} w_{u,v}^{(i)}. \end{aligned}$$

In order to apply Lemma 3.2 we first need to bound the expectation:

$$\begin{aligned} |\mathbf{E}[W]| &= \left| \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \mathbf{E} \left[ \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} \right] e_{u,v} \mathbf{P}^{[i+1, t_2]} e_w \right| \\ &\leq \max_{i=1}^{t_1} \max_{(u,v) \in M^{(i)}} \left| \mathbf{E} \left[ \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} \right] \right| \\ &\quad \cdot \left| \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} e_{u,v} \mathbf{P}^{[i+1, t_2]} e_w \right|. \end{aligned}$$

As we explained before, the last term is at most  $t_1$ , and for any  $i \in [1, t_1]$  and  $(u, v) \in M^{(i)}$ ,  $\mathbf{E} \left[ \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} \right] \leq (1/2 - \alpha)$ . Thus,

$$|\mathbf{E}[W]| \leq t_1 (1/2 - \alpha).$$

Similarly,  $|\mathbf{E}[W^+]| \leq t_1 (1/2 - \alpha)$  and  $|\mathbf{E}[W^-]| \leq t_1 (1/2 - \alpha)$ .

To apply Hoeffdings bound on  $W$ , we bound the sum of the squared ranges of the involved random variables as follows:

$$\begin{aligned} &\sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \mathbf{Range} \left[ \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} \cdot e_{u,v} \mathbf{P}^{[i+1, t_2]} e_w \right]^2 \\ &\leq \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} (e_{u,v} \mathbf{P}^{[i+1, t_2]} e_w)^2 =: \gamma. \end{aligned}$$

Now by Hoeffdings Bound,

$$\begin{aligned} \Pr [|W| \geq |\mathbf{E}[W]| + \varepsilon] &\leq \Pr [|W - \mathbf{E}[W]| \geq \varepsilon] \\ &\leq 2 \exp \left( -2\varepsilon^2 / \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \mathbf{Range} \left[ \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} (e_{u,v} \mathbf{P}^{[i+1, t_2]} e_w) \right]^2 \right) \\ &\leq 2 \exp (-2\varepsilon^2 / \gamma). \end{aligned}$$



Choosing  $\varepsilon = \sqrt{\gamma \log n}$  we get  $\Pr[|W| \geq |\mathbf{E}[W]| + \sqrt{\gamma \log n}] \leq 2n^{-2}$ . Hence by Lemma 3.2 we obtain with  $\delta = |\mathbf{E}[W]| + \sqrt{\gamma \log n}$  that

$$\begin{aligned} \Pr[|W_{\text{Odd}}| \geq |\mathbf{E}[W]| + 2 \max\{|\mathbf{E}[W^+]|, |\mathbf{E}[W^-]|\} + \sqrt{\gamma \log n}] \\ \leq 16 \Pr[|W| \geq |\mathbf{E}[W]| + \sqrt{\gamma \log n}] \\ \leq 32n^{-2}. \end{aligned}$$

As  $t_1(1/2 - \alpha)$  is an upper bound on each of  $|\mathbf{E}[W]|$ ,  $|\mathbf{E}[W^+]|$  and  $|\mathbf{E}[W^-]|$ , this readily implies

$$\Pr[|W_{\text{Odd}}| \geq 3(1/2 - \alpha)t_1 + \sqrt{\gamma \log n}] \leq 32n^{-2}.$$

Finally, taking the union bound, we conclude by equation (3) that for all vertices  $w \in V$

$$\Pr\left[\bigwedge_{w \in V} \left(|\xi_w^{(t_2)} - x_w^{(t_2)}| \leq 3(1/2 - \alpha)t_1 + \gamma\sqrt{\log n}\right)\right] \geq 1 - 32n^{-1}.$$

**3.3. Proof of Lemma 3.2.** Before we begin the proof of Lemma 3.2, we require the following two technical lemmas.

**Lemma 3.3.** *For  $0 \leq \alpha \leq 1/2$ , arbitrary  $w_i \in \mathbb{R}$ , and independent random variables  $X_i$  that are  $-1$  with probability  $\alpha$  and  $1$  otherwise, let  $X := \sum_{i=1}^n w_i X_i$ . Then,*

- *If all  $w_i \geq 0$ , then  $\Pr[X < 0] \leq 1/2$ ,  
and for any number  $\delta \geq 0$ ,  $\Pr[X > \delta] \geq \Pr[X < -\delta]$ .*
- *If all  $w_i \leq 0$ , then  $\Pr[X > 0] \leq 1/2$ ,  
and for any number  $\delta \geq 0$ ,  $\Pr[X < -\delta] \geq \Pr[X > \delta]$ .*

*Proof.* Note that it suffices to prove the statement with all  $w_i \geq 0$ , as the case with  $w_i \leq 0$  follows from the first by considering  $X' := \sum_{i=1}^n w'_i X_i$  with  $w'_i := -w_i$ . Let  $Y := \sum_{i=1}^n w_i Y_i$  where each  $Y_i$  is an independent and uniform random variable taking values in  $\{-1, +1\}$  (corresponding to  $X$  with  $\alpha = 1/2$ ). As  $Y$  is a sum of symmetrical distributed random variables, we have  $\Pr[Y < 0] \leq 1/2$ . Since for any  $\alpha \leq 1/2$ ,  $X$  stochastically dominates  $Y$ , we obtain  $\Pr[X < 0] \leq \Pr[Y < 0] \leq 1/2$ , and the first claim of the lemma follows.

The second claim is proven similarly by observing that

$$\Pr[X > \delta] \geq \Pr[Y > \delta] = \Pr[Y < -\delta] \geq \Pr[X < -\delta]. \quad \square$$

The following lemma bounds the probability that a sum  $X := \sum_{i=1}^n w_i X_i$  deviates by more than a factor two from its expectation.

**Lemma 3.4.** *For  $0 \leq \alpha \leq 1/2$ , arbitrary  $w_i \in \mathbb{Q}$ , and  $X_i$  independent random variables that are  $-1$  with probability  $\alpha$  and  $1$  otherwise, let  $X := \sum_{i=1}^n w_i X_i$ . Then,*

- (i) *If all  $w_i \geq 0$ , then  $\Pr[X > 2\mathbf{E}[X]] \leq 7/8$ .*

(ii) *If all  $w_i \leq 0$ , then  $\Pr[X < 2\mathbf{E}[X]] \leq 7/8$ .*

*Proof.* As before, it suffices to prove the first statement. Moreover, we may focus on the case  $\mathbf{E}[X] \neq 0$ , as otherwise  $X$  is symmetrically distributed around 0. Finally, we also assume the weights to be integral, that is,  $w_i \in \mathbb{Z}_{\geq 0}$ . Rational weights  $w_i \in \mathbb{Q}_{\geq 0}$  can be easily reduced to integral weights by multiplying all weights with their least common multiple of the denominators and applying the bound for the integral case.

For the sake of contradiction, suppose that  $\Pr[X > 2\mathbf{E}[X]] > 7/8$ . As  $0 \leq \alpha \leq 1/2$ , we have  $\mathbf{E}[X] = \sum_{i=1}^n w_i(1 - 2\alpha) \geq 0$  and hence  $\Pr[X \leq 0] \leq \Pr[X \leq 2\mathbf{E}[X]] < 1/8$ . Let  $k > 2\mathbf{E}[X]$  be such that  $\Pr[X \geq k] \geq 1/8$ , and  $\Pr[X > k] < 1/8$ .

As we assumed the weights  $w_i$  to be integral, we can use the following two well-known counting tricks:

$$\begin{aligned} \sum_{x \geq k} x \Pr[X = x] &= \sum_{x \geq k} \sum_{y=1}^x \Pr[X = x] \\ &= \sum_{y \geq 1} \sum_{x \geq \max(k, y)} \Pr[X = x] \\ &= \sum_{x \geq 1} \Pr[X \geq \max(k, x)], \end{aligned}$$

and similarly,  $\sum_{x \leq -k} -x \Pr[X = x] = \sum_{x \geq 1} \Pr[X \leq -\max(k, x)]$ .

We use both to obtain

$$\begin{aligned} \sum_{x \geq k} x \Pr[X = x] &= \sum_{x \geq 1} \Pr[X \geq \max(k, x)] \\ &= \sum_{1 \leq x < k} \Pr[X \geq k] + \sum_{x \geq k} \Pr[X \geq x] \\ &> \sum_{1 \leq x < k} (\Pr[X \leq 0] + \Pr[X \geq k] - 1/8) \\ &\quad + \sum_{x \geq k} \Pr[X \geq x]. \end{aligned}$$

By Lemma 3.3 we now get

$$\begin{aligned} \sum_{x \geq k} x \Pr[X = x] &\geq k(\Pr[X \geq k] - 1/8) + \sum_{1 \leq x < k} \Pr[X \leq -x] \\ &\quad + \sum_{x \geq k} \Pr[X \leq -x] \\ &= k(\Pr[X \geq k] - 1/8) + \sum_{x \geq 1} \Pr[X \leq -x] \\ &= k(\Pr[X \geq k] - 1/8) + \sum_{x \leq -1} -x \Pr[X = x] \end{aligned}$$

Plugging this in the definition of  $\mathbf{E}[X]$ , we get

$$\begin{aligned} \mathbf{E}[X] &= \sum_x x \Pr[X = x] \\ &= \sum_{x \leq -1} x \Pr[X = x] + \sum_{0 \leq x < k} x \Pr[X = x] + \sum_{x \geq k} x \Pr[X = x] \\ &\geq k (\Pr[X \geq k] - 1/8) + \sum_{0 \leq x < k} x \Pr[X = x]. \end{aligned}$$

Using above assumptions on  $k$  and  $\Pr[X \geq 2\mathbf{E}[X]] > 7/8$  we now arrive at the desired contradiction,

$$\begin{aligned} \mathbf{E}[X] &> 2\mathbf{E}[X] (\Pr[X \geq k] - 1/8) + \sum_{2\mathbf{E}[X] \leq x < k} x \Pr[X = x] \\ &\geq 2\mathbf{E}[X] (\Pr[X \geq k] - 1/8 + \Pr[2\mathbf{E}[X] \leq X < k]) \\ &\geq 2\mathbf{E}[X] (\Pr[X \geq 2\mathbf{E}[X]] - 1/8) \\ &\geq 3/2\mathbf{E}[X]. \end{aligned} \quad \square$$

We are now ready to prove Lemma 3.2. In the following, we will use subsums of  $W$  denoted as  $W_{\text{Even}}^+, W_{\text{Even}}^-, W_{\text{Odd}}^+, W_{\text{Odd}}^-$  which are defined by combining the previous definitions in a natural way, e. g.,

$$W_{\text{Odd}}^+ := \sum_{i=t_1}^{t_2} \sum_{\substack{(u,v) \in M_{\text{Odd}}^{(i)} \\ \Phi_{u,v}^{(i)} w_{u,v}^{(i)} > 0}} \Psi_{u,v}^{(i)} \Phi_{u,v}^{(i)} w_{u,v}^{(i)}.$$

Let  $M^{[t_1, t_2]} = \bigcup_{i=t_1}^{t_2} M^{(i)}$  be the set of all matching edges in the given time span. Let  $M_{\text{Odd}}^{[t_1, t_2]}$  be the set of all odd ones. Moreover, let us simply write  $\mathcal{M}$  for a particular assignment of  $\text{Odd}(\cdot)$  for all balancers in  $M^{[t_1, t_2]}$ . We now begin by bounding  $W_{\text{Odd}}$  in terms of  $W$ . By the definition of conditional probability and the law of total probabilities, it follows for arbitrary  $\delta_1, \delta_2 \in \mathbb{R}$ ,

$$\begin{aligned} \Pr[W < \delta_2 \mid W_{\text{Odd}} \geq \delta_1] &= \frac{\Pr[W < \delta_2 \wedge W_{\text{Odd}} \geq \delta_1]}{\Pr[W_{\text{Odd}} \geq \delta_1]} \\ &\leq \frac{\Pr[W - W_{\text{Odd}} < \delta_2 - \delta_1 \wedge W_{\text{Odd}} \geq \delta_1]}{\Pr[W_{\text{Odd}} \geq \delta_1]} \\ &= \sum_{\mathcal{M}} \frac{\Pr[M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \Pr[W_{\text{Even}} < \delta_2 - \delta_1 \wedge W_{\text{Odd}} \geq \delta_1 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}]}{\Pr[W_{\text{Odd}} \geq \delta_1]}. \end{aligned}$$

Note that for fixed  $\mathcal{M}$ , the ranges of  $W_{\text{Odd}}$  and  $W_{\text{Even}}$  are determined and therefore the probability spaces are disjoint. This implies that  $W_{\text{Odd}}$  and  $W_{\text{Even}}$  are independent conditioned on  $M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}$ . Using this observation,

we get

$$\begin{aligned}
 & \Pr[W < \delta_2 \mid W_{\text{Odd}} \geq \delta_1] \\
 (4) \quad & \leq \sum_{\mathcal{M}} \Pr[M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \overbrace{\Pr[W_{\text{Even}} < \delta_2 - \delta_1 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}]}^{(*)} \\
 & \quad \cdot \Pr[W_{\text{Odd}} \geq \delta_1 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] / \Pr[W_{\text{Odd}} \geq \delta_1].
 \end{aligned}$$

By plugging  $\delta_2 := \delta + 2 \mathbf{E}[W^-]$  and  $\delta_1 := \delta$  in  $(*)$  we obtain

$$\begin{aligned}
 & \Pr[W_{\text{Even}} < 2 \mathbf{E}[W^-] \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \\
 & = 1 - \Pr[W_{\text{Even}} \geq 2 \mathbf{E}[W^-] \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \\
 & \leq 1 - \Pr[W_{\text{Even}}^+ \geq 0 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \\
 & \quad \cdot \Pr[W_{\text{Even}}^- \geq 2 \mathbf{E}[W^-] \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}].
 \end{aligned}$$

Observing that  $\mathbf{E}[\Psi_{u,v}] = \mathbf{E}[\Psi_{u,v} \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}]$  for  $(u, v) \in M_{\text{Even}}^{(i)}$ , we get  $\mathbf{E}[W^-] \leq \mathbf{E}[W_{\text{Even}}^- \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}]$  and hence by Lemmas 3.3 and 3.4,

$$\begin{aligned}
 & \Pr[W_{\text{Even}} < 2 \mathbf{E}[W^-] \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \\
 & \leq 1 - \Pr[W_{\text{Even}}^+ \geq 0 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \cdot \\
 & \quad \Pr[W_{\text{Even}}^- \geq 2 \mathbf{E}[W_{\text{Even}}^- \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \\
 (5) \quad & \leq 1 - \frac{1}{2} \cdot \frac{1}{8} = \frac{15}{16}.
 \end{aligned}$$

Plugging this into equation (4) yields

$$\begin{aligned}
 & \Pr[W < \delta + 2 \mathbf{E}[W^-] \mid W_{\text{Odd}} \geq \delta] \\
 & \leq \frac{15}{16} \sum_{\mathcal{M}} \frac{\Pr[M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \Pr[W_{\text{Odd}} \geq \delta_1 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}]}{\Pr[W_{\text{Odd}} \geq \delta_1]} = \frac{15}{16},
 \end{aligned}$$

and hence

$$\begin{aligned}
 & \Pr[W \geq \delta + 2 \mathbf{E}[W^-]] \\
 & \geq \Pr[W \geq \delta + 2 \mathbf{E}[W^-] \wedge W_{\text{Odd}} \geq \delta] \\
 & = \Pr[W_{\text{Odd}} \geq \delta] \cdot \Pr[W \geq \delta + 2 \mathbf{E}[W^-] \mid W_{\text{Odd}} \geq \delta] \\
 (6) \quad & \geq \frac{1}{16} \cdot \Pr[W_{\text{Odd}} \geq \delta].
 \end{aligned}$$

It remains to lower bound the deviation of  $W_{\text{Odd}}$  in terms of  $W$  in a similar fashion. As before, we derive

$$(7) \quad \begin{aligned} & \Pr[W > \delta_2 \mid W_{\text{Odd}} \leq \delta_1] \\ & \leq \sum_{\mathcal{M}} \Pr[M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \overbrace{\Pr[W_{\text{Even}} > \delta_2 - \delta_1 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}]}^{(*)} \\ & \quad \cdot \Pr[W_{\text{Odd}} \leq \delta_1 \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] / \Pr[W_{\text{Odd}} \leq \delta_1]. \end{aligned}$$

We now choose  $\delta_2 := -\delta + 2 \mathbf{E}[W^+]$ , and  $\delta_1 := -\delta$ . As before in equation (5), we can now use Lemmas 3.3 and 3.4 to get for (\*) that

$$\Pr[W_{\text{Even}} > 2 \mathbf{E}[W^+] \mid M_{\text{Odd}}^{[t_1, t_2]} = \mathcal{M}] \leq \frac{15}{16}.$$

Plugging this into equation (7) yields

$$\Pr[W > -\delta + 2 \mathbf{E}[W^+] \mid W_{\text{Odd}} \leq -\delta] \leq 15/16$$

and hence

$$(8) \quad \begin{aligned} & \Pr[W \leq -\delta + 2 \mathbf{E}[W^+]] \\ & \geq \Pr[W \leq -\delta + 2 \mathbf{E}[W^+] \wedge W_{\text{Odd}} \leq -\delta] \\ & = \Pr[W_{\text{Odd}} \leq -\delta] \cdot \Pr[W \leq -\delta + 2 \mathbf{E}[W^+] \mid W_{\text{Odd}} \leq -\delta] \\ & \geq \frac{1}{16} \cdot \Pr[W_{\text{Odd}} \leq -\delta]. \end{aligned}$$

Combining equation (6) and equation (8), we have shown for any  $\delta \geq 0$ ,

$$\begin{aligned} & \Pr[|W_{\text{Odd}}| \geq \delta] \\ & = \Pr[W_{\text{Odd}} \geq \delta] + \Pr[W_{\text{Odd}} \leq -\delta] \\ & \leq 16 \Pr[W \geq \delta + 2 \mathbf{E}[W^-]] + 16 \Pr[W \leq -\delta + 2 \mathbf{E}[W^+]] \\ & \leq 16 \Pr[|W| \geq \delta - 2 \max\{|\mathbf{E}[W^+]|, |\mathbf{E}[W^-]|\}]. \end{aligned}$$

Adding  $2 \max\{|\mathbf{E}[W^+]|, |\mathbf{E}[W^-]|\}$  to  $\delta$  gives the assertion of the lemma.

#### 4. DERIVING THE UPPER BOUND FOR THE PERIODIC CASE

Consider the network after  $\frac{2 \log(Kn)}{1-\lambda(\mathbf{Q})}$  (periodic) repetitions of the balancing network (each repetition consists of  $d$  rounds, so the network consists of a total of  $T = d \cdot \frac{2 \log(Kn)}{1-\lambda(\mathbf{Q})}$  matchings).  $\mathbf{Q}$  is the so-called round matrix corresponding to the application of  $d$  consecutive matchings (recall that we use the following abbreviations  $\mathbf{P}^{[t_1, t_2]} = \prod_{i=t_1}^{t_2} \mathbf{P}^{(i)}$ ). This is indeed a special case of our general scheme where the matrices  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(T)}$  are applied, but now  $\mathbf{Q} = \mathbf{P}^{[1, d]} = \mathbf{P}^{[d+1, 2d]}$  and so on. Recall the notation that  $e_{u,v}$  denotes the row-vector with  $+1$  at  $u$ -th column and  $-1$  at  $v$ -th column. In Theorem 1.1 we plug  $t_2 = T = \frac{2d \log(Kn)}{1-\lambda(\mathbf{Q})}$  and  $t_1 = T - \frac{2d \log \log(n)}{1-\lambda(\mathbf{Q})}$ . Hence,

$$\Lambda_1 = \max_{w \in V} \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} (e_{u,v} \mathbf{P}^{[i+1,t_2]} e_w)^2}.$$

Let  $v_1^{(i)}, v_2^{(i)}, \dots, v_n^{(i)}$  be the eigenvectors of  $\mathbf{P}^{[i+1,t_2]}$  forming an orthogonal basis of  $\mathbb{R}^n$  and let  $\lambda_1^{(i)}, \lambda_2^{(i)}, \dots, \lambda_n^{(i)}$  be the corresponding eigenvalues. Since  $\mathbf{P}^{[i+1,t_2]}$  is a stochastic matrix,  $v_1^{(i)} = \mathbf{1}$ . Write  $e_w = \sum_{j=1}^n c_j^{(i)} v_j^{(i)}$ . Using this, we rewrite  $\Lambda_1$  as follows (for short we drop the “ $\max_w$ ” part as our final result will not depend on  $w$ ),

$$\begin{aligned} \Lambda_1 &= \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \left( e_{u,v} \mathbf{P}^{[i+1,t_2]} \sum_{j=1}^n c_j^{(i)} v_j^{(i)} \right)^2} \\ &= \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \left( e_{u,v} \sum_{j=1}^n c_j^{(i)} \lambda_j^{(i)} v_j^{(i)} \right)^2}. \end{aligned}$$

Observe that  $e_{u,v}$  is orthogonal to  $v_1^{(i)}$ . Therefore  $c_1^{(i)} = 0$  for every  $i$  which gives

$$\Lambda_1 = \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} \left( e_{u,v} \sum_{j=2}^n c_j^{(i)} \lambda_j^{(i)} v_j^{(i)} \right)^2}.$$

Define a vector  $z^{(i)} := \sum_{j=2}^n c_j^{(i)} \lambda_j^{(i)} v_j^{(i)}$ . The latter is then

$$\begin{aligned} \Lambda_1 &= \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} (e_{u,v} z^{(i)})^2} \\ &= \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} (z_u^{(i)} - z_v^{(i)})^2} \\ &\leq \sqrt{\log n \sum_{i=1}^{t_1} \sum_{(u,v) \in M^{(i)}} 2(z_u^{(i)})^2 + 2(z_v^{(i)})^2}. \end{aligned}$$

Since  $M^{(i)}$  is a matching, for each  $i$  each vertex is counted only at most once, thus

$$(9) \quad \Lambda_1 \leq 2 \sqrt{\log n \sum_{i=1}^{t_1} \|z^{(i)}\|^2}.$$

By standard calculation (cf. Section 3.1), we have

$$\|z^{(i)}\|^2 \leq \lambda(\mathbf{P}^{[i+1,t_2]})^2 \cdot \|e_w\|^2 = \lambda(\mathbf{P}^{[i+1,t_2]})^2.$$

Plugging this into (9) and using the facts that for any two stochastic matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,  $\lambda(\mathbf{AB}) \leq \lambda(\mathbf{A})$  and for any integer  $k$ ,  $\lambda(\mathbf{A}^k) = \lambda(\mathbf{A})^k$ ,

$$\begin{aligned} \Lambda_1 &\leq 2 \sqrt{\log n \sum_{i=1}^{t_1} \lambda(\mathbf{P}^{[i+1,t_2]})^2} \\ &\leq 2 \sqrt{\log n \sum_{i=1}^{t_1} \lambda(\mathbf{Q}^{\lfloor (t_2 - (i+1))/d \rfloor})^2}. \end{aligned}$$

Regrouping the matrices according to the periods the latter reduces to

$$\Lambda_1 \leq \sqrt{\log n \sum_{i=1}^{t_1/d} d \lambda(\mathbf{Q})^{2(t_2/d - (i+1))}}.$$

With  $\rho = ((t_2 - t_1)/d) - 1$ , we can upper bound  $\Lambda_1$  by

$$2\sqrt{\log n d \sum_{i=\rho}^{\infty} \lambda(\mathbf{Q})^{2i}} \leq 2\sqrt{\log n d \frac{\lambda(\mathbf{Q})^{2\rho}}{1 - \lambda(\mathbf{Q})^2}}.$$

Plugging in our choices for  $t_1$  and  $t_2$  we end up with

$$\begin{aligned} \Lambda_1 &\leq 2\sqrt{d \log n \frac{\lambda(\mathbf{Q})^{-1+4 \log \log n/(1-\lambda(\mathbf{Q}))}}{1 - \lambda(\mathbf{Q})^2}} \\ &\leq 2\sqrt{d \log n \frac{\exp(-3 \log \log n)}{1 - \lambda(\mathbf{Q})^2}} \\ &\leq 2 \frac{1}{\log n} \sqrt{d \frac{1}{1 - \lambda(\mathbf{Q})^2}} \\ &\leq \frac{2\sqrt{d}}{\log n} \cdot \frac{1}{1 - \lambda(\mathbf{Q})}. \end{aligned}$$

With the same arguments, we get

$$\Lambda_2 \leq \lambda(\mathbf{P}^{[1, t_2]}) \sqrt{n} K \leq \sqrt{n} K / (Kn)^2.$$

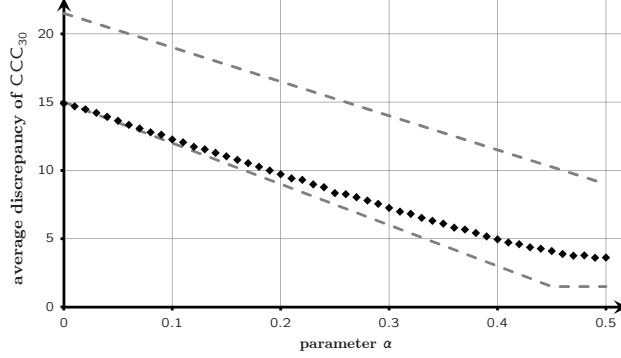
Plugging all this into Theorem 1.1 shows that at step  $T = \frac{2 \log(Kn)}{1 - \lambda_2(\mathbf{P})}$  the discrepancy is at most

$$\begin{aligned} &\frac{2d \log \log n}{1 - \lambda(\mathbf{Q})} + \frac{3(1/2 - \alpha) \cdot 2d \log(Kn)}{1 - \lambda(\mathbf{Q})} + \frac{2\sqrt{d}}{\log n} \cdot \frac{1}{1 - \lambda(\mathbf{Q})} + \frac{\sqrt{n} K}{(Kn)^2} \\ &= \mathcal{O}\left(\frac{d \log(Kn)}{1 - \lambda(\mathbf{Q})} \cdot \left(\frac{1}{2} - \alpha\right) + \frac{d \log \log n}{1 - \lambda(\mathbf{Q})}\right). \end{aligned}$$

## 5. DERIVING THE UPPER BOUND FOR $\text{CCC}_n$

In Theorem 1.1 we choose  $t_2 = \log n$  and  $t_1 = \log n - \log \log n$ . For matrix multiplication we use the following abbreviated form:  $\mathbf{P}^{[i, j]} := \mathbf{P}^{(i)} \mathbf{P}^{(i+1)} \dots \mathbf{P}^{(j)}$ . First we observe that

- $\mathbf{P}_{u,v}^{[k, t_2]} = 2^k/n$  if wires  $u$  and  $v$  are at distance  $\leq \log n - k$  and differ only in the last  $\log n - k$  bits (in their binary representation),
- otherwise,  $\mathbf{P}_{u,v}^{[k, t_2]} = 0$ .



**Figure 2.** Discrepancy for various  $\alpha$ -values of  $\text{CCC}_{2^{30}}$  with random uniformly distributed input from  $\{0, 1, 2, \dots, 2^{30} - 1\}$ .  $\alpha = 0$  corresponds to the adversarial model while  $\alpha = 1/2$  is the completely random model. The dotted line describes the experimental results, the broken lines are our theoretical lower and upper bounds.

In particular this shows that  $\mathbf{P}^{[1, t_2]}$  is equal to the all- $(\frac{1}{n})$  matrix and thus  $\lambda(\mathbf{P}^{[1, t_2]}) = 0$  (that is,  $\Lambda_2 = 0$ ). Moreover, for any fixed wire  $w$ ,

$$\begin{aligned}
 \Lambda_1 &= \sqrt{\log n \sum_{i=1}^{\log n - \log \log n} \sum_{(u,v) \in M^{(i)}} (e_{u,v}(\mathbf{P}^{[i+1, t_2]}) e_w)^2} \\
 &= \sqrt{\log n \sum_{i=1}^{\log n - \log \log n} \sum_{(u,v) \in M^{(i)}} (\mathbf{P}_{u,w}^{[i+1, t_2]} - \mathbf{P}_{v,w}^{[i+1, t_2]})^2} \\
 &\leq \sqrt{\log n \sum_{i=1}^{\log n - \log \log n} \frac{n}{2^i} \left(\frac{2^{i+1}}{n}\right)^2} \\
 &= 4 \sqrt{\log n \sum_{i=1}^{\log n - \log \log n} \frac{2^i}{n}} \\
 &\leq 4 \sqrt{\log n \cdot \frac{2}{\log n}} \\
 &= \mathcal{O}(1).
 \end{aligned}$$

## 6. EXPERIMENTAL RESULT

We examined experimentally how well a CCC balances a random input. We implemented a  $\text{CCC}_{2^{30}}$  consisting of roughly one billion wires and thirty billion balancers. The input was chosen independently uniformly at random from  $\{0, 1, 2, \dots, 2^{30} - 1\}$ . For initial directions of the balancers all up and different values  $\alpha$  between 0 and  $1/2$  we measured the resulted discrepancy.

Figure 2 presents the average over 100 runs, together with theoretical lower and upper bounds. As the input is random, the so-far presented bounds can be tightened. Following the same lines of proof, one can easily show the following slightly better bounds on the *expected* discrepancy  $\Delta$  in the random-input case:



- $\Delta \leq (\frac{1}{2} - \alpha) \cdot (\log n - \lceil \log \log n \rceil) + \lceil \log \log n \rceil + 4,$
- $\Delta \geq \max\{(1/2 - \alpha) \log n, 1/2(1 - \frac{1}{n})(\lceil \log \log n \rceil - 1)\}.$

As these bounds are only used for visualization in Figure 2 and the proofs are very similar to the above, they are omitted.

## REFERENCES

- [1] D. Arthur and S. Vassilvitskii. Worst-case and smoothed analysis of the icp algorithm, with an application to the k-means method. In *47th IEEE Symp. on Found. of Comp. Science (FOCS'06)*, pages 153–164, 2006.
- [2] J. Aspnes, M. Herlihy, and N. Shavit. Counting networks. *J. of the ACM*, 41(5):1020–1048, 1994.
- [3] D. Bertsekas and J. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [4] T. Bohman, A. Frieze, and R. Martin. How many random edges make a dense graph hamiltonian? *Random Structures and Algorithms*, 22(1): 33–42, 2003.
- [5] A. Coja-Oghlan, U. Feige, A. M. Frieze, M. Krivelevich, and D. Vilenchik. On smoothed  $k$ -CNF formulas and the walksat algorithm. In *20th ACM-SIAM Symp. on Discrete Algorithms (SODA'09)*, pages 451–460, 2009.
- [6] M. Dowd, Y. Perl, L. Rudolph, and M. Saks. The periodic balanced sorting network. *J. of the ACM*, 36(4):738–757, 1989.
- [7] U. Feige. Refuting smoothed 3CNF formulas. In *48th IEEE Symp. on Found. of Comp. Science (FOCS'07)*, pages 407–417, 2007.
- [8] A. Flaxman and A. Frieze. The diameter of randomly perturbed digraphs and some applications. *Random Structures and Algorithms*, 30: 484–504, 2007.
- [9] T. Friedrich and T. Sauerwald. Near-perfect load balancing by randomized rounding. In *41st Annual ACM Symposium on Theory of Computing (STOC'09)*, pages 121–130, 2009.
- [10] T. Friedrich, T. Sauerwald, and D. Vilenchik. Smoothed analysis of balancing networks. In *36th International Colloquium on Automata, Languages, and Programming (ICALP'09)*, volume 5556 of *Lecture Notes in Computer Science*, pages 472–483. Springer, 2009.
- [11] M. Herlihy and N. Shavit. *The Art of Multiprocessor Programming*. Morgan Kaufmann, 2008.
- [12] M. Herlihy and S. Tirthapura. Randomized smoothing networks. *J. Parallel and Distributed Computing*, 66(5):626–632, 2006.
- [13] M. Herlihy and S. Tirthapura. Self-stabilizing smoothing and counting networks. *Distributed Computing*, 18(5):345–357, 2006.
- [14] M. Krivelevich, B. Sudakov, and P. Tetali. On smoothed analysis in dense graphs and formulas. *Random Structures and Algorithms*, 29(2): 180–193, 2006.

- [15] B. Manthey and R. Reischuk. Smoothed analysis of binary search trees. *Theoret. Computer Sci.*, 378(3):292–315, 2007.
- [16] M. Mavronicolas and T. Sauerwald. The impact of randomization in smoothing networks. In *27th Annual ACM Principles of Distributed Computing (PODC'08)*, pages 345–354, 2008.
- [17] Y. Rabani, A. Sinclair, and R. Wanka. Local divergence of Markov chains and the analysis of iterative load balancing schemes. In *39th Annual IEEE Symposium on Foundations of Computer Science (FOCS'98)*, pages 694–705, 1998.
- [18] H. Röglin and B. Vöcking. Smoothed analysis of integer programming. *Math. Program.*, 110(1):21–56, 2007.
- [19] D. Spielman and S. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. of the ACM*, 51(3):385–463, 2004.
- [20] R. Vershynin. Beyond hirsch conjecture: Walks on random polytopes and smoothed complexity of the simplex method. In *47th IEEE Symp. on Found. of Comp. Science (FOCS'06)*, pages 133–142, 2006.